8-2014

Opinion Mining of Sociopolitical Comments from Social Media

Swapna GOTTIPATI
Singapore Management University, SWAPNAG@smu.edu.sg

Opinion Mining of Sociopolitical Comments from Social Media

by

Swapna Gottipati

Submitted to School of Information Systems in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Information Systems

<u>Dissertation Committee:</u>

Assistant Professor Jing Jiang(Supervisor / Chair)
Singapore Management University

Professor Ee-Peng Lim (Co-supervisor)
Singapore Management University

Assistant Professor Zhu Feida
Singapore Management University

Associate Professor Min-Yen Kan
National University of Singapore

Singapore Management University
2013

Opinion Mining of Sociopolitical Comments from Social Media

Swapna Gottipati

Abstract

Opinions are central to almost all human activities by influencing greatly the decision making process. In this thesis, we present the problems of mining issues, extracting entities and suggestive opinions towards the entities, detecting thoughtful comments, and extracting stances and ideological expressions from online comments in the sociopolitical domain. This study is essential for opinion mining applications that are beneficial for policy makers, government sectors and social organizations. Much work has been done to try to uncover consumer sentiments from online comments to help businesses improve their products and services. However, sociopolitical opinion mining poses new challenges due to complex topic and sentiment expressions.

We first present the problem of issue extraction from sociopolitical comments for which we propose an unsupervised approach based on latent variable methods for identifying and extracting the issues in the comments, and linking comments to the issues in the associated article. We evaluate our approach on political speeches and associated comments from social media.

In the sociopolitical domain, users express their sentiments on the entities such as individuals or organizations. These sentiments are not only in the form of positive and negative expressions, but also in the form of suggestive opinions towards the entities. We present a new problem of extracting the entities and associated suggestive opinions. We propose a two-stage approach based on conditional random fields (CRF) and clustering for extracting and normalizing the entities and the associated suggestive opinions from the users.

A key feature of social media is that it enables anyone to freely express his/her

opinions. As a result of the large amount of online comments, there is an urge for extracting opinions which are highly valuable. In terms of thoughtful comment extraction, we study the task of extracting valuable comments from social media. We propose a supervised approach based on natural language processing and linguistics techniques to identify and extract valuable comments in the sociopolitical domain from social media.

Users take positions/stances and express opinions towards controversial sociopolitical issues. We present the problem of extracting the topics, stances, and ideological expressions of users from their comments on ideological debates related to sociopolitical domain. We propose an unsupervised approach based on latent variable methods and evaluate on Debatepedia for identifying and extracting the positional words and entities associated with the issues.

In summary, this thesis identifies a number of key problems in mining sociopolitical comments and proposes appropriate solutions to these problems.

Table of Contents

List of Figures

List of Tables

Acknowledgments

First, it is a pleasant task to express my respects to numerous people including my well wishers, friends, colleagues and various institutions for their continuous help and support in all stages of my doctoral study at Singapore Management University.

First and foremost, I owe a great deal of gratitude to my advisor, Assistant Professor Jing Jiang. Her immense knowledge and unsurpassed vision and mission in research which conjured me have become the cornerstone of this thesis. Her irrevocable guidance, encouragement and fruitful cooperation have helped me to shape my interest and ideas. Jing's valuable comments directed me to sharpen my text and ideas, and were of great support for my thesis. Moreover, she was one of the best advisors and remains my best role model for a scientist, mentor and teacher. The thesis would not have come to a successful completion without the unequivocal support throughout, I received from Jing for which mere thanks likewise does not suffice.

Besides my mentor, I would like to blue print my heartful thanks to the blessed group, the thesis committee, Professor Lim Ee-Peng, Assistant Professor Zhu Feida and Associate Professor Kan Min-Yen for their research vision and commitment. Their constructive comments and suggestions made this thesis possible and held this work to a high standard. I am extremely indebted to Prof Kan from National University Singapore who provided invaluable comments for shaping the thesis.

I gratefully acknowledge Assistant Professor David Lo, Assistant Professor Rajesh Balan, Assistant Professor Debin Goh, Assistant Professor Narayan Ramasubbu and Professor Venky Shankararaman for their irreparable encouragement and insightful feedback on my research. I extend my thanks to Prof Steve Fienberg (Director, Living Analytics Research Centre), Prof Lim Ee-Peng (Director, Living Analytics Research Centre) and Prof Noah Smith from Carnegie Mellon University for giving me an opportunity to be a part of CMU exchange program.

I am very thankful to Living Analytics Research Centre(LARC) for the generous financial support in the form of LARC graduate scholarship throughout my pursuance of CMU exchange program. I would like to express my deep thanks to Noah for his great support and energetic guidance for my research during CMU exchange program. The Debatepedia work presented in Chapter 6 benefited from guidance by Noah and helpful discussions from Yanchuan Sim from CMU.

I have been blessed with a cheerful group of colleagues and friends at LARC; Minghui Qiu, Diao Qiming, Payas Gupta, Yang Liu, Karthik Muralidharan and Hoang Tuan Anh. I would like to thank my amazing friends from School of Information Systems (SIS), Lucia, Yazhe, Shaoying, Kar way, Sougata and Yueqiang with whom I shared my precious graduate life. I have been benefited a lot from them through research collaborations within as well as out of my research area.

Finally, I am indebted to my husband, parents, in-laws, siblings and friends whose moral and emotional support, and unconditional love has helped me to complete this venture. I would not have made it this far without them. Thanks and loving appreciation to Swetha, Kavita and Sam. My darling kids Anish and Nidhi have been always with me. They ensure to wish me good luck for my exams, talks, conferences and all important events of my graduate journey. This thesis would not have been possible without their cooperation and is dedicated to them.

Above all, I owe it all to Almighty God for granting me, the wisdom, health and strength to undertake this work and enabling me to its completion.

to Anand, my parents and in-laws

Chapter 1

Introduction

With the emerging Web 2.0 technologies and Web 2.0 sites such as forums, blogs, and online social networks, there has been a tremendous amount of interest and efforts to utilize user-generated content on the Web. One of the major striking differences of Web 2.0 from Web 1.0 is the abundance of ordinary people's opinions on various topics expressed in social media. The availability of huge volumes of opinionated data has opened room for opinion mining, a field of computational study that analyzes people's opinions or sentiments.

In particular, people express their opinions on sociopolitical issues, and these opinions are very important for policy makers, government sectors and social organizations. This thesis is about mining users' comments from social media on sociopolitical issues. Much work has been done to try to uncover consumer sentiments from online reviews, blogs, etc., to help businesses improve their products and services. In contrast, less attention has been paid to the extraction and summarization of public opinions on social and political issues. The existing work on opinion mining is insufficient for sociopolitical opinion mining due to complex topics and sentiment expressions in the sociopolitical domain. In this introduction chapter, we present some background about opinion mining, followed by the motivation that inspired the research in this thesis and finally our major contributions.

1.1 Background

In the real world, individual consumers and domain experts always want to know the consumers' or public's opinions towards products, services, organizations, individuals, issues, events, topics and their attributes for their decision making process [132, 79]. An individual would like to know others' opinions on a specific product before purchasing that product. Domain experts utilize the public feedback for improving their products or implementing new policy ideas. Hence mining opinions became central to several decision making systems [87, 82, 22, 125].

Opinion mining is a computational study of people's opinions, sentiments, evaluations, attitudes and emotions [99]. Since early 2000, opinion mining has grown to be one of the most active research areas in natural language processing (NLP), data mining, Web mining, and text mining. With the exponential growth of opinionated documents on the Web, individuals and organizations have been using the content for their decision making. To process such enormous data, automated systems are desired. Such automated systems are referred to as opinion mining systems. Opinion mining systems have found their applications in several business and social domains [82, 22, 25].

To examine the main sub-tasks involved in the opinion mining area, let us study the product review (consumer business) mining problem that aims at mining reviews which express positive or negative sentiments towards the products and their features. To illustrate the problem, let us observe the following review about a Nikon camera.

"I bought a Nikon 5230 three months ago(1). It is a wonderful camera and I love it(2). The picture quality is very good(3). The battery life is long(4). However, I feel it is too heavy(5)".

We notice the following points about the review. Sentence 2 is about sentiment regarding the camera. Sentence 3, 4 and 5 are about camera features, namely picture, battery and weight respectively. Sentence 2, 3 and 4 express positive sentiments. In contrast, Sentence 5 expresses negative sentiments.

Opinion mining of product reviews consists of the following most important sub-tasks [61, 45, 65, 100]:

1. Identifying and extracting features in evaluative texts.
2. Determining sentiment polarities on the features of products.
3. Contrastive opinion mining.
4. Opinion quality and spam detection.
5. Opinion Summarization.

The feature extraction task aims at identifying the features of the product from the reviews [61]. The feature extraction task is also sometimes referred to as aspect extraction [79]. Sentiment classification aims at classifying the documents to positive or negative opinion [99] where the task is defined on regular opinions. Other types of opinions are comparative, sarcastic, ironic and spam opinions. We shall explain these in detail during our literature survey in Chapter 2.

1.2 Motivation

Social and political data are much harder to analyse due to complex topic and sentiment expressions [79]. In this section, we show our motivation for opinion mining of sociopolitical comments and the associated challenges. Mining social comments is critically important to build applications which aid policy makers, social organizations or government sectors in decision making[1]. For example, a politician might want to know the response of the public on his speech, and which topics of his speech had a major impact on the public. Mining comments of such a speech helps in quickly zooming into the most important social and political problems in a society. An individual might want to know others' opinions on politicians before making a voting decision. A congress member would like to see how the public

[1] `http://public.dhe.ibm.com/common/ssi/ecm/en/gvs03044usen/GVS03044USEN.PDF`

responded to a new immigration bill before making any amendment decisions. A ruling government might want to get the citizens' opinions towards a change in policy/s. Traditionally, such information has been collected through opinion polls or surveys. However, this kind of feedback channels requires a large amount of manpower support [97]. Also only a small fraction of the entire population can be reached, which may lead to biased information collection.

Online social networks and forums, on the other hand, are now getting increasingly popular and have attracted a large number of users, with many of them expressing and discussing their opinions on sociopolitical topics through these websites [137]. Users post their opinions on social topics in the form of posts or comments. With the availability of dedicated sites that discuss sociopolitical problems, collecting comments from the social media is relatively easy. Comments in the sociopolitical domain usually consist of issues, entities and opinions associated with the entities. We also observe that users' comments vary greatly in quality and not all comments are valuable to a domain user. Further, the users in the sociopolitical domain possess underlying ideological beliefs and they take positions while commenting/arguing on the social issues. Mining such data is useful but also a very challenging task. To illustrate the problem, let us examine the sample comments from Table 1.1.

C1	You want to really drive innovation, job growth and entrepreneurs? Make education, health care and retirement less of a burden on the average family, adopt more socialist policies like Norway (paid for by higher taxes, especially on the rich), and watch our standard of living rise at last
C2	The government should lift diplomatic immunity of the ambassador.
C3	..Oh, so Obama "compromised" on the tax cuts for the wealth"
C4	Low taxes arent helping the vast middle and working class and aren't creating more jobs, its a policy that only benefits the rich.
C5	I am pro-choice simply because I don't believe that the government should have control over this. It's ridiculous. Women should have control over what happens to their body.

Table 1.1: Sample sociopolitical comments from social media

To introduce the terminology in sociopolitical domain, and to better analyze and understand comments such as the ones shown in Table 1.1, we define a few concepts as follows:

Issues: Issues refer to the social, political and economical problems such as education, job, abortion, etc. Comment **C1** is a response to U.S. President Barack Obama's State of the Union Address. In his speech, Obama addressed eight social and political issues. The author of C1 commented on issues related to *education, economy* and *innovation*. Unlike product reviews which focus on the specific product, the social comments target at various issues, and hence it is important to identify and extract these issues from the social comments.

Entities: Similar to product review mining, where users express opinions on features or aspects, in a sociopolitical domain, users express opinions on entities. Entities refer to individuals or organizations associated with an issue. Comment **C2** is a response to a major accident in Singapore caused by a Romanian ambassador. In this comment, the user posted his/her opinion on an entity, *the Government*, associated with this issue.

Suggestive Opinions: Suggestive opinions are actions suggested for an entity. In comment **C2**, the user expressed his/her suggestive opinion towards the entity, *the Government*. The user suggests that the "*diplomatic immunity of the ambassador should be lifted*" as a response to the issue. Extracting the entities and the suggestive opinions is an important subtask of mining sociopolitical comments.

Thoughtful Comments: A valuable/attentive comment is one that provides not only an opinion but also a justification for the commented issue. Finding attentive comments that provide some reasoning is highly valuable in understanding the user's opinion. Comment **C3** depicts an opinion towards *tax*. In this case, the comment doesn't provide any justification and doesn't provide any insights to the users opinion. In contrast, comment **C4** which is also on *tax*, elaborates on the users' viewpoint towards this issue. Such comments are valuable to the domain experts, and we treat such comments as thoughtful comments.

Ideological Positions/Stances: Users express their opinions on controversial social issues with an underlying ideological belief. They take positions, on ideological debates to post their arguments. Comment **C5** is an argument on *abortion* and the user exhibits a supporting/pro position towards it. The users' positions are referred to as ideological stances. The sentiment expression in the comment, *pro-choice*, depicts the user's opinion on abortion. We refer to such expressions or opinion words as ideological expressions.

In brief, the problem of mining sociopolitical comments involves a fine grained analysis of issues, entities, suggestive-opinions, quality, stances, and ideological expressions. We aim to automatically extract fine-grained information from sociopolitical comments which is critically important to build the applications related to opinion mining in the sociopolitical domain.

1.3 Limitations of Existing Studies

With the aforementioned subtasks and new challenges in mining sociopolitical comments, let us now turn to existing opinion mining techniques to see whether they are sufficient to address these new challenges. We argue that they are not. The reasons are the following. Traditional opinion mining techniques and sentiment lexicons are often developed for consumer business such as products and services. Further, the previous research focussed mainly on product reviews from dedicated review sites. Comparatively, very few studies focussed on the opinions from blogs and forums largely due to the complexity of blog posts and annotation challenges [90]. However for sociopolitical issues, the comments are expressed mainly in blogs, forums, debate sites, twitter etc.. Moreover, the existing product review mining studies benefit from the assumptions that the product reviews are focussed for a given product and the features of the product are mostly fixed [80, 60]. For example, in the case of mobile phones such as Nokia and iPhone, usually the forum/review page will be dedicated for a specific model on which users can post reviews. Their opinions are expressed on the associated features such as battery, size, etc. In contrast, in the sociopolitical domain, first an article is posted online for which users can post opinions. In many cases, these sociopolitical articles consist of several issues. For example, a US President's political speech covers many social problems like jobs, healthcare, education, military, etc. Users might comment on one or more of them as shown in comment **C1** and the first challenge is to extract the issues from users' comments. Current methods don't cater for such needs and therefore the first challenge is not only to extract issues, but also to align them with the topics in the associated article.

Second, the sentiment expressions on products are either polarised or ranked. Similar to product sentiments, for sociopolitical issues the users provide polarised or ranked sentiments on the entities related to the issues and the overall sentiment of a sociopolitical comment can be discovered using the current sentiment techniques. Nevertheless, the sentiments in sociopolitical comments are much more than simply polarised or ranked. Overall, on sociopolitical data users express sentiments in various ways; on entities associated with the issues in the form of polarised expressions, on entities associated with the issues in the form of suggestions, on issues in the form of stances/stance expressions, and on the aspects of issues in the form of polarised expressions. While sentiment polarity classification in the sociopolitical domain can still benefit from existing techniques, we also require new techniques for other types of sentiments.

In this study we focus on suggestion expressions and stance expressions which are described in the following paragraphs. To study suggestion sentiments, we are interested in extracting the suggestive opinions on the entities from the users' comments. The sentiments expressed towards product features vary from the sentiment expressed towards the social issues. For example, on products users might comment that *"The call quality is good"* or *"I like the screen"*, where *call quality* and *screen* are the product features [100, 99], and *good* and *like* are sentiments. In contrast, for social issues users comment on the entities (individuals or organizations) associated with the issue. These comments can be in the form of likes or dislikes or in the form of suggestive opinions as shown in comment **C2**. For example, a user might comment that *"The Government should adopt stricter policies for immigrants"*. In this example, *Government* is an entity and *adopt stricter policies for immigrants* is a suggestive opinion towards the entity. This urges the need for extracting the entities and the associated suggestive opinions, which is not supported by current opinion mining techniques.

Third, traditional quality assessment techniques are designed for student essays or product reviews, and hence are not suitable for social comments due to the different nature of high quality comments in the sociopolitical domain. For example, a high quality product review is regarded as the one which elaborates on all features of the product to help the readers [71, 47]. In contrast, a valuable comment on a social issue is the one that provides the justification or reasoning or insightful ideas to the issue commented upon [120] as shown in Comment **C4**. Further, to be useful to domain experts, the comment doesn't have to describe all the issues with respect to the social article. Therefore there is need for capturing the justification component together with the relevance of the issue to the document, which is not handled by current quality assessment techniques.

Another unique property of sociopolitical data is that the users take stances on the issues with an underlying ideological belief. Such stances and the expressions can be treated as the sentiments of the users on the controversial issues. An ideological belief is expressed with opinion words which are specific to the social problem as shown in Comment **C5**. For example, a person who is against *abortion* uses sentiment expressions such as *pro-life, birth control*, etc. Current sentiment lexicons comprises of lists of sentiment words that express positive or negative sentiments [134] on product features. For example, *good, wonderful* and *amazing* are positive sentiment words and in contrast, *bad, poor* and *terrible* are negative sentiment words. Although using such lexicons is important for sentiment classification in product review mining, they still pose some limitations in cases such as: opposite orientation in different domains, sarcastic sentences, ironies and implicit opinionated sentences. Therefore, the fourth challenge is that, for sociopolitical data, these existing lexicons are insufficient to capture the sentiment expressions that are required for discovering the ideological stances and expressions.

1.4 Objective

Our objective of this thesis is to study fine grained opinion mining that aims to extract and summarize people's opinions related to sociopolitical aspects. In this thesis, we present a comprehensive study of mining issues, entities, suggestive opinions, quality, stances and ideological expressions related to sociopolitical domain. The study involves four basic tasks:

1. Identify and extract the issues in comments and the associated article.
2. Extract and normalize the entities and suggestive opinions towards the entity.
3. Identify and extract valuable (quality) comments.
4. Identify and extract the ideological positions/stances and ideological expressions towards the issues.

A most important question to be answered prior to reading this thesis is, why these four studies are important and what other tasks are important in the sociopolitical opinion mining research. First, extracting only issues or only sentiments from the comments is insufficient to know the pulse of the citizens/commenters. For example, extracting only sentiment words might let us know the sentiment polarity of the user but not on which issue. Hence, we need to extract the issues as well as the sentiments on those issues. Quality studies help us to filter the noise, which is one of the major challenges in social media, and to extract the useful comments for the domain experts to make decisions. Second, most of the opinion mining studies have focussed on consumer business such as product reviews and very little research has been done in sociopolitical context. Therefore, there is a need to frame the problem and the tasks under this umbrella to direct the research in the future on sociopolitical data. Finally, these four studies are not the exhaustive list of tasks under this research. In the conclusion chapter, we present a framework that integrates these tasks for fine grained sociopolitical opinion mining. We further discuss other related problems that can be integrated into this framework to implement advanced opinion mining applications.

1.5 Contribution

To summarize, the following contributions have been made in this thesis:

- **Issue Extraction:** We first study the problem of issue extraction from sociopolitical comments. We propose an unsupervised approach based on latent variable methods for identifying and extracting the issues in the comments, and linking comments to the issues in the associated article. For example, in response to Obama's State of the Union address, a user might comment on two issues, *healthcare* and *jobs*. Hence it is desired to extract the issues existing in each comment and align them to the issues in the document. Extracting issues is the first task in our thesis objective and we study this task in Chapter 3.

- **Entity-Suggestive Opinion Extraction:** Second, we present a new problem of extracting the entities and associated suggestive opinions from sociopolitical comments. For extracting and normalizing the entities and suggestive opinions from the users, we propose a two-stage approach based on conditional random fields and clustering. Suggestive opinion extraction is the second task of our objective and we present this study in Chapter 4.

- **Thoughtful Comment Extraction:** Valuable comments are useful for decision making and high quality summarization. We then present the task of detecting thoughtful comments in the sociopolitical domain. We propose a supervised approach based on computational linguistics techniques to identify and extract valuable comments in the sociopolitical domain from social media. Our third task of the thesis objective is thoughtful comment detection and we present this work in Chapter 5.

- **Ideological Position and Expression Extraction:** Finally, we study the problem of extracting the stances and ideological expressions of users from their comments on ideological debates related to sociopolitical domain. We propose an unsupervised approach based on latent variable methods for identifying and extracting the positional words/ideological expressions and entities associated with the issues. For this study, we model issues, positional words, entities, and users' stances in a principled way using topic models. For our evaluation, we use arguments in debates as they provide a platform for users to express their positional opinions/sentiments on the issues debated. Our final task of the thesis objective is the study on the debates to extract ideological stances and expressions, and we present this study in Chapter 6.

The studies presented in this thesis were originally reported in Gottipati et al. [51, 52, 50, 49]. The thesis gives a more thorough exposition, relating the problems and solutions to other work, and presents more experimental results and error analysis.

1.6 Road Map

This dissertation document is structured as follows. In Chapter 2, we present a comprehensive literature review of opinion mining research and the research on sociopolitical data. Chapter 3 presents the model, methods and experimental findings for issue extraction task from sociopolitical comments. Chapter 4 describes the problem of extracting entity-suggestive opinions from users' comments, where we present our method and experimental findings. In Chapter 5, we present our method and experimental findings for detecting thoughtful comments in the sociopolitical corpus. Chapter 6, presents the problem of identifying positions and ideological expressions from the users arguments on ideological debates. We discuss the motivation, solution model, data set and evaluation results of this study. Finally, in Chapter 7, we conclude our work and point out future directions to explore. Appendix A gives additional details on the annotation and datasets used for experimentation. Appendix B gives Gibbs sampling equations for implementing the inference for the model in Chapter 6 and additional experiment results.

Chapter 2

Literature Review

This dissertation draws inspiration from various different tasks which fall under the umbrella of opinion mining as well as the gaps of data mining in sociopolitical domain. Here, we provide literature review of both these areas:

1. Opinion mining
2. Sociopolitical data mining

2.1 Opinion Mining

Opinion mining is a well studied research topic for the past ten years mainly focusing on the opinion extraction, sentiment classification, opinion quality, complimentary opinion mining tasks and applications in real world. Opinion mining found its roots in many real-life applications and several application-oriented research studies have been published. Product reviews were exploited by [87] to rank products and merchants. To predict the sales performance, Liu et al. [82] proposed a sentiment model based on pLSA using blog data. Studies on Twitter, where Bollen et al. [22] applied the Profile of Mood States (POMS) to Twitter updates on stock market, Tumasjan et al. [125] used Twitter political sentiment to predict political results and O'Connor et al. [97] used tweets sentiments together with public polls to predict election results are some examples of opinion mining application-oriented research on Twitter corpus.

The major tasks and the complimentary tasks studied in opinion mining research area are depicted in the Figure 2.1. In this section, we study the research progress at both levels. We first provide a survey of research related to major tasks as they are the building blocks for opinion mining as well as our thesis focus. We also survey the research on complimentary tasks which are very close to our thesis focus. For simplicity, other tasks related to opinion mining such as sarcasm, emotions, fake reviews, user interactions, sentiment lexicon generation, opinion trend tracking, opinion search, geo-location based opinions, time-based opinions etc., are not shown in the figure [79]. Readers wishing more details are advised to read books [99, 79] and latest updates on opinion mining research [1].

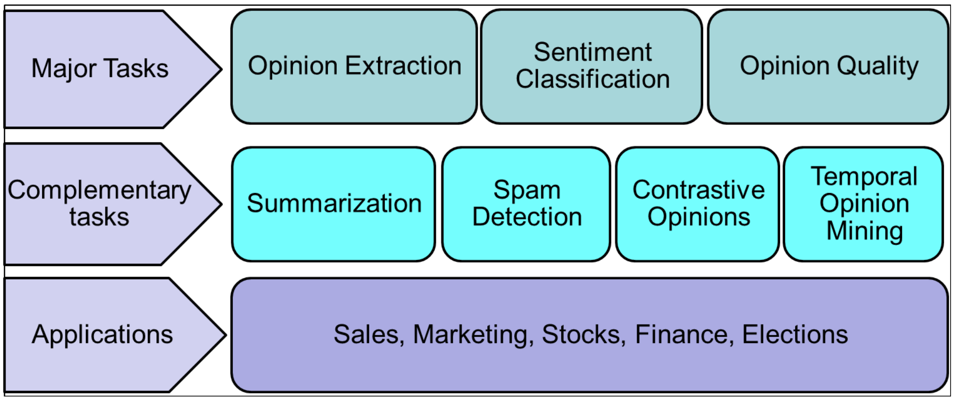

Figure 2.1: General overview of Opinion Mining

2.1.1 Opinion Extraction

Opinion extraction aims at automatically finding attitudes or opinions about specific targets, such as named entities, consumer products or public events [33, 78, 15, 140, 86, 63]. However, according to Hu et al. [60], for many applications opinion extraction is insufficient, and a fine-grained opinion mining and analysis is highly effective. Typical tasks in fine-grained opinion mining include feature identification by Popescu et al. [105], and linking opinions to features by Lin et al. [76].

More sophisticated problems in fine-grained opinion analysis include: opinion

[1] http://www.cs.uic.edu/~liub

holder extraction [69], opinion expression extraction [66, 23] and opinion target extraction [70, 136]. Opinion polarity classification is also an important subtask but will be discussed in the next subsection.

General techniques developed to solve the above tasks include supervised methods such as sequence labeling algorithms [66, 29, 28] and unsupervised methods such as generative LDA-based models [123, 95].

Supervised Methods: Using maximum entropy ranking algorithm Kim et al. [69] achieved 64% of accuracy for identifying the holder of opinions. For expression identification, [23] used linear-chain conditional random field and achieved expression-level performance within 5% of the human interannotator agreement. Using conditional random fields, Johansson et al. [66] discovered that joint model of expression extraction and polarity labeling significantly improves over the sequential approach. Combining multiple techniques several researchers proposed architectures/frameworks for opinion extraction; framework based on semantic role labeling for opinion extraction by exploiting the sematic structure of sentences [70] and framework based on tree kernel by exploiting phrase dependency parsing [136] are some examples.

Unsupervised Methods: Generative topic models have been successfully implemented in opinion extraction tasks such as feature identification [123], entity topic extraction [95], mining contentious expressions and interactions [92] and specific aspect-opinion word extraction from labeled data [141]. Using experiments of perplexity and KL-Divergence Mukherjee et al. [92] showed that the topic models fit the data better and discover more distinctive topics and contention-agreement expressions from the debates. Multi grain topic models [123] are used for extracting the ratable aspects of objects from online user reviews. These models not only extract ratable aspects, but also cluster them into coherent topics. Modeling interrelationship between words and entities in the text using LDA, Newman et al. [95] managed to extract entities and answered questions as who, where and what. Extending topic models and combining with maximum entropy method Zhao et al. [141], discovered aspects and aspect-specific opinion words.

Suggestive Opinion Extraction: Users provide two types of sentiments in their reviews. Either they might provide their sentiments as positive or negative towards the product/entity or they might provide suggestions to improve the product. The first case is handled by sentiment classification study and the second is called suggestive opinion study or actionable knowledge study. Extracting suggestions, also referred to as actionable information [116] from user generated content is of growing interest recently. Actionable knowledge research is relatively new and very few studies attempted to address this problem. Zhang et al. [139] attempted to discover the diagnostic knowledge. Their work is more focussed towards manufacturing applications in which the problems are identified to aid the designers in product design improvements. Simm et al. [116] analyzed actionable knowledge in on-line social media conversation and the concept of actionability is defined as request or suggestion. Ferrario et al. [41] aims at discovering aspects of actionable knowledge in the social media. Their objective is more towards investigating the dynamic aspect of the language the people use to express actionable knowledge. They conducted their study on Twitter and discovered user language aspects using simple heuristics. In our thesis, we focus on opinion extraction task in sociopolitical domain. We study the problem of extracting issues, entities and suggestive opinions from the comments. In particular, our study is close to fine grained opinion analysis focussing on opinion target extraction and opinion expression extraction tasks.

2.1.2 Sentiment Classification

Subjective text classification [133, 135] leads to opinion mining tasks such as sentiment classification. Sentiment classification aims at classifying the data into positive or negative polarities [100] using supervised methods or unsupervised methods. Similar to opinion extraction, fine grained sentiment analysis is desired as it is highly effective to understand the pulse of the consumers at feature level. The task of sentiment target detection [60] aims at extracting the sentiment targets in the reviews using multiple heuristic techniques. Theories of lexical cohesion motivate the representation used by [34] for sentiment polarity classification of financial news.

Supervised models such as Naïve Bayes, maximum entropy, Support Vector Machines (SVMs) etc., were applied by [100, 16]. Unsupervised models such as pattern matching, associated rule mining, lexicon-based methods and topic models have be applied in [80, 126, 88, 76].

Supervised Methods: Pang et al. examined several supervised machine learning methods like SVM and Bayes classification for sentiment classification of movie reviews and showed that classifiers performed poorly on sentences as sentences contains less information [100]. Besplov et al. proposed a method based on supervised latent n-gram analysis [16] and achieved superior performance in comparison to the state of the art methods.

Unsupervised Methods: Using rule mining methods and pattern discovery, Liu et al. developed an application for analysis and visualization of the opinion polarity comparison [80]. Using syntactic pattern based algorithm based on mutual information between document phrases, Turney [126] achieved an accuracy of 66% for movie reviews for classification. LDA-based topic models have been proposed by various researchers for extraction and classification of sentiments. Mei et al. [88] proposed joint sentiment mixture model which models aspects together with positive and negative sentiments learned with some training data. Multigrain topic models use global variable for global topics and local variable for discovering aspects [123] but without any sentiment detection. Jointly modeling aspects and sentiments by Lin et al. [76] doesn't separate aspects and sentiments effectively. Using MaxEnt-LDA, Zhao et al. [141] discovered aspects and aspect-specific opinion words.

In our thesis, we do not work explicitly on sentiment classification problem, but we use the findings of the above studies for extracting ideological expressions. Moreover, motivated by the efforts to exploit prior knowledge [76, 141], we exploited sentiment lexicons for ideological expression extraction from debates.

2.1.3 Opinion Quality

One of the key features of social media is that it enables anyone to freely express opinions from any part of the world. This property enables to capture highly valuable unbiased opinions. However, it comes with a price. First, not all comments are of high quality or useful. Second, it allows users/companies to post fake reviews. This urges a need for detecting valuable comments from the social media. Many recent studies examined the challenges on the quality of comments. Kim et al. [71] studied how to predict the helpfulness of product reviews. They found that a helpful review should describe the features of the products and the pros/cons of the features. A more elaborative review that provides the complete details of the product is more likely to be considered high quality. Another study by Ghose et al. [47] on review helpfulness looked into factors related to the reviewer, such as reviewer characteristics and reviewer history.

Work on measuring quality of social media content considers not only the quality of the content itself but also its authority in the social network through the author's authority, its popularity, etc., [59]. Several researchers explored the social network together with the content of the reviews to predict the review quality. Bian et al. [17] proposed a mutual reinforcement learning framework to simultaneously predict content quality and user reputation. In contrast, Lu et al. [83] proposed a linear regression model with various social contexts for review quality prediction. They combined textual and social context information to evaluate the quality of individual reviewers and to assess the quality of the reviews. Similar line of work can be seen by Chen at al. [27], Liu et al. [82] and Bian et al. [17].

Spam Detection: A complimentary task of opinion mining and close to opinion quality is the spam detection problem. Posting fake opinions has become one of the a major issues in social media and the field of spam detection aims at discovering fake opinions or opinion spammers. The problem can be seen as a classification problem with two classes, spam and non-spam. Duplicate reviews were used by Jindal et al. [65] as spam reviews to detect the fake reviews. Using reviewer behavior pattern to discover the spammers is exploited by Lim et al. [75]. Mukherjee et al. [93] studied the problem of detecting groups of spammers using graph model.

In our thesis, we focus on opinion quality task in sociopolitical domain. We study the problem of extracting thoughtful comments in sociopolitical domain. The main difference is that in traditional sense, high quality text should be grammatical, coherent and readable. For sociopolitical comments, we focus more on the insightfulness or thoughtfulness of comments.

2.1.4 Contrastive Opinions and Stances

Some studies have specifically analyzed contrastive viewpoints or stances in general discussion text. Ganapathibhotla et al. [45] and Paul et al. [101] developed an unsupervised method for summarizing contrastive opinions from customer reviews. Abu-Jabra et al. [2] and Dasigi et al. [32] developed techniques to address the problem of automatically detecting subgroups of people holding similar stances in a discussion thread. Somasundran et al. [117] and Anand et al. [10] were interested in ideological content in debates, relying on discourse structure and leveraging sentiment lexicons to recognize stances. [117] and [10] proposed supervised learning methods for stance classification and tested on debatepedia.org and convinceme.net respectively. In our thesis, we focus on contrastive opinions in sociopolitical domain. The debates on controversial issues exhibit a contrastive opinion behavior, and we aim to discover topics, entities, positions/stances and ideological expressions from the arguments on the debateable issues.

2.1.5 Temporal Opinion Mining

Temporal opinion mining is the process of monitoring and detecting possible changes to specific opinions over a given period of time. The timeline can be presented by the predominant polarity [73] or as a graph based on sentiment value [44]. Fukuhara et al. [44] considered news and blog articles and produced two sets of graphs: a topic graph and an emotion graph. The topic version graphed out topics associated with a certain sentiment. With a specific sentiment, it was possible to see when certain events were highly associated with that sentiment. Das et al. [31] developed a prototype system based on conditional random fields to create visualizations of opinions over time and track changes, focussing on temporal relations between events associated with sentiments. Similar to this task is mood tracking. Mishne et al. [91] developed a system that tracks the mood of blogs hosted by LiveJournal. The system continuously downloads updates from thousands of blogs. The mood tracker follows moods in real time and creates graphs based on these time series. Opinions on sociopolitical issues or entities can change over time and combining the temporal techniques with our work is an interesting study and we leave it for the future research.

2.1.6 Opinion Summarization

Summarization is a study that attempts to generate a concise and digestible summary of a large number of opinions [79]. Current research aims at two types of summarization: aspect-based summarization and non-aspect-based summarization. Aspect-based summarization divides input texts into aspects, which are also called features, and generates summaries of each aspect [60, 61, 88]. The non-aspect-oriented summaries either assume that the opinion text has been pre-segmented by aspects or simply produce a generalized summary without consideration of aspects [84, 46].

Combining the existing opinion summarization techniques and the outputs from our work, one can generate sociopolitical opinion summaries. For example, the output of the entity-suggestive opinion extraction study can be exploited with summarization techniques to generate the summary of comments on an article by the entities related to the article. Another possibility is that, the output of model used for debate study can be used to generate summaries of debates by topics, entities and stances. We leave such sociopolitical opinion summarization as a future research.

2.2 Mining Sociopolitical Data

The Web is an enormous repository of data related to sociopolitical domain in the form of news articles, blogs, editorial articles, forums, political speeches and so on. At the same time, with the dramatic rise of text-based social media, millions of people broadcast their thoughts and opinions on a great variety of topics related to social and political issues. There has been a large body of research that explored various research problems related to sociopolitical data and we categorized the most popular ones in Figure 2.2.

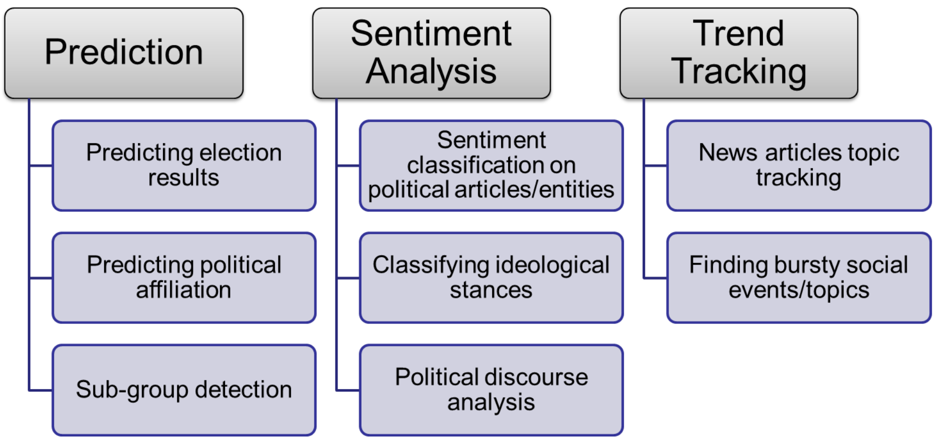

Figure 2.2: General overview of Sociopolitical Data Research

In the sociopolitical domain several studies are dedicated for prediction tasks. Some studies focussed on discovering political affiliations of informal web-based contents like news articles [143], political speeches [30] and web documents [38, 37, 39]. Subramanyan et al. proposed statistical modeling approach for predicting election results [97] using tweets. Closer to these studies is subgroup detection [3, 56, 21, 124]. For group detection tasks, these studies exploit content and corpus specific properties such as hashtags, social networks etc. [3] proposed to profile discussants by their attribute towards other targets and use standard clustering (K-Means) to cluster discussants for sub-group detection. Hassan et al. [56] used correlation clustering to partition the signed network such that positive intra-group links and negative inter-group links are dense for sub group detection. In a separate piece of work which is not part of this thesis, we studied collaborative filtering technique for predicting user's political party affiliation [53].

Users rely on social media for expressing their sentiments on political leaders or controversial sociopolitical issues. Durant et al. [37] proposed Naïve Bayes classifier coupled with a forward feature selection technique to predict a posting's sentiment. Political datasets such as debates and tweets are explored for classifying user stances [129, 118]. Somasundaran et al. [118] took a supervised learning approach by exploiting the discourse structure for stance classification. [77] presented a statistical model for political discourse that incorporates both topics and ideologies; they used debates on the Israeli-Palestinian conflict.

Tracking the topics from posts, comments or tweets aids in understanding the topics/events which are popular over time. Diao et al. [36] studied Twitter for public reactions on major events which might also include social or political events. They proposed solution based on topic models for detecting bursty topics from microblogs to reveal what events have attracted the most online attention. A task of aligning the aspects of the event to the public feedback in Twitter classifies the tweets as episodic or general [62]. Hu et al. [62] proposed a joint statistical model that models topical influences between an event and the tweets around it to detect the tweets and associated the topics. Dyut et al. [114, 113] proposed bag of words approach for reading news articles and comments together by aligning the comments to the segments of the article.

In our thesis, we focus on sociopolitical domain. In particular, we study the task of opinion mining of users' comments on sociopolitical issues focusing on extraction and classification tasks. Though there has been extensive work on opinion mining of product reviews, very little is studied on sociopolitical domain for these problems. Current opinion mining techniques are insufficient as social and political data are much harder to analyse due to the complex expressions [79].

2.3 Aim of the Thesis

Major tasks of opinion mining research are extensively studied for products and reviews. Similarly, the complimentary tasks were also studied mostly on the products. In contrast, we observe that there has been very little research on mining opinions in sociopolitical domain. Mining sociopolitical opinions is a harder problem due to the complexity of the data and hence it is the focus of our thesis. We study the task of sociopolitical opinion mining where we explore users' comments from social media. In our thesis, we focus on opinion extraction, opinion quality and ideological stance classification tasks, which is also very close to contrastive opinion detection problem in the opinion mining umbrella. In particular, we study the problems of mining issues, extracting entities and suggestive opinions towards the entities, detecting thoughtful comments, and extracting ideological stances and expressions from comments in the sociopolitical domain.

2.4 Chapter Summary

In this chapter, we reviewed related works for opinion mining and tasks under its umbrella. We focussed on the studies which are very closely related to the problems we study in this thesis. We also provided survey of research in the sociopolitical domain which is the focus of our thesis. We shall use some of these studies for baseline comparisons and evaluate our models and techniques proposed for the respective research problems.

Chapter 3

Issue Extraction

In this chapter, we study the problem of extracting issues from the public feedback/comments. Analyzing public opinion on social and political issues as well as government policies is of particular importance to policy makers. Given a segmented sociopolitical article with multiple issues and public comments towards the article, the problem aims to extract issues in the comments. We propose an unsupervised approach based on latent variable methods for identifying and extracting the issues in the comments. We evaluate our method quantitatively with a comment linking task. The task aims to link the comments to the relevant issues in the article. We compare our model with state-of-the-art methods and generate promising results. We study the problem on two different data sets. The empirical results on both data sets show that the proposed approach is effective in extracting issues in the comments and linking comments to the issues of the article.

3.1 Introduction

Analyzing public opinion on social and political issues as well as government policies is of particular importance to policy makers. Traditionally, policy makers rely on opinion polls and surveys to capture feedback from the public. However, this kind of feedback channels requires a large amount of manpower support [97]. Also only a small fraction of the entire population can be reached, which may lead to biased information collection. For example, the Singapore government commissioned a telephone survey after the Prime Minister's National Day Rally Speech in 2010 [111]. Online social networks and forums, on the other hand, are now getting increasingly popular and have attracted a large number of users, with many of them expressing and discussing their opinions on government policies through these websites. Collecting and analyzing opinions of these online users can effectively complement, if not entirely replace, telephone surveys. Indeed, many governments have started reaching out to the netizens through e-governance portals[1].

Analyzing public feedback/comments on social issues is a challenging problem and involves many subtasks like sentiment extraction, comment quality detection, opinion target extraction, etc., as shown in Figure 2.1. In this work, we study a particularly interesting problem of mining online comments to political speeches in order to gain insight into public feedback by extracting the issues and linking the feedback to the speech. For the problem we study, we use a political speech and set of comments from users on this speech as an input to our model. Examples of such political speeches include Obama's State of the Union Speech and Singapore Prime Minister's National Day Rally Speech. There are several reasons that this task is important for policy makers. First of all, these speeches usually touch on a wide range of social issues such as economy, education and immigration, with a focus on the most sensitive issues that the public is concerned with. Mining comments of such a speech helps quickly zoom into the most important social and political problems in a society. Second, a public speech as an actual event, generally attracts a sudden surge of interest and comments online. It is therefore relatively easy to gather a large amount of comments within a short period of time. Third, since most of the relevant comments mention the title of the speech, it is also relatively easy to gather relevant comments through keyword search.

However, this task still poses a number of challenges. In this chapter, we par-

[1] www.reach.gov.sg

ticularly focus on the following two challenges. First, *Issue extraction*. Second, *Comment linking*.

Issue extraction from comments: Users write comments in response to the issues in the document/speech. We would like to extract the issues in the comments that are inline with the issues to the speech. A major challenge we face in this case is that the feedback from users consists of not only the issues but also other concerns related to the issues. For example, in response to the issue, *Economy*, the user might post his concerns on *job security*. Hence, using only the comments might generate noisy/incoherent issues. Leveraging the speech is therefore essential to extract the issues.

Comment linking: Since a speech typically has several issues and a user's comment usually only touches on a single issue, it is desired to link the user comments to specific issues. In cases where a comment combines two or more issues, we would also like to allow this kind of "one-to-many" linking. Linking comments to the issues aids in summarization tasks. A major challenge we face in this task is that in comments users tend to use more colloquial terms and abbreviations, which makes text matching hard. For example, in Singapore, instead of using the official term "foreign talent" to refer to highly-skilled immigrants, online users often use the initial "ft".

Topic models are widely used techniques to generate topics from the large data corpus into human interpretable topics. Supervised approaches need annotated data and in this case every speech may need different training data which can be very expensive. Moreover, issues in sociopolitical domain can be classified as "ABORTION_related" or "ECONOMY_related" etc., with the probability of generating various coherent words and thus are similar to topics in topic model algorithms. Hence for our solution we choose topics models for issue extraction problem. In our case, the topics are issues and we propose a solution based on topic models to extract issues from comments. To align the comments to the various issues of the speech, we score the relevance of a comment with respect to each issue. Here we propose to use a semi-supervised feedback topic model which is based on the widely used Latent Dirichlet Allocation (LDA) models [20]. This semi-supervised approach allows us to automatically pick up related terms in the comments and therefore alleviate the vocabulary mismatch problem.

We evaluated our method on two speeches, one by the U.S. President Barack Obama and the other by Singapore's Prime Minister Lee Hsien Loong. For issues extraction task, our model generated more interpretable issues(topics) compared to standard LDA. For comment linking task, our model performed well for precision@K, $K \in \{5, 10, 15, 20\}$ when compared to LDA and TF-IDF based bag-of-word methods.

The major contributions of our work can be summarized as follows:

1. We proposed a novel opinion mining problem: extracting issues from social comments and linking comments/feedback to social issues mentioned in a public speech.

2. We proposed a method based on topic models to extract issues and link comments to the various segments of a speech. The proposed JSC-LDA model, jointly models speech and comments is an extension of standard LDA model. In our proposed model, we supervised the topics using the segments of the speech, and this approach aided in extracting coherent topics from the comments and achieving high precision for comment linking task.

3. We showed in our experiments on two datasets, that our method could outperform two baselines for our problem.

The rest of the chapter is organized as follows. In Section 3.2, we formally define our problem and give an overview of our solution. We present the solution details in Section 3.3 where we first describe the standard LDA model, and then we describe our model which is an extension of LDA. The data sets and annotation details are described in Section 3.4. Experiments and results are presented in Section 3.5. We present our discussions in Section 3.6 and finally we conclude with some related work analysis in Section 3.7.

3.2 Problem Definition and Solution Overview

Generally speaking, the goal of our task is to extract issues from the comments and link comments to the issues mentioned in a speech. For example, if a speech touches on education and war, which are two completely different issues, we would like to extract issues and the feedback related to each issue separately. The first task is similar to extracting coherent topics from documents. The second task is similar to finding relevant documents given a query.

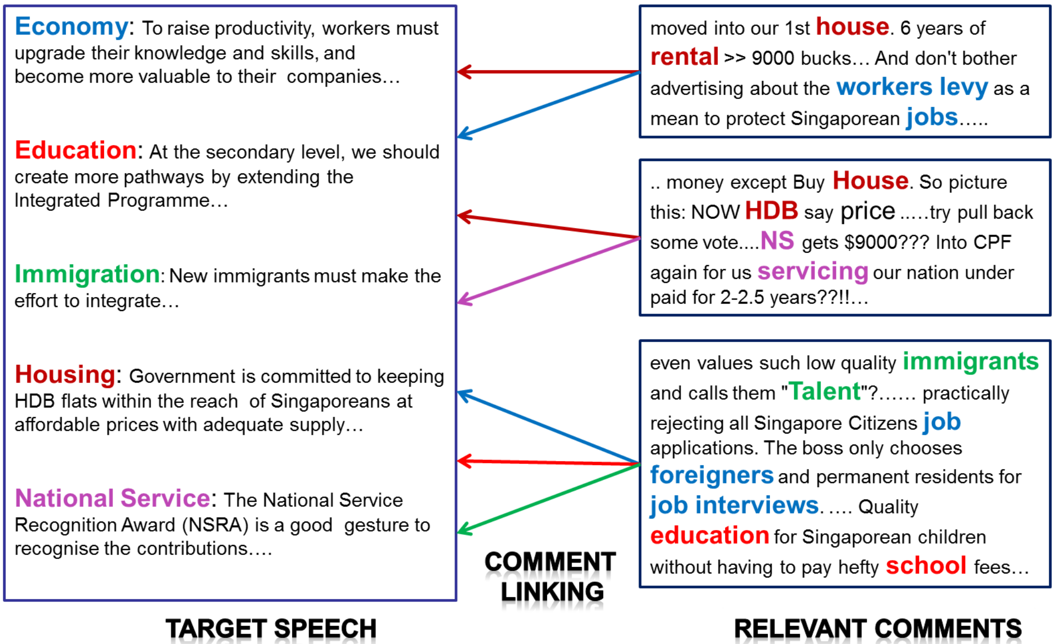

Figure 3.1: An example speech-comment framework to illustrate the definitions. The bolded text in the speech refers to the issues. The bolded text in the comments indicate the words that are semantically coherent to the issue in the speech. The arrows between the speech and comments indicate the linking task (relevant comments). The colors demonstrate the existence of multiple topics in speech, comments and links

.

To formally define our problem, we first introduce a few basic concepts. Figure 3.1 shows an example speech-comment framework (PM's speech from our dataset) to illustrate these concepts visually.

Definition 1 (TARGET SPEECH)**.** *We refer to the political speech about which the comments are analyzed as the target speech.*

Definition 2 (ISSUE)**.** *We assume that the target speech has been pre-segmented and each segment is about a single sociopolitical issue such as war, education, immigration etc., which indicates that the number of issues in the speech is given as an input.*

Definition 3 (RELEVANT COMMENT)**.** *A comment is said to be relevant with respect to a given issue if the comment is to mainly address the issue. At times a comment may address more than one issue and therefore can be a relevant comment with respect to multiple issues.*

Definition 4 (COMMENT LINKING)**.** *Our problem of linking comments is defined as follows: Given a target speech A, which contains T issues, and a set of comments on the speech, $D = \{d_1, d_2, \ldots, d_n\}$, our goal is to to derive a ranking of D with respect to each issue t based on relevance. We treat this task as a comment ranking problem.*

Given an issue t, to rank the comments in D, we use a scoring function. $R_s(t, d_i)$, a score that measures the relevance of document, d_i to issue t.

To define the relevance function, there are several state-of-the-art techniques like TF-IDF similarity function or the Jaccard similarity function. In our work, we use statistical topic modeling approach because it has been shown to be effective for a number of tasks.

3.3 Solution Details

For our solution, we adopt topic modeling techniques to extract issues and score the comments based on their relevance. We use a semi-supervised approach where we seed the topics with words from the speech and at the same time fix the issues for linking.

In the following sub-sections, we first describe the standard topic model, latent Dirichlet allocation (LDA) and then our solution model in detail.

3.3.1 Latent Dirichlet Allocation

LDA-style topic models [20] have been successfully applied to various text mining tasks including summarization, text classification, etc. Latent Dirichlet Allocation is a generative probabilistic model which can be applied to a corpus of text documents in the form of bag-of-words. The graphical representation of LDA is shown in Figure 3.2 The model assumes that we have a collection of D documents. Each document has N_d words, document-topic distribution, θ_d and hidden topics, z. Let us use $z_{d,n}$ to denote latent topic label for nth-word in d-th document, $w_{d,n}$ denote the n-th word in d-th document. The full generative process is as follows:

1. For each topic $t = 1, \ldots, T$, sample ϕ_t from Dirichlet(β)
2. For each document $d = 1, \ldots, D$,
   - (a) Choose θ_d from Dirichlet(α)
   - (b) For each word $w_{d,n}$ in document d
     - i. Choose a topic $z_{d,n}$ from Multinomial(θ_d)
     - ii. Choose a word $w_{d,n}$ from Multinomial($\phi_{z_{d,n}}$)

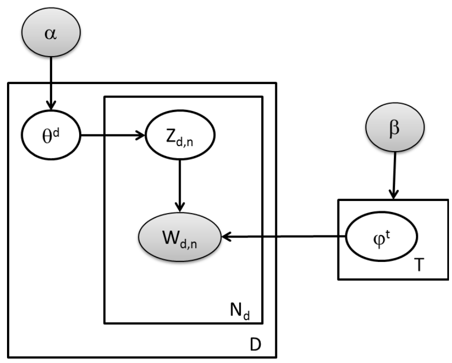

Figure 3.2: Graphical Model Representation of LDA Model

Inference and Parameter Estimation: For parameter estimation, we use Gibbs sampling [55] which is an approximate inference algorithm and a Markov Chain Monte Carlo algorithm. Gibbs sampler is used to produce a sample from a joint distribution when only conditional distributions of each variable can be efficiently computed. In Gibbs sampling, variables are sequentially sampled from their distributions conditioned on all other variables in the model. Such a chain of model states converges to a sample from the joint distribution. For LDA, using Gibbs sampler, new values for $z_{d,n}$ are iteratively sampled for each word $w_{d,n}$ from the posterior probability conditioned on the previous state of the model (i.e., the current values for all other tokens). The Gibbs sampling equation for new variable assignment is as follows:

$$\begin{aligned} & p(z_{d,n} = t | \boldsymbol{z}_{\neg d,n}, \boldsymbol{w}; \vec{\alpha}, \vec{\beta}) \\ \propto \; & \frac{p(z_{d,n} = t, \boldsymbol{z}_{\neg d,n} | \vec{\alpha})}{p(\boldsymbol{z}_{\neg d,n} | \vec{\alpha})} \times \frac{p(\boldsymbol{w} | z_{d,n} = t, \boldsymbol{z}_{\neg d,n}; \vec{\beta})}{p(\boldsymbol{w}_{\neg d,n} | \boldsymbol{z}_{\neg d,n}; \vec{\beta})} \\ = \; & \frac{\alpha + n(t, \boldsymbol{z}_{\neg d,n})}{T\alpha + n(\boldsymbol{z}_{\neg d,n})} \times \frac{\beta + n(t, \boldsymbol{w}_{\neg d,n}, \boldsymbol{z}_{\neg d,n})}{V\beta + n(t, \boldsymbol{z}_{\neg d,n})} \end{aligned} \tag{3.1}$$

where $n(t, \boldsymbol{z}_{\neg d,n})$ is the number of times topic t is assigned to the document, d and without considering the current word and $n(t, \boldsymbol{w}_{\neg d,n}, \boldsymbol{z}_{\neg d,n})$ is the number of times current word is assigned to topic t as observed from $\boldsymbol{w}_{\neg d,n}$ and $\boldsymbol{z}_{\neg d,n}$.

The topic-document distribution, θ_d for a document is given by the following equation:

$$\theta_{d,t} = \frac{\alpha + n(t, \boldsymbol{z}_{d,n})}{T\alpha + n(\boldsymbol{z}_{d,n})} \tag{3.2}$$

where $n(t, \boldsymbol{z}_{d,n})$ is the number of times topic t is assigned to the document, d and $n(\boldsymbol{z}_{d,n})$ is total number of topics assigned.

3.3.2 Our Model

To extract the issues and to determine the relevance of the comment to a given issue in a speech, we propose an extended LDA model that jointly models the target speech and the user comments. We call this model Joint Speech Comment-

Latent Dirichlet Allocation (JSC-LDA)[2]. The graphical representation of JSC-LDA is shown in Figure 3.3. JSC-LDA assumes that each of the segments of the target speech corresponds to a unique topic, which is also an issue. From our input, we have a speech article which is divided into sections based on the issues that the speaker is addressing. We also observed that for better flow of the speech, the topics discussed are confined to one segment each. The given input and this observation drove us to define single topic to each segment. In other words, each issue in the target speech corresponds to a topic. Let us assume that there are S issues in the speech. We therefore have also S topics in our model. Hence, we have $T = S$ number of topics with N_s words in each segment. Let us use z_s to denote the index for the topic of segment s and $w_{s,n}$ to denote the n-th word in segment s.

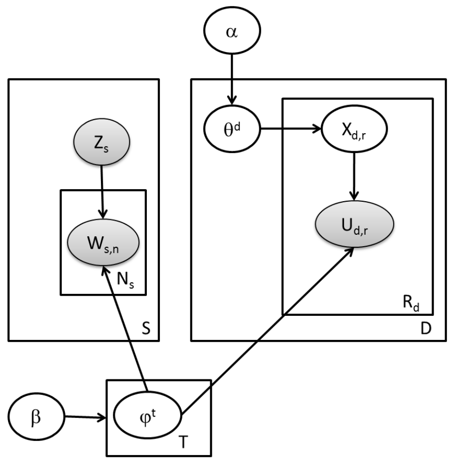

Figure 3.3: Graphical Model Representation of our Model, JSC-LDA.

Furthermore, the model assumes that we have a collection of D comments. Each comment has R_d words, a topic distribution denoted as θ_d and hidden topics. Let us use $x_{d,r}$ to denote the topic of the r-th word in document d and $u_{d,r}$ to denote the r-th word in document d. Finally, one draws a word from a distribution over words associated with topic $t \in T$. The formal definition of the generative process is as follows:

1. For each topic $t = 1, \dots, T$, sample ϕ_t from Dirichlet(β)

[2]In the original paper [51], this model is referred to as SLDA.

2. For each comment $d = 1, \ldots, D$,

    (a) Choose a distribution $\theta_d \sim$ Dirichlet(α)

    (b) For each word $u_{d,r}$ in comment d

        i. Choose a topic $x_{d,r} \sim$ Multinomial(θ_d)

        ii. Choose a word $u_{d,r} \sim$ Multinomial($\phi_{x_{d,r}}$)

Inference and Parameter Estimation:

Exact inference of the posterior distribution of the hidden variables is intractable. We will instead approximate this using Gibbs sampling. For JSC-LDA, using Gibbs sampler, new values for $x_{d,r}$ are iteratively sampled for each token $u_{d,r}$ from the posterior probability conditioned on the previous state of the model (i.e., the current values for all other tokens).

The Gibbs sampling equation for new variable assignment is as follows:

$$\begin{aligned} & p(x_{d,r} = t |, \boldsymbol{x}_{\neg d,r}, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{u}; \vec{\alpha}, \vec{\beta}) \\ \propto \quad & \frac{p(x_{d,r} = t, \boldsymbol{x}_{\neg d,r} | \vec{\alpha})}{p(\boldsymbol{x}_{\neg d,r} | \vec{\alpha})} \times \frac{p(\boldsymbol{u}, \boldsymbol{w} | x_{d,r} = t, \boldsymbol{x}_{\neg d,r}, \boldsymbol{z}; \vec{\beta})}{p(\boldsymbol{u}_{\neg d,r}, \boldsymbol{w} | \boldsymbol{x}_{\neg d,r}, \boldsymbol{z}; \vec{\beta})} \\ = \quad & \frac{\alpha + n(t, \boldsymbol{x}_{\neg d,r})}{T\alpha + n(\boldsymbol{x}_{\neg d,r})} \times \frac{\beta + n(t, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{u}_{\neg d,r}, \boldsymbol{x}_{\neg d,r})}{V\beta + n(t, \boldsymbol{z}, \boldsymbol{x}_{\neg d,r})} \end{aligned} \tag{3.3}$$

$n(t, \boldsymbol{w}, \boldsymbol{z}, \boldsymbol{u}_{\neg d,r}, \boldsymbol{x}_{\neg d,r})$ is the number of times current word is assigned to topic t as observed from $\boldsymbol{w}, \boldsymbol{z}, \boldsymbol{u}_{\neg d,r}, \boldsymbol{x}_{\neg d,r}$. $n(t, \boldsymbol{z}, \boldsymbol{x}_{\neg d,r})$ is the number of times we see topic t in $\boldsymbol{z}$ and $\boldsymbol{x}_{\neg d,r}$. $n(t, \boldsymbol{x}_{\neg d,r})$ is the number of times t as observed from $\boldsymbol{x}_{\neg d,r}$. $n(\boldsymbol{x}_{\neg d,r})$ is the total number of topics assigned excluding the current topic assignment.

To estimate topic-document distributions θ_d for a document, we use the following equation:

$$\theta_{d,t} = \frac{\alpha + n(t, \boldsymbol{x}_{d,r})}{T\alpha + n(\boldsymbol{x}_{d,r})} \tag{3.4}$$

3.3.3 Relevance Score

Intuitively, the relevance function of the comment to a given issue is given by the probability of topic t given comment d, where topic t corresponds to the given issue.

$$R_s(t,d) = p(t|d) = \theta_{d,t} \tag{3.5}$$

The above equation is used to rank the comments for the comment linking task.

3.4 Dataset

We first acquired the following two political speeches:

1. Singapore Prime Minister's National Day Rally Speech in 2010. (*PM's speech*). We use PM to represent Prime Minister.

2. US President's State of the Union address in 2011. (*Obama's speech*)

We further manually broke down each speech into several segments based on topical boundaries (Obama's speech) or sub-headings (PM's speech). We observed that in many cases the sub-headings are embedded in political speeches and when such data is unavailable, the speeches are well organized with clear boundaries to aid the segmentation. The segments/issues discussed in each speech are listed in Table 3.1. In case of PM's speech, the labels were available in the speech and for Obama's speech, the issues were manually labeled to aid the reader.

Speech	**Issues**
PM's Speech	Immigration, National Service, Housing, Economy, Congestion, Education, Founding Fathers, Singapore Spirit, Youth Olympics
Obama's Speech	Economy, Innovation, Education, Rebuild, Taxes, Debts, Military

Table 3.1: Issues discussed in both data sets.

3.4.1 Data Processing

To collect an unbiased sample of comments for each speech, we use two search queries ("national day rally speech 2010" and "president state union address 2011") and Google API to obtain a list of URLs. We further manually selected those URLs from online forums and blogs. After cleaning the data and removing short comments (with no more than two words), we finally obtained 550 comments for the first speech and 800 comments for the second speech. We randomly choose 150 comments for each data set (1350 pairs of issue-comments for PM's Speech and 1150 pairs of issue-comments for Obama's speech) to conduct our quantitative study for the comment linking task. For PM's speech, the average number of words across all comments is around 139 and for Obama's speech, the average number of words across all comments is around 66, shorter than comments to PM's speech.

3.4.2 Annotation

We engaged two human annotators for each speech to judge the comments we had collected. Our annotators for PM's speech are Singaporeans who are familiar with the economy and the social issues of Singapore. At the same time, they are also familiar with the local language aspects (e.g."ft" for foreign talent). For Obama's speech, we have two annotators, an American and an immigrant, who are familiar with American economy and social issues. All the judges are above 25 and are working professionals. To judge the relevance of a comment to an issue, each judge should refer to the segments of the speech and look for the following:

(a) Is the comment relevant to any issue of the speech? For example, consider the following comment: "Presidents message missed very subtle points like closing loop holes so that a more fair Corporate Tax can be created where all Business is taxed properly..You need to stop Fighting Last Years War. The Law Passed will not be repealed that War ended like Iraq is ending". The first sentence shows that the comment is related to the issue, *Taxes*.

(b) Is the comment relevant to more than one issue of the speech? In the above example, the comment is also related to another issue, *Military*.

We calculated the inter-annotator agreement level using Cohen's kappa. On relevance judgement the kappa is 0.717 and 0.754 for PM's speech and Obama's

Issue	**Top Words**
Economy	*ft*, government, good, jobs, job, *money*, time, *pay*, working, *bad*, oil, country, workers, employers, *simple*
Immigration	foreign, workers, jobs, citizens, foreigners, chinese, talent, economy, immigrants, understand, world, *local, foreigner*
National Service	foreigners, *salary, country, govt, people*, ns, *nsmen, vote, election*, pay, *policies, send, lower, family, private, sporeans*, service
Congestion	time, job, *change*, hours, line, *work*, problem, *place, talented*, trains, people, working, *foreigners*, coming, run, *bad*
Housing	hdb, flats, *live*, foreigners, people, local, housing, property, long, *high, poor*, time, work, *clear, afford, fw, stop*, population
Education	good, students, school, schools, education, programme, poly, universities, work, government, academic, normal, university, *overseas*
Singapore Spirit	*fts*, work, people, *happy*, things, *citizens, boss, kind*, proper, *lives, project*, spirit, fellow, *ge, case, teach, pledge, action*
Founding Fathers	goh, *policy*, education, dr, future, generation, *party*, time, world, building, *voice*, people, *issue*, young, *hard, quality*
Youth Olympics	*pap, talents, living, life*, people, world, *trash, lies, food*, start, point, times, *stress, towkays, maids, businessmen*

Table 3.2: Top words from PM's speech using JSC-LDA model.

speech respectively. We use the annotations from the stricter judge for our experiments. Interested readers kindly refer to Appendix A for more details on annotation process and sample examples. To encourage comparative work, we have made the annotation available for download[3].

3.5 Experiments

The main objective of this work is to extract issues from the comments. We first present a qualitative discussion on the issue extraction results. For the comment linking task, recall that in Section 3.2, we defined this problem as to rank all the comments with respect to a given issue in the target speech in terms of relevance. To evaluate our proposed approach, we need to quantify the performance of the ranking results compared with the human annotations. For our proposed model,

[3]https://sites.google.com/site/swapnagotipati/datasets

Issues	**Top Words**
Education	good, students, school, education, schools, programme, universities, work, poly, government
Singapore Spirit	work, time, young, things, people, place, feel, spirit, happy, lives
Housing	hdb, flats, people, long, home, private, housing, issue, live, property
Founding Fathers	country, salary, goh, generation, quality, dr, education, families, service, times
Immigration	foreign, workers, talent, world, economy, immigrants, chinese, society, countries, important
National Service	people, ns, pap, change, election, vote, govt, send, talented, cpf
Economy	foreigners, jobs, citizens, job, pay, foreign, locals, back, residents, policy
Congestion	government, time, problem, coming, grow, means, fact, working, job, run
Youth Olympics	ft, living, talents, fts, life, local, high, family, bad, lower

Table 3.3: Top words from PM's speech using LDA model. Issues are manually labeled. The last two rows shows the conflated topics.

we run the Gibbs sampler for 500 iterations using standard settings α = 0.1, β = 0.01 [20].

3.5.1 Issue Extraction Task

In this section we present a qualitative discussion of model's performance in issue extraction for both data sets. In topical influence study, we would like to see how the model aids in discovering the impact of speech on the public. We compare the feedback behavior of public on each of the topics associated with the speech using JSC-LDA and LDA models. First, we explain the details of experimental setup for LDA. We begin the results and analysis with PM's speech followed by Obama's data set.

LDA: For this baseline, we use standard LDA discussed in Section 3.3. We learnt the hidden parameters in the model using Gibbs sampling [55]. We run the sampler for 500 iterations using standard settings α = 0.1, β = 0.01 and the number of topics

T is set to the number of issues in the speech. For preliminary experiments, we observed that using comments alone in the corpus produces topics with low quality. Hence we combined the speech segments to the corpus and this helped to generate coherent (top words are coherently related to each other representing a semantic concept/topic together) topics. The topics of LDA model are manually labeled by observing the top words in each topic. We observed that the conflated (top words are not coherently related to each other and together represents more than one semantic concept/topic) topics made it difficult to label some topics and we will discuss it in detail when presenting the results.

PM's Speech Analysis using JSC-LDA Model: The topics from LDA model for PM's speech are shown in Table 3.3. Table 3.3 presents the top probability words for each topic. The issues are manually labeled by observing the top words and look for a semantic concept formed by them. We observe that the last two topics are very difficult for a human to label as these topics, congestion, economy, and youth olympics are conflated.

The topical results of JSC-LDA model for PM's speech are shown in Table 3.2. Table 3.2 presents the highest probability words for PM's speech. The issues are labeled using the model through the speech segments. The words in black are the speech words and those words highlighted in blue are the feedback words from the comments, i.e. words that are not observed from the speech. We observe that modeling the speech together with the comments is useful for forming high quality topics where top words indicate the topic concept.

In PM's speech, for example, given the issue *Economy*, the feedback word *ft* represents foreign talent and the corresponding user's comment is "The company I am working now has 70% FT & FW where are all our best people?...". This shows that given the issue *Economy*, foreign talent is one of the users' top concerns. These results support the hypothesis that the observed topics of speech in the model aid in generating coherent topics as well as more interpretable speech and feedback representations.

Further, for all the issues, the top words are coherent to the issue in the segment except for *Singapore Spirit* and *Youth Olympics*. We observed that this is due to very low public response to these topics. In our data set we have 13.33% and 10.67% comments related to these topic respectively. Table 3.4 shows the distribution of the comments breakdown by the issue as predicted by JSC-LDA. The table is ranked by issue influence (number of comments) on the public.

Topical influence: In this study, we would like to see how the model aids in discovering the impact of speech on the public. This can be answered by studying the feedback behavior of public on each of the topics associated with the speech. For PM's speech, Table 3.4 shows that the issues *Immigration*, *National Service*, *Housing* and *Economy* have strong influence on the public. To study this behavior empirically, we used human judgement, i.e. the relevance annotations as explained previously, to validate this observation. Figure 3.4 presents the comparison of human judgement to the JSC-LDA and LDA prediction for PM's speech.

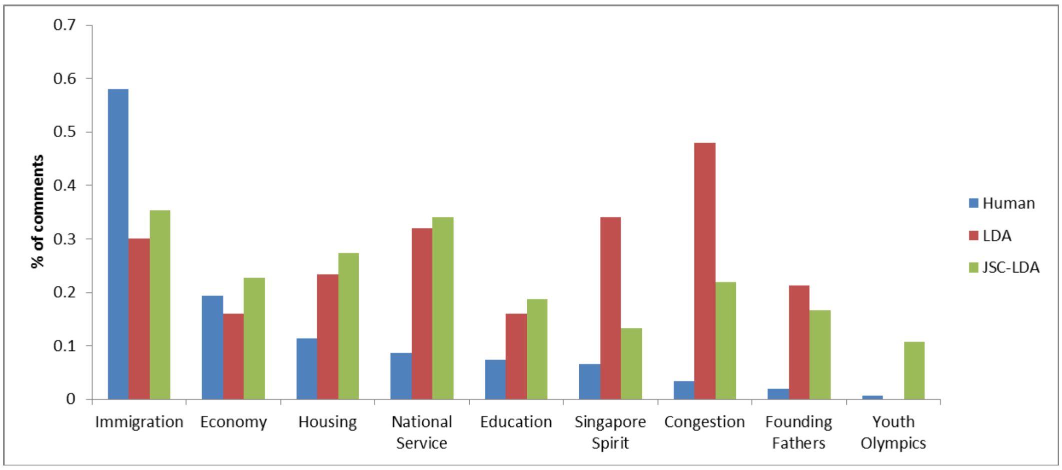

Figure 3.4: PM's speech: Ranking of the issues by influence on public (Human Vs LDA Vs JSC-LDA)

From Figure 3.4, we observe that human judgement is similar to the JSC-LDA prediction in ranking the issues by public influence. We used Spearman's correlation coefficient to measure the rankings generated by our model and LDA as compared to human ranking. For LDA, ρ=0.08 and for JSC-LDA, ρ=0.85. LDA shows congestion as one of the top issues and in our analysis we found that youth olympics, economy and congestion are conflated. Human judgment shows that *Immigration*, *Economy* and *Housing* are among the top ranked topics. According to the survey conducted by the Singapore government [111], the public is more concerned by the *Immigration* and *Housing* and conforms to our results. These results indicate that JSC-LDA model captures the topical distribution of the comments with promising results.

Issue	**# Comments**
Immigration	35.33%
National Service	34.00%
Housing	27.33%
Economy	22.67%
Congestion	22.00%
Education	18.67%
Founding Fathers	16.67%
Singapore Spirit	13.33%
Youth Olympics	10.67%

Table 3.4: PM's speech: Comments statistics on public response breakdown by issue (JSC-LDA).

Obama's Speech Analysis using JSC-LDA Model: The topics from LDA model for Obama's speech are shown in Table 3.7. Table 3.7 presents the top probability words for each topic. The issues are manually labeled by observing the top words and look for a semantic concept formed by them. We observe that the last three topics are very difficult for a human to label as these topics; debts, taxes and rebuild are conflated.

The results of JSC-LDA model for Obama's speech are shown in Table 3.6. Table 3.6 presents the highest probability words for Obama's speech. The issues are labeled using the model through the speech segments. The words in black are the speech words and those words highlighted in blue are the feedback words from the comments. We have the similar observations as PM's speech that modeling the speech together with the comments is useful for forming high quality topics where top words indicate the topic concept.

For Obama's speech, given the issue *Taxes*, one of the feedback words is *food* and the corresponding user's comment is "The only difference this time is the poor people will live shorter lives have no access to health care, wont be able to buy healthy food will have no access to education. The middle class pays the bulk of social security, the wealthy reach the limit in the first month of every year and don't have to contribute any more. Why don't you eliminate the the limit and create a more fair tax law." This shows that given the issue *Taxes*, *food for poor* is one of the user's concerns. Table 3.5 shows the distribution of the comments breakdown by the issue as predicted by JSC-LDA. The table is ranked by issue influence on the public.

Topical influence: For Obama's corpus, we would like to see how the model aids in discovering the impact of speech on the public. Similar to previous evaluation, we study the feedback behavior of public on each of the topics associated with Obama's speech. For Obama's speech, Table 3.5 shows that the issues *Taxes*, *Economy* and *Debts* have high influence on the public and to study this behavior empirically, we further used human judgement to validate this observation.

Figure 3.5 presents the comparison of human judgement to the JSC-LDA model prediction for Obama's Speech. We observed that human judgement is similar to the JSC-LDA and LDA prediction in ranking the issues by public influence. We used Spearman's correlation coefficient to measure the rankings generated by our model and LDA as compared to human ranking. For LDA, ρ=0.71 and for JSC-LDA, ρ=0.86. For the topic *Military*, LDA predicted higher influence and in our analysis we found that the issues such as *taxes*, *debts* and *rebuild* were incoherent.

Summary: Top terms for each issue generated by our model not only shows the issues but also feedback terms. Such representation of data can aid the system users for better visualization of issues and the corresponding feedback/sentiments from the users as shown in Table 3.2 and Table 3.6.

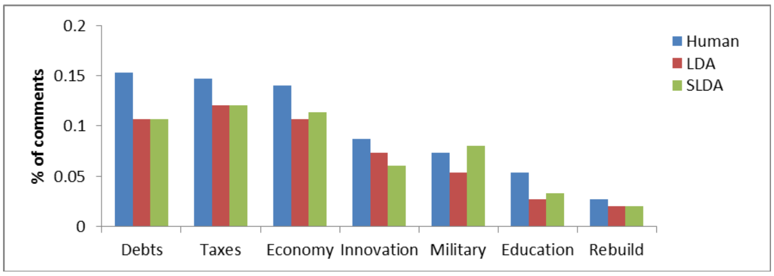

Figure 3.5: Obama speech: Ranking of the issues by influence on public (Human Vs LDA Vs JSC-LDA)

Issue	# Comments
Taxes	12.00%
Economy	11.33%
Debts	10.67%
Military	8.00%
Innovation	6.00%
Education	3.33%
Rebuild	2.00%

Table 3.5: Obama's speech: Comments statistics on public response breakdown by issue (JSC-LDA).

3.5.2 Experiments on Comment Linking Task

Recall that our second challenge in analysing the feedback/comments is comment linking. We now evaluate our model on comment linking. Intuitively, the comment linking task deduces to the ranking problem where for a given topic/segment in the speech, we rank the comments by the ranking score to evaluate how well our model links the comments to segments of the speech. To evaluate the performance of comment linking task, we use the ground truth from our annotation task and the ranked list from Equation (3.5) to evaluate the results. We use precision@K as a metric for evaluation where we set K to 5, 10, 15 and 20.

We now study the ranking performance of JSC-LDA on both data sets against two baselines. We compare the ranking performance of JSC-DLA and baselines in linking comment to the issues in the speech. First, we explain the details of experimental setup for baselines and JSC-LDA. We then present the qualitative evaluation of all three models.

Issues	**Top Words**
Economy	jobs, *private, real*, change, technology, future, world, *china, big*, research, idea, innovation, good, work, *pretty*, changed, *security*
Innovation	energy, *ryan*, oil, *voted, money*, science, *stop*, work, *government*, clean, *stimulus, problems, lot, million, communities, paul, head, gop*
Education	education, people, *spend*, schools, *good, family, math*, school, college, top, *achievement, man, born, responsible, politics*, race, high
Rebuild	*job*, good, people, *hope, solutions, sound, simple*, time, unemployment, *buy, political, economy, investment, student, goods*, highspeed
Taxes	jobs, law, *middle*, business, back, democrats, *food*, create, *care*, health, *taxes, rich*, companies, *move, act*, support, small, republicans, corporate
Debts	spending, tax, cut, deficit, cuts, *republicans*, government, care, *defense*, dollars, cutting, *medicare*, means, *billion*, health, federal, *trillion*
Military	people, *great*, united, *start, open*, troops, afghan, country, *free, democrats, care, past, vote*, iraq, war, tonight, part, government, love, *states*

Table 3.6: Top words from Obama's speech using JSC-LDA model.

We explain the two baselines as follows:

LDA: For this baseline, we use standard LDA discussed in Section 3.3 to compute the relevance score of each comment to each issue of the speech. We learnt the hidden parameters in the model using Gibbs sampling [55]. We run the sampler for 500 iterations using standard settings $\alpha = 0.1$, $\beta = 0.01$ and the number of topics T is set to the number of issues in the speech. For preliminary experiments, we observed that using comments alone in the corpus produces topics with low quality. Hence we combined the speech segments to the corpus and this helped to generate coherent topics. After Gibbs sampling, we can estimate the topic distribution for a comment d as $\theta_{d,t}$. The relevance function of the comment to a given issue is given by Equation (3.2). The topics of LDA model are manually labeled by observing the top words in each topic. We observed that the conflated topics made it difficult to label some topics and we will discuss it in detail in the results analysis discussions.

BOW: The second baseline uses a bag-of-word representation with TF-IDF weight-

Issues	Top Words
Economy	jobs, world, work, future, technology, tonight, energy, research, companies, ago
Innovation	work, people, business, innovation, money, big, economy, dream, working, plan
Education	education, country, high, school, spend, stop, race, science, child, back
Military	people, time, good, states, war, private, men, man, troops, afghan
Debts	spending, tax, deficit, government, care, cut, congress, health, cuts, defense
Taxes	president, union, party, ryan, job, taxes, state, voted, thing, middle
Rebuild	republicans, law, job, class, back, words, programs, presidents, stimulus, politics

Table 3.7: Top words from Obama's speech using LDA model. Issues are manually labeled. The last three rows shows the conflated topics.

ing for the speech segments and the comments, and measures the relevance of a comment (relevance score) to a segment using cosine similarity. We refer to this second baseline as BOW.

JSC-LDA: For our proposed model, we run the Gibbs sampler for 500 iterations using standard settings α = 0.1, β = 0.01. We calculated relevance score $R_s(t, d_i)$ using Equation (3.5).

PM's Speech: Average precision@K across all issues is calculated and the results are shown in Table 3.8 for PM's speech.

Model	p@5	p@10	p@15	p@20
BOW	0.1333	0.1556	0.1400	0.1222
LDA	0.2222	0.2222	0.2071	0.1889
JSC-LDA	**0.3444**	**0.3183**	**0.2886**	**0.2889**

Table 3.8: PM's speech: Average precision@K across all issues.

We observe that topic models perform significantly better than BOW (under a Wilcoxon signed rank test $p < 0.05$). This can be anticipated as topics models tend to categorize similar words in the latent space and handles vocabulary issues. JSC-LDA has higher precision@K for all K in $\{5, 10, 15, 20\}$ compared to both the models- bolded figures in the Table 3.8. JSC-LDA has better performance than LDA due to the supervision of the issues (under a Wilcoxon signed rank test $p < 0.05$). Further, the model generates high relevance score for exact match of the words in the speech topic to comment. LDA model performed better than BOW model as topic models are capable of capturing the linguistic notions such as synonymy and polysemy [20]. Further, Figure 3.6 shows the performance of each model breakdown by the issues from PM's speech.

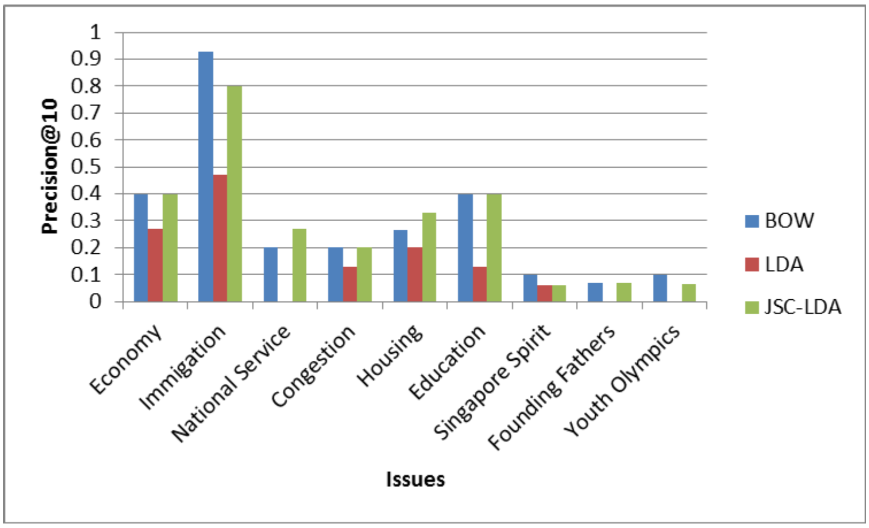

Figure 3.6: PM's speech: Precision@10 break down at issues level.

From Figure 3.6, the precision@K for JSC-LDA model outperformed the other models for most of the issues in the speech. For some of the topics, there are very few comments and all the models have given very low performance. For example, on *Singapore Spirit*, the comments were very few, 13% and all models have p@10 less than 0.1. The issue *Immigration* has the highest performance for all models. This also indicates that *Immigration* is one of the top concerns from the public.

Analysis: In our analysis, to understand the advantages of JSC-LDA, we study the comments that are scored high by JSC-LDA but low by BOW and LDA. For PM's speech, we observed several cases where LDA and JSC-LDA scores are high for comments with colloquial language and abbreviations. Topic models are capable of finding the relationships in the data and dealing with abbreviations or misspellings. For example, in the case of the issue *National Service*, users tend to use informal words like *nsmen* and *nsman*. BOW model scores are ranked lower for such comments, whereas LDA and JSC-LDA scored high for such comments. In the case of LDA, the topics on economy, congestion and youth olympics were conflated and degraded the performance of LDA compared to JSC-lDA. Table 3.10 shows some sample comments that are linked correctly and scored high using JSC-LDA model but scored low by BOW and LDA models.

Obamas's Speech: Similarly, for Obama's speech we use the ranked list from Equation (3.5) to evaluate the results against the ground truth. Average precision@K across all issues is calculated and the results are shown in Table 3.9 for Obama's speech.

Model	**p@5**	**p@10**	**p@15**	**p@20**
BOW	**0.2571**	0.2429	0.2333	0.2000
LDA	**0.2571**	0.2429	0.2429	0.1929
JSC-LDA	**0.2571**	**0.2571**	**0.2667**	**0.2357**

Table 3.9: Obama's speech: Average precision@K across all issues.

From the results, notice that the results are mostly the same for all the models and JSC-LDA still performed better. Under Wilcoxon test, the results are not statistically significant. This can be due to smaller and focused comments, and all models have performed almost similarly in linking the comments to the issues in article. Hence a better design of the experiments can be formulated with larger datasets. JSC-LDA has higher precision@K (bolded) for all K in $\{5, 10, 15, 20\}$ compared to the other two models. In our analysis, we observed that unlike for PM's speech where commenters tend to use different words such as abbreviations in comments, comments on Obama's speech mostly use the words directly from the speech. This behavior shows that the users language aspects impact the comment linking process. It is interesting to study how the linguistics (language aspects) of users impact the relevance score calculation. We leave it to the future studies. Furthermore, Figure 3.7 shows the performance of each model breakdown by the issues from Obama's speech.

Figure 3.7 shows that precision@10 for JSC-LDA model performed the best for

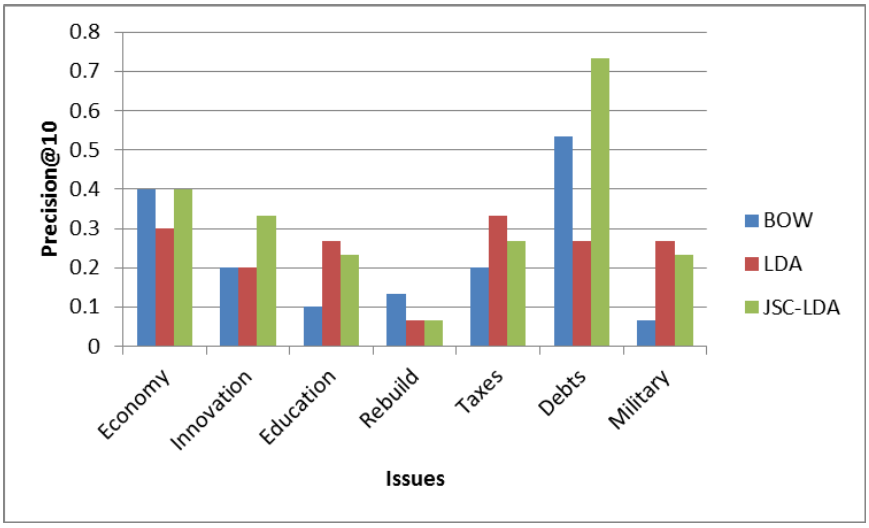

Figure 3.7: Obama's speech: Precision@10 for each issue.

most of the issues in the speech. Similar to previous results, BOW model sometimes performed better than topic models. The issue *Debts* has the highest performance for all the models. This suggests that *Debts* is one of the top concerns from the public.

Analysis: Similar to PM's speech, we analysed the results to understand the advantages of JSC-LDA, by studying the comments that are scored high by JSC-LDA but low by BOW and LDA. For Obama's speech, the observations are very similar to the PM's speech. However, the comments share the same vocabulary as the speech and are shorter than comments on the PM's speech and benefited BOW. In case of LDA, topics for the issues, *taxes*, *debts* and *rebuild* were incoherent, resulting in its lower performance than JSC-LDA. Similarly, Table 3.11 shows some sample comments that are linked correctly and scored high using JSC-LDA model but scored low by BOW and LDA models.

Summary: Our model performed better than other baselines for comment linking. All models have performed well on Obama's speech compared to PM's speech. We observe that the comments which share vocabulary similar to the article are easier to handle even by simple techniques like BOW, but for comments with abbreviations, wide feedback vocabulary and spelling problems, topic models have better performance over the simple text matching technique as they are capable to discover the unobserved relationships in the data. Our model performed better than LDA, as it has been designed to incorporate the segments of the speech in a principled manner to generate coherent topics.

Issue	Comment
Housing	I am living in the west area 80% of people I see around my area are mostly foreign it is really that bad here. I doesn't feel that I am living in HDBflat in Singapore much more like foreign hostel. All thanks to those people who love to rent out ..
Economy	I am an engineer ... the shrinking manufacturing industry coupled with the policy of letting in cheap engineers because they are SKILLED AND TALENTED! Guess i should not have listen to them in uni when they advertise about needing more engineers. Oh well time to learn how to deal cards at the IRs .. oh wait .... should i listen to them again?..
National Service	$9000 for NSMen is like taking away 2.5 years off your personal freedom + 13 years active reservist (not sure but like maybe 40 days a year x 10). So maybe like 3.5 years in total.

Table 3.10: Sample comments on PM's speech scored high by JSC-LDA model but scored low by BOW and LDA

3.6 Discussions

Our experiments demonstrate the benefit of modeling speech together with the comments in issue extraction. While the results are very promising for issues extraction and comment linking tasks, one threat to validity in interpreting results corresponds to generalizing our results. We evaluated our model on two different speeches. We admit that this is a small set. However, to minimize the impacts, we choose the speeches from two different countries where there is huge linguistic gap between the citizens of these two countries - for Obama's speech, the commenters are mostly the native English speakers. At the same time, we ensured larger dataset for the comment pairs for each corpus to conduct model evaluations (1350 pairs of issue-comments for PM's Speech and 1150 pairs of issue-comments for Obama's speech). Though we used unsupervised approach for our solution, we admit that comparing our solution with a supervised model would give us an idea on how the fundamental supervised approaches perform. Another threat is interpreting the topics discovered by the baseline and our solution. In our analysis, we used a simple visualization to interpret the resulted topics using the top words in each topic [20]. A principled method of interpreting the topics is suggested by Chang et al [26] and requires more resources to evaluate.

Our work can be extended in several directions. First, since not all comments are of high quality and useful to the domain users, it is interesting to study the problem of extracting high quality comments. Second, segmenting the comments based on the issues discussed would be desirable to understand correlated issues. Third, it should be worth extending this work to understand the polarization of the comment to generate the overall sentiment of the public on the topic. The challenge here is that the sentiment of the comment is not expressed by standard sentiment words, but is often implied with opinion expressions related to the sociopolitical context. Capturing hierarchial issues is an interesting subtask of this problem. For example, given a main issue, *women*, the examples of sub issues are *women rights*, *religion*, *health* etc,. To capture such hierarchial interpretation of the data, JSC-LDA can be extended to join the topics together in a hierarchy by using the nested Chinese restaurant process [18]. Finally, our model currently doesn't capture the phrasal content from the comments. This task can be achieved by extending the model using n-grams or with key phrase extraction supervision mechanisms. Moreover, it should be very interesting to study the abstractive summarization problem to generate a summary of the feedback.

3.7 Related Work

The problem we study is new and therefore there is no existing work addressing the same problem. There are, however, a few lines of work that are closely related to our problem.

Issue	Comment
Debts	besides defense, Obama's move indicates a willingness to cut discretionary-spending. Democrats have repeatedly tried to reduce tax expenditures. In the 9/11 responders health care bill, the funding mechanism was removing the loophole that allowed companies to offshore money and avoid taxes. ..
Education	As for education, achievement in K-12 science tracks math achievement closely, and US K-12 math achievement is lagging. This is where modern education fails. Math achievement is very much a case of "practice makes perfect", but this is neither interesting nor fun, and current K-12 dogma tries to make all learning "fun".
Innovation	We can have the benefits of technology as well as agrarianism. Why not In-fill the neighborhoods and repurpose all this wasted urban space for things that actually serve the community's basic human needs of food, water, shelter, and education.

Table 3.11: Sample comments on Obama's speech scored high by JSC-LDA model but scored low by BOW and LDA.

Opinion Mining: Opinion mining research is currently active in mining opinions related to social, economical and political aspects. Some work has focused on understanding what voters are thinking [94], while others [44] analyze the public sentiment on social events. Most of this research is based on data from news or blogs [48].

Comments Linking: A task of aligning the aspects of the event to the public feedback in Twitter classifies the tweets as episodic or general [62]. The assumption made in their work is that, every segment in the speech is a mixture of topics, while we treat each segment as a single topic and align comments to the segment. They focus on detecting topics but in our work, we align the comments to the segment of the articles. There has been some recent work on linking comments to the different segments of a news article or social speeches [114, 113]. Our work is also very close to reading news articles and comments together [114, 113]. Their task is to align the comments to the segments of the article. In their study, they found that using the original bag-of-word representation produces the best results compared to LDA and semi-supervised probabilistic latent semantic analysis (pLSA) based on pLSA [58].

In our work, we are not only linking the comments but also extracting the issues from the comments. Inspired by previous works where topics are extracted from unstructured texts [62], we propose a method based on statistical modeling and the evaluations show that our method performs better than bag-of-words technique.

3.8 Chapter Summary

In this chapter, we presented a study of extracting issues and linking social comments to the topics discussed in a document/speech. It aims to extract issues from the social comments in response to an article and link comments to the relevant issues in the article. We proposed a topic modeling approach to solve the problem. We experimented our proposed solution on two data sets and compared against two baselines: bag of words and standard topic model, LDA. Experiment results show that our proposed model performed better in terms of precision@K compared to both baseline models.

Chapter 4

Entity and Suggestive Opinion Extraction

The previous chapter helps identify major issues, but lacks in extracting fine-grained opinions such as the entities and the sentiments towards the entities in a comment. In the sociopolitical domain, users express their sentiments towards individuals or organizations who are related to a social event. We refer to such individuals or organizations as *entities*. The users express their sentiments not only in the form of positive and negative expressions, but also in the form of suggestive opinions towards the entities. Such *suggestive opinion comments* are useful for domain experts. Expressions that contain a request or a suggestion that can be acted upon are referred to as *actionable content* [41]. In this chapter, we study the problem of extracting entities, and suggestive opinions towards the entities from the users' comments in the sociopolitical domain. The problem aims at extracting and normalizing the entity-action pairs. In this work, we use the terms action and suggestive opinion (suggestion) interchangeably. We propose a two-stage approach based on conditional random fields and clustering technique for extracting and normalizing the entities and the associated suggestive opinions from the users.

4.1 Introduction

Extracting actionable knowledge from online social media has attracted a growing interest from both academia and the industry. Users' intention to write comments on non-product issues like social, economical and political problems is to express positive sentiments or negative sentiments or suggestions towards the issue. In this chapter, we focus on comments that contain sentiment expressions in the form of suggestions. Following the work by Whittle et al. [131] and Ferrario et al. [41], we define *actionable comments* as expressions that contain the requests or suggestions that can be acted upon. While motivating our task based on the previous work, we further extend the definition of an *actionable comment* as an *expression with an entity such as person or organization and a suggestion that can be acted upon*. For example, in the comment, "the government should tighten immigration rules," "the government" is an entity and "tighten immigration rules" is an action (suggestive opinion) expression.

Detecting actionable comments is an important subtask for various problems. First, actionable knowledge detection opens a new perspective to opinion mining such that it taps into the aspect of suggestion generation process currently missed by traditional content analysis approaches. Second, this task aids in finding the public's suggestive sentiment towards the entity by exploiting the individual value of an opinion and aids domain experts [41]. Third, when users intend to get the gist of the comments, this task aids in generating such well structured entity-based summaries on suggestive opinions as shown in Table 4.1.

Finding a piece of actionable knowledge in social media typically involves extensive human inspection, which is labor-intensive and time-consuming. In this work, we focus on automatically extracting such knowledge from social media. At first glance, the task looks like a typical information extraction problem, where entity and action expressions are to be extracted from comment sentences. However, in addition to entity-action pair extraction, the task also involves normalization of these expressions. Social media users tend to refer to the same entity with various forms and this triggers the need for entity normalization. Similarly, similar actions are rephrased and this triggers the need for action normalization to eliminate the redundancy. Therefore, part of our task is to normalize both entity mentions and action expressions.

To illustrate the nature of the task, let us examine the following examples:

E1 "*The government should lift diplomatic immunity of the ambassador.*"

E2 "*Govt must inform the romanian government of what happened immediately.*"

E3 "*SG government needs to cooperate closely with romania in persecuting this case.*"

E4 "*Hope the government help the victims by at least paying the legal fees.*"

E5 "*I believe that goverment will help the victims for legal expenses.* "

The above comments are in response to the news about a car accident in Singapore that involved a Romanian diplomat and hence raised immense public concern. First, all the above sentences consist of actions and the corresponding entities who should take/act upon the actions. Second, users tend to express the actions in various sentence structures, and hence extracting entities and actions is desired and challenging as well. Third, we observe that entities in all the above sentences refer to the same entity, Government, but expressed in various canonical forms. This drives the need for normalizing the entities. Finally, similar actions are expressed differently as shown in the last two sentences, *E4* and *E5*, which drives the need for normalizing the actions. The normalization should handle the redundancy of actions. We treat all the above expressions as actionable comments, and here we study how to extract and normalize entities and actions from users' comments. Table 4.1 gives an example output of our task. The last two sentences *E4* and *E5* are redundant and hence only one sentence is used in the output and the other is discarded. We shall explain in detail how the redundancy is handled and the representative sentence is chosen in our solution section 4.4.

Entity	Action
government	lift diplomatic immunity of the ambassador.
government	inform the romanian government of what happened immediately.
government	cooperate closely with romania in persecuting this case
government	help victims by at least paying the legal fees.

Table 4.1: Sample output of actionable comments extraction and normalization task.

The two main challenges that we address in this task are *entity-action pair extraction* and their *normalization*. We take a principled approach to tackle these challenges. We define the first part of the task as an information extraction problem and use sequence tagging techniques to solve the task. Using syntactic, semantic, positional and dependency features, we train a linear-chain conditional random fields [74] to identify the entity-action pairs. The second part of the problem is treated as a clustering task. We adopt clustering techniques [64] combined with external knowledge to tackle this problem. Our idea is to expand the entities with additional features and to cluster them using hierarchial clustering. After clustering the same entities in to a single group, we can normalize them by the representative of the cluster as explained in the Section 4.4. We expand an entity using the Google search engine snippets together with the dependency tree semantic-similarity sieves adopted from the Stanford coreference resolution algorithm [109]. For action normalization, one can approach this problem using similarity matching techniques to find redundancies.

Our study has the following contributions:

1. Based on actionable knowledge studies, we define a novel problem of entity-suggestive opinion pair extraction and normalization from user generated content. We study the problem on the dataset we obtained from users' comments on news articles.

2. We propose a principled approach to motivate and design the solution to detect actionable comments with entities and actions. We use the standard conditional random fields (CRF) techniques for extraction and clustering techniques for normalization.

3. We conduct comprehensive experiments to evaluate the techniques proposed in our study. We further compare the model with standard techniques like pattern matching. Experimental results show that the CRF model and agglomerative clustering model performs well. The system identifies exactly matched entities with 75.1% F-score and exactly matched actions with 76.43% F-score. Complete link measure outperforms single link measure for all articles in normalizing task with an average precision of 81.15%.

The rest of the chapter is organized as follows. In Section 4.2, we assesses the English language usage for motivating and designing the solution by characterizing how actionable comments are expressed in text. Section 4.3 describes the task definition. In Section 4.4, we describe the solution for extraction and normalization of the actionable comments. In Section 4.5 and Section 4.6, we describe the dataset and experimental results. We present our discussions in Section 4.7 and finally we conclude with some related work analysis in Section 4.8.

4.2 The Nature of Actionable Comments

How are actionable comments expressed in English sentences? In this section, we study the language aspects of actionable comments at sentence level and at phrase level. As mentioned, this study is important for motivating and designing our solution. At the sentence level, we show that many actionable comments are consistently expressed using a compact set of keywords, and quantify their frequency based on a sample of 500 sentences selected at random from a news forum. This observation helps us understand the statistics of actionable comments and explain the need for the data pre-processing to aid our entity-action extraction solution. At the phrase level, our observations justify the motivation of framing this problem as extraction and normalization task.

4.2.1 Sentence Level Study

First, to understand how frequently a user writes an actionable comment, we randomly selected 500 sentences from AsiaOne.com[1], a news forum site. These sentences are from users' comments and each comment contains one or more sentences. These sentences are not used for our experiments. We manually labeled these sentences as actionable comments or non-actionable comments. Our first observation is that 68 sentences, which is 13.6% of the sentences, are actionable comments. This is a very small set of candidates and hence justifies the need for detecting actionable comments. Second, to understand how to filter the comments that are non-actionable using some patterns, we further analyzed actionable comments using a different set of 458 actionable comments at sentence level and our second observation is that, 88.3% of the actionable comments use the keywords listed in Table 4.2. These findings are very similar to [41], in which the authors observed that actionable knowledge in Twitter is expressed using a set of keywords such as *should*, *need*, *take* and *better*. Table 4.2 also shows the frequency of keywords in the actionable comments from the dataset we used for our experiments. The most common patterns which are frequently used by the users are, *noun phrase* + *keyword* + *verb phrase* and *keyword* + *noun phrase* + *verb phrase*.

We also observed that the remaining comments do not show any strong pattern or keyword nature or they might appear very rarely in the corpus. For example, consider the comment "*Its time the SPF open up the case and be transparent to the public on the punishment faced by the abusers*." *Its time* represents a keyword, but this appears only once in 458 sentences. Based on Lexipedia synonym structure[2], we also added the following keywords to enhance the list: *require*, *better*, *why cant*, *how about*, *expect*, *please* and *why not*. The full keyword list can be used for candidate selection process and aids in the following extraction step.

Using the above keywords we now study the accuracy of identifying the action-

[1] www.asiaone.com

[2] `www.lexipedia.com/english/"keyword"`

Keyword	Frequency	Keyword	Frequency
should	54.24%	hope	8.47%
may be	5.08%	have to	5.08%
to be	3.39%	suggest	3.39%
need to	3.39%	must	3.39%
believe	3.39%	advise	3.39%
suppose to	1.69%	request	1.69%
ought	1.69%	needs to	1.69%

Table 4.2: Keywords and their relative frequencies in actionable comments (458) from the dataset used for our experiments.

able comments. We randomly extracted 550 sentences with the actionable keywords defined in Table 4.2 and traced for actionable comments. We identified that 458, which is 83.4% of the comments, are actionable and others are non-actionable comments. These observations justify the need for filtering the user comments using the keywords and generating the candidate set of sentences. The relative frequencies shown in the Table 4.2 are from the actionable comments of this dataset which has been used for our experiments. For our solution, we rely on data pre-processing by leveraging on these language dynamics.

4.2.2 Phrase Level Study

Intuitively, given an actionable comment, the entities can be treated as noun phrases and actions as verb phrases. To illustrate the nature of the problem at the phrase level, consider the examples below:

C1 *Police must follow up immediately as you have a duty in this criminal act.*

C2 *The diplomat should be banned by the government.*

C3 *Govt should tighten immigration rules.*

C4 *They should learn to understand Singapore's culture.*

C5 *Hope goverment can come out some law to protect owners as well.*

C6 *The government should bring in laws to protect the owners.*

We observe the following challenges in extracting actionable comments:

Entity extraction: Users can express suggestions in either active or passive voice. Therefore, the subject of the sentence is not always the entity we want to extract. In C1, *Police* is the agent (entity) of the action and is expressed in active voice. In C2, *government* is the agent (entity) of the action but is expressed in passive voice. The first challenge is to identify the correct entity in an actionable comment.

Normalization: People may refer to the same entity using different expressions. For example, the Singapore government may be referred to as "*govt*" as in C3 or by "*government*," "*Singapore Govt*," "*SG government*," etc. Ideally we should normalize these different expressions. C4 shows an example actionable opinion where "*they*" refers to foreigners and implies that pronouns should be normalized as well. We manually read the full comment to disambiguate the entities for the above examples. The second challenge is to normalize the entity mentions to their canonical form.

Redundancy: Very similar actions can be expressed differently. For example, C5 and C6 above suggest the same actions but different expressions are used. This motivates the normalization of actions to aid the reduction of redundancy. The third challenge is to normalize similar actions to aid in redundancy elimination.

Overall, the first challenge motivates us to detect entity-action pairs as an information extraction task and the last two challenges motivate normalizing the entity-action pairs as a normalization task. We design our solution based on some standard techniques. We apply sequence tagging techniques to solve the extraction task and leverage on the above observations to define the feature set. We take a clustering approach to solve the normalization task. In the next sections, we describe our task and solution formally.

4.3 Task Definition

The goal of our task is to extract and normalize actionable comments from user generated content in response to a news article. The actionable comments will be represented as an entity-action pair. To formally define our problem, we first introduce some basic concepts.

Entity Mention: We assume that the news article is about an issue that has generated wide interest in public. Entity mentions are the people or organizations mentioned in the users' comments. For example, "*people in the crime scene*," "*authorities*," "*they*," "*victims' families*," etc., are the entity mentions.

Action Expression: When users read a news article, they comment with an intention to suggest what should be done. We refer to the suggested action as an action expression. For example, "*ban the maid*," "*bring in new laws*," etc., are the action mentions.

Normalized Entity: In free language, an entity is mentioned in various ways. A normalized entity is the actual entity that an entity mention refers to. For example, "*President*," "*Obama*," "*he*," "*Barack*," "*Barack Obama*," etc., may all refer to the current U.S. president Barack Obama. The process of selecting of representative for normalized entity is explained in the Section 4.4.

Normalized Action: We assume that users tend to suggest similar actions in various ways. For example, "*protect the owners*" and "*owners are to be protected*" refer to same action "*owners' protection*," which is the normalized action. The process of selecting of representative for normalized action is explained in the Section 4.4.

Actionable Comment: An actionable comment is an entity-specific suggestion that consists of an entity and action pair. For example, $\{$*government, protect the owners*$\}$ and $\{$*govt, owners are to be protected*$\}$ are actionable comments.

Normalized Actionable Comment: A normalized actionable comment can be defined as an actionable comment that contains the normalized entity and normalized action. In the above example, $\{$*government, protect the owners*$\}$ is a normalized actionable comment.

Our problem of detecting normalized actionable comment is defined as follows: Given a news article A and corresponding candidate comments $C = \{c_1, c_2, \ldots, c_n\}$ extracted using the keywords, our goal is to detect pairs of $\{ne_i, na_i\}$ where ne_i is a normalized entity and na_i is a normalized action.

We tackle this task in two steps: In the first step we detect entity mentions and action expressions from the comments, and in the second step we normalize the entities and actions. For the first step, we use conditional random fields (CRF) and for the second step, we use clustering techniques combined with part of the Stanford coreference algorithm [109]. The details of the solution are given in the next section.

4.4 Solution

In this section, we first describe our solution for entity-action extraction using CRF model that can be trained to learn how actionable comments are expressed in English. We then describe our normalization model based on the clustering techniques for entity and action normalization.

4.4.1 Entity-action Extraction

The entity-action extraction problem can be treated as a sequence labeling task. Let $x = (x_1, x_2, \ldots, x_n)$ denote a comment sentence where each x_i is a single token. We need to assign a sequence of labels or tags $y = (y_1, y_2, \ldots, y_n)$ to x. We define our tag set as {BE, IE, BA, IA, O}, following the commonly used BIO notation [110], where E stands for entity and A stands for action. The detailed description of CRF technique can be found in [74]. Our task can be reduced to finding the best label sequence $\hat{y}$ among all possible label sequences for x.

Linear-chain CRF model

For our sequence tagging problem, we create a liner-chain CRF based on an undirected graphical model in which the conditional probability of a label sequence y given the observations x is

$$p(y|x, \Lambda) = \frac{\exp\left(\sum_i \sum_k \lambda_k f_k(y_{i-1}, y_i, x)\right)}{Z(x, \Lambda)}, \tag{4.1}$$

where Λ is the set of model parameters, f_k is an arbitrary feature function defined over two consecutive labels and the whole observation sequence, k is the index of a feature, λ_k denotes weight assigned for the corresponding feature, and

$$Z(x, \Lambda) = \sum_{y'} \exp \left(\sum_i \sum_k \lambda_k f_k(y'_{i-1}, y'_i, x) \right), \tag{4.2}$$

is the normalization constant for each x.

Given a set of training instances x_j, y_j^*, i.e. a set of sentences paired with their correct labels, we can learn a best model parameters $\hat{\Lambda}$ as follows,

$$\hat{\Lambda} = \arg\min_{\Lambda} \left(- \sum_j \log p(y_j^* | x_j, \Lambda) + \beta \sum_k \lambda_k^2) \right), \tag{4.3}$$

where $\beta \sum_k \lambda_k^2$ is a regularization term.

Features

We now describe the features used in the linear-chain CRF model. To develop features, we consider three main properties of actionable comments. First, the entities of the actionable pairs are mostly nouns or pronouns. Second, the entities display the positional properties with respect to the keywords. For example, given the keyword "*should*," entities generally precede the keyword as in "*The police should check on all agencies regularly*." Third, the entities should be grammatically related to the actions. For example, a verb in the action phrase is related to the subject which is an entity of the actionable comment. Consider the following comment: "*The government should ban the maid from entering the country*." The nominal subject for the verb "*ban*" is "*government*."

With these properties in mind, we define the following set of features for each word x_i in an input candidate sentence.

Parts-of-speech features: To capture the first property, we classify each word x_i into one of the POS tags using the Stanford POS tagger[3]. We combine this feature

[3]http://nlp.stanford.edu/software/tagger.shtml

with the POS features of neighboring words in [-2, +2] window.

Positional features: To capture the second property, we find the position of each word, x_i with respect to the keyword in the given sentence. The feature is represented as positive numbers for words preceding the keyword and negative numbers for words succeeding the keyword in the sentence. We do the same for neighboring words in [-2, +2] window. For example, in the sentence "*This diplomat should be banned by the Government*," the position of "*diplomat*" is 1 since it is one word before the keyword *should*, whereas the position of "*banned*" is -2 as it appears after the keyword.

Dependency tree features: To capture the third property, for each word x_i, we check if it is nominal subject in the sentence and represent it by *nsubj*. When the actionable comment is phrased in a passive voice as in "*Diplomat should be banned by the Government*," the entity "*Government*" is an *agent* in the sentence. The dependency tree features can be extracted using Stanford dependencies tool[4]. We use all the dependency features for our experiments.

The output of this task is $S = \{e_i, a_i\}$, a set of entity-action pairs. The next task is to normalize S which is described below.

4.4.2 Entity-action Normalization

Given $S = \{e_i, a_i\}$, a set of entity-action pairs, the goal is to generate $NS = \{ne_i, na_i\}$, a set of normalized entity-action pairs.

Entity Normalization

Entity normalization can be defined as a task that identifies a canonical unambiguous referent for entities. This task aims at grouping or clustering the various forms of entities. One of the techniques suitable to solve this problem is clustering. We use hierarchial clustering technique together with expanding the entity with the features

[4]http://nlp.stanford.edu/software/stanford-dependencies.shtml

from Google and Semantic-Similarity Sieves adopted from Stanford coreference algorithm [109]. We first explain the clustering model, then the features and also an approach to choosing the representative entity for the group of entity mentions.

1. Agglomerative clustering: This is a hierarchical clustering method which works bottom-up [98]. The distance between two clusters can be computed by using several measures. Two measures are widely used: single link and complete link.

Single link measure is given by:

$$d_{min}(C_i, C_j) = min_{e_i \in C_i, e_j \in C_j} d(e_i, e_j) \tag{4.4}$$

Complete link measure is given by:

$$d_{max}(C_i, C_j) = max_{e_i \in C_i, e_j \in C_j} d(e_i, e_j) \tag{4.5}$$

where d is the distance function such as Tf-IDF or Jaccard, C_i is the i-th cluster, C_j is the jth cluster, e_i is the entity member of C_i and e_j is the entity member of C_j.

To measure the distance between two entity members, we expand each entity mention with features developed as explained next.

2. Features: Two types of features are used to expand an entity mention: first from Google and second from the parse tree structure.

a. Alias features: This sieve addresses name aliases, which are detected as follows: Given an entity mention, it is first expanded with the title of the news article and this query is fed to the Google API. Table 4.5 shows the titles of the articles used for our experiments. Google outputs the ranked matching outputs. One option is to use the entire snippet as the features. Another option is to use partial snippet. Google returns snippets that has bolded aliases as shown in Table 4.3. We can use them as the alias features for a given entity mention. For example, the alias features for "Ionescu + *title*" are *Dr.Ionescu, Silvia Ionescu, Romanian Diplomat Ionescu,* etc. This sieve also aids in solving the spell problems. For example, for the query "goverment + *title*", Google returns one of the alias features, *Government*. In our experiments we use top 10 Google snippets.

Entity Mention	Sample Google Snippet
Ionescu	**Romanian diplomat Ionescu** was the driver in fatal hit-and-run ... of Romanian embassy car which was involved in ...
MFA	Romanian Diplomat Escapes Singapore After Double Hit and Run ... The **Singapore Ministry of Foreign Affairs (MFA)** claimed ...

Table 4.3: Sample snippets from Google.

b. Semantic-similarity features: We execute the following steps from the relaxation algorithm from Stanford coreference resolution tool: (a) remove the text following the mention head word; (b) select the lowest noun phrase (NP) in the parse tree that includes the mention head word; (c) use the longest proper noun (NNP*) sequence that ends with the head word; (d) select the head word. For example, the entity mention "*The President Bill Clinton*" is changed to "*president Bill Clinton*" then "*Bill Clinton*" and finally "*Clinton*." These individual features are added to the feature set of a given entity mention.

For un-named entities:(a) select the parse tree of the entity; (d) remove the determinants (DT) associated; (b) remove the modifiers (amod) or remove noun compound modifiers (nn); (c) finally select the head. For example, the entity mention "*The old man*" is changed to "*old man*" and then "*man*." Table 4.4 shows the feature set generated using alias and semantic-similarity method for an entity mention "*SMRT CEO Ms Saw*." The terms shown are from both the features and the overlapped terms are used only once.

SMRT CEO Ms Saw, CEO Ms Saw, Ms Saw, Saw, SMRT's, Ms, SMRT Security, MRT Breach, SMRT's CEO, security breaches, Vandals, Saw's Security Breach, SMRT CEO Saw, CEO, Mrs, SMRT depot

Table 4.4: Feature set for entity mention, "SMRT CEO Ms Saw".

3. Representative entity: Usually a centroid represents the label for a given cluster and it is calculated by taking the average of the vector representations of the members in the cluster. Since we need an entity or member from the cluster itself to represent the cluster, one approach is to find an entity mention that has the maximum similarity to all the other entity mentions. Hence the representative of a cluster, n_e is chosen to be the entity mention who has the largest average similarity distance (Jaccard similarity) from the other entity mentions in the cluster. For example, in the above case, "CEO" is the representative entity.

Action Normalization

The main objective of normalizing the actions is to remove the redundant actions. This task can be treated as a clustering problem, where we group similar actions together and choose one representative action from the group. We choose hierarchial clustering same as above to normalize the actions associated with the same normalized entity. The feature set for this task is simply bag-of-words with stop word removal. The representative action is also chosen similar to the approach taken for finding the representative entity.

4.5 Dataset

Since the task of actionable comment extraction is new, we gathered and annotated our own dataset for evaluation. Our dataset consists of 5 contentious news articles and the corresponding comments from Asiaone.com, an online forum. Table 4.5 shows the topics/titles of the 5 articles and the corresponding numbers of user comments.

Article	Title	#Comments
A1	Graffiti vandal breaches security at MRT depot	622
A2	Jack Neo's press conference	664
A3	Maid abuses child, caught on video	285
A4	Romanian Embassy Diplomatic Car Hit & Run	822
A5	Wrongly caned prisoner's letter to family withheld by prison	379

Table 4.5: Topics of articles and the corresponding number of user comments.

4.5.1 Pre-processing

Recall the need for the candidate selection and candidate extraction method explained in Section 2. For the dataset preparation, we use the keywords listed in Table 4.2 to extract the candidate sentences from all the comments where each comment may have more than one sentences. We randomly use 110 candidate sentences from each article and in total 550 candidates for experiments to evaluate our method. We explain the details of annotation in our experiments in Section 4.6. As such, to encourage comparative work, we have made the resulting annotation available for download[5].

4.6 Experiments

We design our experiments to answer the following questions:

(Q1) How does the CRF technique perform in identifying actionable comments? We can answer this question by treating this as a classification task and measure the performance against ground truth using F-score.

(Q2) How does the CRF technique perform in extracting the entities and actions from the actionable comments? We can answer this question by comparing it with some baseline techniques and compare the accuracy between the models against the ground truth.

(Q3) Between single link and complete link, which technique is more suitable for this problem? We can answer this question by evaluating the results of entity normalization against ground truth using Purity, Entropy and F-score.

(Q4) How does the clustering-based solution perform in normalizing the entity-action pairs? We can answer this question by measuring the precision against human judgement.

[5]https://sites.google.com/site/swapnagotipati/datasets

4.6.1 Experiments on Entity-action Extraction

To evaluate the entity-action extraction, we prepare the ground truth using the dataset described in Section 4.5. We first answer (Q1), how well the model performs in identifying actionable comments. We then evaluate the entity and action extraction from the actionable comments to answer (Q2). We experimented with various combinations of features for CRF model and combined feature set gives the best results. We perform 10-fold cross validation for all our experiments.

Annotation

To prepare the ground truth, we engaged two annotators to label 550 candidate sentences for suggestion, entity and action. One of the annotators is a graduate student and another is a working professional. Both are above 25 years old, not the authors and are familiar with Singapore social and political affairs. For this annotation task, the annotator should do the following:

1. Check if the sentence is a suggestion from the commenter. If no, label the comment as 0.

2. Look for the person(s) or organization(s) who should execute the suggestion, and label the entity with BE (beginning of an entity) and IE (inside an entity).

3. Look out for the action that should be performed by the entity, and label it as an action: BA (beginning of an action), IA (inside an action). The others are labeled as O (other).

4. If both entity and action are found, sentence is a valid suggestion. Label it as 1. Otherwise, label it as 0.

We calculated the inter-annotator agreement level using Cohen's kappa. Cohen's kappa on actionable comments is 0.7679, which displays a strong agreement between the annotators. We use the annotations from the stricter annotator for our experiments. Interested readers kindly refer to Appendix A for more details on annotation process and sample examples. The statistics of actionable comments for each article are shown in Table 4.6.

Article	Actionable Comments
A1	80.52%
A2	89.21%
A3	83.19%
A4	79.63%
A5	84.48%

Table 4.6: Statistics of actionable comments in our testbed.

Actionable Knowledge Detection Results

Recall that detecting actionable comments can be treated as a classification task. Essentially, if CRF successfully extracts both an entity and an action from a comment, then this comment is classified as an actionable comment. Our model achieved precision of 88.26%, recall of 93.12% and F-score of 90.63% in classifying actionable comments and that answers our Q1. In our analysis, we observed that the model failed in detecting the actionable comments when the sentences have poor grammatical structure. For example, "*Dont need to call the helpline.*.", has a poor grammatical structure. It is not an actionable comment but is mistakenly identified as one by our CRF, where *Dont* is the entity and *call the helpline* is the action.

Baseline

In Section 4.2, we described that the actionable comments are expressed in English language using a compact set of keywords. This might trigger us to use the pattern-matching technique to extract the action-entity pairs. Hence we use this pattern matching technique as a baseline and compare with our approach. For our baseline model, we further added some basic rules on adjectives, nouns and pronouns for entity mention detection, and verb phrases for action mention detection. It is likely possible to engineer more sophisticated models such as rule-based or graph-model methods for this task, but the goal of our work is to see how well we can extract actionable comments using standard state-of-the-art information extraction techniques. Our goal is not to compare rule-based vs. statistical methods for this task.

Entity Extraction Results

For entity extraction, we evaluate the performance using two different criteria: overlap match and exact match. Overlap match is a more relaxed criterion: if the entity mention overlaps with the true entity (i.e. at least one token is common), it is considered correct. The results are shown in Table 4.7. From the table, we notice that the baseline, which is the pattern matching technique, has high recall for exact match. In this task, the precision together with recall determines the effectiveness of the model. The baseline outperformed the CRF model on the overlap F-score and this is due to the relax mode of the overlap. However, for the exact match CRF has high F-score of 75.09% which is relatively 6.67% higher than the baseline. CRF has performed significantly better than baseline under a Wilcoxon signed rank test $p < 0.05$. We observed that for cases like "the sgreans", which actually refers to "The Singaporeans", the baseline captures "the" as an entity and performs high for overlap measure and CRF captures nothing. For sentences such as "It is right for the CEO to step down on her own accord", baseline fails to capture the entity "the CEO". On the other hand, CRF captures both entity and action correctly. This answers our (Q2) for entity extraction evaluation.

	Exact Match		**Overlap Match**	
Metrics	**Baseline**	**CRF**	**Baseline**	**CRF**
Recall	**0.8799**	0.8352	0.9032	**0.9306**
Precision	0.5866	**0.6849**	**0.9597**	0.8578
F-score	0.7039	**0.7509**	**0.9306**	0.8927

Table 4.7: Entity Extraction Results

Action Extraction Results

For action extraction, we evaluate the performance using two different criteria: head match and exact match. Head match is used because most actions start with a verb. Head match is a more relaxed criterion: if the action mention begins with the head of the true action, it is considered correct. The results of our experiments are shown in Table 4.8. From the table, we notice that the baseline, which is the pattern matching technique, has high recall for both exact match and head match. However, for both exact match and head match CRF has high F-score of 76.43% and 82.7%, respectively, which is relatively 11.9% and 0.03% higher than the baseline. The results are significant under a Wilcoxon signed rank test $p < 0.05$. Head match has generally high performance for both due to the property that an action is expressed as a verb. This answers our (Q2) for action extraction evaluation.

	Exact Match		**Head Match**	
Metrics	**Baseline**	**CRF**	**Baseline**	**CRF**
Recall	**0.8947**	0.8944	**0.9200**	0.9169
Precision	0.5519	**0.6741**	0.7468	**0.7544**
F-score	0.6827	**0.7643**	0.8244	**0.8270**

Table 4.8: Action Extraction Results

4.6.2 Experiments on Entity-action Normalization

Recall that our second task is to normalize the entity-action pairs which has been framed as a clustering problem. We now study how our proposed approach generates the normalized clusters of entity-action pairs. In Section 4.2.1, we mentioned that cluster distance can be computed by using single link or complete link. First, we compare the performance of these two measures on entity normalization task to answer (Q3). We then use the results of this study to normalize actionable comments and then measure the performance of normalized entity-action pairs to answer (Q4).

Single Link Vs Complete Link

Annotation: To answer Q3, we build the ground truth using entities from all the articles and compare the complete vs single link techniques. The human annotator is given a set of entities from each article and asked to first group the similar entities together and then assign a label to each group. This annotated set is used to measure the performance of complete link vs single link in normalizing the entities.

Results: For grouping entities using the solution described in Section 4.2.1, we set K empirically using inter-cluster distance and cluster the entities. We use three different measures: Purity, Entropy [142] and F-score to evaluate our results. The results of Purity and Entropy are shown in Table 4.9.

	Purity		**Entropy**	
Article	**Single Link**	**Complete link**	**Single Link**	**Complete Link**
A1	**0.96**	0.9	**0.39**	0.43
A2	**1**	0.86	**0.23**	0.3
A3	**0.89**	0.79	**0.34**	0.38
A4	**0.82**	0.79	**0.48**	0.53
A5	**0.94**	0.91	**0.25**	0.28

Table 4.9: Purity and Entropy results comparison between single link and complete link

The Purity and Entropy results show that the single link measure has better performance when compared to the complete link. High Purity is easy to achieve when each entity mention gets its own cluster. We observed that for single link measure, most of the entity mentions take up their own cluster and at the same time, number of clusters in single link and complete link are not the same. This may cause Purity and Entropy to be biased and thus they are not ideal metrics for evaluating this task. Hence we prefer to use F-score which balances both precision and recall to evaluate this task. As shown in Table 4.10, even though the precision for single link is high, complete link out performs significantly single link on recall and F-score (under Wilcoxon signed rank test $p < 0.05$) and answers our Q3.

For example, "*the ceo*" and "*ceo, smrt ceo ms saw*" are grouped into single cluster using complete link. Where as, for single link cluster, "*smrt ceo ms saw*" is

	Single Link			Complete Link		
Article	**Precision**	**Recall**	**F-Score**	**Precision**	**Recall**	**F-Score**
A1	0.5161	0.5039	0.5100	**0.8462**	**0.6929**	**0.7619**
A2	**1.0000**	0.3333	0.5000	0.7143	**0.5238**	**0.6044**
A3	**0.7368**	0.3218	0.4480	0.5664	**0.7356**	**0.6400**
A4	**0.6258**	0.4567	0.5280	0.5328	**0.6689**	**0.5931**
A5	**0.9661**	0.4560	0.6196	0.7282	**0.6000**	**0.6579**

Table 4.10: F-score results comparison between single link and complete link

a false negative. Hence Purity is high for single link but F-score is low.

Entity-Action Normalization Results

Annotation: We first normalize entities and actions using complete link measure based on the previous experiments. To measure the performance of the normalized entity-action data and answer our Q4, we asked an annotator to validate the normalized entity-action pairs. Only if both entity and action are normalized (entity should be in canonical form and action should be non-redundant), the pair is labeled as valid, else it is an invalid pair. If we obtain (e1, a1), (e2, a2), and a1 and a2 refer to the same action, we label one of them as invalid.

Results: We use precision to evaluate the performance. The results of the entity-action normalization are shown in the Figure 4.1.

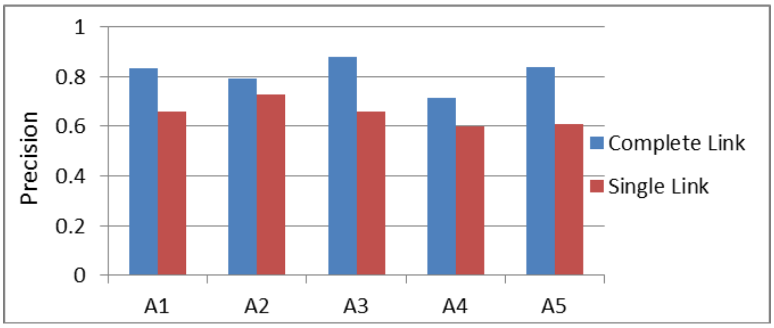

Figure 4.1: Entity-action normalization results

From Figure 4.1 we notice that on all articles, the precision is high for complete link measure. This can be justified due to high F-score from complete link measure. In our analysis, we observed that for single link, the entities like *he*, *they* are not normalized into the correct clusters resulting in the lower precision which is one of the drawbacks of single link where clusters are forced together due to single elements being close to each other, even though many of the elements in each cluster may be very distant to each other. Complete link measure significantly outperforms single link measure for all articles in normalizing task with an average precision of 81.15% (under Wilcoxon signed rank test $p < 0.05$) and that answers our Q4. We further analyzed the results for complete link and observed that the model fails if the feature set is not distinctive. For example, given article A4 and the normalized entity *Ionescu*, the actionable comments have entity mentions like *asshole*, *dog*, etc., which could not be normalized due to non-distinctive feature set.

4.7 Discussion

For the entity-action extraction task, we described patterns in the comments for discovering the actionable comments. In our study, we used conditional random fields (CRFs) to solve this sequence labeling problem. We admit that other methods such as rule-based algorithms or hidden markov models are also useful to solve this task. However, our focus in this study is to define a new problem, suggestive opinion extraction task, and to show that using principled machine learning techniques, one can solve this important problem.

Extracting actionable comment is very challenging problem and though our model performed with promising results for extraction and normalization, we observed some invalid or false positives such as "Compensation details should be disclosed since it involves the people money otherwise the government can conveniently hide the damage from its mistake from the people" or "Justice have to be administered fairly and accurately". In the above cases entities are abstract and do not refer to any individual or organization. To minimize the impact of such false positives, it is important to summarize the actionable comments with a threshold on number of entity-actions pairs. For example, we can set the threshold to two to capture non-abstract entities by filtering out the entities who have less than two entity-action pairs.

We also observed the plural forms of entities that refer to the same entity; Authority/Authorities, Ministry/Ministers. To handle the plural forms, the lexical features of terms can be implemented in the clustering model. For our solution, we proposed agglomerative clustering, as one of the benefits of this technique is that the number of clusters need not be specified upfront and thus helped us to focus more on the overall problem and its evaluations. The model results can be improved by strengthening the evaluations with measures such as B-cubed precision [9]. For parameter settings, we used the test data and not a development dataset. Though this might impact the results, we believe that the impact is minimal. For cross validation, we ensure that we test on random sets of articles.

4.8 Related Work

Mining comments for opinions and knowledge extraction has been of growing interest. Our work can be compared to such problems.

Opinion Mining: Most sophisticated problems in fine-grained opinion analysis include: opinion holder extraction [69] and opinion target extraction [70]. Our task is very close to [85], where their task aims at extraction of the named entities(targets) in the articles. Similarly, we extract entities (named as well as unnamed) in opinion comments. In addition to the entity extraction, actions to be acted upon are extracted as well.

Actionable content: [139] attempted to discover the diagnostic knowledge and defined diagnostic data mining as, "a task to understand the data and/or to find causes of problems and actionable knowledge in order to solve the problems". Their work is more focussed towards manufacturing applications in which the problems are identified to aid the designers in the product design improvements. [116] analysed actionable knowledge in on-line social media conversation and the concept of actionability is defined as request or suggestion. [41] work aims at discovering aspects of actionable knowledge in the social media. Their objective is more towards investigating the dynamic aspect of the language the people use to express actionable knowledge. They conducted their study on Twitter and discovered user language aspects. We too focus on the actionable knowledge and our task is to extract the actionable knowledge. In addition, we also aim to extract the entities associated with the action content. Moreover, solving the challenge of entity-action normalization is important in this task due to ambiguous references and redundant user generated content. To the best of our knowledge, our problem of extracting and normalizing entity-action pairs from users' comments is not studied.

Techniques: Our task can be projected as information extraction and normalization problems. In terms of models and techniques, we use pattern-based techniques [96] as our baseline and standard linear-chain CRF [74] as our proposed approach, both which are widely popular among many information extraction problems. Normalization techniques that use external knowledge combined with coreference resolution algorithms [109] are proven to be successful in disambiguating the entity mentions. Moreover, the state-of-the-art clustering techniques proved to be successful in grouping the similar entity mentions together [14]. In our study we use agglomerative clustering model [98] to normalize the entities and actions.

4.9 Chapter Summary

In this chapter, we proposed a novel problem in the line of opinion mining of sociopolitical comments. Extracting and normalizing entity and suggestive pairs is a complimentary feature of opinion mining process. We proposed a principled approach using CRF model and clustering techniques. Our comprehensive experiments show that the CRF has better performance in entity-action extraction. The model detects exactly matched entities with 75.1% F-score and exactly matched actions with 76.43% F-score. Clustering with complete-link measure performs well in the entity-action normalization with an average precision of 81.15% for all articles.

Chapter 5

Valuable Comment Extraction

The previous two chapters are close to opinion extraction tasks where we studied the problems of issues extraction and entity-suggestions extraction from social comments. In this chapter we tackle the problem of opinion quality in sociopolitical comments. Online user comments contain valuable user opinions. Comments vary greatly in quality and detecting high quality comments is a subtask of opinion mining and summarization research. Finding attentive comments that provide some reasoning is highly valuable in understanding the users opinion particularly in sociopolitical opinion mining and aids policy makers, social organizations or government sectors in decision making. In this chapter, we study the problem of detecting thoughtful comments in the sociopolitical domain. We empirically study various textual features, discourse relations and relevance features to predict thoughtful comments. We propose supervised approach for classifying the comments and test on two datasets related to sociopolitical domain.

5.1 Introduction

In recent years sentiment analysis and opinion mining has been extensively studied in natural language processing [99], largely because of the availability of a huge amount of opinionated text in online product reviews, blogs, social networking sites, forums, etc. When we go beyond review mining and consider the general problem of opinion mining from social media, many other subtasks and challenges arise. One of them is how to assess the quality of online comments and select high quality ones for further analysis and summarization.

Social media enables anyone to freely express the opinions and that urges a need for extracting opinions which are highly valuable. Consider the problem of mining the comments found in online social media towards a political speech such as Obama's State of the Union address. By restricting the search space to politically active blogs and forums and by using queries such as "*State of the Union*," likely we are able to retrieve users' comments to the speech. However, not every comment contains valuable insight into the public's opinions regarding the sociopolitical issues addressed in the speech. Comments such as "*innovate and innovation appeared 10 times*" and "*To him..investment = more deficit spending*" are subjective but lack thoughtful explanations to support their claims. In comparison, comments like "*You want to really drive innovation, job growth and entrepreneurs? Make education, health care and retirement less of a burden on the average family, adopt more socialist policies like Norway (paid for by higher taxes, especially on the rich), and watch our standard of living rise at last!*" provide much more insightful reasoning that government policy makers may find highly valuable in understanding the general public's sentiment. [120] used thoughtful comments to study the human social behavior in online commenting.

Following Sukumaran et al. [120] we define thoughtfulness as insightful ideas and reasoning with relevance to the issues discussed in the article. So, detection of the comments with reasoning or justification is the focus of our task. Thoughtfulness is assessed only for relevant comments. This problem of finding *thoughtful* comments from social media is what we study in this work. Formally, a *thoughtful comment* is relevant to the target document and has a justification or an argument to the issue(s) in the target document [120]. It is particularly important for sociopolitical opinion mining because of the complexity of sociopolitical issues. Sukumaran et.al [120] proposed degree of thoughtfulness in their study, but in our work we deal with binary data and degree can be a complimentary task for our work.

Intuitively, finding thoughtful comments is related to measuring text quality. There has been a large body of previous work on text quality prediction, but the methods are usually applied to student essays [11] and news articles [122], (TREC novelty track 2003 and 2004). In social media mining, there have also been a number of studies on finding high quality reviews (e.g. [71, 4]), but the focus in not on finding thoughtful comments, which requires us to look for reasoning in text. Presumably, a thoughtful comment should be logically well organized and coherent. We therefore hypothesize that discourse relations such as comparison, expansion and contingency [107] will play an important role in finding thoughtful comments. A well organized comment is not always thoughtful. Comments such as, "*He is a great speaker as he writes the speech by himself and also delivers it very confidently*" are justified but are not relevant to the issues discussed in the article. Hence we hypothesize that relevance factors play an important role in detecting thoughtful comments.

Inspired by [102] for similar problem, we adopt a supervised learning approach and consider a diverse set of factors ranging from lexical usage to discourse relations, all derived from the textual content of comments. Many of the factors we consider are based on the study by Pitler et al. [102]. In addition, we also consider a relevance feature because of the nature of our problem. We construct two data sets to evaluate the various factors, one based on Singapore Prime Minister's National Day Rally speech[1] and the other on the US President's State of the Union address[2].

Empirical evaluation reveals that discourse relations and relevance scores together with the standard textual features aid in better prediction of thoughtful comments. We achieve a prediction score of 79.37% and 73.47% in terms of F-measure on the two data sets, respectively. We further tested our model across data collec-

[1]http://www.pmo.gov.sg/

[2]http://www.whitehouse.gov/

tions. Our test result shows that the model with combined textual, discourse and relevance features still performs better than textual features alone.

The rest of the chapter is organized as follows. In Section 5.2, we formally define our problem and give an overview of our solution. We present the various features we consider in Section 5.3. The data set details are presented in Section 5.5. Evaluation and results are presented in Section 5.6. We present our discussions in Section 5.7 and finally we conclude with some related work analysis in Section 5.8.

5.2 Problem Definition and Overview of Solution

We assume the following general definition of the task of finding thoughtful comments: Given a comment c made with respect to a *target* document d, we would like to determine whether c is a thoughtful comment. We will explain in Section 5.5 how we instruct the human annotators to label thoughtful comments. Generally speaking, a thoughtful comment is relevant to the target document and has a justification or an argument to the issue(s) in the target document [120] .

While the task defined above is certainly not trivial, and theoretically speaking one would need a deep understanding of both the target document and the comment as well as relevant world knowledge to be able to judge whether a comment is thoughtful. Here we take an empirical approach and test whether features defined at lexical, syntactic, discourse levels and relevance factors have correlations with the thoughtfulness of comments and whether they can be used to achieve decent prediction accuracy. A large portion of the linguistic features we consider are inspired by existing work on measuring text quality. Indeed, at first glance our problem may appear to be the same as measuring text quality. News articles and student essays are formal and usually lengthier, whereas online comments are usually much shorter and less formal. In traditional sense, high-quality text should be grammatical, coherent and readable. Our problem seems to be text quality assessment which is defined as above, but we are not looking for grammatical correctness and readability. Instead we look for insightful reasoning with relevance to the article, as mostly user comments are not formal in social media.

We adopt a supervised learning approach to our problem. Specifically, we assume that we have a set of N training examples $\{(d_i, c_i, y_i)\}_{i=1}^N$, where d_i is a target document, c_i is a comment on d_i, and y_i is a binary label indicating whether c_i is a thoughtful comment with respect to d_i. With a set of feature functions, we can represent (d_i, c_i) by a feature vector $\mathbf{x}(d_i, c_i)$ (which we refer to as $\mathbf{x}_i$). We can then use standard classification algorithms to learn a classifier from $\{(\mathbf{x}_i, y_i)\}_{i=1}^N$. This classifier can be used to predict y for any unseen pair of d and c. In the following sections, we will explain in detail the features we consider and the classification algorithm we use.

5.3 Features

There have been many studies on measuring text quality and many features have been proposed to capture text quality. As mentioned previously our methodology is based on existing work on this topic. In particular, we follow the work by Pitler et al. [102]. They conducted a systematic study on text quality using various linguistic features and Wall Street Journal articles. Based on the major findings of their study, we take the following features as our starting point.

5.3.1 Structural Features

Structural features are generated from the comment structure. Pitler et al. [102] tested various structural features including the average number of characters per word, the average number of words per sentence, the maximum number of words per sentence, and article length. According to their findings, article length was the only significant factor with good correlation with text quality. Hence, we define our first feature F_1 as the number of words in the comment.

5.3.2 Lexical Features

Lexical features aim to capture the lexical usage of a piece of text compared to some reference corpus. [102] used a lexical feature based on unigram language models, which provide a principled way to statistically model text. Specifically, it is assumed that there is a reference corpus that represents high quality text, e.g. a corpus of Wall Street Journal articles. A unigram language model, denoted as θ_r, can be estimated from this reference corpus. The lexical feature is defined as the log likelihood of the comment based on θ_r, calculated as:

$$\sum_w n(w, c) \log P(w|\theta_r), \tag{5.1}$$

where $P(w|\theta_r)$ is the probability of word type w according to θ_r, and $n(w, c)$ is the number of times word type w appears in comment c. We call this feature F_2.

5.3.3 Syntactic Features

Pitler et al. [102] examined various syntactic features including the average parse tree height, the average number of noun phrases per sentence, the average number of verb phrases per sentence and the average number of subordinate clauses per sentence. They found that the average number of verb phrases per sentence was a useful feature with high correlation with text quality. So, the third feature F_3 we use for our study is the average number of verbs per comment.

We also experimented with other syntactic features like average number of noun phrases and noun to verb ratio calculated from the user's comments. We found that the average number of verbs per comment had the highest correlation with comment quality, and therefore we do not consider these other syntactic features in our experiments.

5.3.4 Discourse Features

The previous study by Pitler et al. [102] found that discourse relations were also correlated with text quality. Discourse relations aim to capture textual structures such as comparison, elaboration, cause-effect explanations and examples. They are considered key for the ability to properly interpret or produce discourse. For the problem of finding thoughtful comments, we hypothesize that discourse relations may play an even larger role because a logical argument will likely rely on coherently connecting textual units through discourse relations.

Discourse relations are divided into four major semantic classes [107]:

Expansion covers those relations where the second argument expands the discourse of the first argument or move its narrative forward.

Comparison relations highlight prominent differences between the two arguments of a relation.

Contingency is marked when one of the situations described in an argument causally influences the other argument.

Temporal relations are marked when the situations described in the arguments are related temporally, either synchronously or sequentially.

It has been found that a large portion of discourse relations can be detected through connectives, i.e. cue words and phrases [103]. We use a list of such connectives compiled by Prasad et al. [107] and study the statistics of our corpus to discover the discourse relations. Table 5.1 shows that the statistics of discourse relations in our dataset.

DR Class	**Singapore**	**US**
COMPARISON	44.50%	45.87%
EXPANSION	16.75%	15.47%
CONTINGENCY	38.75%	38.66%
TEMPORAL	5.70%	5.72%

Table 5.1: Discourse relations statistics in our corpus

Table 5.1 shows that the frequency of temporal relations is low in our corpus. It is not surprising because for many online comments the arguments are not temporal. Hence, we ignore the temporal class for the rest of our work, and restrict our attention to only the other three major classes, namely, expansion, comparison and contingency. At the same time, many relations are explicit and can be discovered using the connectives/words as used in other applications [112].

The full list of the phrases for each class are shown in Table 5.2. This list is collated from [107]. We observed that some words are ambiguous: 'if', 'and', 'but', 'as' etc. In our study, such words are counted only once while combining the classes for the feature generation. For each class, if we see the same word multiple times in a sentence it is only counted once.

In the earlier work by Pitler et al. [102], the Penn Discourse Treebank was used for computing the discourse features. For us, we take a simpler approach and count the number of discourse relations in a comment[3]. This becomes the F_4 in our experiments.

5.3.5 Relevance Features

One of the important differences between our problem and standard text quality assessment is that the quality of a comment also relies on its relevance to the target of the comment. In our problem definition, the target is also a piece of text. For example, consider comments made to Obama's State of the Union speech. A comment such as "We are very lucky to live in the USA. I always have and always will support our president" is not directly related to any issue addressed by Obama in his speech, and therefore is not considered to be a thoughtful comment. Hence for thoughtful comment prediction, we also consider a relevance feature in addition to text quality features. There are several ways to measure relevance, and here we choose KL-divergence score, a principled measure for relevance commonly used in information retrieval tasks.

The KL-divergence score between a comment c and a target document d is defined as the KL-divergence between the unigram language models θ_c and θ_d esti-

[3]Note that we only use the explicit discourse relations in this study.

Class	Phrases
COMPARISON	although, as though, but, by comparison, even if, even though, however, nevertheless, on the other hand, still, then, though, while, yet, and, meanwhile, in turn, next, ultimately, meantime, also, as if, even as, even still, even then, regardless, when, by contrast, conversely, if, in contrast, instead, nor, or, rather, whereas, while, yet, even after, by contrast, nevertheless, besides, much as, as much as, whereas, neither, nonetheless, even when, on the one hand indeed, finally, in fact, separately, in the end, on the contrary, while
EXPANSION	accordingly, additionally, after, also, although, and, as, as it, as if besides, but, by comparison, finally, first, for example, for one thing, however, in addition, in fact, in other words, in particular, in response, in sum, in the end, in turn, incidentally, indeed, instead, likewise, meanwhile, nevertheless, on the one hand, on the whole, overall, plus, separately, much as, whereas, ultimately, as though, rather, at the same time, or, then, if, in turn, furthermore, in short, turns out, while, yet, that is, so, what's more as a matter of fact, further, in return, moreover, similarly, specifically,
CONTINGENCY	and, when, typically, as long as, especially if, even if, even when, if, so, when if only, lest,once, only if, only when, particularly if, at least partly because, especially as, especially because, especially since, in large part because, just because, largely because, merely because, not because, not only because, particularly as, particularly because, particularly since, partly because, because, simply because, since, then, after, one day after, reportedly after, consequently, mainly because, for, thus, apparently, in the end, in turn, primarily because, largely as a result, as, because, therefore, only because, particularly, when, so that, thereby, presumably, hence, as a result, if and when, unless, until, in part because, now that, perhaps because, only after, accordingly,

Table 5.2: Discourse relations

mated from c and d, respectively:

$$Div(\theta_c||\theta_d) = \sum_{w \in \mathcal{V}} p(w|\theta_c) \log \frac{p(w|\theta_c)}{p(w|\theta_d)}, \tag{5.2}$$

where $\mathcal{V}$ is the vocabulary.

KL-divergence using only nouns: We hypothesize that the topical relevance between a comment and its target relies more on the overlap of nouns in the two pieces of text. Hence, we consider another KL-divergence measure using only nouns in c and d. Specifically, we use the unigram language models that are defined over nouns only. Let θ_c^N and θ_d^N denote the two language models. We have

$$Div(\theta_c^N||\theta_d^N) = \sum_{w \in \mathcal{V}} p(w|\theta_c^N) \log \frac{p(w|\theta_c^N)}{p(w|\theta_d^N)}. \tag{5.3}$$

We define the fifth feature that uses only nouns, F_5=-$Div(\theta_c^N||\theta_d^N)$ and the sixth feature which is based on all words, F_6=-$Div(\theta_c||\theta_d)$.

KL-divergence between comment and average comment: Lu et al. [83] proposed conformity features in which the comment c is compared with other comments by looking at the KL-divergence between the unigram model of the comment c, and unigram model of an "average" comment that contains the text of all comments for an article. We did a preliminary analysis to study the impact of conformity on the quality. We found that this KL-divergence score has low correlation with comment quality on our data sets and therefore we do not consider it in the rest of this work.

5.4 Logistic Regression

So far we have introduced six features, which are summarized in Table 5.3.

Feature Set	**Description**
F_1	Comment length
F_2	Comment likelihood
F_3	Average number of verbs
F_4	Number of discourse relations
F_5	KL-divergence score using nouns only
F_6	KL-divergence score using all words

Table 5.3: Full feature set for comment representation

Once features are defined, we can use a classification algorithm to learn a model from the training data and apply the model to unseen data for thoughtful comment prediction. To classify the comments, we need a classifier that can combine multiple features and we chose logistic regression as it is capable of not only combining features but also map the scores to a probability distribution.

As we have pointed out earlier, a comment c together with its target document d can be represented by a feature vector $\mathbf{x}$. A logistic regression classifier models the probability of observing a discrete label y for a given $\mathbf{x}$ as follows:

$$p(y|\mathbf{x};\mathbf{w}) = \frac{1}{Z(\mathbf{x},\mathbf{w})}\exp(\mathbf{w}_y^T\mathbf{x}), \tag{5.4}$$

where

$$Z(\mathbf{x},\mathbf{w}) = \sum_{y\in\mathcal{Y}}\exp(\mathbf{w}_y^T\mathbf{x}).$$

Here $\mathbf{w}$ is a weight matrix and $\mathbf{w}_y$ is the weight vector corresponding to class y, and $\mathcal{Y}$ is the set of class labels.

Given training data $\{\mathbf{x}_i, y_i\}_{i=1}^N$, we learn a weight matrix by minimizing the following objective function:

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}}\left[\lambda\|\mathbf{w}\|^2 - \frac{1}{N}\sum_{i=1}^{N}\log p(y_i|\mathbf{x}_i;\mathbf{w})\right], \tag{5.5}$$

where $\|\mathbf{w}\|^2 = \sum_{y\in\mathcal{Y}}\|\mathbf{w}_y\|^2$ and λ is a regularization parameter that is empirically set.

5.5 Dataset

5.5.1 Data Collection

Our objective is to study how the thoughtfulness of a comment is reflected in the various linguistic factors including discourse relations and relevance factors to the article. As we have mentioned earlier, measuring the thoughtfulness of a comment is especially important for sociopolitical opinion mining. We therefore collected two data sets in this domain for our evaluation. We first acquired the following two political speeches:

1. Article 1 : Singapore PM's National Day Rally Speech. [4]

[4]http://www.pmo.gov.sg/

2. Article 2 : US Presidents' State Union Address. [5]

We further manually broke down each speech into several segments based on topical boundaries. For PM's speech, sub-headings are embedded in political speeches and when such data is unavailable (Obama's speech), the speech is well organized with clear boundaries to aid the segmentation. The topics of each speech are listed in Table 5.4.

Speech	Topics
Singapore	Economy, Immigration, Congestion, Housing, Education, National Service (NS), Singapore Spirit, Founding Fathers, Youth Olympics
US	Economy, Innovation & Research, Education, Rebuild, Taxes, Debts, Military

Table 5.4: Topics in the two speeches.

To collect an unbiased sample of comments for each speech, we use two search queries ("national day rally speech 2010" and "president state union address 2011") and Google API to obtain a list of top 50 URLs. We further manually selected URLs from online forums and blogs. We cleaned the data by removing short comments with no more than two words.

For F_2, the lexical feature, we need a suitable reference corpus. For the Singapore data set, we collected 1200 news articles from AisaOne.com to form our reference corpus. For the US data set, we used a set of 1358 New York Times articles to form the reference corpus.

We show some sample comments in Table 5.5 for both datasets.

5.5.2 Annotation

We engaged two human annotators to judge the comments we had collected. Our annotators for PM's speech are Singaporeans who are familiar with the economy and the social issues of Singapore. For Obama's speech, we have two annotators,

[5] http://www.whitehouse.gov/

Topic	Quality	User Comment
Innovation	Thoughtless	I love these plans on energy, but alas, the energy secretary appears to be asleep.
	Thoughtful	You want to really drive innovation, job growth and entrepreneurs? Make education, health care and retirement less of a burden on the average family, adopt more socialist policies like Norway
Taxes	Thoughtless	..Oh, so Obama ”compromised” on the tax cuts for the wealthy
	Thoughtful	Low taxes aren't helping the vast middle and working class and aren't creating more jobs, it's a policy that only benefits the rich.
Housing	Thoughtless	By the way did anybody count the no of flags on a HDB flat. believe me 95% of the time u will take less than 10 sec to do it
	Thoughtful	I am glad that to hear more HDB houses to be built. But do I got a taste of this pie? What about those who are genuine to upgrade their existing 3 room flat but not 1st timer?..
National Service	Thoughtless	i'm still waiting before the budget and erection. otherwise i'll vote oppo. 9k is ?
	Thoughtful	Just 9000 for NSman. Those foreign scholar in NUS NTU got tution non-subsidize fee alone is 20000 one year. That even exclude lodging and return ticket fully paid by PAP.

Table 5.5: Sample comments: First two topics are from US and last two are from Singapore

an American and an immigrant who are also familiar with American economy and social issues. All the judges are above 25 and are working professionals. The annotators were asked to judge (1) whether a comment was relevant to each segment of its corresponding speech, and if so, (2) whether the comment was a thoughtful one. In other words, we treat each segment of a speech as a target document. For each pair of a comment c and a target document (i.e., a speech segment) d, we obtained two binary labels: a label z that indicates whether c is relevant to d, and a label y that indicates whether c is a thoughtful comment with respect to d.

To judge whether a comment was thoughtful, the annotators were asked to use the following criteria:

1. Is the comment a mere repetition or a rephrase of the speech text? For example, “PM says that we should stay open for the foreigners” is a repetition of the text from the article in passive voice. Such comments are relevant to the article, but not insightful.

2. Does the comment contain opinions of the commenter? For example, “Eliminating the deficit-Im sure this makes Mitch happy” is about the topic “Spending and Taxes” but without any insightful opinion. Such comments are relevant but not insightful.

3. Does the commenter provide argument to support her opinion? For example, “All this deficit crap reminds me of when Reagan ran for president ; how the deficit was terrible etc, etc, and after he got elected he ran up the biggest deficit ever. This was mostly due to spending on the military and tax cuts for the rich. This was even after he slashed domestic spending. If the US would wake up to the fact that we can’t afford the wars, we might be able to move forward.” Such comments are relevant as well as insightful.

In total, the human judges annotated 1350 pairs of issue-comments for PM speech and 1150 pairs of issue-comments for Obama Speech. Since the annotation is still subjective, we calculated the inter-annotator agreement level using Cohen’s Kappa coefficient. Cohen’s kappa on quality is 0.8965 for all comments. On relevance the kappa is 0.7355 for all comments. We use the judgment from the judge who is stricter as our ground truth. The statistics of the labeled data are shown in Table 5.6. Interested readers kindly refer to Appendix A for more details on annotation process and sample examples.

Comment type	**Singapore**	**US**
Thoughtless	63.35%	68.25%
Thoughtful	36.65%	31.75%

Table 5.6: Comment statistics for both articles

5.6 Experiments

To check whether the features we have defined correlate with the thoughtfulness of comments based on human judgement, we first compute the Pearson correlation coefficients for all the features summarized in Table 5.3. The results are shown in

Table 5.7. We observe that all features are positively correlated with the thoughtfulness of comments.

Feature	Singapore	US
F_1	0.3744	0.3594
F_2	0.3782	0.3755
F_3	0.3639	0.3911
F_4	0.3913	0.3554
F_5	0.1606	0.2437
F_6	0.1191	0.2146

Table 5.7: Pearson correlation coefficients between the features and the thoughtfulness of comments.

In the remaining of this section, we show our experimental results that answer the following questions:

RQ1: Does the KL-divergence relevance score based on nouns work better than the KL-divergence score based on all words?

RQ2: Which discourse relations have bigger impact on the performance?

RQ3: Which combination of various features gives the best prediction of thoughtfulness?

For all the experiments below, we use the standard precision, recall and F-score as our performance measures.

5.6.1 Relevance Model

To answer our RQ1, we first tested the performance on finding relevant comments on the Singapore dataset for both KL methods discussed in Section 5.3.5. For this evaluation, we used only the labels z from the human judgment, i.e. the relevance judgment. We tested both relevance models: KL-divergence using all words and KL-divergence using only nouns. The results are shown in Table 5.8. We used the F-measure to evaluate the results. If score is greater than $\tau >$ -2.2 (set empirically), the comment is relevant to the topic. We observe that using nouns to compute the KL-divergence score works better. So, for the succeeding experiments we use F_5 which is the feature based on KL-divergence score between a comment and a target speech segment using nouns only.

Feature	Model	F-1
F_5	KL-divergence using nouns only	**0.634**
F_6	KL-divergence using all words	0.611

Table 5.8: Comparison between the two KL-divergence scores on the Singapore dataset.

5.6.2 Discourse Relations

To answer our RQ2, we studied the influence of various discourse relations on the F-measure of the comment thoughtfulness using the logistic regression model. For this evaluation, we used the labels y from the human judgment, i.e. the thoughtful comment. Table 5.9 shows the comparison of all three classes of discourse relations (Comparison, Expansion and Contingency) on comment quality. We notice that for both data sets, when *comparison* relations are used, the accuracy is the highest for both data sets. For the subsequent experiments, we use only the *comparison* relations to form our discourse feature, i.e. F_4 is set to be the number of *comparison* relations in a comment.

DR-Level	Singapore	US
All	0.6186	0.6464
Comparison	**0.6313**	**0.6538**
Expansion	0.5824	0.6111
Contingency	0.6213	0.6309

Table 5.9: Comparison of different classes of discourse relations using F-measure.

5.6.3 Thoughtful Comment Study

To answer RQ3, we conducted a detailed analysis on all the feature combinations we summarized in Table 5.3. We tested the thoughtfulness of the comments for a given article using the logistic regression model. The results of our experiments are shown in Table 5.10 for Singapore and in Table 5.11 for US. In both the tables, the feature combination with the best performance is shown with the bolded scores on the evaluation measures. For all our experiments we performed 5-fold cross validation and with all the combinations of the features. For better analysis, we show only the most important combinations in the results.

Feature Set	**Recall**	**Precision**	**F-1**
F_1+F_2+F_3	0.7097	0.7586	0.7333
F_1+F_2+F_4	0.8065	0.7143	0.7576
F_1+F_3+F_4	0.8387	0.6667	0.7429
F_2+F_3+F_4	0.8065	0.6944	0.7463
F_1+F_2+F_3+F_4	0.7742	0.7273	0.7500
F_1+F_2+F_3+F_5	0.7419	0.7667	0.7541
F_1+F_2+F_4+F_5	0.8065	0.7813	**0.7937**
F_1+F_3+F_4+F_4	0.7419	0.7931	0.7667
F_2+F_3+F_4+F_5	0.7742	0.7500	0.7619
F_1+F_2+F_3+F_4+F_5	0.7742	0.8000	0.7869

Table 5.10: Prediction results of thoughtful comments for Singapore using various feature combinations.

For the Singapore data set, using linguistic features alone (F_1+F_2+F_3) leads to a F-score of 73.33%. Our hypothesis is that discourse relations play important role in detecting thoughtful comments. (F_1+F_2+F_3+F_4) is a standard baseline for predicting quality [102]. The results confirm that using discourse relations together with linguistic features yields (F_1+F_2+F_3+F_4) a 75% F-score. This due to the language aspects of the comment where the users' justification is expressed by the phrases from the Table 5.2. However, the model performs slightly better without syntactic features (F_1+F_2+F_4)) with 75.76%, which is a 0.76% increase over combined features and 2.43% higher than the linguistic features. In our analysis, we observed that grammatical correctness lacks in the thoughtful comments and this explains why syntactic features do not play major role in the classification task. Our second hypothesis is that relevance factors play an important role in detecting thoughtful comments. The results confirm that using relevance scores together with linguistic features and discourse relations (F_1+F_2+F_3+F_4+F_5) leads to 78.69% F-score, which is a 3.69% increase compared to linguistic together with discourse relations. Here again, we notice that the model has better performance without syntactic features (F_1+F_2+F_4+F_5) with F-score of 79.37%, which is a 4.37% increase compared to linguistic together with discourse relations and 6.04% higher than linguistic features alone.

Feature Set	Recall	Precision	F-1
F_1+F_2+F_3	0.6522	0.6818	0.6667
F_1+F_2+F_4	0.7826	0.6429	0.7059
F_1+F_3+F_4	0.7391	0.6296	0.6800
F_2+F_3+F_4	0.7391	0.5862	0.6538
F_1+F_2+F_3+F_4	0.7826	0.6207	0.6923
F_1+F_2+F_3+F_5	0.7391	0.6296	0.6800
F_1+F_2+F_4+F_5	0.7826	0.6923	**0.7347**
F_1+F_3+F_4+F_4	0.7826	0.6429	0.7059
F_2+F_3+F_4+F_5	0.7391	0.6538	0.6939
F_1+F_2+F_3+F_4+F_5	0.7826	0.6667	0.7200

Table 5.11: Prediction results of thoughtful comments for US using various feature combinations.

For experiments on US data set, using linguistic features (F_1+F_2+F_3) alone leads to the F-measure of 66.67%. Discourse relations together with linguistic features (F_1+F_2+F_3+F_4) yields 69.23% F-measure which is a 2.56% increase over features without discourse relations. Here, we also notice that the model performs slightly better without syntactic features (F_1+F_2+F_4) with F-score of 70.59% which is 1.36% increase over combined features. Using relevance scores together with linguistic features and discourse relations (F_1+F_2+F_3+F_4+F_5) leads to 72% F-measure which is 2.77% increase compared to linguistic together with discourse relations. We also notice that the model has better performance with out syntactic features (F_1+F_2+F_4+F_5) with F-score of 73.47, which is 4.24% increase compared to linguistic together with discourse relations - answers our RQ3.

During our analysis, we observed that the US data is less verbose compare to Singapore data. Another thing we also noticed is that the US data users focus more on the speech delivery rather than the actual speech issues. At the same time, they tend to discuss mostly one issue in each comment where as, Singapore users tend to combine many issues in their comments. So, even if one issue is justified in the comment, the comment is treated as thoughtful comment. This explains the performance differences in the two data sets. It will be interesting to study more fine grained opinion analysis at the comment level and we leave it for future work.

5.6.4 Cross Collection Experiments

We further perform cross data collection experiments to test the performance of our model. We tested Singapore using USs' 5-feature model and vice versa. To compare the cross data collection results with original results, we depict Table 5.12 which shows the F-measure prediction on thoughtful comments. Singapore performed with a quality prediction F-measure lowered by 4.49%, whereas the US performance decreased by 5.11% compared to actual model.

		train	
		Singapore	**US**
test	**Singapore**	0.7937	0.7488
	US	0.6836	0.7347

Table 5.12: Cross data collection comparison. F-Measure for thoughtful comments.

5.6.5 Parameter Sensitivity

The regularization parameter λ in Equation (5.5) is set empirically. We study the optimal value and tuned it by regular cross validation. Figure 5.1 shows our experiments for both Singapore and US datasets. We get optimum results when we set λ to 0.1 or 1. We choose 0.1 for both datasets as it generates higher prediction performance in general.

5.7 Discussion

Features used in our study are very fundamental and the results can be improved with better feature engineering. In particular the syntactic (verbs) and discourse (explicit) features used in our evaluation can be further studied to improve the results. In case of syntactic features, though our preliminary results show low corre-

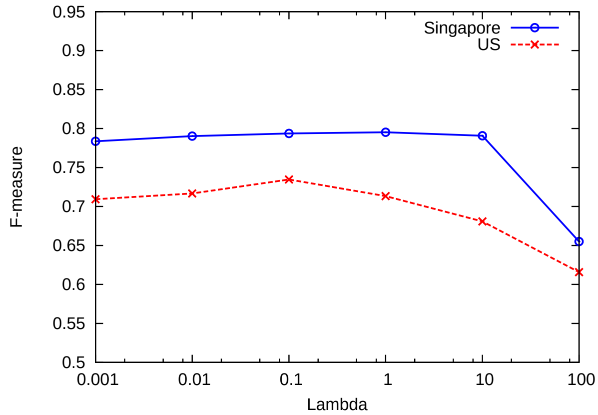

Figure 5.1: Regularization parameter sensitivity study

lations for other POS features, studying them in detail with various combinations is an interesting complimentary study to our problem. Similarly, implicit discourse features also play an important role in discovering the thoughtful comments. Such extensive feature engineering aids in improving the results and we leave it to the future studies. In this chapter, we focus on motivating a new problem in sociopolitical domain and solve it with a principled approach.

Results from our cross collection experiments are not surprising as it is normal for the model to perform with a lower accuracy when trained on different dataset and tested on the other dataset. We believe that the accuracy can be improved by considering a tweak in lexical features. Currently, we use only specific reference corpus for Singapore and US respectively. Combining the reference corpus may aid in improving the cross collection experiments. Currently, we use KL divergence to compute the similarity but users tend to use abbreviations for some words and this impacts the performance of KL-divergence scores. It is interesting to implement other similarity techniques based on topic modeling and enhance the relevance performance. Extending the problem to identify the sentiment orientation is another useful subtask of opinion mining which is another future direction. In the future, we would like to extend our work to application base, and investigate the usage of thoughtful opinions in opinion summarization.

5.8 Related Work

Our work is related to a large body of literature on measuring text quality in NLP, but our problem has some essential differences. The main difference is that in traditional sense, high-quality text should be grammatical, coherent and readable. For online comments, we focus more on the insightfulness or thoughtfulness of comments.

Opinion Mining: Our work is also related to opinion retrieval [140, 63, 86], which aims at automatically finding attitudes or opinions about specific targets, such as named entities, consumer products or public events. In most existing work on opinion retrieval, only relevance and subjectivity are considered, whereas we propose that quality in terms of thoughtfulness is also an important factor.

Product Reviews: Many recent studies examined the challenges on the quality of comments. [71] studied how to predict the helpfulness of product reviews. They found that a helpful review should describe the features of the products and the pros/cons of the features. A more elaborative review that provides the complete details of the product is more likely to be considered high quality. Our problem is more general and the comments are not necessarily about products. Moreover, comments on sociopolitical articles need not elaborate on all the issues in the article. Therefore products and their features are not relevant to our problem. Another study by [47] on review helpfulness looked into factors related to the reviewer, such as reviewer characteristics and reviewer history. In our work, we focus on features observed from the text only. Other social factors such as a commenter's profile or past behavior are complementary to our method.

High quality content study is also important for question answering services. [4] proposed a classification framework of estimating answer quality. They studied content-based features and usage-based features derived from question answering communities.

Social Context: Work on measuring quality of social media content considers not only the quality of the content itself but also its authority in the social network through the author's authority, its popularity, etc. [59]. Several researchers explored the social network together with the content of the reviews to predict the review quality. [17] proposed a mutual reinforcement learning framework to simultaneously predict content quality and user reputation, whereas [83] proposed a linear regression model with various social contexts for review quality prediction. They combined textual and social context information to evaluate the quality of individual reviewers and to assess the quality of the reviews. We do not consider these factors as we want to focus on textual cues first. These additional features can be factored in as an independent step. Similar line of work can be seen by [27, 81, 17].

Our work is similar to [8] where they introduced argument analysis together with opinion. In their task, properties of a person or product (honesty, rigor, friendliness, etc.) are treated as arguments. The task is oriented towards aggregating features related to the product and supporting arguments to detect polarity. The task we address in this paper is quite different from their work in two main aspects. Firstly, for sociopolitical issues, the policy makers look for insightful reasoning text to understand the public sentiment in which case, properties are insufficient. Secondly, we study the attentiveness of the comments but not the polarity. Polarity orientation is a separate task which can be studied individually.

5.9 Chapter Summary

In this chapter, we studied the problem of extracting attentive comments that provide some reasoning which is highly valuable in understanding the user's opinion in sociopolitical opinion mining. We perform an empirical study using syntactic, vocabulary, discourse and relevance features for prediction and combination of all is substantially better than the baseline surface features. Moreover, through our cross data collection experiments, we show that prediction using our approach achieves competitive performance.

Chapter 6

Ideological Positions and Expressions Extraction

In the previous three chapters we studied two major tasks of opinion mining in sociopolitical domain; Extraction and Quality. In this chapter, we study the problem of controversy (contrastive opinions) in social comments, a complimentary task of opinion mining. Debates on controversial sociopolitical issues provide room for the netizens to take positions and post opinions towards those issues. In this chapter, we study a problem of extracting the stances and ideological expressions of users from their comments on ideological debates related to sociopolitical domain. We propose an unsupervised approach based on latent variable methods for identifying and extracting the positional words and entities associated with the issues. For this study, we model issues, positional words/ideological expressions, entities and the stances of the users in a principled way using topic models. For our evaluation, we use arguments in debates from Debatepedia[1] as they provide a platform for users to express their positional opinions on the issues debated.

[1] `http://dbp.idebate.org`

6.1 Introduction

Debatepedia is an online, community-authored encyclopedia of debates. Many of these debates are centered around contemporary political and social issues, and they link together various arguments expressed by different people in a structured and organized manner. Users take positions on these debates with an underlying ideological beliefs. Mining stances and the ideological expressions is an interesting task that can be useful for political domain experts in campaigning, advertising and recommending systems. To achieve this goal, we seek to discover a low-dimensional, human-interpretable representation of the space of sociopolitical debates. To this end, we conduct a model-driven exploration of Debatepedia, an online, community-authored encyclopedia of debates. Debatepedia's debates focus on sociopolitical issues, and they link together various arguments expressed by different people in a structured and organized manner. In the domain of political discourse, one frequently finds opinionated texts such as editorials and blog posts, the interpretation of which may be best carried out in the context of major debates of the day. Debatepedia gives structure to the debate space, organizing 1,300 debates with questions and more than 33,000 arguments into sides ("yes" or "no"). These arguments link externally to opinionated texts.

We draw inspiration from Lin et al [77] and Ahmed et al [6], who used generative models to infer **topics**—distributions over words—and other word-associated variables representing perspectives or ideologies. We view topics as lexicons, and propose that grounding a topic model with evidence beyond bags of words can lead to more lexicon-like representations. Specifically, our generative topic model grounds topics using the hierarchical organization of arguments within Debatepedia. Further, we use named entity recognition as a preprocessing step, an existing sentiment lexicon to construct an informed prior, and we incorporate a latent, discrete **position** variable that cuts across debates.[2] There are no existing solutions that

[2]This variable might serve to cluster debate sides according to "abstract beliefs commonly shared by a group of people," sometimes called *ideologies* [127].

model issues, positional words/ideological expressions, entities and the stances of the users. The resultant model can be used for various tasks such as stance detection, article-argument attachment. In our solution we modeled all the components of a debate in a principled way using topic models.

The rest of the chapter is organized as follows. In Section 6.2, we describe the dataset, its structure and the properties of the debates. We present our solution model in Section 6.3. Finally, we present our experiments and results in Section 6.4. We present our discussions in Section 6.5 and finally we conclude with some related work analysis in Section 6.6.

6.2 Data

Debatepedia, like Wikipedia, is constructed by volunteer contributors and has a system of community moderation. Many of the debate issues covered are controversial and salient in current public discourse. Because it is primarily expressed as text, it is a *corpus* of debate topics, but it is organized hierarchically, with multiple issues in each debate topic, questions within each issue, and arguments on two sides of each question. An important element of the corpus is the widespread quotation and linking to external articles on the Web, including news stories, blog postings, wiki pages, and social media forums; here we use these external articles in evaluation.

Table 6.1 shows excerpts from a debate page[3] from Debatepedia. Each debate contains "questions," which reflect the different aspects of a debate. In this particular debate, there are 13 questions (2 shown), ranging from economic benefits to enforceability to social impacts. For each question, there are two distinct sides, each with its own set of supporting arguments. Many of these arguments also contains links to online articles where the quotes are extracted from (articles are not shown in Table 6.1). For example, in the second argument on the "No" side, there is an

[3] http://dbp.idebate.org/en/index.php/Debate:_Gun_control

<table>
<tr><td colspan="2">Debate: Gun control; should laws be passed to limit gun ownership further?</td></tr>
<tr><td colspan="2">Question: Self-defense – Is self-defense a good reason for gun ownership?</td></tr>
<tr><td>Side: Yes</td><td>Side: No</td></tr>
<tr><td>Argument: A citizen has a “right” to guns as a means to self-defense: Many groups argue that a citizen should have the “right” to defend themselves, and that a gun is frequently the …</td><td>Argument: The protection of property is not a good justification for yielding a lethal weapon. While people have a right to their property, this should not justify wielding a lethal …</td></tr>
<tr><td>Argument: Gun restrictions and bans disadvantage citizens against armed criminals. Citizens that are not allowed to carry guns are disadvantaged against lawless criminals that …</td><td>Argument: Robert F. Drinan, Former Democratic US Congressman, “Gun Control: The Good Outweighs the Evil”, 1976 – “These graphic examples of individual instances of …</td></tr>
<tr><td colspan="2">Question: Economic benefits – Is gun control economically beneficial?</td></tr>
<tr><td>Side: Yes</td><td>Side: No</td></tr>
<tr><td>Argument: Lax gun control laws are economically costly. The Coalition for Gun Control claims that, “in Canada, the costs of firearms death and injury alone have been estimated at …</td><td>Argument: Gun sports have economic benefits. Field sports bring money into poor rural economies and provide a motivation for landowners to value environmental protection.</td></tr>
</table>

Table 6.1: An example of a Debatepedia debate on the topic “Gun control.”

inline link to the article written by Congressman Drinan.[4]

Within a debate topic, the sides cut across different questions, aligning arguments together. In general, the questions are phrased so that a consistent “pro” and “con” structure is apparent throughout each debate, aligned to a high-level question (i.e., the “Yes” sides of all the questions are consistent with the same side of the larger debate). The example of Table 6.1 deviates from this pattern, with the self-defense “Yes” arguing “no” to the high-level debate question—*Should laws be passed to limit gun ownership further?*—and the economic “Yes” arguing “yes” to the high-level question. Table 6.2 presents statistics of our corpus.

Debates	1,303
Arguments	33,556
Articles linked by exactly one argument	3,352
Tokens	1,710,814
Types (excluding NE mentions)	59,601
Person named entity mentions	9,496

Table 6.2: Debatepedia corpus statistics. Types and tokens include unigrams, bigrams and person named entities.

[4] `http://www.saf.org/LawReviews/Drinan1.html`

6.2.1 Preprocessing

We scraped the Debatepedia website and extracted the debate, question, argument and side structure of the debate topics. We crawled the external Web articles that were linked from the Debatepedia arguments. For the Web articles, we extracted the main text content (ignoring boilerplate elements such as navigation and advertisments) using Boilerpipe [72].[5] We tokenized the text and filtered stopwords.[6] We considered both unigrams and bigrams in our model, keeping all unigrams and removing bigram types that appeared less than 5 times in the corpus. Although our modeling approach ultimately treats texts as bags of terms (unigrams and bigrams), one important preprocessing step was taken to further improve the interpretability of the inferred representation: recognizing named entity mentions of persons. We identified these mentions of persons using Stanford NER [42] and treated each person mention as a single token.

6.3 Model

Our model defines a probability distribution over terms[7] that are observed in the corpus. Each term occurs in a context defined by the tuple $\langle d, q, s, a \rangle$ (respectively, a *debate*, a *question* within the debate, a *side* within the debate, and an *argument*). At each level of the hierarchy is a different latent variable:

- Each question q within debate d is associated with a distribution over topics, denoted $\boldsymbol{\theta}_{d,q}$.
- Each side s of the debate d is associated with a position, denoted $i_{d,s}$ and we posit a global distribution ι that cuts across different questions and arguments. In our experiments, there are two positions, and the two sides of a debate are constrained to associate with opposing positions. As illustrated by Table 6.1,

[5] `http://code.google.com/p/boilerpipe`
[6] `www.ranks.nl/resources/stopwords.html`
[7] Recall that our model includes bigrams. We treat each unigram and bigram token (after filtering discussed in Section 6.2.1) as a separate term.

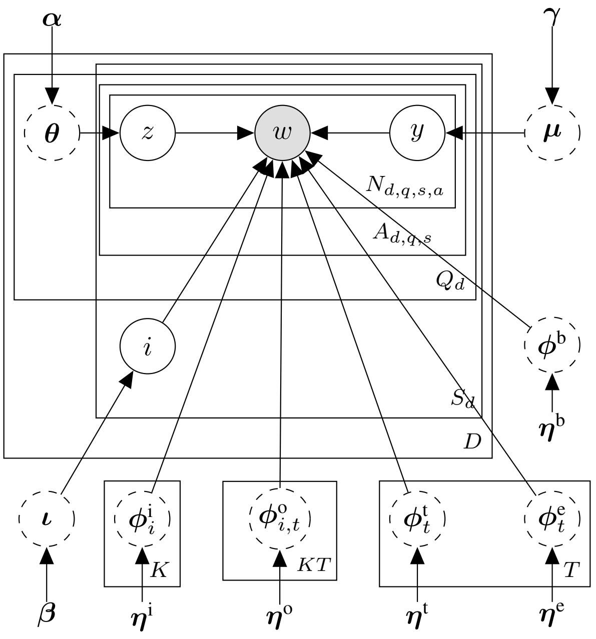

Figure 6.1: Plate diagram. K is the number of positions, and T is number of topics. The shaded variables are observed and dashed variables are marginalized. $\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}$ and all $\boldsymbol{\eta}$ are fixed hyperparameters (§6.3.1).

this assumption is not always correct, though it tends to hold most of the time.

- Each term $w_{d,q,s,a,n}$ (n is the position index of the term within an argument) is associated with one of five *functional term types*, denoted $y_{d,q,s,a,n}$. This variable is latent, except when it takes the value "entity" (e) for terms marked as named entity mentions. When it is not an entity, it takes one of the other four values: "general position" (i), "topic-specific position" (o), "topic" (t), or "background" (b). Thus, every term w is drawn from one of these 5 types of bags, and y acts as a switching variable to select the type of bag.
- For some term types (the ones where $y \in \{\mathsf{o}, \mathsf{t}\}$), each term $w_{d,q,s,a,n}$ is associated with one of T discrete topics, as indexed by $z_{d,q,s,a,n}$.

Figure 6.1 illustrates the plate diagram for the graphical model underlying our approach. The generative story is given in Figure 6.2.

1. $\forall$ topics t, draw topic-term distribution $\boldsymbol{\phi}^{\mathsf{t}}_t \sim \mathrm{Dirichlet}(\boldsymbol{\eta}^{\mathsf{t}})$ and topic-entity distribution $\boldsymbol{\phi}^{\mathsf{e}}_t \sim \mathrm{Dirichlet}(\boldsymbol{\eta}^{\mathsf{e}})$.
2. $\forall$ positions i, draw position-term distribution $\boldsymbol{\phi}^{\mathsf{i}}_i \sim \mathrm{Dirichlet}(\boldsymbol{\eta}^{\mathsf{i}})$.
3. $\forall$ topics t, $\forall$ positions i, draw topic-position term distribution $\boldsymbol{\phi}^{\mathsf{o}}_{i,t} \sim \mathrm{Dirichlet}(\boldsymbol{\eta}^{\mathsf{o}})$.
4. Draw background term distribution $\boldsymbol{\phi}^{\mathsf{b}} \sim \mathrm{Dirichlet}(\boldsymbol{\eta}^{\mathsf{b}})$.
5. Draw functional term type distribution $\boldsymbol{\mu} \sim \mathrm{Dirichlet}(\boldsymbol{\gamma})$.
6. Draw position distribution $\boldsymbol{\iota} \sim \mathrm{Dirichlet}(\boldsymbol{\beta})$.
7. $\forall$ debates d:
   a. Draw $i_{d,1}, i_{d,2} \sim \mathrm{Multinomial}(\boldsymbol{\iota})$, assigning each of the two sides to a position.
   b. $\forall$ questions q in d:
      i. Draw topic mixture proportions $\boldsymbol{\theta}_{d,q} \sim \mathrm{Dirichlet}(\boldsymbol{\alpha})$.
      ii. $\forall$ arguments a under question q and term positions n in a:
         A. Draw topic label $z_{d,q,s,a} \sim \mathrm{Multinomial}(\boldsymbol{\theta}_{d,q})$.
         B. Draw functional term type $y_{d,q,s,a} \sim \mathrm{Multinomial}(\boldsymbol{\mu})$.
         C. Draw term $w_{d,q,s,a} \sim \mathrm{Multinomial}\left(\boldsymbol{\phi}^{y_{d,q,s,a}} \mid i_{d,1}, i_{d,2}, z_{d,q,s,a}\right)$.

Figure 6.2: Generative story for our model of Debatepedia.

6.3.1 Priors

Typical probabilistic topic models assume a symmetric Dirichlet prior over its term distributions or apply empirical Bayesian techniques to estimate the hyperparameters. Motivated by past efforts to exploit prior knowledge [141, 76], we use the OpinionFinder sentiment lexicon [134][8] to construct $\boldsymbol{\eta}^{\mathsf{i}}$ and $\boldsymbol{\eta}^{\mathsf{o}}$. Specifically, terms w in the lexicon were given parameters $\eta^{\mathsf{i}}_w = \eta^{\mathsf{o}}_w = 0.01$, and other terms were given $\eta^{\mathsf{i}}_w = \eta^{\mathsf{o}}_w = 0.001$, capturing our prior belief that opinion-expressing terms are likely to be used in expressing positions. 5,451 types were given a "boost" through this prior.

Information retrieval has long exploited the observation that a term's document frequency (i.e., the number of documents a term occurs in) is inversely related its usefulness in retrieval [67]. We encode this in $\boldsymbol{\eta}^{\mathsf{b}}$, the prior over the background term distribution, by setting each value to the logarithm of the term's argument frequency. The other priors were set to be symmetric: $\eta^{\mathsf{e}} = 0.01$ (entity topics), $\eta^{\mathsf{t}} = 0.001$ (topics), $\alpha = 50/T = 1.25$ (topic mixture coefficients), $\beta = 0.01$ (positions), and $\gamma = 0.01$ (functional term types). Preliminary tests showed that final topics are relatively insensitive to the values of the hyperparameters.

[8] http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/

6.3.2 Inference and Parameter Estimation

Exact inference of the posterior distribution of the model is intractable. Instead, we approximate it using a Monte Carlo Markov Chain method known as Gibbs sampling. As we used conjugate priors for our distributions, we can easily integrate out the dotted variables in Figure 6.1. To sample $i_{d,s}$ for each debate d, side s, we need to consider those position words and general position words inside. The notable deviations from typical uses of collapsed Gibbs sampling are: (i) we jointly sample $i_{d,1}$ and $i_{d,2}$ to respect the constraint that they differ; and (ii) we fix the priors, in some cases to be asymmetric, as discussed in Section 6.3.1. We perform Gibbs sampling for 2,000 iterations over the dataset, discarding the first 500 iterations for burn-in, and averaging over every 10th iteration thereafter to get the MAP estimates for our term distributions. Kindly refer to Appendix B for the details of Gibbs sampling.

6.4 Experiments

We estimated our model on the Debatepedia debates (not including hyperlinked articles). We used T = 40 topics and K = 2 positions. Recall that the aim of this work is to infer a low-dimensional representation of debate text. We estimated our model on the Debatepedia debates (not including hyperlinked articles), and conducted several evaluations of the model, each considering a different aspect of the goal. We exploit external articles hyperlinked from Debatepedia described in Section 6.2 as supporting texts for arguments, treating each one's association to an argument as variable to be predicted. Firstly, we evaluate our model on the article associating task. Secondly, we evaluate our model on the position prediction task. The average length of all arguments is around 109 words and the external articles is around 1617 words. More specifically, the average number of words in arguments are 108 and 110 for positions, one and two respectively. Then, we compare our model's positional assignment of arguments to human annotated clusterings. Finally, we present qualitative discussion. As such, to encourage comparative work, we have made the resulting corpus and judgments available for download[9].

We reiterate that our aim is to use Debatepedia to infer a low-dimensional, interpretable representation of the domain, not to match an existing golds tandard annotation for a particular task. We therefore consider a suite of evaluations of different facets of the representation inferred by our method.

6.4.1 Quantitative Evaluation

Topics

As described in Section 6.2, our corpus includes 3,352 articles hyperlinked by Debatepedia arguments.[10] Our model can be used to infer the posterior over topics associated with such an article, and we compare that distribution to that of the Debatepedia article that links to it. Calculating the similarity of these distributions, we get an estimate of how closely our model can associate text related to a debate with the specific argument that linked to it. We compare with LDA [20], which ignores sentiment, and the joint sentiment topic (JST) model [76], an unsupervised model that jointly captures sentiment and topic.[11] KL-Divergence is one directional and hence in this study, we use Jensen-Shannon divergence to measure the similarity between two distributions. Using Jensen-Shannon divergence (JS), we find that our approach embeds these pairs significantly closer than LDA and JST (also trained with 40 topics), under a Wilcoxon signed rank test ($p < 0.001$). Figure 6.3 shows the histogram of divergences between our model, JST, and LDA.

Associating external articles. More challenging, of course, is *selecting* the argument to which an external article should be associated. We used the Jensen-Shannon

[9] https://sites.google.com/site/swapnagotipati/datasets

[10] We consider only those articles linked by a single Debatepedia argument.

[11] JST multiplies topics out by the set of sentiment labels, assigning each token to both a topic and a sentment. We use the OpinionFinder lexicon in JST's prior in the same way it is used in our model.

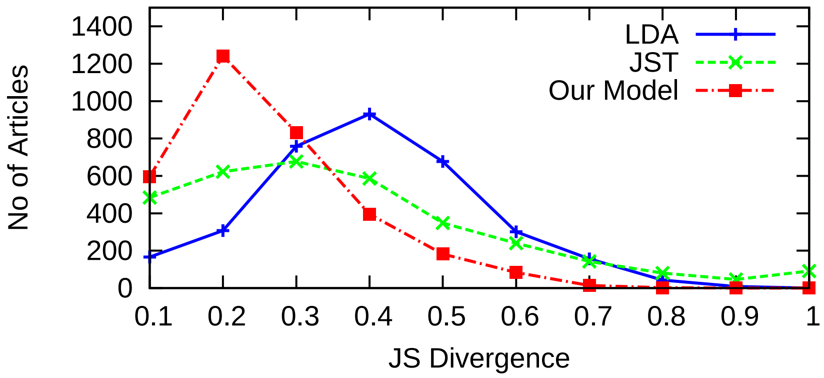

Figure 6.3: The distribution over Jensen-Shannon divergences between a hyperlinked article and the corresponding Debatepedia argument, $n = 3,352$.

divergence between topic distributions of articles and arguments to rank the latter, for each article. The mean reciprocal rank scores [128] for LDA, JST, and our model were 0.1272, 0.1421, and 0.1507, respectively; the difference is significant (Wilcoxon signed rank test, $p < 0.001$). We found the same pattern for MRR@k, $k \in \{5, 10, 15, 20, 25, \infty\}$, as shown in Figure 6.4.

It is likely possible to engineer more accurate models for attaching articles to arguments, but the attachment task is our aim only insofar as it contributes to an overall assessment of an inferred representation's quality.

Positions

Positional distance by topic. We next consider the JS divergences of position term distributions by topic; for each topic t, we consider the divergence between inferred values for $\phi^{\circ}_{1,t}$ and $\phi^{\circ}_{2,t}$. Figure 6.5 shows these measurements sorted from most to least different; these might be taken as evidence for which issue areas' arguments are more *lexically* distinguishable by side, perhaps indicating less common ground in discourse or (more speculatively) greater controversy. For example, our model suggests that debates relating to topics like presidential politics, foreign policy, teachers, abortion, religion, and Israel/Palestine are more heated (within the Debatepedia community at the time the debates took place) than those about the minimum wage, Iran as a nuclear threat, or immigration.

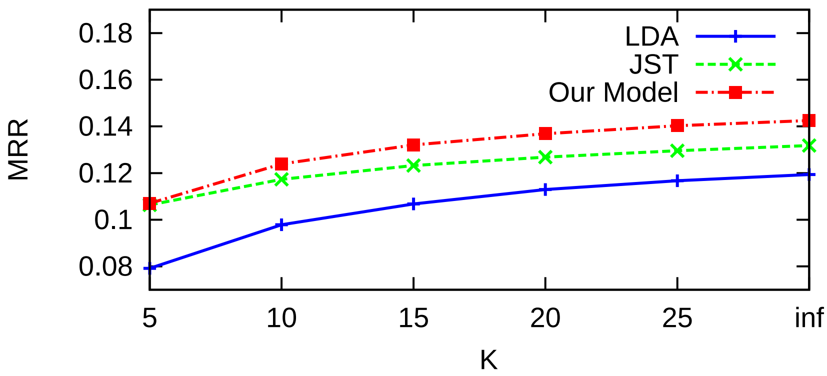

Figure 6.4: Mean reciprocal ranks for the association task.

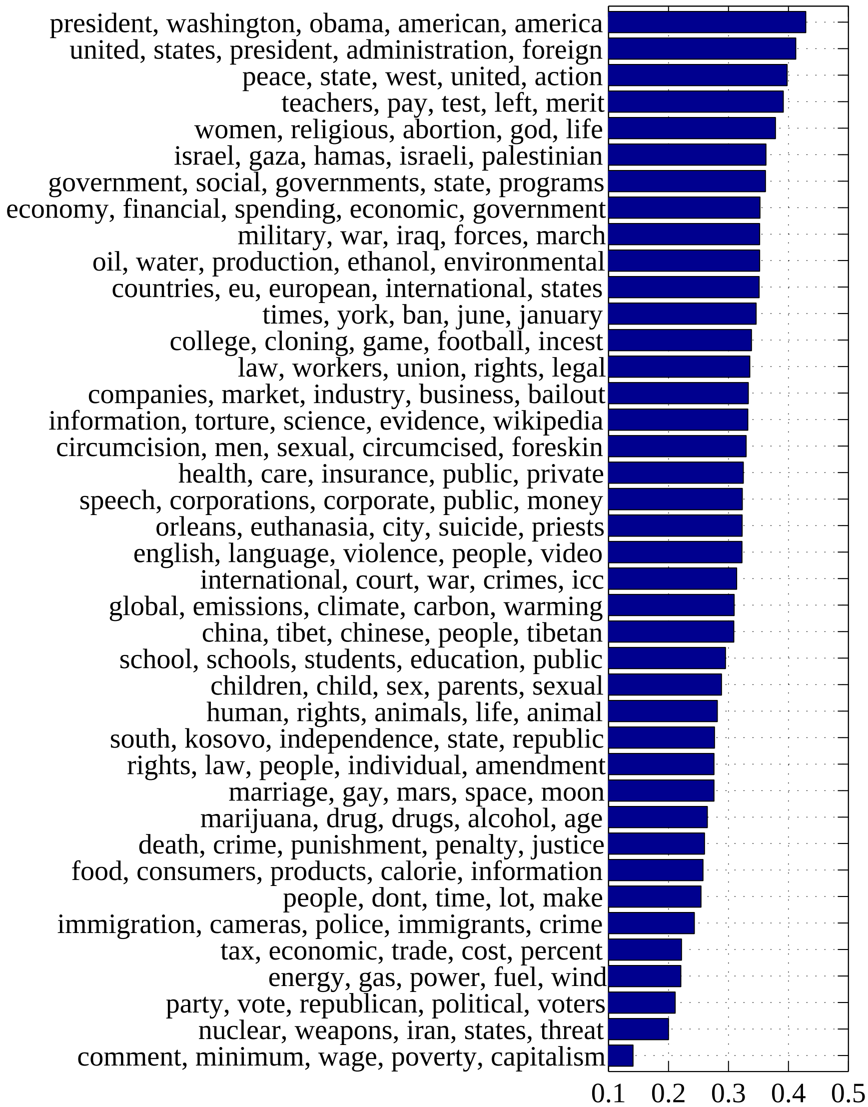

Figure 6.5: Jensen-Shannon divergences between topic-specific positional term distributions, for each topic. Topics are labeled by their most frequent terms from ϕ^t.

Predicting positions for arguments. We tested our model's ability to infer the positions of arguments. In this experiment (only), we held out 3,000 arguments during parameter estimation. The held-out arguments were selected so that every debate side maintained at least one argument whose inferred side could serve as the correct answer for the held-out argument. We then inferred i for each held-out argument from debate d and side s, given the parameters, and compared it with the value of $i_{d,s}$ inferred during parameter estimation. The model achieved 86% accuracy (Table 6.3 shows the confusion matrix). Note that JST does not provide a baseline for comparison, since it does not capture debate sides.

	$i = 1$	$i = 2$
$i^* = 1$	1,272	216
$i^* = 2$	199	1,313

Table 6.3: Confusion matrix for position prediction on held-out *arguments*.

Predicting positions for external articles. We can also use the model to predict the position adopted in an external text. For articles linked from within Debatepedia, we have a gold standard: from which side of a debate was it linked? After using the model to infer a position variable for such a text, we can check whether the inferred position variable matches that of the argument that links to it. Table 6.4 shows that our model does not successfully complete this task, assigning about 60% of both kinds of articles $i = 1$.

	$i = 1$	$i = 2$
$i^* = 1$	1,042	623
$i^* = 2$	1,043	644

Table 6.4: Confusion matrix for position prediction on hyperlinked *articles*.

Genre. We manually labeled 500 of these articles into six genre categories. We had two annotators for this task (Cohen's $\kappa = 0.856$). These categories, in increasing order of average Jensen-Shannon divergence, are: blogs, editorials, wiki pages, news, other, and government. Figure 6.6 shows the results. While the only

	A1 (11)	A2 (5)	A3 (16)
Model (2)	3.21	2.58	3.45
A1 (11)		2.15	2.15
A2 (5)			2.63

Table 6.5: Variation of information scores for each pairing of annotators and model.

difference between the first and last groups are surprising by chance, we are encouraged by our model's suggestion that blogs and editorials may be more "Debatepedia argument-like" than news and government articles.

Note that our model is learned only from text *within* Debatepedia; it does not observe the text of external linked articles. Future work might incorporate this text as additional evidence in order to capture effects on language stemming from the interaction of position and genre.

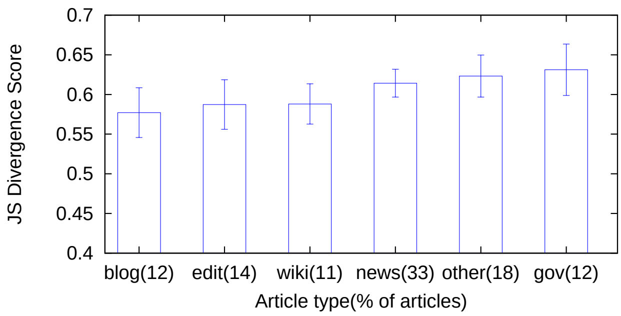

Figure 6.6: Position prediction on 500 hyperlinked *articles* by genre.

Comparison to Human Judgments of Positions

We compared our model's inferred positions to human judgments. For each of the 11 topics in Table 6.8, we selected two associated debates with more arguments than average (24.99). The debates were provided to each of three human annotators,[12] who were instructed to group the 44 sides of the debates. The instructions stated:

> Our goal is to see what you think about how the different sides of different debates can be lined up. You might find it convenient to think of these in terms of political philosophies, contemporary political party platforms, or something

[12] All were native English-speaking American graduate students not otherwise involved in this research. Each is known by the authors to have basic literacy with issues and debates in American politics.

	“Israel-Palestine”	“Same-sex marriage”	“Drugs”	“Healthcare”	“Death penalty”	“Abortion”
i_1	pre emptive israeli palestinian open and shut	same sex long term second class	hands free performance enhancing in depth	single payer so called self sustaining	anti death non violent african american	pro choice pro life non muslim
i_2	two state long term self destructive	opposite sex well intentioned day time	long term high speed short term	government run government approved high risk	semi automatic high profile hate crime	would be full time late term

a. Our model: topic-specific position bigrams associated with six selected topics.

–	war assault disproportionate	large possibility problems	illegal abuse high	support force threat	death penalty murder	power limit civil
+	peace independence self-determination	civil rights affirmative	disease nature potential	care universal uninsured	power clean waste	care suicide death

b. JST: sentiments associated with six selected topics manually aligned to our model’s topics.

Table 6.6: Terms associated with selected topics. The labels and alignments between the two models’ topics were assigned manually. (a.) Our model: topic-specific position bigrams which are ranked by comparing the log odds conditioned on the position and topic: $\log \phi^o_{i_1,t,w} - \log \phi^o_{i_2,t,w}$. We show the top three terms for each position (b.) JST: we show the top three terms for each sentiment (negative and positive).

> else. Any of these is fine; we want you to tell us the grouping you find most reasonable.

Kindly refer to Appendix A for more details on the annotation of debates and some examples. All three annotators (hereafter denoted A1, A2, and A3) used fairly involved labeling schemes; the annotators used 37, 30, and 16 unique labels, respectively.[13] A1 used keyword lists to label items; we coarsened his labels manually by removing or merging less common keywords (resulting in: *Republican*, *Democrat*, *science/environment*, *nanny*, *political reform*, *fiscal liberal*, *fiscal conservative*, *libertarian*, *Israel*, *Palestine*, and one unlabeled side). A2 provided a coarse annotation along with each fine-grained one (*liberal*, *conservative*, *?*, and two unlabeled sides). We used 100 samples from our Gibbs sampler to estimate posteriors for each $i_{d,s}$;

[13]In a small number of cases, an annotator declined to label a side. Each unlabeled item received its own cluster.

Topic	$i = 1$	$i = 2$
None (ϕ^i)	vice president, c sections, twenty four, cross pressures, pre dates, anti ballistic, cost effectiveness, anti landmine, court appointed, child poverty	cross examination, under runs, hand outs, half million, non christians, break down, counter argument, seventy five, co workers, run up
"Israel-Palestine"	pre emptive, israeli palestinian, open and shut, first time, hamas controlled, democratically elected	two state, long term, self destructive, secretary general, right wing, all out, near daily, short term
"Same-sex marriage"	same sex, long term, second class, blankenhorn rauch, wrong headed, self denial, left handed	opposite sex, well intentioned, day time, planet wide, day night, child rearing, low earth, one way, one third
"Drugs"	hands free, performance enhancing, in depth, hand held, best kept, non pharmaceutical, anti marijuana	long term, high speed, short term, peer reviewed, alcohol related, mind altering, inner city, long lasting
"Healthcare"	single payer, so called, self sustaining, public private, for profit, long run, high cost, multi payer	government run, government approved, high risk, two tier, government appointed, low cost, set up
"Death penalty"	anti death, non violent, african american, self help, cut and cover, heavy handed, dp equivalent	semi automatic, high profile, hate crime, assault weapons, military style, high dollar, self protective
"Abortion"	pro choice, pro life, non muslim, well educated, anti abortion, much needed, church state, birth control	would be, full time, late term, judeo christian, life style, day to day, non christian, child bearing

Table 6.7: General position (first row) and topic-specific position bigrams associated with six selected topics. Kindly refer to Appendix B for more results.

these were always 99% or more in agreement, so we mapped each debate side into its single most probable cluster. Recall that the two sides of each debate must be in different clusters.

Table 6.5 shows the variation of information measure [89] for each pairing among the three annotators and our model. The model agrees with A2's coarse clustering most closely, and in fact is closer to A2's clustering than A2 is to A3's; it also agrees with A2's coarse clustering better than A2's coarse and fine clusterings agree (3.36, not shown in the table). This is promising, but we do not have confidence that the positional dimension is being captured especially well in this model; for those debate-sides labeled *liberal* or *conservative* by A2, the best match of our two positions was still only in agreement only about 60% of the time, and agreement with each human annotator is within the interval of what would be expected if each debate's sides were assigned uniformly at random to positions.[14]

[14]This was determined using a Monte Carlo simulation with 1,000 samples.

Topic	Terms	Person entity mentions
"Israel-Palestine"	israel, gaza, hamas, israeli, palestinian	Benjamin Netanyahu, Al Jazeera, Mavi Marmara, Nicholas Kristoff, Steven R. David
"Same-sex marriage"	marriage, gay, mars, space, moon	Buzz Aldrin, Andrew Sullivan, Moon Base, Scott Bidstrup, Ted Olson
"Drugs"	marijuana, drug, drugs, alcohol, age	Four Loko, Evo Morales, Toni Meyer, Sean Flynn, Robert Hahn
"Healthcare"	health, care, insurance, public, private	Kent Conrad, Paul Hsieh, Paul Krugman, Ezra Klein, Jacob Hacker
"Death penalty"	death, crime, punishment, penalty, justice	Adam Bedau, Thomas R. Eddlem, Jeff Jacoby, John Baer, Peter Bronson
"Abortion"	women, religious, abortion, god, life	Ronald Reagan, John Paul II, Sara Malkani, Mother Teresa, Marcella Alsan

Table 6.8: For 6 selected topics (labels assigned manually), top terms (ϕ^{t}) and person entities (ϕ^{e}). Bigrams were included but did not rank in the top five for these topics. The model has conflated debates relating to same-sex marriage with the space program. Kindly refer to Appendix B for more results.

6.4.2 Qualitative Analysis

Of the $T = 40$ topics our model inferred, we subjectively judged 37 to be coherent; a glimpse of each is given in Figure 6.5. We review some of the key results here and refer the interested reader to Appendix B for additional results and analysis. We manually selected six of the most interpretable topics for further evaluation in this section.

As a generative modeling approach, our model was designed for the purpose of reducing the dimensionality of the sociopolitical debate space, as evidenced by Debatepedia. It is like other topic models in this regard, but we believe that some effects of our design choices are noteworthy. Table 6.6 compares the positional bigrams of our model to the sentiments inferred by JST. We observe the benefit of our model in identifying terms associated with positions on social issues, while JST selects more general sentiment terms.

Table 6.7 shows bigrams most strongly associated with general position distributions ϕ^{i} and selected topic-position distributions ϕ°.[15] We see the potential benefit of multiword expressions. Although we have used frequent bigrams as a poor man's approximation to multiword expression analysis, we find the topic-specific positions

[15]For more topics, please refer to the supplementary notes in Appendix B.

terms to be subjectively evocative. While somewhat internally coherent, we do not observe consistent alignment across topics, and the general distributions ϕ^{i} are not suggestive.

The separation of personal name mentions into their own distributions, shown for some topics in Table 6.8, gives a distinctive characterization of topics based on relevant personalities. Subjectively, the top individuals are relevant to the subject matter associated with each topic (though the topics are not always pure; same-sex marriage and the space program are merged, for example).

6.5 Discussion

Within debates and within topics, the model uses the position variable to distinguish sides well. For external text, the model performs well on articles such as blogs and editorials but on others the positional categories do not seem meaningful, perhaps due to the less argumentative nature of other kinds of articles. Noting the vast literature focusing on ideological positions expressed in text, we believe this failure suggests (i) that broad-based positions that hold across many topics may require richer textual representations (see, e.g., the "syntactic priming" of Greene and Resnik), [54] or (ii) that an alternative representation of positions, such as the spatial models favored by political scientists [104], may be more discoverable. Aside from those issues, a stronger theory of positions may be required. Such a theory could be encoded in a more informative prior or weaker independence assumptions across debates. Finally, exploiting explicitly ideological texts alongside the moderated arguments of Debatepedia might also help to identify textual associations with general positions [115]. We leave these directions to future work.

6.6 Related Work

Insofar as debates are subjective, our study is related to **opinion mining**. Subjective text classification [133] leads to opinion mining tasks such as opinion extraction [33], positive and negative polarity classification [100], sentiment target detection [61, 45], and feature-opinion extraction [136]. The above studies are conducted mostly on product reviews, a domain with a simpler opinion landscape and more concrete rationales for those opinions, compared to sociopolitical debates.

Generative **topic models** have been successfully implemented in opinion mining tasks such as feature identification [123], entity-topic extraction [95], mining contentious expressions and interactions [92] and specific aspect-opinion word extraction from labeled data [141]. Most relevant to this research is work on feature-sentiment extraction [76, 88]. [88] built on PLSI, which is problematic for generalizing beyond the training sample. The JST model of [76] is an LDA-based topic model in which each word token is assigned both a sentiment and a topic; they exploited a sentiment lexicon in the prior distribution. Our model is closely related, but introduces a switching variable that assigns *some* tokens to positions, some to topics, and some to both. Unlike Lin and He's sentiments, our model's positions are associated with the two sides of a debate, and we incorporate topics at the level of questions within debates.

Some studies have specifically analyzed **contrastive viewpoints** or **stances** in general discussion text. [5] used graph mining based method to classify authors in to opposite camps for a given topic. [101] developed an unsupervised method for summarizing contrastive opinions from customer reviews. [2] and [32] developed techniques to address the problem of automatically detecting subgroups of people holding similar stances in a discussion thread.

Several prior studies have considered **debates**. [24] developed a system based on argumentation theory which recognizes the entailment and contradiction relationships between two texts. [12] used a debate corpus, Debatepedia as a seed for extracting person-opinion-topic tuples from news and other web documents and in later work classified the quotations to specific topics and polarity using language models [13]. [117] and [10] were interested in ideological content in debates, relying on discourse structure and leveraging sentiment lexicons to recognize stances. [117] used Debatepeida corpus for evaluating their approach for stance detection.

Closer to the methodology we describe, [77] presented a statistical model for political discourse that incorporates both topics and ideologies; they used debates on the Israeli-Palestinian conflict. [43] showed that it is possible to isolate a subset of terms from media content that are informative of a news organization's bias towards a particular issue. [6] introduced multi-level latent Dirichlet allocation, and [40] introduced sparse additive generative models, both conceived as extensions to well-established probabilistic modeling techniques [20]; these were applied to debates and political blog datasets. Our approach builds on these models (especially the switching variables of Ahmed and Xing). We go farther in jointly modeling text across *many* debates evidenced by the structure of Debatepedia, thus grounding our models more solidly in familiar sociopolitical issues, and in making extensive use of existing NLP resources.

6.7 Chapter Summary

In this chapter, we studied the problem of extracting the stances and ideological expressions of users from their comments on ideological debates related to sociopolitical domain. We proposed an unsupervised approach based on latent variable methods for identifying and extracting the ideological expressions/positional words and entities associated with the issues. Using Debatepedia, we inferred topics and position term lexicons in the domain of sociopolitical debates. Our approach brings together tools from information extraction and sentiment analysis into a latent-variable topic model and exploits the hierarchical structure of the dataset. Our qualitative and quantitative evaluations show the model's strengths and weaknesses.

Chapter 7

Conclusions

In this chapter, we summarize the findings of this thesis, present an integrated framework for sociopolitical opinion mining applications, and point out some future research directions.

7.1 Summary

Online social networks and forums are now getting increasingly popular and have attracted a large number of users, with many of them expressing and discussing their opinions on sociopolitical topics through these websites. Mining social comments is critically important to build applications which aid policy makers, social organizations or government sectors in their decision making process. Social and political data are much harder to analyse due to complex topics and sentiment expressions. A comment in the sociopolitical domain is usually a collection of issues, entities and opinions associated with the entities. Therefore, there is great need in techniques that can mine the comments for discovering low-dimensional, human-interpretable representation of the space of sociopolitical comments.

In this thesis, we studied the tasks of mining issues, extracting entities and suggestive opinions towards the entities, detecting thoughtful comments and extracting ideological stances and expressions in the sociopolitical domain. We first studied the problem of issue extraction from sociopolitical comments in Chapter 3 for which, we proposed an unsupervised approach based on latent variable methods for identifying and extracting issues in the comments and linking comments to the issues in the associated article [51].

Second, we studied a new problem of extracting the entities and associated suggestive opinions [49] in Chapter 4. In the sociopolitical domain, users express their sentiments on the entities such as individuals or organizations. These sentiments are not only in the form of positive and negative expressions but also in the form of suggestive opinions towards the entities.

Since social media enables anyone to freely express the opinions, there is an urge for extracting opinions which are highly valuable. Third, we studied the problem related to quality data extraction from social media in Chapter 5. In terms of thoughtful comment extraction task, we studied the problem of extracting valuable comments from social media [50].

Debates on controversial sociopolitical issues provide room for the netizens to take positions and post opinions towards those issues. Finally, we studied the problem of extracting the topics, stances and ideological expressions of users from their comments on ideological debates related to sociopolitical domain [52] in Chapter 6.

7.2 Conceptual Sociopolitical Opinion Mining Framework

One of the purposes of this thesis is to develop and gain an understanding of the concepts, functions and uses of sociopolitical opinion mining for decision making applications in a broader sense. It is also natural to question the integration aspects of the four tasks described in this thesis to know how these tasks work together for a common goal. To answer this we propose a conceptual framework for sociopolitical opinion mining. However, this thesis does not attempt to evaluate an end-to-end system. It is our hope that this framework will inspire further research at the intersection of sociopolitical data and opinion mining, and will produce more applications in the industry. In this section, we first present a conceptual framework of sociopolitical opinion mining, and then discuss the general challenges and possible extensions of this framework.

7.2.1 Conceptual Framework

A general methodology/framework for sociopolitical opinion mining is given in Figure 7.1. Systems implemented based on the framework would help to acquire new knowledge through operations and analytics, and aid in decision making process for the domain experts. The aim of this integrated framework is to automatically detect and summarize the public opinions towards sociopolitical problems. As the figure shows, starting from e-citizens' opinions in social media, we should first retrieve relevant posts using appropriate retrieval methods. With the retrieved posts, next we can perform extraction to identify issues and entities that are opinion targets. Once these targets are identified, other techniques such as opinion extraction and classification can be used to identify and characterize the opinions expressed by e-citizens on these opinion targets. These opinions may include polarized sentiments, suggestions, stances, etc. Finally, to make the extracted information easy to digest by the end users, a summarization component can be applied to organize the extracted information into structured, human-readable form such as textual summaries and graphs. These summaries can be consumed by the end users such as government bodies and eventually help them make decisions. Note that this does not end the process, because as new policies are executed, the process will repeat itself as shown in the figure.

Comment Retrieval: Considering that people voice out their opinions in social media, we need methods that gather the citizens' feedback from social media sites, e.g., blogs, forums, etc. Posts represent blog posts, forum posts, threads or even

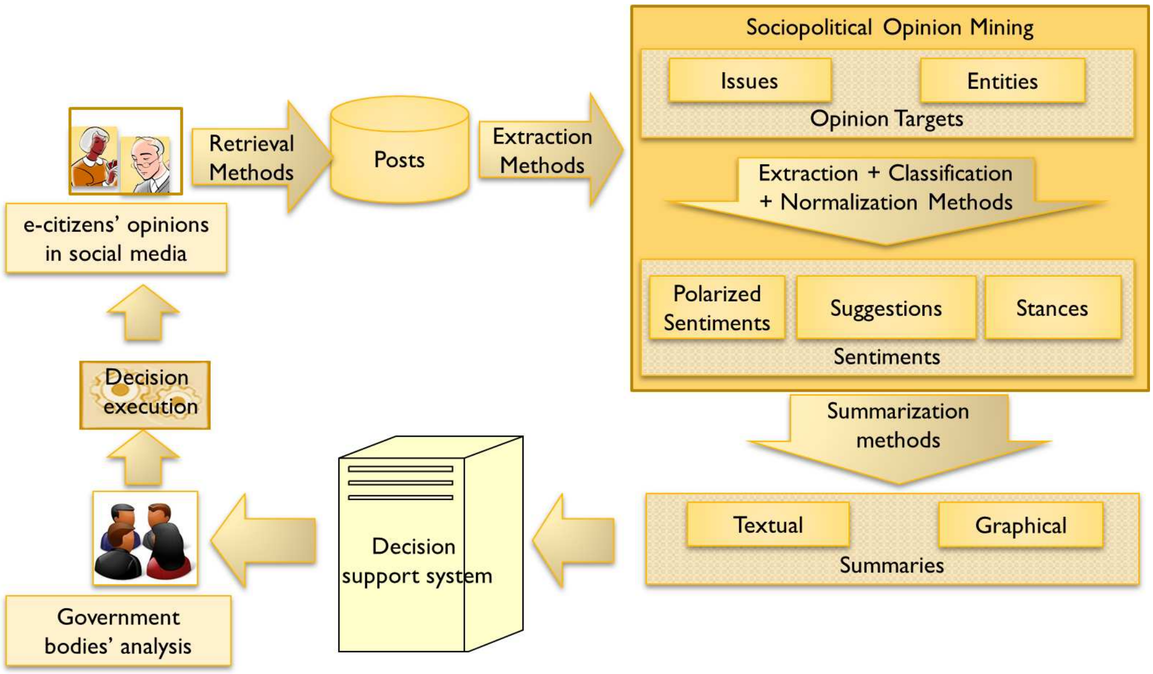

Figure 7.1: Conceptual Sociopolitical Opinion Mining Framework

microblog posts. Such content can be harvested either manually by domain experts, who indicate the data sources that need to be collected, or automatically via information retrieval techniques [140, 63], or via trained focused crawler application to specific portals. After retrieval, the comments represent the candidates for opinions on the social article/document/event.

Sociopolitical Opinion Mining: The next task is to apply the opinion mining techniques to extract and classify the opinions/sentiments from the comments. In case of social comments, the targets of the opinions are issues and entities. Note that issues can be hierarchial in nature. For example given the issue, *immigration*, the entities can be *immigration officers*, *minister of international affairs* etc., and sub-issues can be *security*, *citizenship*, *workers*, etc., Hierarchial issue extraction is an interesting direction for fine grained opinion mining. However, this thesis does not answer this problem and we leave it for future work. The first step in this stage is to extract these opinion targets using opinion extraction techniques [51, 49]. Once the opinion targets are known, the sentiment classification or discovery techniques aid in discovering the sentiments on the issues or entities or sub-issues [49, 52]. However, the opinions often suffer with quality problems, as social media users may not always generate highly useful comments. Filtering of such noisy comments and extracting useful comments aids in generating high quality summaries during the next stage. The output of this stage is a low-dimensional human interpretable representations of the huge corpus of comments in the form of various types of sentiments on the targets of sociopolitical opinions. Such representations can be combined with summarization techniques for generating textual or graphical summaries.

Summarization: Based on the output of the above stage, we may not only capture the citizen's polarized sentiments, suggestions to entities and stances on sociopolitical issues/entities, but also be able to measure the influence of the issues/events on the citizens and the society. The abundance of opinions poses challenges in digesting all the massive information. For example, some controversial issues or contentious events may get hundreds or even thousands of comments. To address this problem, we need to summarize the comments in human interpretable forms [60, 61]. The goal of this stage is to process the output from opinion mining stage and generate textual summaries or graphical summaries. For example, by exploiting summarization techniques [88, 84, 46] we can generate issue-based short textual summaries. However, to generate entity-based and ideological summaries there is a need for new techniques to produce textual and graphical outputs.

Decision Support Applications: The domain experts use the citizens feedback to understand the pulse of citizens on sociopolitical issues. For example, the feedback can be used for taking decisions on the new policies or amending existing policies. Traditionally, the telephone surveys were the major feedback channels for the domain users. However with the growing participation of citizens in the social media, using this framework the feedback of the citizens from social media can be analysed by the domain experts through the decision support systems which are fed with the summaries. Moreover, such opinion analytics is a cyclic process and after the execution of the decisions (e.g. new/amended policies), the feedback from the citizens can be extracted and analysed iteratively.

7.2.2 Challenges

Social Media data sets are very voluminous and hence pose three main challenges; noise, size and dynamism. The above framework is affected with these three challenges. In this section, we describe them in detail and present the possible solutions to tackle these challenges. Note that this thesis has only addressed some of these challenges to some extent. The purpose of this section is to identify the major challenges in a broader sense and hence motivate future research in this direction.

Noise: In blogs and forum discussions both issues and entities are not easily recognized and there are high levels of insignificant data which constitute noise. User generated contents in social media tend to be less grammatically correct and they are informally written. These texts often make use of emoticons or abbreviations or spelling mistakes or unorthodox casing or malformed sentences. Relevant content on webpages is usually surrounded by irrelevant elements like spams, mutual discussions and diversions to non-relevant topics [35]. All these forms of noises make the detection and processing of opinions a complicated task. In this thesis, we handled some forms of noise such as using topic models to merge informal words (e.g. ”ft”) into the same topic as formal words (e.g. ”foreign talent”) and using supervised learning to identify high quality comments. However, there are also many other forms of noise that we have not addressed. Techniques such as normalization of text [1] can be incorporated to our methods to further handle noise. Spam detection techniques [93] can be incorporated into the framework for high quality comment extraction.

Scalability Social media corpora are very voluminous and require automated information processing for analysing it within a reasonable time. Many of our solutions are based on machine learning methods, some of which may be time-consuming to run on large data sets. While in this thesis we have not addressed the scalability problem, we would like to point out that scalability can be regarded as an orthogonal problem to socio-political opinion mining. As many of our underlying techniques such as conditional random fields and topic models are fundamental, we could improve the scalability of our algorithms by tapping into recent advances in the machine learning and data mining communities that study more efficient and scalable versions of standard algorithms. Examples include fast semantic analysis [130, 106] and map reduce framework [138]. Models proposed in Chapter 3 and Chapter 6 are capable to tap into these new advances in the machine learning studies to handle scalability. In addition, for our solutions that are supervised in nature, training the model can be done offline, and once a model is trained, classification and prediction generally can be done very efficiently.

Dynamism One of the key challenges with social media is that it is a highly dynamic platform. For example, every time a new entry is detected on a blog or forum, a large amount of text mining techniques have to be launched again. The dynamism in social media corpora causes it to evolve rapidly over time and fortunately data mining techniques are versatile in handling the new (unobserved) issues or sentiments. Topic models are capable of handling such dynamism when new topics are added to the social media content as there exist techniques for incrementally training topic models. In the recent advances of machine learning research, some studies focussed on the new topic generation aspects and online streaming aspects. Dynamic topic models are designed to handle this dynamism of new topic generation [19] and extensions to standard LDA model were developed by the researchers for online streaming [57]. Our models proposed in Chapter 3 and Chapter 6 are capable of tapping into these new advances in the machine learning studies to handle dynamism in the social media context.

7.2.3 Extendibility

In this thesis, we presented techniques for opinion extraction of issues and entities. The framework can be further extended with other related opinion mining tasks for implementing smart decision supporting systems. For example opinions change over time and temporal opinion mining [31] can be achieved through this framework by including one more component, time. Subsequently, to generate the opinion summaries based on time, temporal summarization techniques [7] can be integrated into the framework. Sentiment lexicons are valuable resources that are useful for classification tasks [134] on products or entities. The framework can be extended to build sentiment lexicons for sociopolitical comments to gain new insights into how sentiments on issues are expressed lexically.

7.3 Contributions

This thesis made the following contributions to mining users' sociopolitical comments from social media:

- **Issue Extraction:** We proposed an unsupervised approach based on latent variable methods for identifying and extracting the issues in the comments and linking the comments to the issues in the associated article. Our solution model extracted issues from the social comments in response to a political speech, and linked users' comments to the relevant issues in the speech. Our experiment results on two data sets show that our proposed model performed better in terms of precision when compared against state-of-the-art methods.
- **Entity-Suggestive Opinion Extraction:** We proposed a two-stage approach based on Conditional Random Fields and clustering method for extracting and normalizing the entities and the associated suggestive opinions from the users. Extracting and normalizing entity and suggestive pairs is an opinion extraction problem. Our comprehensive experiments show that CRF model has better performance in entity-action extraction. Agglomerative clustering method for entity-action normalization using complete-link measure performs well with an average precision of 81.15% for all articles.
- **Thoughtful Comment Extraction:** Valuable comments are useful for deci-

sion makers and high quality summarization. The study of extracting attentive comments that provide some reasoning is highly valuable for understanding the users' opinions in sociopolitical opinion mining process. We proposed a supervised approach based on NLP and linguistics techniques to identify and extract valuable comments in the sociopolitical domain from social media. We performed an empirical study using syntactic, vocabulary, discourse, and relevance features for prediction and combination of all features is substantially better than the baseline surface features.

- **Ideological Position and Expression Extraction:** We proposed an unsupervised approach based on latent variable methods for identifying and extracting the positional words/ideological expressions and entities associated with the issues. We modeled issues, positional words, entities, and the stances of the users in a principled way using topic models. Using text from Debatepedia, we inferred topics and position term lexicons in the domain of sociopolitical debates. Our qualitative and quantitative evaluations show the model's strengths and weaknesses.

- **Conceptual Sociopolitical Opinion Mining Framework:** We proposed an integrated and extendible framework of sociopolitical opinion mining that not only shows the possible integration of the four tasks described in this thesis, but also describes how these tasks work together for a common objective. We hope that this framework will inspire further research at the intersection of sociopolitical data and opinion mining, and will produce more applications in the industry.

- **Corpus:** Last but not least, as an important outcome of this research, I constructed and released the corpus desgined for research on mining sociopolitical comments. This corpus consists of three major datasets: two political speeches and corresponding comments from the users, five news articles and corresponding actionable comments, and 1303 debates and corresponding arguments and linked external articles. Furthermore, the annotations on this data for all the tasks discussed in this thesis are provided. The details of the corpus and the annotations can be found in Appendix A.

A key contribution of the thesis is that now, principled techniques can be applied for mining sociopolitical comments for opinion mining applications, whereas, before there was little work that uses principled techniques to mine the sociopolitical comments which are more complex in terms of topics and sentiment expressions (the existing techniques of opinion mining are dedicated mostly for product reviews and less attention has been paid for opinion mining of sociopolitical posts).

7.4 Future Directions

In the future, we plan to further optimize the current solutions as well as to study a few new directions related to opinion mining of sociopolitical content in social media.

- **Generating an ideology phrase lexicon:** The words "good", "long", "quality" usually indicate sentiment when describing an electronic device or service in restaurant [121]. However, to voice out opinions on the sociopolitical issues, the citizens mostly rely on ideology specific words. In fact, these ideology words are specific to the topic of the issue and the ideology of the user. For example, on *abortion*, the ideology of the person can be liberal or conservative. Depending on the ideology, the user uses the opinion phrases such as, "unethical","immoral", "pro-choice", "freedom", "age appropriate", "human life" etc., to depict his/her stance/sentiment towards the issue. Such topical-ideology lexicons can be useful in various applications to the social or political scientists. Human annotation of such lexicons is very expensive. One of the future directions one can explore is the an auto-extraction of ideology lexicons. Given a collection of ideological debates, the task is to generate the ideological phrases.

Some existing approaches aid to solve this tasks such as rule-based and keyphrase extraction techniques [108]. However, with the popularity of topical models in extracting topics, we could leverage on graphic models for this task. For example, one sensible solution is to use the solution model from Chapter 6 to generate the seed list of the topic specific ideology expressions, and then apply bootstrap techniques to expand the ideology lexicon. In order to generate the bigram and trigram phrases, the model can be extended to n-gram topic model.

- **Predicting user profile using ideological stances and social networks:** Predicting users' attributes such as political party from social media has important impacts on many real world applications such as targeted advertising, recommendation and personalization. Another future direction one can explore is to exploit users' ideological positions on controversial issues to predict political party of online users. Studies on American politics demonstrate the fact that the political parties (Democrats, Republicans etc.,) take positions towards critical policies and sociopolitical issues, which can ultimately lead to great differences in philosophies and ideal [119]. Hence users' political affiliation is largely dependent on his/her ideological stances on the major social and political issues [68]. For example, a user who supports abortion and is against gun rights is more likely a Democrat. His/her other stances on issues like *gay marriage*, *health care*, *tax*, *death penalty*, etc. can aid in detecting his/her party affiliation with high accuracy. The social network also plays an important role on user's attribute affiliation. A sensible solution is to use clustering techniques to group the users with similar ideological beliefs and friendships, and measure their average proximity to the party beliefs to predict the party affiliation of each cluster.

- **Exploiting ideological texts:** Within debates and topics, the model from

Chapter 6 uses the position variable to distinguish sides well but for external text/articles, the model performs well on the articles such as blogs and editorials and fails on the articles related to news, wiki and governments. This suggests that broad-based positions that hold across many topics may require richer textual representations. Aside from those issues, a stronger theory of positions may be required. Such a theory could be encoded in a more informative prior or weaker independence assumptions across debates. Exploiting explicit ideological texts alongside the moderated arguments of Debatepedia might help to identify textual associations with general positions.

- **Actionable knowledge from microblogs:** Actionable content extraction is a new direction in opinion mining process with many opportunities and challenges. With the increasing user generated content in micro blogs, detecting actionable knowledge in such media is an interesting as well as challenging problem. For example, during Obama's State of the Union address, apart from political and news forums, the public was asked to express opinions on Twitter using specific hashtags such as "#SOTU"[1], "#jobs" etc[2]. This triggers the need for gathering actionable content in micro blogs. In the same line, diagnostic opinion detection that talks about what could have happened, who should be blamed, etc., is another very important problem in social sciences research.

With the rapid growth of user generated content on the web, there are always new directions for opinion mining and new challenges that trigger urgent and critical tasks under its umbrella.

[1] `http://www.whitehouse.gov/state-of-the-union-2012`
[2] `https://blog.twitter.com/2012/follow-the-state-of-the-union-on-twitter`

Bibliography

[1] Normalization of non-standard words. *Computer Speech & Language*, 15(3):287–333, 2001.

[2] A. Abu-Jbara, P. Dasigi, M. Diab, and D. Radev. Subgroup Detection in Ideological Discussions. In *ACL '12: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 399–409, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[3] A. Abu-Jbara and D. Radev. Subgroup Detector: A System for Detecting Subgroups in Online Discussions. In *ACL '12: Proceedings of the ACL 2012 System Demonstrations*, pages 133–138, Jeju Island, Korea, July 2012. Association for Computational Linguistics.

[4] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne. Finding High-Quality Content in Social Media. In *WSDM '08: Proceedings of the international conference on Web search and web data mining*, pages 183–194, New York, USA, 2008. ACM.

[5] R. Agrawal, S. Rajagopalan, R. Srikant, and Y. Xu. Mining Newsgroups Using Networks Arising From Social Behavior. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*, pages 529–535, 2003.

[6] A. Ahmed and E. P. Xing. Staying informed: supervised and semi-supervised multi-view topical analysis of ideological perspective. In *EMNLP '10: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

[7] J. Allan, R. Gupta, and V. Khandelwal. Temporal Summaries of New Topics. In *SIGIR '01: Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*.

[8] L. Amgoud, F. Bannay, C. Costedoat, P. Saint-Dizier, and C. Albert. Introducing Argumention in Opinion Analysis: Language and Reasoning Challenges. In *SAAIP '11: Sentiment Analysis where AI meets Psychology*, pages 28–34, Chiang Mai, Thailand, November 2011. Asian Federation of Natural Language Processing.

[9] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo. A comparison of extrinsic clustering evaluation metrics based on formal constraints. *Journal of Information Retrieval*, Volume 12(4):461–486, Aug. 2009.

[10] P. Anand, M. Walker, R. Abbott, J. E. F. Tree, R. Bowmani, and M. Minor. Cats rule and dogs drool!: Classifying stance in online debate. In *WASSA '11: Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis*.

[11] Y. Attali and J. Burstein. Automated Essay Scoring With E-rater V.2. *The Journal of Technology, Learning and Assessment (JTLA)*, 4(3), February 2006.

[12] R. Awadallah, M. Ramanath, and G. Weikum. OpinioNetIt: understanding the opinions-people network for politically controversial topics. In *CIKM '11: Proceedings of the 20th ACM international conference on Information and knowledge management*.

[13] R. Awadallah, M. Ramanath, and G. Weikum. PolariCQ: polarity classification of political quotations. In *CIKM '12: Proceedings of the 21st ACM international conference on Information and knowledge management*.

[14] A. Baron and M. Freedman. Who is Who and What is What: Experiments in Cross-Document Co-Reference. In *EMNLP '08: Proceedings of the Conference on Empirical Methods in Natural Language Processing*.

[15] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. S. Subrahmanian. Sentiment Analysis: Adjectives and Adverbs are Better than Adjectives Alone. In *ICWSM'07: Proceedings of the International Conference on Weblogs and Social Media*, 2007. Short paper.

[16] D. Bespalov, B. Bai, Y. Qi, and A. Shokoufandeh. Sentiment Classification Based on Supervised Latent N-gram Analysis. In *CIKM '11: Proceedings of the 20th ACM International Conference on Information and Knowledge Management*.

[17] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *WWW '09: Proceedings of the 18th International Conference on World Wide Web*, pages 51–60. ACM, 2009.

[18] D. M. Blei, T. L. Griffiths, M. I. Jordan, and J. B. Tenenbaum. Hierarchical Topic Models and the Nested Chinese Restaurant Process. In *NIPS' 04: Advances in Neural Information Processing Systems*, 2004.

[19] D. M. Blei and J. D. Lafferty. Dynamic Topic Models. In *ICML '06: Proceedings of the 23rd International Conference on Machine Learning*.

[20] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent Dirichlet Allocation. *The Journal of Machine Learning Research*, (3):993–1022, Mar. 2003.

[21] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, Vol. 2008, No. 10, 2008.

[22] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *Journal of Computational Science*, abs/1010.3003, 2010.

[23] E. Breck, Y. Choi, and C. Cardie. Identifying Expressions of Opinion in Context. In *IJCAI'07: Proceedings of the 20th International Joint Conference on Artifical Intelligence*.

[24] E. Cabrio and S. Villata. Combining textual entailment and argumentation theory for supporting online debates interactions. In *ACL '12: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*.

[25] M. Castellanos, U. Dayal, M. Hsu, R. Ghosh, M. Dekhil, Y. Lu, L. Zhang, and M. Schreiman. LCI: A social channel analysis platform for live customer intelligence. In *SIGMOD '11: Proceedings of the 2011 ACM SIGMOD International Conference on Management of data*.

[26] J. Chang, J. Boyd-Graber, C. Wang, S. Gerrish, and D. M. Blei. Reading Tea Leaves: How Humans Interpret Topic Models. In *NIPS' 09: In Advances in Neural Information Processing Systems 21*, 2009.

[27] B.-C. Chen, J. G. 0002, B. L. Tseng, and J. Y. 0015. User Reputation in a Comment Rating Environment. In *KDD '11: Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 159–167. ACM, 2011.

[28] Y. Choi, E. Breck, and C. Cardie. Joint Extraction of Entities and Relations for Opinion Recognition. In *EMNLP '06: Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*.

[29] Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Identifying sources of opinions with conditional random fields and extraction patterns. In *HLT '05: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*.

[30] M. Dahllf. Automatic prediction of gender, political affiliation, and age in Swedish politicians from the wording of their speeches - A comparative study of classifiability. *The Journal Literary and Linguistic Computing, LLC*, 27(2):139–153, 2012.

[31] D. Das, A. K. Kolya, A. Ekbal, and S. Bandyopadhyay. Temporal analysis of sentiment events: A visual realization and tracking. In *CICLing '11: Proceedings of the 12th international conference on Computational linguistics and intelligent text processing - Volume Part I*.

[32] P. Dasigi, W. Guo, and M. Diab. Genre independent subgroup detection in online discussion threads: A pilot study of implicit attitude using latent textual semantics. In *ACL '12: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*.

[33] K. Dave, S. Lawrence, and D. M. Pennock. Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *WWW '03: Proceedings of the 12th international conference on World Wide Web*.

[34] A. Devitt and K. Ahmad. Sentiment Analysis in Financial News: A Cohesion-based Approach. In *ACL '07: Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pages 984–991, 2007.

[35] L. Dey and S. K. M. Haque. Opinion mining from noisy text data. In *AND '08: Proceedings of the Second Workshop on Analytics for Noisy Unstructured Text Data*.

[36] Q. Diao, J. Jiang, F. Zhu, and E.-P. Lim. Finding bursty topics from microblogs. In *ACL '12: Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*.

[37] K. T. Durant and M. D. Smith. Mining Sentiment Classification from Political Web Logs. In *WebKDD '06: Proceedings of Workshop on Web Mining and Web Usage Analysis of the 12th SIGKDD*, 2006.

[38] K. T. Durant and M. D. Smith. Predicting the political sentiment of web log posts using supervised machine learning techniques coupled with feature selection. In *WebKDD '06: Proceedings of the 8th Knowledge discovery on the web international conference on Advances in web mining and web usage analysis*, pages 187–206, 2006.

[39] M. Efron. Using cocitation information to estimate political orientation in web documents. *Journal of Knowledge and Information Systems*, 9(4), 2006.

[40] J. Eisenstein, A. Ahmed, and E. P. Xing. Sparse additive generative models of text. In *ICML '11: Proceedings of International Conference on Machine Learning*, pages 1041–1048, 2011.

[41] M. A. Ferrario, W. Simm, J. Whittle, P. Rayson, M. Terzi, and J. Binner. Understanding Actionable Knowledge in Social Media: BBC Question Time and Twitter, a Case Study. In *ICWSM '12: Proceedings of the International AAAI Conference on Weblogs and Social Media*, 2012.

[42] J. R. Finkel, T. Grenager, and C. Manning. Incorporating non-local information into information extraction systems by Gibbs sampling. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*.

[43] B. Fortuna, C. Galleguillos, and N. Cristianini. Detecting the bias in media with statistical learning methods. In *Text Mining: Classification, Clustering, and Applications*, pages 27–50. Chapman & Hall/CRC, 2009.

[44] T. Fukuhara, H. Nakagawa, and T. Nishida. Understanding sentiment of people from news articles: Temporal sentiment analysis of social events. In *ICWSM '07: Proceedings of the International Conference on Weblogs and Social Media*, 2007.

[45] M. Ganapathibhotla and B. Liu. Mining Opinions in Comparative Sentences. In *COLING '08: Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1*.

[46] K. Ganesan, C. Zhai, and J. Han. Opinosis: A Graph Based Approach to Abstractive Summarization of Highly Redundant Opinions. In *COLING '10: Proceedings of the 23rd International Conference on Computational Linguistics*, 2010.

[47] A. Ghose and P. G. Ipeirotis. Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics. *Journal of IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512, 2011.

[48] N. Godbole, M. Srinivasaiah, and S. Skiena. Large-Scale Sentiment Analysis for News and Blogs. In *ICWSM '07: Proceedings of the International Conference on Weblogs and Social Media*, 2007.

[49] S. Gottipati and J. Jiang. Extracting and Normalizing Entity-Actions from Users' Comments. In *COLING '12: Proceedings of the international conference on Computational Linguistics, (Posters)*, pages 421–430, 2012.

[50] S. Gottipati and J. Jiang. Finding Thoughtful Comments from Social Media. In *COLING '12: Proceedings of the international conference on Computational Linguistics*, pages 995–1010, 2012.

[51] S. Gottipati and J. Jiang. What's Public Feedback? Linking High Quality Feedback to Social Issues Using Social Media. In *SocialCom '12: Proceedings of the 2012 ASE international conference on Social Computing, SocialCom/PASSAT*, 2012.

[52] S. Gottipati, M. Qiu, Y. Sim, J. Jiang, and N. A. Smith. Learning Topics and Positions from Debatepedia. In *EMNLP '13: Proceedings of the 2013 conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October 2013.

[53] S. Gottipati, M. Qiu, L. Yang, F. Zhu, and J. Jiang. Predicting User's Political Party using Ideological Stances. In *SocInfo '13; Proceedings of the 5th International Conference on Social Informatics*, 2013.

[54] S. Greene and P. Resnik. More than Words: Syntactic Packaging and Implicit Sentiment. In *NAACL '09: Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 503–511, Boulder, Colorado, June 2009. Association for Computational Linguistics.

[55] T. L. Griffiths and M. Steyvers. Finding Scientific Topics. *Journal of the National Academy of Sciences of the United States of America*, 101:5228–5235, 2004.

[56] A. Hassan, A. Abu-Jbara, and D. Radev. Detecting subgroups in online discussions by modeling positive and negative relations among participants. In *EMNLP-CoNLL '12: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 2012.

[57] M. D. Hoffman, D. M. Blei, and F. R. Bach. Online learning for latent dirichlet allocation. In J. D. Lafferty, C. K. I. Williams, J. Shawe-Taylor, R. S. Zemel, and A. Culotta, editors, *NIPS '10: Advances in Neural Information Processing Systems*, pages 856–864. Curran Associates, Inc., 2010.

[58] T. Hofmann. Probabilistic Latent Semantic Indexing. In *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*.

[59] C.-F. Hsu, E. Khabiri, and J. Caverlee. Ranking Comments on the Social Web. In *CSE '09: Proceedings of the 2009 International Conference on Computational Science and Engineering - Volume 04*, pages 90–97. IEEE Computer Society, 2009.

[60] M. Hu and B. Liu. Mining Opinion Features in Customer Reviews. In *AAAI'04: Proceedings of the 19th national conference on Artifical intelligence*.

[61] M. Hu and B. Liu. Mining and Summarizing Customer Reviews. In *KDD '04: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177, 2004.

[62] Y. Hu, A. John, D. D. Seligmann, and F. Wang. What were the Tweets about? Topical Associations between Public Events and Twitter Feeds. In *ICWSM '12: Proceedings of the 6th international AAAI conference on Weblogs and Social Media*, 2012.

[63] X. Huang and W. B. Croft. A unified relevance model for opinion retrieval. In *CIKM '09: Proceedings of the 18th ACM conference on Information and knowledge management*.

[64] A. Jain, M. Murty, and P. Flynn. Data Clustering: A Review. *Journal ACM Computing Surveys (CSUR)*, 31(3):264–323, 1999.

[65] N. Jindal and B. Liu. Opinion Spam and Analysis. In *WSDM '08: Proceedings of the 2008 International Conference on Web Search and Data Mining*.

[66] R. Johansson and A. Moschitti. Extracting Opinion Expressions and Their Polarities: Exploration of Pipelines and Joint Models. In *HLT '11: Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*.

[67] K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

[68] M. Killian and C. Wilcox. Do Abortion Attitudes Lead to Party Switching? *Journal Political Research Quarterly Vol. 61, No. 4*, pages 561–573, 2008.

[69] S.-M. Kim and E. Hovy. Identifying Opinion Holders for Question Answering in Opinion Texts. In *AAAI '05: Proceedings of the AAAI Workshop on Question Answering in Restricted Domains*, 2005.

[70] S.-M. Kim and E. Hovy. Extracting Opinions, Opinion Holders, and Topics Expressed in Online News Media Text. In *SST '06: Proceedings of the workshop on Sentiment and Subjectivity in Text*, Sidney, AUS, 2006.

[71] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti. Automatically Assessing Review Helpfulness. In *EMNLP '06: Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 423–430, Sydney, Australia, July 2006. Association for Computational Linguistics.

[72] C. Kohlschütter, P. Fankhauser, and W. Nejdl. Boilerplate detection using shallow text features. In *WSDM '10: Proceedings of the third ACM international conference on Web search and data mining*.

[73] L.-W. Ku, L.-Y. Lee, and H.-H. Chen. Opinion extraction, summarization and tracking in news and blog corpora. In *AAAI-CAAW '06: Proceedings of AAAI-2006 spring symposium on Computational Approaches to Analyzing Weblogs*, 2006.

[74] J. D. Lafferty, A. McCallum, and F. C. N. Pereira. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *ICML '01: Proceedings of the Eighteenth International Conference on Machine Learning*.

[75] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw. Detecting Product Review Spammers Using Rating Behaviors. In *CIKM '10: Proceedings of the 19th ACM international conference on Information and knowledge management*, 2010.

[76] C. Lin and Y. He. Joint Sentiment/Topic Model for Sentiment Analysis. In *CIKM '09: Proceedings of the 18th ACM conference on Information and knowledge management*.

[77] W.-H. Lin, E. Xing, and A. Hauptmann. A Joint Topic and Perspective Model for Ideological Discourse. In *ECML PKDD '08: Proceedings of the European conference on Machine Learning and Knowledge Discovery in Databases*.

[78] B. Liu. Sentiment Analysis and Subjectivity. *Handbook of Natural Language Processing*, 2nd ed, 2010.

[79] B. Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

[80] B. Liu, M. Hu, and J. Cheng. Opinion Observer: Analyzing and Comparing Opinions on the Web. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*.

[81] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou. Low-Quality Product Review Detection in Opinion Summarization. In *EMNLP-CoNLL '07: Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 334–342, 2007. Poster paper.

[82] Y. Liu, X. Huang, A. An, and X. Yu. ARSA: A sentiment-aware model for predicting sales performance using blogs. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*.

[83] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi. Exploiting Social Context for Review Quality Prediction. In *WWW '10: Proceedings of the 19th international conference on World wide web*, pages 691–700. ACM, 2010.

[84] Y. Lu and C. Zhai. Opinion integration through semi-supervised topic modeling. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*.

[85] T. Ma and X. Wan. Opinion Target Extraction in Chinese News Comments. In *COLING '10: Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, pages 782–790, Beijing, China, August 2010.

[86] C. Macdonald, I. Ounis, and I. Soboroff. Overview of the TREC 2009 Blog Track. In *TREC '09: Proceedings of Text REtrieval Conference*, 2009.

[87] M. McGlohon, N. Glance, and Z. Reiter. Star Quality: Aggregating Reviews to Rank Products and Merchants. In *ICWSM '10: Proceedings of Fourth International Conference on Weblogs and Social Media*, 2010.

[88] Q. Mei, X. Ling, M. Wondra, H. Su, and C. Zhai. Topic sentiment mixture: modeling facets and opinions in weblogs. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*.

[89] M. Meila. Comparing Clusterings by the Variation of Information. *Learning Theory and Kernel Machines*, 2777:173–187, 2003.

[90] P. Melville, W. Gryc, and R. D. Lawrence. Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. In *KDD '09: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[91] G. Mishne and M. de Rijke. MoodViews: Tools for Blog Mood Analysis. In *AAAI-CAAW '06: Proceedings of AAAI Symposium on Computational Approaches to Analysing Weblogs*, pages 153–154, 2006.

[92] A. Mukherjee and B. Liu. Mining Contentions from Discussions and Debates. In *KDD '12: Proceedings of the 18th ACM SIGKDD international conference on Knowledge Discovery and Data mining*.

[93] A. Mukherjee, B. Liu, and N. Glance. Spotting fake reviewer groups in consumer reviews. In *WWW '12: Proceedings of the 21st international conference on World Wide Web*.

[94] T. Mullen and R. Malouf. A Preliminary Investigation into Sentiment Analysis of Informal Political Discourse. In *AAAT-CAAW '06: Proceedings of AAAI spring symposium Computational Approaches to Analyzing Weblogs*, pages 159–162. AAAI, 2006.

[95] D. Newman, C. Chemudugunta, and P. Smyth. Statistical Entity-Topic Models. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*.

[96] T. Nie, D. Shen, Y. Kou, G. Yu, and D. Yue. An Entity Relation Extraction Model Based on Semantic Pattern Matching. In *WISA '11: Proceedings of 8th international conference Web Information Systems and Applications*, volume 0, pages 7–12, Los Alamitos, CA, USA, 2011. IEEE Computer Society.

[97] B. O'Connor, R. Balasubramanyan, B. R. Routledge, and N. A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *ICWSM '10: Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, 2010.

[98] C. F. Olson. Parallel algorithms for hierarchical clustering. *Journal of Parallel Computing*, 21:1313–1325, 1995.

[99] B. Pang and L. Lee. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.

[100] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *EMNLP '02: Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing*, pages 79–86, July 2002.

[101] M. J. Paul, C. Zhai, and R. Girju. Summarizing contrastive viewpoints in opinionated text. In *EMNLP '10: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

[102] E. Pitler and A. Nenkova. Revisiting Readability: A Unified Framework for Predicting Text Quality. In *EMNLP '08: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 186–195. ACL, 2008.

[103] E. Pitler, M. Raghupathy, H. Mehta, A. Nenkova, A. Lee, and A. K. Joshi. Easily Identifiable Discourse Relations. In *COLING '08: 22nd International Conference on Computational Linguistics,(Posters)*, pages 87–90, 2008.

[104] K. Poole and H. Rosenthal. Patterns of Congressional Voting. *American Journal of Political Science*, pages 118–178, 1991.

[105] A.-M. Popescu and O. Etzioni. Extracting Product Features and Opinions from Reviews. In *HLT '05: Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing*, October 2005.

[106] I. Porteous, D. Newman, A. Ihler, A. Asuncion, P. Smyth, and M. Welling. Fast Collapsed Gibbs Sampling for Latent Dirichlet Allocation. In *KDD '08: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

[107] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. K. Joshi, and B. L. Webber. The Penn Discourse TreeBank 2.0. In *LREC '08: Proceedings of the sixth international conference on Language Resources and Evaluation*. European Language Resources Association, 2008.

[108] G. Qiu, B. Liu, J. Bu, and C. Chen. Expanding domain sentiment lexicon through double propagation. In *IJCAI '09: Proceedings of the 21st international jont conference on Artifical intelligence*.

[109] K. Raghunathan, H. Lee, S. Rangarajan, N. Chambers, M. Surdeanu, D. Jurafsky, and C. Manning. A Multi-Pass Sieve for Coreference Resolution. In *EMNLP '10: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

[110] L. A. Ramshaw and M. P. Marcus. Text Chunking using Transformation-Based Learning. *CoRR*, cmp-lg/9505040, 1995.

[111] REACH. Reactions to pm's national day rally 2010, `http://www.reach.gov.sg/portals/0/MediaRelease/Reactions_to_NDR_2010.pdf`. 2010.

[112] M. Saito, K. Yamamoto, and S. Sekine. Using Phrasal Patterns to Identify Discourse Relations. In *HLT-NAACL '06: Proceedings of the Human Language Technology Conference of the NAACL*, pages 133–136, New York City, USA, June 2006. Association for Computational Linguistics.

[113] D. K. Sil, S. H. Sengamedu, and C. Bhattacharyya. Supervised matching of comments with news article segments. In *CIKM '11: Proceedings of the 20th ACM international conference on Information and knowledge management*.

[114] D. K. Sil, S. H. Sengamedu, and C. Bhattacharyya. ReadAlong: reading articles and comments together. In *WWW '11: Proceedings of the 20th International Conference on World Wide Web*, pages 125–126. ACM, 2011.

[115] Y. Sim, B. Acree, J. H. Gross, and N. A. Smith. Measuring ideological proportions in political speeches. In *EMNLP '13: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, Seattle, WA, October 2013.

[116] W. Simm, M. A. Ferrario, S. S. Piao, J. Whittle, and P. Rayson. Classification of Short Text Comments by Sentiment and Actionability for VoiceYourView. In *SocialCom/PASSAT '10 : Proceedings of the 2010 IEEE second international conference on Social Computing*, pages 552–557, 2010.

[117] S. Somasundaran and J. Wiebe. Recognizing stances in online debates. In *ACL '09: Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

[118] S. Somasundaran and J. Wiebe. Recognizing stances in ideological on-line debates. In *CAAGET '10: Proceedings of the NAACL HLT 2010 Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 116–124, 2010.

[119] R. W. Speel. The Evolution of Republican and Democratic Ideologies. *Journal of Policy History*, 12:413–416, 7 2000.

[120] A. Sukumaran, S. Vezich, M. McHugh, and C. Nass. Normative Influences on Thoughtful Online Participation. In *CHI '11: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

[121] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. Lexicon-based methods for sentiment analysis. *Journal Computational Linguistics*, 37(2):267–307, June 2011.

[122] R. Tang, K. B. Ng, T. Strzalkowski, and P. B. Kantor. Automatically Predicting Information Quality in News Documents. In *HLT-NAACL '03: Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*, 2003.

[123] I. Titov and R. McDonald. Modeling online reviews with multi-grain topic models. In *WWW '08: Proceedings of the 17th international conference on World Wide Web*.

[124] V. Traag and J. Bruggeman. Community detection in networks with positive and negative links. *Physical Review E*, 80(3):036115, 2009.

[125] A. Tumasjan, T. O. Sprenger, P. G. Sandner, and I. M. Welpe. Predicting elections with twitter: What 140 characters reveal about political sentiment. In *ICWSM '10: roceedings of the Fourth International AAAI Conference on Weblogs and Social Media*, pages 178–185, 2010.

[126] P. D. Turney. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*.

[127] T. A. Van Dijk. *Ideology: A Multidisciplinary Approach*. Sage Publications Limited, 1998.

[128] E. M. Voorhees et al. The TREC-8 question answering track report. In *TREC '99: Proceedings of Text REtrieval Conference*, volume 8, pages 77–82, 1999.

[129] M. A. Walker, P. Anand, R. Abbott, J. E. F. Tree, C. Martell, and J. King. That is your evidence?: Classifying stance in online political debate. *Journal Decision Support Systems*, 53(4):719–729, 2012.

[130] Y. Wang. Distributed Gibbs Sampling of Latent Topic Models: The Gritty Details. Technical report, 2008.

[131] J. Whittle, W. Simm, M. A. Ferrario, K. Frankova, L. Garton, A. Woodcock, B. Nasa, J. Binner, and A. Ariyatum. VoiceYourView: Collecting Real-time Feedback on the Design of Public Spaces. In *UbiComp '10: Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, pages 41–50, 2010.

[132] J. Wiebe. Instructions for annotating opinions in newspaper articles. Department of Computer Science Technical Report TR-02-101, University of Pittsburgh, 2002.

[133] J. Wiebe and E. Riloff. Creating Subjective and Objective Sentence Classifiers from Unannotated Texts. In *CICLing '05: Proceedings of the 6th International Conference on Intelligent Text Processing and Computational Linguistics*, pages 486–497, Mexico City, Mexico, 2005.

[134] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. OpinionFinder: a system for subjectivity analysis. In *HLT-Demo '05: Proceedings of HLT/EMNLP on Interactive Demonstrations*.

[135] T. Wilson, J. Wiebe, and R. Hwa. Just how mad are you? Finding strong and weak opinion clauses. In *AAAI '04: Proceedings of the 19th national conference on Artifical intelligence*.

[136] Y. Wu, Q. Zhang, X. Huang, and L. Wu. Phrase Dependency Parsing for Opinion Mining. In *EMNLP '09: Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, pages 1533–1541, August 2009.

[137] S. Yu and S. Kak. A Survey of Prediction Using Social Media. *Journal Computing Research Repository*, abs/1203.1647, 2012.

[138] K. Zhai, J. Boyd-Graber, N. Asadi, and M. L. Alkhouja. Mr. LDA: A Flexible Large Scale Topic Modeling Package Using Variational Inference in MapReduce. In *WWW '12: Proceedings of the 21st International Conference on World Wide Web*.

[139] L. Zhang, B. Liu, J. Benkler, and C. Zhou. Finding Actionable Knowledge via Automated Comparison. In *ICDE '09: Proceedings in IEEE 25th International Conference on Data Engineering*, pages 1419–1430, 2009.

[140] M. Zhang and X. Ye. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*.

[141] W. X. Zhao, J. Jiang, H. Yan, and X. Li. Jointly modeling aspects and opinions with a MaxEnt-LDA hybrid. In *EMNLP '10: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*.

[142] Y. Zhao and G. Karypis. Evaluation of Hierarchical Clustering Algorithms for Document Datasets. In *CIKM '02: Proceedings of the eleventh international conference on Information and knowledge management*, pages 515–524. ACM Press, 2002.

[143] D. X. Zhou, P. Resnick, and Q. Mei. Classifying the Political Leaning of News Articles and Users from User Votes. In *ICWSM 11: Proceedings of 5th International AAAI Conference on Weblogs and Social Media*, 2011.

Appendix A

Annotation

In this appendix, we discuss the annotation process for the corpus used in the thesis with some examples. To encourage comparative work, the complete datasets and the corresponding annotations are released to the public download[1].

A.1 Entity and Actions

The objective of this annotation is to label the comments for suggestive opinions on entities. Table A.1 shows sample annotation for comments news articles. The human judge labels each cell of the first column with 1 (suggestion) or 0 (not suggestion). Each cell in the second column is annotated with "E" - entity and "A" - action. The other words are annotated automatically as "O" - other. This annotation is used by experiments in Chapter 4.

Labeling task: Annotators should look for the comments which are suggestions or solutions or tactics for the problem. They should read the articles before labeling the comments. Labeling steps are as follows:

1. First to check if the sentence is a suggestion from the commenter. If No , label as 0.
2. Look for the person or organization who should execute the suggestion and label the entity as shown in the samples sheet. eg: "government should immediately ban these" - The entity is "Government". eg: "He should be arrested by the police" - The entity is "Police".
3. Look out for the actions that should be performed by the entity and label it as action. E[Jack Neo] should A[not have gotten his wife to speak out at the press conference].
4. If either of them are not found then it is a ambiguous suggestion. Hence label as 0.
5. If the suggestion is very general like we, us etc., then it is also ambiguous and hence label as 0.

A.2 Issues and Quality

The objective of this annotation is to label the comments for linking(relevance) to the issues in the speech and whether they are insightful or not. Table A.2 shows a sample annotation

[1] https://sites.google.com/site/swapnagotipati/datasets

Suggestion?	Annotation
1	hope E[government] can A[come our some law to protect owners as well]
1	E[They] should A[just let the art work stay]
1	A[More strict rules should be brought in] by the E[Government]
0	But we must ensure that he DO NOT get away with murder

Table A.1: Annotation sample for comments on news articles used in entity-action extraction task (Chapter 4).

for comments on PM speech. The human judge labels each cell with relevance and quality. "Related" indicates whether the comment is related to the article (PM speech) where "1" indicates "Yes" and "0" indicates no. For each issues, the judge has to label Y/Y (relevant and thoughtful) or Y/N (relevant and thoughtless) or N/Y (irrelevant and thoughtful) or N/N (irrelevant and thoughtless). This annotation is used by experiments in Chapter 3 and Chapter 5.

Labeling task: Annotators should look for the comments which are relevant to the specific issue in the speech. Once they identify it as relevant or not, they should look out if the comment is attentive. They should read the speech document before labeling the comments.

A.3 Debates and positions

The objective of this annotation is to categorize the debates into political clusters. Table A.3 shows a sample annotation for debates from Debatepedia.com website. The human judge labels each cell of the pro and con columns with terms of political philosophies, contemporary political party platforms, or something else. This annotation is used by experiments in Chapter 6.

Labeling task: We shared a spreadsheet with the annotators that lists 22 URLs to debates in random order. Each debate has two sides, labeled "pro" and "con". The goal is to see what the annotators think about how the different sides of different debates can be lined up. The annotators might find it convenient to think of these in terms of political philosophies, contemporary political party platforms, or something else. Any of these is fine; the annotators should tell the grouping they find most reasonable. The term "ideologies" is used very loosely here; a better term might be "political perspectives". Labeling steps are as follows:

1. For each instance, go to the URL and read through the debates on both sides.
2. Assign labels to the two sides (using the spreadsheet) such that the labels are reused as much as possible. It's okay if some sides don't match nicely to any group.
3. At the end, annotator produce clustering of 44 items into hopefully fewer than 44 clusters, with hopefully not too many singleton clusters.

Sample comments for human labeling	Related	Economy	Immigration	Goh Keng Swee and Founding Fathers	YOG	Others
I agree with Financial Adviser that words such as 'Foreign Trash' is too harsh. Not being able to speak perfect English does not mean someone is worthless. The use of such words make this article seem very subjective and biased. The writer did successfully manage to point out one very important issue - PM is missing the point. =(.	0		Y/N			
Some jaws dropped when the 3 to 1 figure was mentioned isn't it? Hush silence. No clapping. Sembawang and Keppel employed a total of 15000 FWs and 5000 Sgeans. 3 to 1. This could be the trend for the future for red dot isn't it? 3 FWs and 1 SGean combination in the total national employment level. Possible? Be prepared. We wait..	1		Y/Y			
I think raising productivity is not people's key concern. Its all about raising salaries. The PAP has tried raising productivity by mass importing foreign workers. If the PAP bothers to listen they will start to rephrase key concerns correctly to citizen voters. Power Cuts Send a private message to Power Cuts ...	1	Y/Y	Y/Y			

Table A.2: Annotation sample on PM speech for relevance and quality for comment linking task (Chapter 3) and thoughtful comment classification (Chapter 5). We show 4 out of 9 issues in the PM speech.

Debate	PRO	CON
http://dbp.idebate.org/en/index.php/Debate:_Arizona_illegal_immigration_law		
http://dbp.idebate.org/en/index.php/Debate:_In_some_cases_juveniles_should_be_tried_as_adults		
http://dbp.idebate.org/en/index.php/Debate:_700_mile_US_Mexico_border_fence		
http://dbp.idebate.org/en/index.php/Debate:_Death_penalty		
http://dbp.idebate.org/en/index.php/Debate:_Israeli_military_assault_in_Gaza		
http://dbp.idebate.org/en/index.php/Debate:_Return_of_Israel_to_pre-1967_borders		
http://dbp.idebate.org/en/index.php/Debate:_Mandatory_ultrasounds_before_abortions		
http://dbp.idebate.org/en/index.php/Debate:_Civil_unions_vs._gay_marriage		
http://dbp.idebate.org/en/index.php/Debate:_Gay_marriage		
http://dbp.idebate.org/en/index.php/Debate:_Vegetarianism		
http://dbp.idebate.org/en/index.php/Debate:_Partial-birth_abortion		
http://dbp.idebate.org/en/index.php/Debate:_Legalization_of_drugs		
http://dbp.idebate.org/en/index.php/Debate:_Earmarks		
http://dbp.idebate.org/en/index.php/Debate:_The_UN_should_prioritize_poverty_over_climate_change		
http://dbp.idebate.org/en/index.php/Debate:_Jay_Inslee_vs_Rob_McKenna_for_WA_State_governor		
http://dbp.idebate.org/en/index.php/Debate:_Artificial_life		
http://dbp.idebate.org/en/index.php/Debate:_Public_health_insurance_option		
http://dbp.idebate.org/en/index.php/Debate:_Trans_fat_ban		
http://dbp.idebate.org/en/index.php/Debate:_Geoengineering		
http://dbp.idebate.org/en/index.php/Debate:_Medical_marijuana_dispensaries		
http://dbp.idebate.org/en/index.php/Debate:_Mandatory_calorie_counts_on_menus		
http://dbp.idebate.org/en/index.php/Debate:_Health_insurance_cooperatives		

Table A.3: Annotation on debates used in position inference task of debates. (Chapter 6).

Appendix B

Supplementary notes for Chapter 6

B.1 Model

Our model defines a probability distribution over words. Each word occurs in a context defined by the tuple $\langle d, q, s, a \rangle$ (respectively, a *debate*, a *question* within the debate, a *side* within the debate, and an *argument*). The details of model are explained in 6.3. Figure 6.1 illustrates the plate diagram for the graphical model underlying our approach.

B.2 Inference

Exact inference of the posterior distribution of the model is intractable. Instead, we approximate it using Gibbs sampling. As we used conjugate priors for our distributions, we can easily integrate out the dotted variables in Figure 6.1.

We refer the interested reader to [55] for details of using collapsed Gibbs sampling for LDA-like topic models.

For positions, we require that two sides of a debate to be associated with different positions. Hence, we define the joint probability $i_{d,1}, i_{d,2}$ for side 1 and side 2 of a debate as follows:

$$p(i_{d,1} = k, i_{d,2} = k' | \boldsymbol{\iota}) \propto \begin{cases} 0 & \text{if } k = k' \\ p(k \mid \boldsymbol{\iota}) p(k' \mid \boldsymbol{\iota}) & \text{if } k \neq k' \end{cases} \tag{B.1}$$

where k and k' are positions.

To sample $i_{d,s}$ for each debate d, side s, we need to consider those position words and general position words inside. We highlight the associated model parameters that we need to consider when sampling $i_{d,s}$ in Figure B.1.

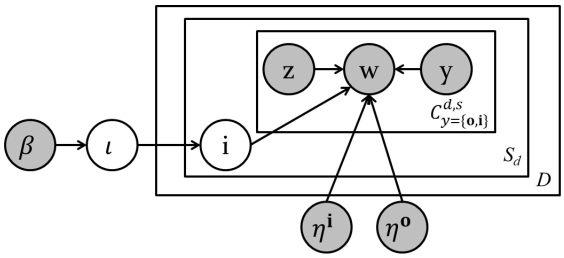

Figure B.1: Model parameters associated with position $i_{d,s}$.

We jointly sample $i_{d,1}$ and $i_{d,2}$ for two sides in debate d according to the following equation:

$$
\begin{aligned}
p(i_{d,1}=k_1, i_{d,2}=k_2 \mid \mathbf{z},\mathbf{y},\mathbf{w},\mathbf{i}_{\neg\{d,s\}},\beta,\boldsymbol{\eta}) \propto \prod_{s=1}^{2} \Bigg(& \frac{C_{k_s}^{(\cdot)}+\beta}{\sum_{i=1}^{K} C_i^{(\cdot)}+K\beta} \cdot \frac{\prod_{w=1}^{V}\prod_{a=0}^{C_{w,y=\mathsf{i}}^{d,s}-1}(C_{w,y=\mathsf{i},k_s}^{\neg\{d,s\}}+\eta_w^{\mathsf{i}}+a)}{\prod_{b=0}^{C_{y=\mathsf{i}}^{d,s}-1}(\sum_{w=1}^{V}(C_{w,y=\mathsf{i},k_s}^{\neg\{d,s\}}+\eta_w^{\mathsf{i}})+b)} \\
& \cdot \prod_{t=1}^{T} \frac{\prod_{w=1}^{V}\prod_{a=0}^{C_{w,y=\mathsf{o},t}^{d,s}-1}(C_{w,y=\mathsf{o},k_s,t}^{\neg\{d,s\}}+\eta_w^{\mathsf{o}}+a)}{\prod_{b=0}^{C_{y=\mathsf{o},t}^{d,s}-1}(\sum_{w=1}^{V}(C_{w,y=\mathsf{o},k_s,t}^{\neg\{d,s\}}+\eta_w^{\mathsf{o}})+b)} \Bigg). \qquad \text{(B.2)}
\end{aligned}
$$

where $C_i^{(\cdot)}$ denotes the number of times position i appears in arguments, $C_{w,y=\mathsf{i},i_{d,s}}^{\neg\{d,s\}}$ is the number of times word w is associated with position $i_{d,s}$ without considering words in debated d and side s, and $C_{w,y=\mathsf{o},i_{d,s},t}^{\neg\{d,s\}}$ is the number of times word w is treated as a opinion word associated with position $i_{d,s}$ and topic t without considering words in debated d and side s.

Let p denotes $\{d,q,s,a,n\}$. For a word w_p in document d, question $q \in \{1,\ldots,Q_d\}$, each side $s \in \{1,2\}$, argument $a \in \{1,\ldots,A_{d,q,s}\}$, and position $n \in \{1,\ldots,N_{d,q,s,a}\}$, we sample its corresponding topic z_p as follows:

$$
\begin{aligned}
p(z_p=t \mid \mathbf{z}_{\neg p},\mathbf{y},\mathbf{w},\mathbf{i},\alpha,\boldsymbol{\eta}) \propto & \frac{C_t^{d,q}+\alpha}{C_{(\cdot)}^{d,q}+T\alpha} \cdot \left(\frac{C_{w_p}^{y_p,t}+\eta_{w_p}^{y_p}}{C_{(\cdot)}^{y_p,t}+\sum_{w=1}^{V}\eta_w^{y_p}}\right)^{\mathbb{I}(y_p\in\{\mathsf{e},\mathsf{t}\})} \\
& \cdot \left(\frac{C_{w_p}^{\mathsf{o},t,i_{d,s}}+\eta_{w_p}^{\mathsf{o}}}{C_{(\cdot)}^{\mathsf{o},t,i_{d,s}}+\sum_{w=1}^{V}\eta_w^{\mathsf{o}}}\right)^{\mathbb{I}(y_p=\mathsf{o})}, \qquad \text{(B.3)}
\end{aligned}
$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Similarly, we sample y_p according to the following equation:

$$
\begin{aligned}
p(y_p=y \mid \mathbf{z},\mathbf{y}_{\neg p},\mathbf{w},\mathbf{i},\gamma,\boldsymbol{\eta}) \propto & \frac{C_y^{(\cdot)}+\gamma}{\sum_{y'\in\{\mathsf{b},\mathsf{a},\mathsf{o},\mathsf{i}\}} C_{y'}^{(\cdot)}+4\gamma} \cdot \left(\frac{C_{w_p}^{\mathsf{b}}+\eta_{w_p}^{\mathsf{b}}}{C_{(\cdot)}^{\mathsf{b}}+\sum_{w=1}^{V}\eta_w^{\mathsf{b}}}\right)^{\mathbb{I}(y=\mathsf{b})} \\
& \cdot \left(\frac{C_{w_p}^{\mathsf{i},i_{d,s}}+\eta_{w_p}^{\mathsf{i}}}{C_{(\cdot)}^{\mathsf{i},i_{d,s}}+\sum_{w=1}^{V}\eta_w^{\mathsf{i}}}\right)^{\mathbb{I}(y=\mathsf{i})} \cdot \left(\frac{C_{w_p}^{\mathsf{t},z_p}+\eta_{w_p}^{\mathsf{t}}}{C_{(\cdot)}^{\mathsf{t},z_p}+\sum_{w=1}^{V}\eta_w^{\mathsf{t}}}\right)^{\mathbb{I}(y=\mathsf{t})} \\
& \cdot \left(\frac{C_{w_p}^{\mathsf{o},z_p,i_{d,s}}+\eta_{w_p}^{\mathsf{o}}}{C_{(\cdot)}^{\mathsf{o},z_p,i_{d,s}}+\sum_{w=1}^{V}\eta_w^{\mathsf{o}}}\right)^{\mathbb{I}(y=\mathsf{o})}. \qquad \text{(B.4)}
\end{aligned}
$$

We do not consider $p(y_p=\mathsf{e} \mid \cdots)$ as we assume all the entities are pre-labeled.

Using Gibbs sampler, new values for $i_{d,s}$, $z_{d,q,s,a,n}$ and $y_{d,q,s,a,n}$ are iteratively sampled for each token $w_{d,q,s,a,n}$ from the posterior probability conditioned on the previous state of the sampler.

After sampling the model, we estimate the parameters as follows:

$$\phi_{i,w}^{\mathsf{i}} = \frac{C_w^{\mathsf{i}}+\eta_w^{\mathsf{i}}}{C_{(\cdot)}^{\mathsf{i}}+\sum_{w=1}^{V}\eta_w^{\mathsf{i}}}. \qquad \text{general position word distribution} \qquad \text{(B.5)}$$

$$\phi_{t,w}^{\mathsf{t}} = \frac{C_w^{\mathsf{t},t}+\eta_w^{\mathsf{t}}}{C_{(\cdot)}^{\mathsf{t},t}+\sum_{w=1}^{V}\eta_w^{\mathsf{t}}}. \qquad \text{topical word distribution} \qquad \text{(B.6)}$$

$$\phi_{t,i,w}^{\mathsf{o}} = \frac{C_w^{\mathsf{o},t,i}+\eta_w^{\mathsf{o}}}{C_{(\cdot)}^{\mathsf{o},t,i}+\sum_{w=1}^{V}\eta_w^{\mathsf{o}}}. \qquad \text{topical-position distribution} \qquad \text{(B.7)}$$

$$\phi_{t,w}^{\mathsf{e}} = \frac{C_w^{\mathsf{e}}+\eta_w^{\mathsf{e}}}{C_{(\cdot)}^{\mathsf{e}}+\sum_{w=1}^{V}\eta_w^{\mathsf{e}}}. \qquad \text{topical-entity distribution} \qquad \text{(B.8)}$$

B.3 Qualitative Analysis

As a generative modeling approach, our model was designed for the purpose of reducing the dimensionality of the sociopolitical debate space, as evidenced by Debatepedia. 37 out of 40 topics were subjectively judged to be coherent; we manually selected eleven of the most interpretable topics for further analysis here.

Topic	Terms	Person entity mentions
"Israel-Palestine"	israel, gaza, hamas, israeli, palestinian	Benjamin Netanyahu, Al Jazeera, Mavi Marmara, Nicholas Kristoff, Steven R. David
"Death penalty"	death, crime, punishment, penalty, justice	Adam Bedau, Thomas R. Eddlem, Jeff Jacoby, John Baer, Peter Bronson
"Global warming"	global, emissions, climate, carbon, warming	Alan Robock, Al Gore, Ken Caldeira, Andrew C. Revkin, George Monbiot
"Human rights"	human, rights, animals, life, animal	Tom Regan, Michael Pollan, Peter Singer, Leonardo Da Vinci, Immanuel Kant
"Healthcare"	health, care, insurance, public, private	Kent Conrad, Paul Hsieh, Paul Krugman, Ezra Klein, Jacob Hacker
"Food"	food, consumers, products, calorie, information	Steve Chapman, Jeff Jacoby, David Kiley, Jacob Sullum, Ezra Klein
"Drugs"	marijuana, drug, drugs, alcohol, age	Four Loko, Evo Morales, Toni Meyer, Sean Flynn, Robert Hahn
"Abortion"	women, religious, abortion, god, life	Ronald Reagan, John Paul II, Sara Malkani, Mother Teresa, Marcella Alsan
"Same-sex marriage"	marriage, gay, mars, space, moon	Buzz Aldrin, Andrew Sullivan, Moon Base, Scott Bidstrup, Ted Olson
"American Congress"	president, washington, obama, american, america	Barack Obama, John McCain, Bill Clinton, George W. Bush, Ronald Reagan
"Immigration"	immigration, cameras, police, immigrants, crime	Ken Garcia, Jan Brewer, Kris Kobach, Edwin S. Rubenstein, Jim Gilchrist

Table B.1: For 11 selected topics (labels assigned manually), top terms (ϕ^{t}) and person entities (ϕ^{e}).

Table B.2 shows bigrams most strongly associated with general position distributions ϕ^{i} and selected topic-position distributions ϕ^{o}. Terms are ranked by comparing the log odds conditioned on the position and topic, e.g., $\log \frac{\phi^{o}_{i_1,t,w}}{\phi^{o}_{i_2,t,w}}$. This table is comprehensive version of Table 6.7 presented in Chapter 6. We assigned labels manually. While these are somewhat internally coherent, we do not observe consistent alignment across topics, and the general distributions ϕ^{i} are not suggestive.

The separation of personal name mentions into their own distributions, shown in Table B.1, gives a distinctive characterization of topics based on relevant personalities. This table is comprehensive version of Table 6.8 presented in Chapter 6. Bigrams were included but did not rank in the top five for these topics. The model has conflated debates relating to same-sex marriage with the space program. Subjectively, the top individuals are relevant to the subject matter associated with each topic (though the topics are not always pure; same-sex marriage and the space program are merged, for example). Our model incorrectly linked some entities (false positives) in the corresponding topic. For example, Ezra Klein is not related to the *food* topic as he is a *Washington Post* journalist specializing in health care and budget policy.

Topic	$i = 1$	$i = 2$
None (ϕ^{i})	vice president, c sections, twenty four, cross pressures, pre dates, anti ballistic, cost effectiveness, anti landmine, court appointed, child poverty	cross examination, under runs, hand outs, half million, non christians, break down, counter argument, seventy five, co workers, run up
"Israel-Palestine"	pre emptive, israeli palestinian, open and shut, first time, hamas controlled, democratically elected, knee jerk	two state, long term, self destructive, secretary general, right wing, all out, near daily, short term, life threatening
"Death penalty"	anti death, non violent, african american, self help, cut and cover, heavy handed, dp equivalent, law breaking	semi automatic, high profile, hate crime, assault weapons, military style, high dollar, self protective, state authorized
"Global warming"	cap and trade, long term, blue ribbon, fossil fuel, sunspot driven, forest based, short lived, anti nuclear	non profit, large scale, half degree, climate change, low carbon, non compliance, human caused, opt in, multi pollutant, inter glacial
"Human rights"	self legislative, life saving, non human, self restricting, auto nomous, self conscious, god given, one another	cost benefit, non animal, cock fighting, bull baiting, self centered, peace loving, non emotional, pan european, state invested, pleasure pain
"Healthcare"	single payer, so called, self sustaining, public private, for profit, long run, high cost, multi payer, government funded	government run, government approved, high risk, two tier, government appointed, low cost, set up, one sixth, draft age
"Food"	health care, health conscious, low cost, point of, reduced fat, time consuming, multi billion, mid range, miracle diet	force fed, trans fat, anti obesity, ill informed, non gm, medium sized, cajun lime, impossible to ignore, well seasoned, fat free
"Drugs"	hands free, performance enhancing, in depth, hand held, best kept, non pharmaceutical, anti marijuana, non toxic, marijuana related	long term, high speed, short term, peer reviewed, alcohol related, mind altering, inner city, long lasting, needle exchange, anti drug
"Abortion"	pro choice, pro life, non muslim, well educated, anti abortion, much needed, church state, birth control, fully informed	would be, full time, late term, judeo christian, life style, day to day, non christian, child bearing, non religious
"Same-sex marriage"	same sex, long term, second class, blankenhorn rauch, wrong headed, self denial, left handed, single parent	opposite sex, well intentioned, day time, planet wide, day night, child rearing, low earth, one way, one third, life bearing
"American Congress"	op ed, state sponsored, fear mongering, on the job, anti earmark, oil rich, lower level, sixty seven, ultra conservative	left wing, smoot hawley, party line, self indulgent, un american, off target, republican controlled, reagan bush
"Immigration"	law abiding, anti social, high profile, american born, one way, hard won, present day, crime solving, high mast	in state, anti crime, low paid, so called, taxpayer funded, out of state, anti immigrant, closed circuit, un american, clear up

Table B.2: General position (first row) and topic-specific position bigrams associated with eleven selected topics.