



MiMuSA—mimicking human language understanding for fine-grained multi-class sentiment analysis

Zhaoxia Wang¹ · Zhenda Hu² · Seng-Beng Ho³ · Erik Cambria⁴ · Ah-Hwee Tan¹

Received: 30 June 2022 / Accepted: 5 April 2023 / Published online: 19 April 2023
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2023

Abstract

Sentiment analysis is an important natural language processing (NLP) task due to a wide range of applications. Most existing sentiment analysis techniques are limited to the analysis carried out at the aggregate level, merely providing negative, neutral and positive sentiments. The latest deep learning-based methods have been leveraged to provide more than three sentiment classes. However, such learning-based methods are still black-box-based methods rather than explainable language processing methods. To address this gap, this paper proposes a new explainable fine-grained multi-class sentiment analysis method, namely MiMuSA, which mimics the human language understanding processes. The proposed method involves a multi-level modular structure designed to mimic human's language understanding processes, e.g., ambivalence handling process, sentiment strength handling process, etc. Specifically, multiple knowledge bases including Basic Knowledge Base, Negation and Special Knowledge Base, Sarcasm Rule and Adversative Knowledge Base, and Sentiment Strength Knowledge Base are built to support the sentiment understanding process. Compared with other multi-class sentiment analysis methods, this method not only identifies positive or negative sentiments, but can also understand fine-grained multi-class sentiments, such as the degree of positivity (e.g., strongly positive or slightly positive) and the degree of negativity (e.g., slightly negative or strongly negative) of the sentiments involved. The experimental results demonstrate that the proposed MiMuSA outperforms other existing multi-class sentiment analysis methods in terms of accuracy and F1-Score.

Keywords Human-like understanding · Fine-grained sentiment understanding · Multi-class sentiment analysis · Sentiment strength · Explainable sentiment understanding · Sarcasm handling · Knowledge base · Multi-level modular structure

1 Introduction

Sentiment analysis is a natural language processing (NLP) task that aims to identify or study sentiments, opinions, subjective information or attitude hidden in human communication [1]. Sentiment analysis has become increasingly important due to a wide range of applications, e.g., to address companies' eagerness in seeking to know about users' sentiments, to collect opinions or attitudes toward various services and products, etc. [2–4]. It is also a branch of affective computing research that aims to classify human communication data, such as text, audio and video into positive or negative polarity [5–7]. It has been applied to different fields with different applications.

Most of sentiment analysis methods merely identify sentiment polarity at the aggregate level, e.g., positive, negative or neutral [8–16]. Some of them even consider sentiment analysis as a mere binary classification problem (positive vs. negative). Compared to aggregate-level sentiment analysis, some previous work proposed a kind of fine-grained sentiment analysis which can yield more specific fine-grained results, such as characterizing sentiments into finer subcategories such as anxiety, sadness and anger for negative sentiments or emotions, and excitement and happiness for positive sentiments or emotions [17]. Such fine-grained sentiment analysis methods are good attempts to identify emotions [17, 18]. However, this is not the kind of fine-grained sentiment analysis that this research aims to address. Wang et al. [19] introduced multi-level fine-scaled sentiment sensing methods;

however, their experimental results were still aggregate-level sentiment analysis. To implement fine-grained multi-class sentiment analysis for more accurate sentiment identification and more extensive application, this research aims to identify the degree of the sentiments involved (e.g., strongly positive or slightly negative).

Deep learning (DL) techniques have been leveraged for sentiment analysis and some of the works consider multi-class sentiment classification, but they are still black-box methods and unexplainable [13, 14]. The dependence on large labeled training data is the other limitation of the applications of deep learning methods especially for classification tasks. Therefore, to address the issue of unexplainability, we develop an algorithm to mimic the human language understanding process and hence improve the explainability of the sentiment analysis models.

In this paper, using conceptual dependency as the theoretical basis for human language understanding process [20–22], we address the gap by proposing a new method—human-like fine-grained multi-class sentiment understanding. It not only overcomes the issues of unexplainability of the learning-based methods, but also implements human-like fine-grained multi-class sentiment understanding through mimicking the processes of how humans understand languages.

The main contributions of this paper are summarized as follows:

1. Novel main algorithm: This paper proposes MiMuSA, a method that mimics the language understanding process of human beings. It is an explainable fine-grained multi-class sentiment analysis method which builds various knowledge bases to overcome the limitation of aggregate-level sentiment analysis, including providing different sentiment strength levels.
2. Multi-knowledge base representations: These knowledge bases are built according to human’s multi-level knowledge acquisition process. These knowledge bases include Basic Knowledge Base, Local Language Knowledge Base, Negation and Special Knowledge, Sarcasm Rule, Adversative Base, Amplifier & Diminisher Knowledge Base, etc.
3. Multi-level modular structure designs: Multi-level modular functional designs are implemented, which mimics human’s language understanding processes, e.g., ambivalence handling process, handling of different sentiment strength.
4. Experiment on the fine-grained ground truth data: Besides leveraging the existing datasets, a new fine-grained multi-class sentiment ground truth data in the transportation domain crawled from Reddit.PRAW is built, through consistent agreement among the human subjects. Such ground truth dataset enriches the multi-

class sentiment dataset and provides a new comparison criterion for other research and researchers.

This paper is organized as follows. Section 2 discusses and analyzes the existing related work. In Sect. 3, the proposed MiMuSA method is presented in details. The datasets used are described in Sect. 4. In Sect. 5, the experiments comparing MiMuSA with the existing methods are presented. Lastly, we conclude the work in Sect. 6.

2 Related work

A fair amount of research work, which claimed multi-class or multi-level sentiment analysis, has been done [8–14, 23–25]. However, if a method produces only 3-class sentiments (such as positive, negative and neutral), or 4-class sentiments (such as positive, negative, ambivalence/mixed and neutral), it is still considered an aggregate-level method, because it basically provides only the polarity. True multi-class sentiment analysis must be able to produce a finer distinction by providing the associated strengths such as strongly positive, slightly positive, neutral, slightly negative and strongly negative. This means that there should be at least five levels of distinguishable sentiment categories.

Liu et al. investigated multi-class sentiment classification comparing feature selection strategies through different machine learning algorithms [10]. The results demonstrated that in terms of classification accuracy, different feature selection algorithms could enhance the performance of different learning-based methods. Such results are consistent with the previous work [11, 12]. However, true multi-class sentiment analysis, such as five or more than five multi-class sentiment identification tasks, was not part of their study.

There are some research works which mentioned multi-class sentiment analysis [26–37]. However, all of them in fact focused on aggregate level of sentiment analysis without considering the strength of the positivity or negativity. They did not handle five or more than five multi-level or multi-class sentiment analysis.

For example, Xiong et al. [38] proposed Twitter sentiment classification methods by using multi-level sentiment-enriched word embeddings. The proposed method is a learning-based method considering word level sentiment and tweet level sentiment in the learning process. It successfully detected the sentiment polarity toward different subtasks, such as expression-level and message-level subtasks. However, their multi-level sentiment analysis is still aggregate-level sentiment analysis without considering the strength of the positivity or negativity.

There are multi-level or multi-class sentiment analysis methods reported for identifying more than three sentiment classes [17–19, 39–43]. Bouazizi and Ohtsuki proposed a pattern-based approach for multi-class sentiment analysis for Twitter data, named SANTA. The method, SANTA, classifies the Twitter texts into one out of seven classes: “love,” “happiness,” “fun,” “neutral,” “hate,” “sadness” and “anger” [18]. Their results are consistent with previous work that showed that sentiments and emotions can be both properly identified [17, 44]. Even though multi-level or multi-class sentiment analysis was conducted in their research, the strength of the positivity or negativity was not considered in their work. Kocon et al. proposed a multi-level sentiment analysis method for the specific dataset, named PolEmo. 2.0 [41]. PolEmo 1.0 is a corpus of consumer reviews from four domains: medicine, hotels, products and school.

DL techniques have also been leveraged for sentiment classification tasks [13, 14, 34, 45, 46] and some research works utilized commonsense reasoning to enhance sentiment analysis tasks [2, 47]. Syaekhoni et al. utilized several popular DL models, such as convolutional neural network (CNN), long short-term memory (LSTM) and multilayer neural network models, and they proved that LSTM performed better than other DL models in their research [13]. Alzanzami et al. [14] built a general multi-class sentiment classifier using Domain-Free Sentiment Multimedia Dataset (DFSMD). They utilized light gradient boosting machine (LGBM) to recognize the sentiments of tweets in handling high-dimensional and imbalanced data. Liang et al. [34] developed a graph convolutional network (GCN) on the basis of SenticNet to exploit the affective dependencies for the specific aspect.

Such learning-based methods have been proved feasible if the large training dataset is available. However, the training datasets are not always available for such multi-level or multi-class sentiment analysis tasks. In addition, such learning methods still represent black-box methods and they are unexplainable due to the unexplainable nature of the DL models [48]. It is believed that the insight into the models provided by the human understandable form of knowledge (e.g., in the form of rules and cases) can bring an extra benefit to the users [49].

Summarizing the existing multi-level or multi-class sentiment analysis, it is found that whether they are learning-based methods (e.g., DL), non-learning-based methods (e.g., lexical-based methods) or hybrid methods, there are gaps and limitations. For learning-based methods, labeled training datasets are required for achieving an acceptable level of performance for sentiment classification problems. Especially, when the number of classes of sentiment is more than 4 (e.g., five classes of sentiments), such black-box learning methods suffer from the shortcoming of

the dependency on training datasets [6]. The existing hybrid methods still share the same shortcomings as the learning-based methods [17–19].

Even though the issue of labeled dataset is not of concern for existing non-learning-based methods, such as the lexicon-based methods, the challenge for these methods is how to conduct human-like explainable sentiment analysis. For example, for negative sentiment understanding, how can machines understand the degree of negativity (strongly or slightly negative) just like what humans do. This is interesting and challenging work, which is what this research aims to address.

This paper addresses the challenges by proposing a new method (MiMuSA): human-like fine-grained multi-class sentiment analysis. MiMuSA not only overcomes the unexplainability issues of the learning-based methods, but also implements fine-grained multi-class sentiment understanding through mimicking the human language understanding process.

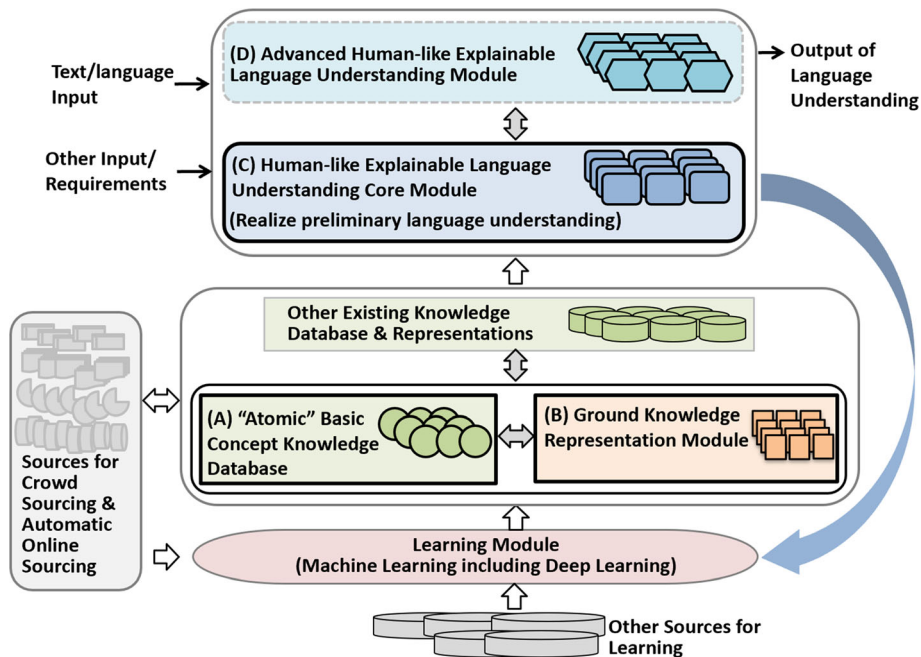
3 Proposed MiMuSA

3.1 The overall design of the proposed MiMuSA

The proposed human-like explainable fine-grained multi-level sentiment analysis method is a subtask of human-like explainable language understanding. The overall framework of the human-like explainable language understanding method is implemented through four important modules/tasks as shown in Fig. 1.

The first module is “Atomic” Basic Concept Knowledge Database Module (module (A) in Fig. 1). The ground concepts such as the basic temporal or sequence information (e.g., ordered sequence information) and spatial information (e.g., location information) are constructed [50]. This knowledge base is constructed through a process of crowd sourcing and automatic online sourcing. The second module is the Ground Knowledge Representation Module (module (B)) [50–52], which is to realize ground concept representation or “atomic” basic concept representation. This knowledge representation scheme is constructed through a process of crowd sourcing and automatic online sourcing. Machine learning including DL methods are used to enhance the knowledge representation. The third module is Human-like Explainable Language Understanding Core Module (module (C)). Based on “atomic” basic concepts and the knowledge representation constructed in the first and second modules, it converts a language-dependent surface sentential structure into a language independent deep-level predicate representation which is related to our physical world [51]. It implements the language understanding processes to realize the

Fig. 1 The overall framework of human-like explainable language understanding



preliminary human-like explainable language understanding methods. The proposed MiMuSA, which is a human-like explainable fine-grained multi-level sentiment analysis method, is specially designed for a sentiment understanding task. It is a subtask and a simplified version of a human-like explainable language understanding method. The fourth module, module (D) in Fig. 1, is the Advanced Human-like Explainable Language Understanding Module. It converts the predicate representation into grounded real-world references and constructs [52]. The implementation of this advanced explainable language understanding process to enable robots to carry out language instructions accordingly is what AI and NLP scientists had wanted to do all along [22, 52].

As discussed by Schank and Abelson, to understand the full story contained within sentences is to mimic what humans do for language understanding [22]. Inspiring by Schank and Abelson’s work [20, 22], the proposed MiMuSA mimics the language understanding processes of human beings for sentiment analysis tasks. A multi-level hierarchical modular design is the main characteristics of the proposed method as shown in Fig. 2.

Figure 2 shows the overall design of the proposed MiMuSA. The main module A, Human-like Fine-grained Multi-level Explainable Sentiment Analysis Module, is a multi-level hierarchical designed including two main sub-modules: A1, Aggregate-Level Sentiment Identification, which is the foundation for submodule, A2, Fine-grained Multi-Level Sentiment Identification, which is the Core Module for Human-like Fine-grained Multi-level Explainable Sentiment Identification.

To support the main module, Human-like Fine-grained Multi-level Explainable Sentiment Analysis Module (A), a Knowledge Base Module (B) which includes different knowledge bases is built. These knowledge bases include B0, Basic Knowledge Base (e.g., Standard English Lexicon Dictionary); B1, Local Language Knowledge Base (e.g., Singlish); B2, Negation and Special Knowledge (e.g., Negation, Special Lexicon); B3, Domain Knowledge Base (e.g., Transport Domain Lexicon); B4, Sarcasm Rule & Adversative Base, and B5, Sentiment Strength Knowledge Base, etc.

Below we summarize the modules discussed above:

A. Human-like Fine-grained Multi-level Explainable Sentiment Analysis Module, which includes the two main submodules:

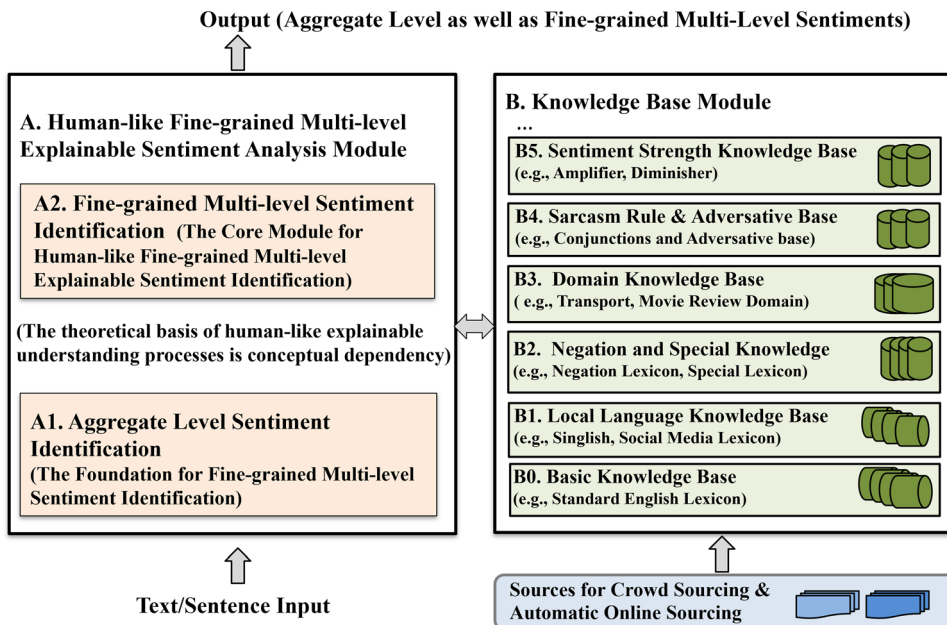
- A1. Aggregate-Level Sentiment Identification.
- A2. Fine-grained Multi-Level Sentiment Identification.

B. Knowledge Base Module, which include six main knowledge bases:

- B0. Basic Knowledge Base
- B1. Local Language Knowledge Base including Social Media Lexicon.
- B2. Negation and Special Knowledge Base
- B3. Domain Knowledge Base
- B4. Sarcasm Rule & Adversative Knowledge Base
- B5. Sentiment Strength Knowledge Base

The modules A1 and A2 reflect the various stages of human reasoning when carrying out the process of sentiment analysis and understanding. They contain submodules of different functions to mimic the language understanding

Fig. 2 The overall design of the proposed MiMuSA—the modules designed captures what humans do for understanding sentiment



processes of human beings, such as ambivalence handling, adversarial sarcasm identification and sentiment strength detection. The different functional modules are the key functional modules for realizing fine-grained multi-class sentiment analysis. These functions are implemented within the framework of the proposed MiMuSA. Not only positive or negative sentiments can be identified, fine-grained multi-class sentiments, such as the degree of positivity (e.g., strongly positive or slightly positive) and the degree of negativity (e.g., slightly negative or strongly negative) of the sentiments involved, can also be identified.

These knowledge bases (B0, B1, B2, B3, B4 & B5) are built according to human’s multi-level knowledge acquisition process. Basic Knowledge Base (B0) contains standard English sentiment words or phrases, and Local Language Knowledge Base (B1) contains sentiment words or phrases of local language. Negation and Special Knowledge Base (B2) covers all the negation words and many special words which represent special meanings. Domain Knowledge Base (B3) contains sentiment words or phrases in the specific domains (e.g., Transport, Movie). Sarcasm Rule & Adversative Knowledge Base (B4) contains Sarcasm Rules and ambivalence indicators for ambivalence handling, while Sentiment Strength Knowledge Base (B5) contains many strength-level indicators (e.g., very, worse and worst) for sentiment strength handling.

3.2 Theoretical basis of human language understanding processes for MiMuSA

To implement human-like explainable sentiment analysis, the proposed MiMuSA mimics what humans do for understanding the sentiment of a piece of text. The theoretical basis of it is conceptual dependency, which is a theory of human-like explainable representation of the meaning of sentences [22]. One of the basic axioms of Schank and Abelson’s theory is “any information in a sentence that is implicit must be made explicit in the representation of the meaning of that sentence” [22].

Therefore, considering sentence sentiment understanding, knowing the meaning (e.g., sentiments) of each element or component of sentence is a necessary step for sentiment identification and understanding, which can be implemented by utilizing the various kinds of knowledge (see Knowledge Base Module in Fig. 2).

Generally, for sentiment understanding or analysis tasks, given a piece of comment, human beings can identify the aggregate-level sentiment meaning first (e.g., positive, negative, neutral). After identifying the aggregate-level sentiment, the degree of the polarity, or fine-grained level (e.g., strongly positive, or slightly positive for a positive comment; strongly negative, or slightly negative for a negative comment) can then be identified accordingly [17].

Therefore, the first step is Aggregate-Level Sentiment Identification, followed by fine-grained sentiment identification to realize multi-class fine-gained sentiment understanding.

3.3 Extending aggregate-level sentiment identification

For a piece of text data including several sentences, sentiment analysis is performed on each opinion sentence. The paragraph level and article level sentiment analysis are carried out through “sum” methods [19]: simply counting the number of positive and/or negative sentences or leveraging on the fuzzy sum based on the adaptive fuzzy inference algorithm [17, 19].

For designing the Human-like explainable Multi-Level Sentiment Identification, the basic concepts used in previous work [19] lay the experimental foundation for the proposed idea in this paper. Typical aggregate-level analysis produces three levels of positive, negative and neutral sentiments. Extended aggregate-level analysis that includes ambivalence sentiment can produce up to four or six levels of sentiments. As shown in Table 1, the extended aggregate-level sentiments are defined and explained [19].

This paper follows Schank and Abelson’ work [22], using sentences or short texts to showcase the procedure. The extended aggregate-level sentiments can be categorized into three classes, four classes and six classes, according to human beings’ language understanding processes. For example, ambivalence is a category, which contains both positive and negative sentiments. Ambivalence category can be categorized into three classes: Mixed-Negative, Mixed-Positive and Mixed-Neutral (or Mixed-Equal) sentiments as shown in Table 1.

Humans can further understand the nature of the ambivalence of Mixed-Positive, Mixed-Negative and Mixed-Neutral as shown in Table 1 in terms of whether they are finally toward positive or negative [17, 19]. Hence, such three ambivalent subcategories can be further categorized into one of common aggregate-level sentiments such as: negative and positive. Mixed-Positive, with a

stronger weight of positive, will be further categorized into positive. Mixed-Equal, which seems to have equal weight of each sentiment polarity, will instead be further categorized into positive or negative, rather than neutral. It is easy for human beings to understand that if there are positive and negative sentiment expressed in a comment, it will never be neutral as we define neutral to mean that in the comment, there is neither positive nor negative sentiment present [17, 19].

The concepts above shown in Table 1 lay the foundation for the human-like fine-grained multi-class sentiment analysis method. For example, further analysis of the three ambivalence categories: Mixed-Positive, Mixed-Negative, and Mixed-Equal. It is easy for human beings to be able to tell that Mixed-Positive, with a stronger weight of positive, should be positive and Mixed-Negative, with a stronger weight of negative, should be negative [19].

Regarding the Mixed-Equal sentiment, which seems to express equal weight of positive and negative information, it should not be treated as neutral sentiment according to the definition of neutral, because neutral implies neither positive nor negative sentiments. In the case of Mixed-Equal, such is not the case—there are no positive and no negative sentiments in the text [19].

The work done above on enhanced Aggregate-Level Sentiment Identification lays the experimental foundation for the proposed human-like fine-grained sentiment analysis method. Based on the theoretical analysis of human-like explainable understanding processes (Sect. 3.2) and the extended Aggregate-Level Sentiment Identification method (Sect. 3.3), fine-grained multi-class sentiment identification is described in the next subsection.

Table 1 Extended aggregate sentiment category, definition, and explanation [19]

Sentiment categories		Definition and explanations according to human’s language understanding process
4 categories	6 categories	
Neutral	Neutral	Neither positive nor negative sentiments. There are no positive and no negative sentiments, only neutral statement in the text.
Negative	Negative	Contains only negative sentiments. There is only negative sentiments and no positive comments in the text.
Positive	Positive	Contains only positive sentiments. There is only positive comments and no negative comments in the text.
Ambivalence	Mixed-Negative	Contains both positive and negative sentiments, but with a stronger weightage of negative sentiments.
	Mixed-Positive	Contains both positive and negative sentiments, but with a stronger weightage of positive sentiments.
	Mixed-Neutral or	Contains both positive and negative sentiments, seems to have equal weightage of each; or difficult to tell which one is stronger before doing deeper analysis.
	Mixed-Equal	

3.4 Implementation of MiMuSA for fine-grained multi-class sentiment identification

Based on the Aggregate-Level Sentiment Identification module in Sect. 3.3 and the theoretical basis of human language understanding processes in Sect. 3.2, a fine-grained multi-class sentiment identification algorithm is designed and implemented. The mathematical description as well as the detailed implementations will be detailed in this section.

3.4.1 The mathematical description

The proposed fine-grained Multi-Level Sentiment Identification algorithm mimics human being’s language understanding process. Such language understanding process enables the machines to answer the questions on whether the different components of the sentence reflect positive or negative sentiment, such as whether the sentiment about

the Actor (Subject) is positive, neutral or negative, whether the sentiment about the Action (Predicate) is positive, neutral or negative, whether the sentiment about the Object is positive, neutral or negative, etc.

This paper details the simplified version of the proposed MiMuSA, which considers the overall sentiment of the whole sentence (in fact, MiMuSA can provide the answers for the sentiment of each of the different components (e.g., Actor (Subject), Action (Predicate), Object or State) separately).

For the simplified version, each piece of text (e.g., a paragraph or an article) can be represented by a series of opinion components. This is represented as a series of vectors, $O = \{o_1, o_2, \dots, o_i, \dots, o_N\}$, where o_i is the i th opinion component. Each opinion component $o_i \in O$ consists of a finite sequence of words, phrases or their abbreviations. The process of fine-grained sentiment identification for each opinion component is shown in Algorithm 1.

Algorithm 1: Fine-grained Multi-class Sentiment Identification

Input: An opinion component (e.g., a sentence)
Output: The final sentiment score vector, C

- 1 After data cleaning, the component vector, $W = \{w_1, w_2, \dots, w_j, \dots, w_K\}$ is obtained;
- 2 **while** $j \leq K$ **do**
- 3 **if** w_j in $B0, B1$ or $B3$ **then**
- 4 Determine the polarity of w_j (-1 for negative or 1 for positive);
- 5 **if** w_j in "Word to Neutral" Knowledge Base **then**
- 6 $wtn_j = -1$
- 7 **end**
- 8 **else**
- 9 $m_j = 0$;
- 10 **end**
- 11 **if** w_j in $B2$. Negation Knowledge Base **then**
- 12 $n_j = 1$ and conduct negation as well as special handling;
- 13 **end**
- 14 **if** w_j in $B4$. Adversative Base **then**
- 15 Determine sa_j is "before CONJ indicator" or "after CONJ indicator" and conduct sarcasm as well as adversative handling ;
- 16 **end**
- 17 **if** w_j in $B5$. Sentiment Strength Base **then**
- 18 Determine whether the strength indicator s_j is α, β or δ , and conduct sentiment strength handling;
- 19 **end**
- 20 **end**
- 21 Obtain the final sentiment score C with Intermediate Sentiment Vector M , Negation Vector N , Sentiment Reverse Vector WTN , Sarcasm and Adversative Vector SA and Sentiment Strength Vector S

Each opinion component (e.g., a sentence), o_i , can be represented by a vector, $W = \{w_1, w_2, \dots, w_j, \dots, w_K\}$, where w_j is the j^{th} word or phrase component. Each component $w_j \in W$ consists of a finite sequence of words, phrases or their abbreviations.

$M = \{m_1, m_2, \dots, m_j, \dots, m_K\}$, represents intermediate sentiment categories, where m_j is the sentiment class of the j^{th} component. Each $m_j \in M$ is one of the sentiment categories. For aggregate-level sentiment analysis, m_j can be positive (1), negative (-1) or neutral (0). M can be obtained through a matching method using the knowledge base we have built.

N represents the negation vector, $N = \{n_1, n_2, \dots, n_j, \dots, n_K\}$, where n_j is the negation category of the j^{th} component. Each $n_j \in N$ is a negation indicator. n_j can be negation (1) or not negation (0).

The WTN (“Word to Neutral”) vector represents the polarity change situation as explained in Sect. 3.4.3 “Sentiment identification with negation as well as special handling.”

$WTN = \{wtn_1, wtn_2, \dots, wtn_j, \dots, wtn_K\}$, where wtn_j is the j^{th} component of the WTN vector. Each $wtn_i \in WTN$ is a polarity change indicator. The value of wtn_j can be -1 or 0. Value “ -1 ” means the polarity of the component is reversed by the negation; “0” means that the meaning is not reversed, but the sentiment polarity will be very weak due to the negation before it.

The wtn value for the component “hate” is 0 (not reversed), and the wtn value for the component “pretty” is -1 (reversed). If a component has a wtn value of -1 , negation will reverse the polarity of the combined component (see examples (1) and (2) below).

If a component has a wtn value of 0, negation will convert the polarity of the combined component to “not positive and not negative, either” (see examples (3) and (4) below).

SA represents sarcasm and ambivalence indicator vector, $SA = \{sa_1, sa_2, \dots, sa_j, \dots, sa_K\}$, where sa_j is indicator of the j^{th} component. Each $sa_j \in SA$ can be “before CONJ indicator” (1), “after CONJ indicator” (-1), or “sarcasm indicator” (0). (“CONJ” means “conjunction component” in the sentence.) The detailed description of these are provided in Sect. 3.4.4 “Sarcasm as well as adversative handling for ambivalence handling” below.

SA can be obtained by leveraging vectors N and M through a negation handling function and a sarcasm handling function, which will be described in Sections 3.4.3 and 3.4.4.

S represents sentiment strength vector of each opinion component of W , $S = \{s_1, s_2, \dots, s_i, \dots, s_K\}$, where s_j is the strength level of w_{ij} . $s_j \in S$ is one of four strength

categories which are predefined, and will be detailed in Sect. 3.4.5, “Sentiment Strength Handling.”

The final sentiment score vector, C , is obtained using the vectors above through mimicking human being’s language processing rules as detailed in Sects. 3.4.2, 3.4.3, 3.4.4 and 3.4.5.

3.4.2 Knowledge setup with data cleaning

According to human being’s language understanding processes, the meaning or the polarity of certain words/phases may be changed when they are compared with the knowledge in the Basic Knowledge Base (B0 in Fig. 2). Therefore, when the proposed MiMuSA is applied to identify the sentiments of the reviews in certain domains, compared to the Basic Knowledge Base (B0 in Fig. 2), the domain knowledge in the Domain Knowledge Base (B3 in Fig. 2) has a higher priority. In other words, if a word or phase in the text is found in both the Basic Knowledge Base and Domain Knowledge Base, the polarity or the meaning from the latter will overwrite that from the former.

For all the dataset, data cleaning is conducted by doing the following: Delete all URLs, email addresses, quotations and tags; Delete all words with “&” or “@” characters; Clean up all “\n” to avoid unnecessary spaces; Replace multiple whitespaces or non-visible characters (such as tabs) with one space; Trim leading and trailing whitespaces.

3.4.3 Negation as well as special handling

Negation and Special Knowledge Base (B2 in Fig. 2) is built to support the negation and special handling function of MiMuSA. When assigning polarity to a word/phase/sentence with negation in the sequence, it can result in two possible outcomes. For examples:

1. She is pretty \rightarrow positive
2. She is not pretty \rightarrow negative
3. I hate this brand \rightarrow negative
4. I do not hate this brand \rightarrow not positive, but not negative either.

where “not pretty” is a negation item followed by a positive item “pretty,” and the phrase is negative. “do not hate” is negation followed by a negative item, but the phrase is not positive.

Besides negation handling, a special handling function is designed to handle the special cases. In fact, such special knowledge is common sense. For example, the two sentences listed below illustrate this function:

5. He like this brand \rightarrow In this case “like” is positive

-
6. He looks like his mother → In this case “like” is neutral

For a sentence component such as “like,” the lexical analysis, such as part of speech (POS) and semantic analysis, is leveraged to support this special handling function. When the POS of “like” is not a verb, the special handler is triggered accordingly.

Other examples of special cases are the misspelling cases. Lexical items, such as “gooooood” and “baaaaaaad,” are treated as “very good” and “very bad” rather than errors or mistakes. They are strengthened forms, compared to base line forms of “good” or “bad.”

Therefore, such special handling functions are also built in the proposed MiMuSA as submodules. When the special cases are detected, the function will be triggered.

In such a way, the overall polarity of the sentence is determined after considering the negation handling as well as special case handling. As shown in Fig. 2, the Negation and Special Knowledge Base (B2) has been built to support the negation handling with special case handling in this research as shown in Algorithm 1.

3.4.4 Sarcasm rule and adversative base for ambivalence handling

Sarcasm handling and adversative handling are important steps for ambivalence handling [53, 54]. This draws knowledge from the Sarcasm Rules and Adversative Base (B4 in Fig. 2)

Sarcasm is commonly used by human language users, and it can be easily detected by using Sarcasm Rules designed in this paper. The general Sarcasm Rules are in the form of the detecting certain sequences of various types of text components. This paper lists two rules here:

Rule 1: [positive, negative, without proper adversative conjunction present]

Rule 2: [negative, positive, without proper adversative conjunction present]

For example:

7. The bad guy broke his arm, he was so lucky. (Sarcasm)
8. The thief is really smart. (Sarcasm)

If any of the above sequence or rules is found in the text, it will trigger sarcasm handling. The polarity of such sarcasm will be negative.

However, another situation must be considered, for example:

9. I like taking the train although it’s crowded. (Slightly positive)
10. The train is a bit delayed but I’m thankful. (Slightly positive)

11. He was so lucky even though he broke his phone. (Not sarcasm, implies positive event and it implies he may get a new phone)

The sentences (9), (10) and (11) above do not satisfy the Sarcasm Rules. Therefore, the adversative conjunction handling or adversative handling (also named CONJ Handling function) is designed to handle such situations.

It is discovered that such ambivalence sentences (including both positive and negative sentiments) usually contain conjunction phases or similar function words such as “although” and “but.”

Two types of conjunction phases (named as “before CONJ” and “after CONJ”) are handled separately, which are illustrated using the examples below:

12. I like taking the train although it’s so crowded (In this case, we name it as “before CONJ” case: the part before the conjunction phase matters more and hence the polarity of the sentence is tending to positive)
13. I like taking the train but it’s so crowded (Whereas in this case, we name it as “after CONJ” case: the part after the conjunction words matters more and hence the polarity of the sentence is tending to negative.)
14. The train is a bit delayed but I’m thankful. (In this case, it is an “after CONJ” situation: the part after the conjunction word matters more and hence the polarity of the sentence is positive.)
15. The train is a bit delayed even though I’m thankful. (Whereas in this case, it is a “before CONJ” case: the part before the conjunction word matters more and hence the polarity of the sentence is tending to negative.)

Hence, identification of the type of “conjunction” enables MiMuSA to determine which part of the sentence should be prioritized to determine the sentiment polarity of the sentence.

A knowledge base (B4. Sarcasm Rule & Adversative Base in Fig. 2) has been built including both types of “conjunction” phrases. If the “conjunction” is type “before CONJ” (see example (12)), MiMuSA will prioritize the polarity of the phrase before the “conjunction.” If the type is “after CONJ” (see example (13)), MiMuSA will prioritize the polarity of the phrase after the “conjunction.”

Conjunction handling (adversative handling) is designed together with sarcasm handling to realize the ambivalence handling function.

3.4.5 Sentiment strength handling

Sentiment strength handling is another core module for the multi-class sentiment identification function (utilizing

knowledge in the Sentiment Strength Knowledge Base—B5 in Fig. 2). Companies or individuals may want to know the intensity of the sentiment (i.e., how positive or how negative the text/sentence is). This requires fine-grained multi-class sentiment analysis that considers the sentiment degree or strength. For examples,

16. He loves this brand → positive
17. He loves this brand very much → strongly positive
18. The film is good → positive
19. The film is damn good → strongly positive

where human beings will identify “love” and “good” as positive sentiment first, and then understand that “love... very much” represents stronger positive sentiment than “love”; and “damn good” represents stronger positive sentiment than “good.”

To enable this capability, sentiment strength handling function is designed. It mimics how human beings understand the text message by using the strength-level indicators (e.g., very, best, worse and worst). An amplifier and diminisher database which contains sentiment strength indicators is built to support the implementation of sentiment analysis, as shown in Table 2. Four types of indicators are defined: α , β , γ and δ , which are designed to modify the sentiment strengths, varying from strongest, stronger, baseline to below baseline.

As shown in Table 2, sentiment strength indicators, e.g., amplifiers, help to strengthen the degree of the sentiments represented in the text, while diminishers weaken the degree of the sentiments. Category α refers to the strongest amplifiers (e.g., “Highest”), β refers to stronger amplifiers (e.g., “Very”), δ refers to diminishers (e.g., “Less”) and γ refers to a case where there are no amplifiers and diminishers in front of a word.

The strength-level indicators described in Table 2 can support the algorithm to understand or identify nine sentiment categories: very strongly negative (− 4), strongly negative (− 3), negative (− 2), slightly negative (− 1), neutral (0), slightly positive (1), positive (2), strongly positive (3) and very strongly positive (4). However, there is no such ground truth dataset available; therefore, this paper leverages the datasets which have five categories: strongly negative, negative, neutral, positive and strongly positive.

4 Datasets

For this research, we use two different datasets in different domains which are available for experiments and comparison. The details are presented in this section.

4.1 Transcomp

TransComp is the Public Web data of transportation domain which we crawled from Reddit. PRAW is leveraged to scrape data from Reddit, which contains query terms like “bus,” “mrtm,” “cab,” “taxi” and “comfort delgro.” Since Reddit data tend to be long stories, the raw data object is broken up into short texts or sentences as this research focuses on short texts or sentences. The data is kept in the initial raw format, which can better test the capability of the proposed MiMuSA for handling the data from real-world data sources.

In order to evaluate MiMuSA as well as the existing methods, pre-labeled data is necessary. Four groups of researchers were invited as volunteer annotators to label the data manually. The annotation results from the four groups were further analyzed and only the data objects for which at least three groups of the annotators provided the same labels were selected to form a set of ground truth data, which contains 1062 data objects.

4.2 Movie review dataset

This paper used the test set of the Stanford Sentiment Treebank dataset¹ focusing on the movie domain, which contains 2210 samples. The data were manually annotated by four volunteers and the data objects for which any three annotators of the four volunteers provided the same labels were selected to form a set of ground truth data. As a result, 1240 samples were obtained. The review sentences in the original dataset contain five different types of labels. The five labels (0, 1, 2, 3 and 4) correspond to the sentiment polarities of strongly negative, negative, neutral, positive and strongly positive, respectively.

5 Experiment, evaluation and discussion

In this section, this paper describes experiments conducted to test different methods, and the details are described in each subsection.

5.1 Parameter setting

For every sentence component, sentiment strength handling submodule searches for sentiment strength indicators: α , β or δ , with α being the highest priority and δ being the least priority, with the immediate next word carrying the same polarity as the overall polarity of the text and the polarity is scaled accordingly. If there are no sentiment strength indicators found in the text, it can be concluded that there

¹ <https://nlp.stanford.edu/sentiment/code.html>.

Table 2 Examples and explanations of strength-level indicators

Sentiment strength indicators	Explanations	Indicator examples
α : strongest indicator	Strongest amplifiers	Highest, biggest, largest, maximum, extremely, super, best
β : stronger indicator	Stronger amplifiers	Pretty, very, fairly, quite, effectively
γ : baseline	No amplifiers or diminishers	There are no indicators, α , β , δ appearing in the text
δ : below-baseline	Diminisher	Slightly, weakly

are no amplifiers and diminishers present, thus the text belongs to group γ .

For this research, the ground truth dataset contains five-categories only (strongly positive, positive, neutral, negative and strongly negative). Therefore, both α and β modify the sentiment to “strongly” level and this setting is consistent with the previous work [19, 55]. It is consistent with the human language understanding process (e.g., “good” is positive, “very good” (with indicator β) and “best” (with indicator α) are strongly positive).

For learning-based models, this paper uses stratified k-fold cross-validation and K is set to 4. In addition, we run the models five times and report the mean value and standard deviation for different learning-based methods. For the proposed MiMuSA, the tests are carried out on the whole dataset since there is no need to split the dataset into train set and test set.

5.2 Comparison of the influences of different knowledge bases

Table 3 shows the influence of different knowledge bases on the 3-class sentiment analysis. It is observed that when the knowledge base becomes richer and richer, the performance of the proposed method becomes better and better. These results are consistent with human being’s capability: the more knowledge we possess, the more powerful we become in solving problems.

5.3 Performance comparison for aggregate-level sentiment analysis

We use the two aforementioned datasets to compare our proposed MiMuSA with four popular sentiment analysis tools, namely Textblob [56], Vader [57], SentiWordNet [58] and SenticNet [59]. Table 4 shows the results of sentence-level sentiment analysis on the two datasets. Since those existing works are only designed for aggregate-level analysis, we conduct the comparison for 3-class sentiment classification task.

As shown in Table 4, MiMuSA with all the knowledge bases performs exceptionally well in the transport domain,

(Accuracy, F1-Score) = (0.9209, 0.9210), and it can also outperform the other four classic non-learning-based methods in the movie domain. These results demonstrate the merit of MiMuSA.

5.4 Performance comparison for fine-grained multi-class sentiment analysis

In order to test the performance of the proposed MiMuSA, various existing multi-class sentiment analysis methods are tested as the baseline models, which are three popular machine learning models including Logistic Regression (LR), Naïve Bayes (NB) and SVM, and two DL models including LSTM and CNN, and two pre-trained language models, namely BERT and SentiBERT [60].

As shown in Tables 5 and 6, MiMuSA significantly outperforms the existing methods with better performance for both 3-class and 5-class sentiment identification on the two datasets.

In addition, comparing the results with the existing multi-class sentiment analysis methods, in terms of both the Accuracy and F1, it can be seen that the proposed MiMuSA performs much better. This demonstrates the merit of the proposed fine-grained multi-class sentiment analysis achieved through the mimicking of human language understanding processes.

In this work, we have conducted 5-class sentiment analysis. However, MiMuSA can be extended to more fine-grained, different strength levels, such as five positive levels and five negative levels. Currently, there has been no such multi-level sentiment analysis datasets or methods/tools available yet.

5.5 An example for explainability

In order to show the explainability function of the proposed MiMuSA, an example is showcased in this section to illustrate the sentiment understanding process.

Consider the sample data item, “I did not like it at beginning, but it is in fact very good I found later.” Firstly, “like” and “good” are identified as Positive sentiment indicators through the Basic Knowledge Base. “not” is

Table 3 Performance of the proposed method with different knowledge bases for aggregate-level sentiment analysis (3 classes) (Transport Domain)

Different knowledge bases	Performance with different knowledge bases		
	Option	Accuracy	F1
B0. Basic knowledge base	B0	0.7401	0.7412
B1. Local language knowledge base	B0, B1	0.7561	0.7573
B2. Negation and special knowledge	B0, B1, B2	0.7976	0.8002
B3. Domain knowledge base	B0, B1, B2, B3	0.8004	0.8029
B4. Sarcasm rule and adversative base	B0, B1, B2, B3, B4	0.9134	0.9152
B5. sentiment strength knowledge base	B0, B1, B2, B3, B4, B5	0.9209	0.9210

Table 4 Performance comparison of MiMuSA with the existing non-learning-based methods for aggregate-level sentiment analysis (3 classes)

Methods	Transport domain		Movie domain	
	Accuracy	F1	Accuracy	F1
Textblob [56]	0.5471	0.5248	0.4665	0.4931
Vader [57]	0.6582	0.6541	0.5335	0.5248
SentiWordNet [58]	0.5452	0.5229	0.5217	0.4797
SenticNet [59]	0.8545	0.8313	0.6774	0.6531
MiMuSA	0.9209	0.9210	0.7629	0.7597

identified as Negation through the Negation Knowledge Base, and the Negation handler function is triggered. According to the WTN vector, the proposed MiMuSA will identify that “not like” as negative since WTN of “like” is “- 1.” Then, the adversative conjunction “but” is identified through the adversative knowledge base, and the ambivalence handling function is triggered. MiMuSA prioritizes the polarity of the phrase after the conjunction “but.” As a result, MiMuSA classifies the sentence at the aggregate level as a positive sentiment. After that, the sentence is further identified as the fine-grained multi-class sentiment—strongly positive due to the strength indicator “very”—is identified to modify “good.”

Table 5 Performance comparison of MiMuSA with the existing learning-based methods for fine-grained multi-class sentiment analysis (Transport Domain)

Methods	3 classes		5 classes	
	Accuracy	F1	Accuracy	F1
LR	0.7034 (± 0.0144)	0.6884 (± 0.0192)	0.5706 (± 0.0061)	0.5436 (± 0.0075)
NB	0.6667 (± 0.0112)	0.6532 (± 0.0052)	0.5292 (± 0.0225)	0.5129 (± 0.0229)
SVM	0.6930 (± 0.0123)	0.6858 (± 0.0114)	0.5697 (± 0.0052)	0.5522 (± 0.0070)
CNN	0.6878 (± 0.0067)	0.6750 (± 0.0072)	0.5533 (± 0.0142)	0.5364 (± 0.0123)
LSTM	0.6904 (± 0.0051)	0.6818 (± 0.0078)	0.5419 (± 0.0183)	0.5247 (± 0.0152)
BERT	0.7203 (± 0.0450)	0.6707 (± 0.0440)	0.5848 (± 0.0205)	0.4936 (± 0.0147)
MiMuSA	0.9209	0.9210	0.6365	0.6444

Table 6 Performance comparison of MiMuSA with the existing learning-based methods for fine-grained multi-class sentiment analysis (Movie Domain)

Methods	3 classes		5 classes	
	Accuracy	F1-Score	Accuracy	F1-Score
LR	0.6218 (± 0.0189)	0.5943 (± 0.0197)	0.4164 (± 0.0216)	0.3692 (± 0.0129)
NB	0.6274 (± 0.0236)	0.6109 (± 0.0239)	0.4011 (± 0.0415)	0.3692 (± 0.0129)
SVM	0.6008 (± 0.0179)	0.5929 (± 0.0163)	0.3863 (± 0.0462)	0.3740 (± 0.0348)
CNN	0.6242 (± 0.0338)	0.6085 (± 0.0312)	0.3944 (± 0.0171)	0.3746 (± 0.0193)
LSTM	0.5452 (± 0.0610)	0.5404 (± 0.0383)	0.3395 (± 0.0667)	0.3189 (± 0.0613)
BERT	0.7484 (± 0.0412)	0.7027 (± 0.0432)	0.4750 (± 0.0215)	0.4364 (± 0.0362)
MiMuSA	0.7629	0.7597	0.5024	0.5043

Table 7 An example of the vector representation of a sentence

Sentence	I did not like it at beginning, but it is very good I found later																
<i>W</i>	I	did	not	like	it	at	beginning	,	but	it	is	very	good	I	found	later	
<i>M</i>	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	
<i>N</i>	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
<i>WTN</i>	0	0	0	-1	0	0	0	0	0	0	0	0	-1	0	0	0	
<i>SA</i>	0	0	0	0	0	0	0	0	-	0	0	0	0	0	0	0	
									1								
<i>S</i>	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
<i>C</i>	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	

Table 7 shows the vector representation of this sample data item. Firstly, through Basic Knowledge Base, Local Language Knowledge Base and Domain Knowledge Base, vector *M*, which represents the intermediate sentiment category of each element (e.g., word) of the sentence, can be obtained, while vector *N* identifies negation words through Negation Knowledge Base. Next, vector *WTN* indicates whether the sentiment polarity of sentiment words would be reversed if negation operation acts on it. For this example, the sentiment of word “like” would be reversed since its *WTN* is “- 1.” Then, the adversative conjunction “but” would be identified through Conjunction and Adversative Base, so we prioritize the polarity of the phrase after the conjunction word “but,” which is denoted by vector *A*. Vector *S* can be obtained through Sentiment Strength Knowledge Base. Finally, the sentiment score vector, *C*, can be obtained using vector *S* as well the vectors above.

5.6 Further analysis and discussion

Sentiment analysis or sentiment understanding problem can be configured as a classification task, and machine learning-based methods are powerful tools for such tasks if huge ground truth training datasets are available. However, such labeled ground truth datasets are not always available, or it is too expensive to obtain the labeled data for solving real-world problem.

For example, each day, the Weibo platform produces thousands of millions of blogs. For machine learning methods, including the DL method, they are black-box methods that require huge amount of training data for classification tasks. Whatever the ratio of training and testing data, e.g., 5:1, 4:1 or 3:1, such training and testing paradigm in fact is fatally unpractical for language understanding tasks including the task of sentiment understanding of simple sentences. This is especially challenging for multi-class sentiment understanding with more categories, e.g., nine sentiment category identification.

This may explain the reason why the latest so-called intelligent robots are still not intelligent enough as there is no true human language understanding processes involved. Such a fact suggests that true language understanding—mimicking the human language understanding process—is the right direction for NLP tasks. Machine learning including DL methods are still powerful tools which can be used to conduct knowledge extraction to enhance true language understanding and other symbolic AI algorithms.

6 Conclusion

In this paper, we proposed MiMuSA, a fine-grained multi-class sentiment analysis method that mimics the human language understanding process. The proposed MiMuSA involves a multi-level modular structure designed to mimic human’s language understanding processes, e.g., ambivalence handling process, sentiment strength handling process, etc. Different knowledge bases including Basic Knowledge Base, Local Language Knowledge Base, Negation and Special Knowledge Base, Adversative Base, Sarcasm Rule and Sentiment Strength Knowledge Base were constructed and used for the proposed sentiment understanding method, in a similar vein as the human’s multi-level knowledge acquisition and understanding process.

In addition, a new set of multi-class sentiment ground truth data in the transportation domain was constructed. The experiments on the ground truth dataset as well as a public dataset—the Stanford Sentiment Treebank dataset—demonstrated better performance of the proposed MiMuSA compared against existing multi-class sentiment analysis methods. The results not only demonstrate the remarkable performance of the proposed MiMuSA across different datasets, but also highlight the gains that can be obtained in implementing and applying interpretable human-like sentiment analysis.

Moving forward, several potential improvements can be made on this research. Aspect or topic-based sentiment

analysis will be considered. More detailed human language understanding processes other than just sentiment understanding will be implemented as part of future work. In addition, more experiments will also be conducted to provide more in-depth analysis on the explainable aspect and various human-like characteristics.

Acknowledgements This research was supported by A*STAR under its Advanced Manufacturing and Engineering (AME) Programmatic Grant (Award No.: A19E2b0098). The authors would like to thank Junru LUO and Siti Salihah Binte Mohamed RILVAN for their contributions to the code development of this research. Many thanks to Jiaxin JIANG, Junru LUO, Siti Salihah Binte Mohamed RILVAN, Aaron LEE, Jaspavan KAUR, Phuong Thao BUI, Swee Hong NG and Jingfeng CUI for their contributions to the knowledge base building and data preparations.

Data availability The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declaration

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this paper.

References

1. Birjali M, Kasri M, Beni-Hssane A (2021) A comprehensive survey on sentiment analysis: approaches, challenges and trends, *Knowl Based Syst* 107:134
2. Dragoni M, Donadello I, Cambria E (2022) OntoSenticNet 2: enhancing reasoning within sentiment analysis. *IEEE Intell Syst* 37:103–110
3. Zhang W, Li X, Deng Y, Bing L, Lam W (2022) A survey on aspect-based sentiment analysis: tasks, methods, and challenges, *IEEE Trans Knowl Data Eng*
4. Verma S (2022) Sentiment analysis of public services for smart society: literature review and future research directions. *Gov Inf Q* 39:101708
5. Grassi M, Cambria E, Hussain A, Piazza F (2011) Sentic web: a new paradigm for managing social media affective information. *Cogn Comput* 3:480–489
6. Cambria E, Wang H, White B (2014) Guest editorial: big social data analysis. *Knowl Based Syst* 69:1–2
7. Ma Y, Nguyen KL, Xing F, Cambria E (2020) A survey on empathetic dialogue systems. *Inf Fus* 64:50–70
8. Lango M (2019) Tackling the problem of class imbalance in multi-class sentiment classification: an experimental study. *Found Comput Decis Sci* 44:151–178
9. Hota S, Pathak S (2018) Knn classifier based approach for multi-class sentiment analysis of twitter data. *Int J Eng Technol* 7:1372–1375
10. Liu Y, Bi J-W, Fan Z-P (2017) Multi-class sentiment classification: the experimental comparisons of feature selection and machine learning algorithms. *Expert Syst Appl* 80:323–339
11. Wang Z, Lin Z (2020) Optimal feature selection for learning-based algorithms for sentiment classification. *Cogn Comput* 12:238–248
12. Wang Z, Tong VJC, Chin HC (2014) Enhancing machine-learning methods for sentiment classification of web data, in: *Asia Information Retrieval Symposium*, Springer, pp. 394–405
13. Syaekhoni MA, Seo SH, Kwon YS (2017) Development of deep learning models for multi-class sentiment analysis. *J Inf Technol Serv* 16:149–160
14. Alzamzami F, Hoda M, El Saddik A (2020) Light gradient boosting machine for general sentiment classification on short texts: a comparative evaluation. *IEEE Access* 8:101840–101858
15. Cambria E, Gastaldo P, Bisio F, Zunino R (2015) An ELM-based model for affective analogical reasoning. *Neurocomputing* 149:443–455
16. Cambria E, Schuller B, Liu B, Wang H, Havasi C (2013) Statistical approaches to concept-level sentiment analysis. *IEEE Intell Syst* 28:6–9
17. Wang Z, Goh SMR, Yang Y (2017) A method and system for sentiment classification and emotion classification, *US Patent App.* 15/523,201
18. Bouazizi M, Ohtsuki T (2017) A pattern-based approach for multi-class sentiment analysis in twitter. *IEEE Access* 5:20617–20639
19. Wang Z, Ho S-B, Cambria E (2020) Multi-level fine-scaled sentiment sensing with ambivalence handling. *Int J Uncertain Fuzziness Knowl Based Syst* 28:683–697
20. Schank RC (1972) Conceptual dependency: a theory of natural language understanding. *Cogn Psychol* 3:552–631
21. Lytinen SL (1992) Conceptual dependency and its descendants. *Comput Math Appl* 23:51–73
22. Schank RC, Abelson RP (2013) *Scripts, plans, goals, and understanding: an inquiry into human knowledge structures*, Psychology Press
23. Homsy MN (2014) Multi-class sentiment analysis using a hierarchical logistic model tree approach, *Maskana* 5
24. Sharma S, Srivastava S, Kumar A, Dangi A (2018) Multi-class sentiment analysis comparison using support vector machine (svm) and bagging technique-an ensemble method, in: *2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE)*, IEEE, pp. 1–6
25. Daudert T (2021) Exploiting textual and relationship information for fine-grained financial sentiment analysis. *Knowl Based Syst* 230:107389
26. Balabantaray RC, Mohammad M, Sharma N (2012) Multi-class twitter emotion classification: a new approach. *Int J Appl Inf Syst* 4:48–53
27. Wang B, He W, Yang Z, Xiong S (2020) An unsupervised sentiment classification method based on multi-level fuzzy computing and multi-criteria fusion. *IEEE Access* 8:145422–145434
28. Liu Y, Bi J-W, Fan Z-P (2017) A method for multi-class sentiment classification based on an improved one-vs-one (ovo) strategy and the support vector machine (svm) algorithm. *Inf Sci* 394:38–52
29. Krawczyk B, McInnes BT, Cano A (2017) Sentiment classification from multi-class imbalanced twitter data using binarization, in: *International Conference on Hybrid Artificial Intelligence Systems*, Springer, pp. 26–37
30. Mukherjee A, Mukhopadhyay S, Panigrahi PK, Goswami S (2019) Utilization of oversampling for multiclass sentiment analysis on amazon review dataset, in: *2019 IEEE 10th Int Conf Aware Sci Technol (iCAST)*, IEEE, pp. 1–6
31. Wang Z, Parth Y (2016) Extreme learning machine for multi-class sentiment classification of tweets, in: *Proceedings of ELM-2015 Volume 1*, Springer, pp. 1–11
32. Farhadloo M, Rolland E, Multi-class sentiment analysis with clustering and score representation, in, (2013) *IEEE 13th Int Conf Data Mining Workshops IEEE* 2013:904–912
33. Attia M, Samih Y, Elkahky A, Kallmeyer L (2018) Multilingual multi-class sentiment classification using convolutional neural networks, in: *Proc Eleventh Int Conf Lang Resour Eval (LREC 2018)*

34. Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. *Knowl Based Syst* 235:107643
35. Oneto L, Bisio F, Cambria E, Anguita D (2016) Statistical learning theory and ELM for big social data analysis. *IEEE Comput Intell Mag* 11:45–55
36. Zaw M, Tandayya P (2018) Multi-level sentiment information extraction using the crbsa algorithm, in: 2018 15th Int Joint Conf Comput Sci Softw Eng(JCSSE), IEEE, pp. 1–6
37. Mehmood Y, Balakrishnan V (2020) An enhanced lexicon-based approach for sentiment analysis: a case study on illegal immigration, *Online Inf Rev*
38. Zhang Z, Wu G, Lan M (2015) Ecnu: Multi-level sentiment analysis on twitter using traditional linguistic features and word embedding features, in: Proc 9th Int Workshop Semant Eval (SemEval 2015), pp. 561–56
39. Li M, Ch'ng E, Chong AYL, See S (2018) Multi-class twitter sentiment classification with emojis, *Ind Manage Data Syst*
40. Kocoń J, Zaśko-Zielińska M, Miłkowski P (2019a) Multi-level analysis and recognition of the text sentiment on the example of consumer opinions, in: Proc Int Conf Recent Adva Nat Lang Proc (RANLP 2019), pp. 559–567
41. Kocoń J, Miłkowski P, Zaśko-Zielińska M (2019b) Multi-level sentiment analysis of polemo 2.0: Extended corpus of multi-domain consumer reviews, in: Proc 23rd Conf Comput Nat Lang Learn (CoNLL), pp. 980–991
42. Xie H, Lin W, Lin S, Wang J, Yu L-C (2021) A multi-dimensional relation model for dimensional sentiment analysis. *Inf Sci* 579:832–844
43. Akhtar MS, Ekbal A, Cambria E (2020) How intense are you? predicting intensities of emotions and sentiments using stacked ensemble. *IEEE Comput Intell Mag* 15:64–75
44. Wang Z, Ho S-B, Cambria E (2020) A review of emotion sensing: categorization models and algorithms. *Multimedia Tools Appl* 79:35553–35582
45. Chaturvedi I, Ong Y-S, Tsang I, Welsch R, Cambria E (2016) Learning word dependencies in text by means of a deep recurrent belief network. *Knowl Based Syst* 108:144–154
46. Kumar AJ, Trueman TE, Cambria E (2021) A convolutional stacked bidirectional lstm with a multiplicative attention mechanism for aspect category and sentiment detection. *Cogn Comput* 13:1423–1432
47. Chaturvedi I, Satapathy R, Cavallari S, Cambria E (2019) Fuzzy commonsense reasoning for multimodal sentiment analysis. *Pattern Recogn Lett* 125:264–270
48. Xue M, Yuan C, Wu H, Zhang Y, Liu W (2020) Machine learning security: threats, countermeasures, and evaluations. *IEEE Access* 8:74720–74742
49. Berka P (2020) Sentiment analysis using rule-based and case-based reasoning, *J Intel Inf Syst* 1–16
50. Ho SB (2012) The atoms of cognition: A theory of ground epistemics, in: Proc Annual Meeting Cognit Sci Soc, volume 34
51. Ho SB, Wang Z (2019a) On true language understanding, in: International Conference on Artificial Intelligence and Security, Springer, pp. 87–99
52. Ho S-B, Wang Z (2019b) Language and robotics: complex sentence understanding, in: International Conference on Intelligent Robotics and Applications, Springer, pp. 641–654
53. Majumder N, Poria S, Peng H, Chhaya N, Cambria E, Gelbukh A (2019) Sentiment and sarcasm classification with multitask learning. *IEEE Intell Syst* 34:38–43
54. Du Y, Li T, Pathan MS, Teklehaimanot HK, Yang Z (2021) An effective sarcasm detection approach based on sentimental context and individual expression habits, *Cogn Comput* 1–13
55. Poria S, Cambria E, Hazarika D, Mazumder N, Zadeh A, Morncy L-P (2017) Multi-level multiple attentions for contextual multimodal sentiment analysis, in: 2017 IEEE Int Conf Data Mining (ICDM), IEEE, pp. 1033–1038
56. Loria S (2020) textblob documentation
57. Hutto C, Gilbert E (2014) Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proc Int AAAI Conf Web Soc Media, volume 8, pp. 216–225
58. Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining, in: Proc Seventh Int Conf Lang Resour Eval (LREC'10)
59. Cambria E, Liu Q, Decherchi S, Xing F, Kwok K (2022) SenticNet 7: a commonsense-based neurosymbolic AI framework for explainable sentiment analysis, in: LREC, pp. 3829–3839
60. Yin D, Meng T, Chang K-W (2020) Sentibert: a transferable transformer-based architecture for compositional sentiment semantics, in: Proc 58th Annu Meeting Assoc Comput Linguist, pp. 3695–3706

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Zhaoxia Wang¹  · Zhenda Hu² · Seng-Beng Ho³ · Erik Cambria⁴ · Ah-Hwee Tan¹

✉ Zhaoxia Wang
zxwang@smu.edu.sg

Zhenda Hu
huzhenda2020@gmail.com

Seng-Beng Ho
hosb@ihpc.a-star.edu.sg

Erik Cambria
cambria@ntu.edu.sg

Ah-Hwee Tan
ahtan@smu.edu.sg

¹ School of Computing and Information Systems, Singapore Management University, Singapore, Singapore

² School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai, China

³ Social and Cognitive Computing Department, Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore

⁴ School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore