10-2021

Integrated discourse analysis & learning skills framework for class conversations

Devyn Wei Hung TAN
Singapore Management University, devyn.tan.2017@sis.smu.edu.sg

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

Kyong Jin SHIM
Singapore Management University, kjshim@smu.edu.sg

SHANKARARAMAN, Venky
Singapore Management University, venks@smu.edu.sg

Integrated Discourse Analysis & Learning Skills Framework for Class Conversations

Devyn Wei Hung Tan, Swapna Gottipati, Kyong Jin Shim, Venky Shakararaman
School of Information Systems
Singapore Management University
devyn.tan.2017,swapnag,kjshim,venks@smu.edu.sg

2021 IEEE Frontiers in Education Conference (FIE) | 978-1-6654-3851-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/FIE49875.2021.9637384

***Abstract*— Constructive interactions through discussion forums allow students to open their horizons and thought processes to acquire more knowledge and develop skills. Thus, discussion forums play an important role in supporting learning. Additionally, the discussion forum provides the content for creating a knowledge repository. It contains discussion threads related to key course topics that are debated by the students. One approach to understanding the student learning experience is through the analysis of the discussion threads. This research proposes the application of discourse analysis and collaborative learning frameworks to discussion forums to gain further insights into the student's learning in a classroom. It is a foray into discourse analysis using in-class discussions. It demonstrates the application of Soller's framework and Penn Discourse Treebank (PDTB) to understand interactions at the discourse and semantic level. It also shows the use of unsupervised automated techniques to diagnose interactions in textual data. In this paper, we present an Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework based on in-class discussions. We describe our experiences of applying IDALS framework and evaluating the solution model in a graduate in-class discussion forum. We also highlight the benefits of using visualizations to present the insights to the instructors.**

Keywords— Collaborative learning analysis, Online Discussion Forums, PDTB framework, Learner's profile

I. INTRODUCTION

Interactions in learning settings enable and empower students to better understand content by drawing connections between the topics of study. Interactions also engage students to build relationships between abstract and concrete concepts. Effective collaboration with peers has proven to be a successful and uniquely powerful learning method [1]. Encouraging each other to partake in questioning, justifying their opinions, clearly articulating reasoning and knowledge, and reflecting on acquired knowledge motivates learners, and it leads to an improved learning experience for all participants [2].

Constructive interactions help students open their horizons and thought processes to acquire more knowledge and develop skills. In the classroom, effective collaborations and interaction with peers have proven to be a successful and uniquely powerful learning method [1]. A key component of interaction that makes them effective is the language used while interacting. The language used by learners and teachers in interactions is therefore considered a rich source of data for understanding the learning effectiveness [3, 4]. The nature and depth of students' interaction in online environments is different from that of face-to-face classrooms [5]. However, regardless of whether the medium is online or offline, the language used is still of the essence in determining the quality of interactions. Therefore, it can be said that the language and quality of interactions play a pivotal role in the quality of learning and thereby the student's learning journey.

To better comprehend different types of interactions, it is important to understand the language used and thereby the interpersonal and behavioral aspects embedded in the discourse [6, 7]. The Penn Discourse Treebank 2.0 (PDTB 2.0) from the University of Pennsylvania is a large-scale resource of annotated discourse relations and their arguments over 1-million-word Wall Street Journal (WSJ) corpus [8, 9]. The PDTB 2.0 dataset is the first lexically grounded approach to annotating discourse relations while supporting the extraction of useful features pertaining to syntax, semantics, and discourse all at once. We have, much like other researchers working in the space of discourse analysis, made use of the PDTB 2.0 resource to guide and train our model.

A large body of research has been dedicated to building frameworks to better understand collaboration and interactions along the different dimensions of behavior. Examples of these frameworks include Walsh's Classroom Interactional Competence (CIC) [10], Hyland's interpersonal aspects of written discourse [11], Aijmer's interpersonal aspects of spoken discourse [12], and Soller's Collaborative Learning Conversation Skill Taxonomy [2]. We will be using Soller's framework in this work.

This paper is a foray into discourse analysis in classroom discussions and makes use of the Soller's framework and the PDTB 2.0 dataset with the aim of understanding interactions at the discourse and semantics level. We propose an Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework based on class discussions. Further, we use unsupervised automated techniques to implement the framework. The outcome of the implementation is a learning profile based on and individual's contributions to in-class discussions that will be made available to both instructors and students. This learning profile is built by mapping the discourse and sense tags to Soller's framework and by examining the behavioral and interpersonal aspects that are mined from the language used in the discourse interactions. Using these learning profiles, the instructor can better plan and focus discussions to enhance and support students to attain the skills they lack. We present a visualization model of collaborative learning at both

individual student and cohort levels that helps to envision the learner's and team's characteristics.

We use data from the online discussion forum of a master's course "Text Analytics and Applications". A study conducted by Burge using students enrolled in a web-based distance program identified challenges associated with handling and managing large quantities of discussion data leading to discussion fragmentation [13]. Therefore, we design the discussion forum using discussion threads which are easily manageable. The discussion threads are created and controlled by the instructor. Thus, allowing the students to participate in structured discussions and gain knowledge from it to prepare for their project and exam.

The main contributions of the paper are as follows. Firstly, it presents an Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework based on class discussions. Secondly, using NLP techniques, it proposes an automated solution design and scoring model to build a learner's profile in a collaborative environment. Thirdly, it describes a visualization model at both individual student and cohort levels that help to envision the learners' aspects in terms of collaborative learning. Fourthly, it shares our experiences on applying IDALS framework and evaluating the solution model in an in-class discussion forum. Finally, this paper also highlights the benefits of using visualizations to present the insights to the instructors and the limitations of the research.

This paper is structured as follows. Section II will review the work about discourse analysis in education and classroom and use of NLP for discussion forum analysis. Section III provides a literature review of theoretical research relevant to the current work. Section IV presents the Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework. Section V presents the solution design of the implementation of IDALS framework using NLP and text analytics techniques. Section VI describes a case study implementation of the solution model using an online discussion forum from a graduate-level information systems course. Finally, conclusions are presented in Section VII.

II. Related Work

Discourse Analysis refers to theories and methods used in educational research that emerged from various disciplines such as linguistics, anthropology, politics, and communications [14]. Discourse analysis is concerned with illustrating how the analysis of spoken or written discourse is informed by analysis of structural social relations [15]. The language is inherently political, and therefore any analysis of language will necessitate a critical approach to understanding how language "uncovers different ways of saying things, doing things, and being things in the world" [6]. By closely examining language in written or spoken text, it offers opportunities to learn how individuals use language to make sense of an activity they are engaged in, what identities they adopt within certain contexts, or how they use it to build relationships with other people [6]. Discourse Analysis aids to improve classroom interaction set out to foster teachers' interactional awareness and improve equity in classrooms [7, 16]. The study of classroom discourse is commonly associated with analysing the language and interaction of teaching and learning [17]. Classroom discourse provides a space and forum for interdisciplinary work to flourish by taking an unrestricted approach to 'classroom' and 'discourse'. Discourse consists of interactional features [18] as well as social issues and phenomena that transcend the immediately unfolding sequential context of lessons and summaries of discussions [19, 20, 31].

Park and Cardie presented a systematic study of features for implicit discourse relation identification and identified feature combinations that optimize F1-score using forward selection [21]. [22] provides the approach that maps discourse senses to their respective discourse relations. Stepanov and Riccardi focus on how to best extract "argument spans" for the task of discourse relation parsing [23]. Chandrasekaran et al. investigated the task of predicting instructor intervention in student posts from discussion forums in Massive Open Online Courses (MOOCs) [24]. This research is similar to the work reported in this paper as it uses discourse relations in discussion forum posts. Their research question is to predict when the instructor should intervene in which MOOC discussion forum threads based on discourse relations whereas our focus is to use discourse relations to analyze collaborative conversation skills of students in discussion forums. Multiple frameworks and approaches are needed to collectively uncover the myriad ways in which discourse is shaped by classroom teaching and learning [25]. There are two works of interest here. The first work [2] discusses how Soller's framework provides the conversation skills taxonomy for collaborative learning, and the second work [26] looks at a practical application of discourse to language usage behavior. Soller's framework enables one to understand the language aspects in discussions and their alignment to student learning. Piurko's work with discourse markers [8] allows one to take a macro view at understanding behavioral aspects of discourse, in specific, how discourse allows the connection of spoken word and written word beyond the level of the sentence to linguistic and non-linguistic behavior [26]. In our research, we adopt Soller's framework and discourse analysis framework [8] to analyse the collaborative learning aspects in online discussion forums.

III. Background

There are broadly two areas of work relevant to this research: 1) understanding discourse relations which is a description of how two segments of discourse are logically connected to each other by examining the behavioral and interpersonal aspects embedded in the language and 2) methods for analyzing conversation behavior in discourse and mapping to effective and ineffective contributions. This helps to gain insights into the collaborative learning aspects and thus discover student's learning profile in a collaborative environment.

A. Discourse Relations Senses Framework

The PDTB 2.0 is a large-scale resource that not only studies the way discourse coherence is expressed, but also how information about discourse commitments or content attribution is conveyed linguistically [8]. According to PDTB 2.0, there are four types of content attribution or discourse relations that exist: explicit, implicit, entity-based coherence, and alternatively-lexicalized. The PDTB 2.0 resource provides the tags or labels for English sentences which indicate these relations.

Discourse Relation Tagging: Explicit connectives are selected from three grammatical classes: subordinating conjunctions (e.g., because, when, etc.), coordinating conjunctions (e.g., and, or, etc.), and discourse adverbials (e.g., instead) [27]. Implicit relations are connections that can be inferred. Most often, by simply inserting a connective expression between the arguments, the connection is made obvious. Entity-based coherence refers to all those relations where there is only an entity-based coherence relation perceived. For example, suppose we have the following two sentences: "Sentence 1: Hale Rogers will be turning 41 years old later this year and is the senior vice president at Samson Electronics Inc. Sentence 2: Mrs. Rogers succeeds Steve Conway who retired last year." The entity-based relation between these two sentences is Mrs. Rogers. Alternatively-Lexicalized (AltLex) relation is described by [9] as "the insertion of an implicit connective to express an inferred relation leading to a redundancy due to the relation being alternatively-lexicalized by some non-connective expression." The last type of relation is the one where there is no perceived discourse relation or entity-based relation between the sentences. Of these five discourse relation tags, we focus only on the explicit relation tags in this paper.

Sense tagging is the automatic assignment of the appropriate sense from some lexicon to each of the words in a text. The tagset of senses is organized hierarchically to consist of three levels: (1) class, (2) type, and (3) subtype. This organization reflects our understanding that there is a small core set of semantic relations that can hold between the situations described in the arguments of connectives. These sense annotations are available for explicit connectives, implicit connectives, and AltLex relations. The first level representing class has four categories: temporal, contingency, comparison, and expansion. For each class, as shown in Table I, a set of types is defined to further refine the sense. The third level of subtype specifies the semantic contribution of each argument which is not shown in Table I.

TABLE I. DISCOURSE RELATION SENSES FRAMEWORK (LEVEL 1 AND LEVEL 2)

Class-Level Senses (Level 1)	Types (Level 2)	Example explicit relations
Comparison	Concession, Contrast	although, as though, but, by comparison, even if, even though, however,
Contingency	Cause, Condition, Purpose	and, when, typically, as long as, especially if, simply because, since, then, after, so, when if only, lest, once,
Expansion	Alternative, Background, Circumstance, Conjunction, Continuation, Exception, Instantiation, Reinforcement, Restatement, Similarity	accordingly, additionally, after, also, although, and, as, as it, as if besides, by comparison, finally, first, for example, for one thing, however, in addition, in fact, ..
Temporal	Asynchronous, Synchronous	once, before, previously, when, then, after, since, previously,

In our proposed framework, IDALS, class-level senses are mapped with the collaborative learning aspects to discover student's learning profile in a collaborative environment. In the next section, we describe the IDALS framework.

B. Collaborative Learning Conversation Skills Taxonomy

Several frameworks specific to conversations in the education domain, have been proposed, that help to analyse interpersonal and behavioural aspects of language. Soller's Collaborative Learning Conversation Skill Taxonomy (CLCST) describes a method of classifying conversation behaviors that can be used to distinguish effective and ineffective contributions to interactions in classroom discussions [2]. The CLCST hierarchically classifies the skills into three different levels. The highest being Level-1 skills consisting of Creative Conflict, Active Learning, and Conversation, which further subsumes 36 Level-3 skills. The CLSCT framework at Level-1 and Level-2 are shown in Table II.

TABLE II. SOLLER'S COLLABORATIVE LEARNING SKILLS & CONVERSATIONS FRAMEWORK.

Conversation skills (Level 1)	Sub-skill (Level 2)	Attributes (Level 3)	Example conversation aspects
Active learning	Request, Inform, Motivate	Encourage, Suggest, Justify, Elaborate,..	Also, let me explain, to elaborate, also, very good, can you, do you know
Conversation	Task, maintenance, Acknowledge	Reject, Appreciate, Accept,..	Let's move on, let me show, would you please, yes, no, thank, sorry
Creative conflict	Argue, mediate	Infer, Agree, Doubt,..	Let's ask teacher, but, because, I agree, if..then, alternatively

As shown in Table II, the framework breaks down each learning conversation skill type into its corresponding subskills (e.g., Request, Inform, Acknowledge), and attributes (e.g., Suggest, Rephrase). Each attribute is assigned a short introductory phrase, or sentence opener, which conveys the appropriate dialogue intention as shown in the examples. "Intention of request" is to ask for help/advice in solving the problem, or in understanding a team-mates comment. "Intention of inform" is to direct or advance the conversation by providing information or advice. "Intension of motivate" is to provide positive feedback and reinforcement. "Intention of task" is to shift the current focus of the group to a new subtask or tool. "Intention of maintenance" is to support group cohesion and peer involvement. "Intention of acknowledge" is to inform

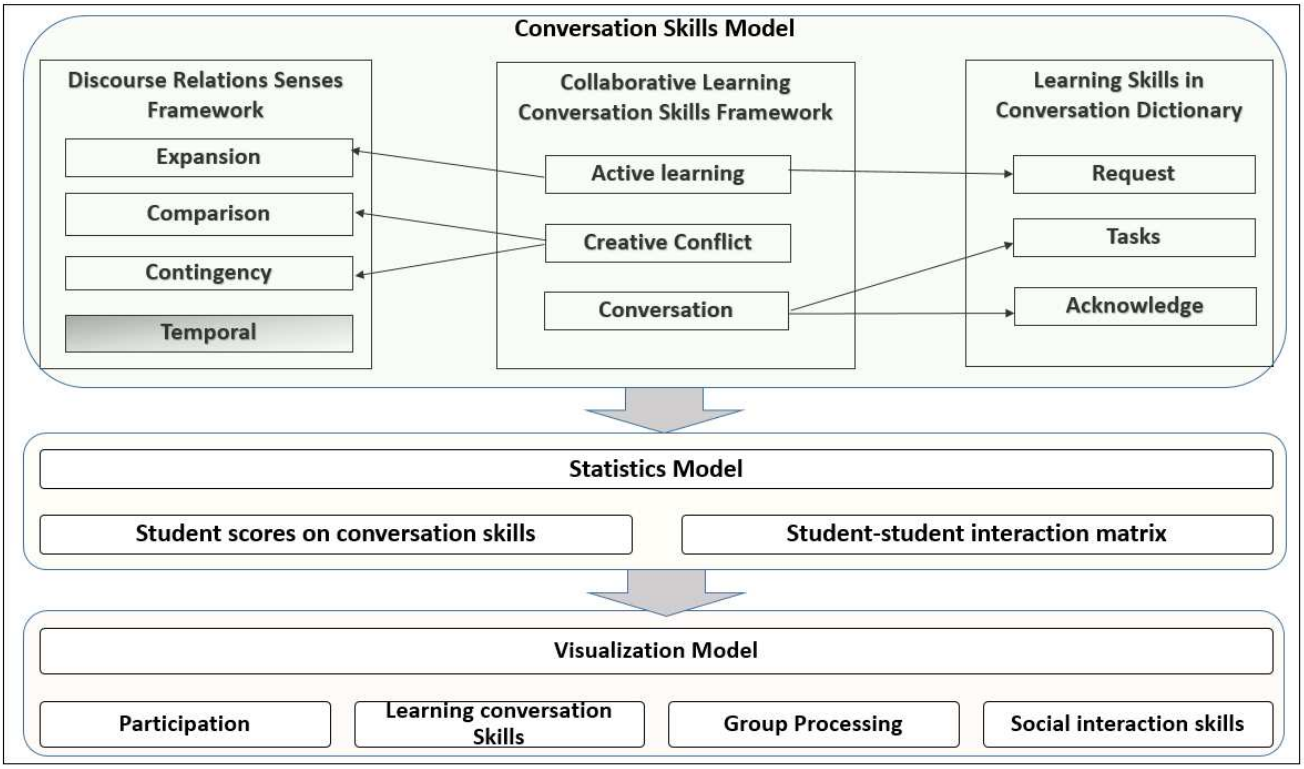


Fig. 1 IDALS framework

peers that one has read and/or appreciates the comments. It also indicates answering yes/no questions. "Intention of argue" is to reason (positively or negatively) about posts made by peer members. "Intention of mediate" is to recommend an instructor intervene to answer a question.

To discover student's learning profile, the discourse aspects of the students in the discussion forum is mapped to the conversation skills, and visualizations are developed to help the instructor analyze the classroom collaborative learning status. In the next section, we present our Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework based on the above frameworks.

IV. INTEGRATED DISCOURSE ANALYSIS AND COLLABORATIVE LEARNING FRAMEWORK (IDALS)

With a careful analysis of both the frameworks namely Discourse Relation Senses and CLCST, we realize an interlink between them in the context of classroom conversations. Based on the mapping, we propose an integrated framework, IDALS, as depicted in Fig. 1.

IDALS consists of three components: conversation skills model, statistics model, and visualization model. We describe each component in this section.

A. Conversation Skills Model

The first part of the framework is skills identification, which is achieved by combining discourse framework and Soller's framework, and where needed, with the use of special dictionaries. The below are some examples of the mapping between the two frameworks.

1. Creative conflict "argue" is about reasoning which can be expressed by providing comparisons or causes or contrast or conditions. Example conversation words such as "nevertheless", "only if" etc., indicate the subskill, "argue". Therefore, the discourse class, "comparison" and "contingency" can be mapped to conversation skill, "creative conflict".

2. Active learning "inform" is about providing more information by expanding the topics or giving more examples or listing more details. Example conversation words such as "accordingly", "besides", etc., indicate the subskill, "inform". Therefore, the discourse class, "expansion" can be mapped to conversation skill, "active learning".

3. We also observe that a few CLCST level-2 behaviors do not fall under discourse relations or classes. For example, conversation tasks are about the actions which are verbs in English, and discourse relations are not meant for the actions. This indicates that certain conversation aspects are not linked, and we need additional dictionaries to link them. Therefore, we created a new dictionary "tasks". It must be noted that the discourse class, "temporal" is not useful in measuring conversation skills, hence it is not mapped.

Based on the above analysis, a complete conversation skills model is created as shown in Fig. 1.

B. Statistics model

The conversation skills model when applied to any classroom discussion will generate scores of each skill for every student who participated in the conversation. These scores can be calculated at Level-1 or Level-2, which includes request, inform, motivate, argue, task, maintenance, and acknowledge.

Apart from the scores, using the network theory, an interaction matrix showing interactions between the students can be generated to analyze the social skills. For example, the social skills of students can be analysed by individual engagement levels and distribution of the student-student interactions [28]. Quoting and referring to other learners in forum posts, complimenting other learners, and greeting the class are examples of social interaction in online discussion forums.

C. Visualization Model

To generate the visuals that depict the collaborative learning insights, the visualization model requires variables and visual representations. The common visual variables for all visuals are color, size, position, shape, etc. IDALS visualization model details are shown in Table III. The data variables are the outputs from the statistics model. In several cases, novel or complex representations which combine multiple coordinated views are used to represent data that includes student's profile and collaborative learning information.

TABLE III. VISUALIZATION MODEL

Characteristics of Collaborative Learning teams	***Data Variables***	***Visual Representation***
Participation	Student posts, date	Summary of student's posts by date
Learning conversation Skills	Student scores on Request, Inform, Motivate, Argue	Summary of student's learning skills by time and comparisons with other students
Group Processing	Student scores on task, maintenance, Acknowledge	Summary of student's learning skills by time and comparisons with other students
Social interaction skills	Student-student interactions matrix	Network diagram of students

V. IMPLEMENTATION - SOLUTION DESIGN FOR IDALS

The solution design of the tool based on text mining and NLP techniques is shown in Fig. 2. It is a four-step process.

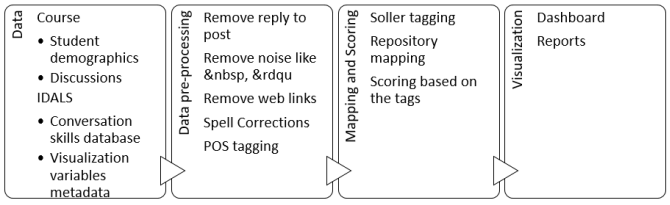


Fig. 2. Solution model for IDALS framework implementation

A. Data

Course data such as student's information and discussion data such as online discussion forums form the first set of input data for the solution model. IDALS framework shown in Fig. 1 can be converted into a repository such as databases or Excel spreadsheets. Such repositories are easy to maintain, and they can be used for future extensions of the current work. In our case study, we created an excel spreadsheet repository for the implementation stage. The metadata for visual variables is also generated as an Excel spreadsheet.

B. Data Pre-processing

The first objective of this stage is the normalization of noise in the discussion forum posts. Posts usually consist of several noise aspects such as reply to posts (repetitive posts content), special characters, web links, and spelling errors. These can be removed by leveraging the regular expressions library in Python. The second objective is to generate Part-of-Speech (POS) Tagging. This stage enables us to identify the language aspects and connect them with the discourse relations in IDALS database. The POS tags can be generated using the pos_tag function available in the NLTK library [29].

C. Mapping and Scoring

The objective of this stage is to generate the Soller tags for each post. Since the "discourse class" deals with conjunctions, for each post, we use the POS tags, which are labelled conjunctions and map them to the Soller tags. Note that each post may contain multiple Soller tags, and hence it is logical to store the count of each tag for each student. Such scoring provides detailed insights into a students' learning profile. An example mapping is depicted in Table IV.

TABLE IV. CONVERSATION SKILLS SCORING FOR AN EXAMPLE POST

Post	***Discourse tag***	***Collaborative Learning Skills***
In the recent years, the text mining and the data analytic application in the educational field is growing, as all the teaching materials and students assessing methods have turned …	Expansion-Conjunction-2 Contingency-Cause-3	Active Learning-Inform-2 Creative Conflict-Argue-3

D. Visualization

The output data from the mapping and scoring stage is stored in spreadsheet format which can be easily integrated into visualization tools such as Tableau or Qlik Sense. Some example visual representation graphs include overall analysis, weekly analysis, student profile, comparison analysis, and cluster analysis.

VI. APPLICATION OF IDALS – CASESTUDY

A. Methodology

1) Research Problem

IDALS framework is applied to discover the insights of student learning skills in using an online discussion forum from Information Systems graduate course. We study the following two research questions.

RQ1: What is the performance of the solution model?

RQ2: What are the insights that can be gained in terms of collaborative learning skills of students from analysing the discussion forum?

2) Online Forum Settings

As part of the text analytics course for graduate students, the instructor designed a weekly question and answering forum in Learning Management System (LMS). We observed that students were reluctant to use the discussion forum if it was a

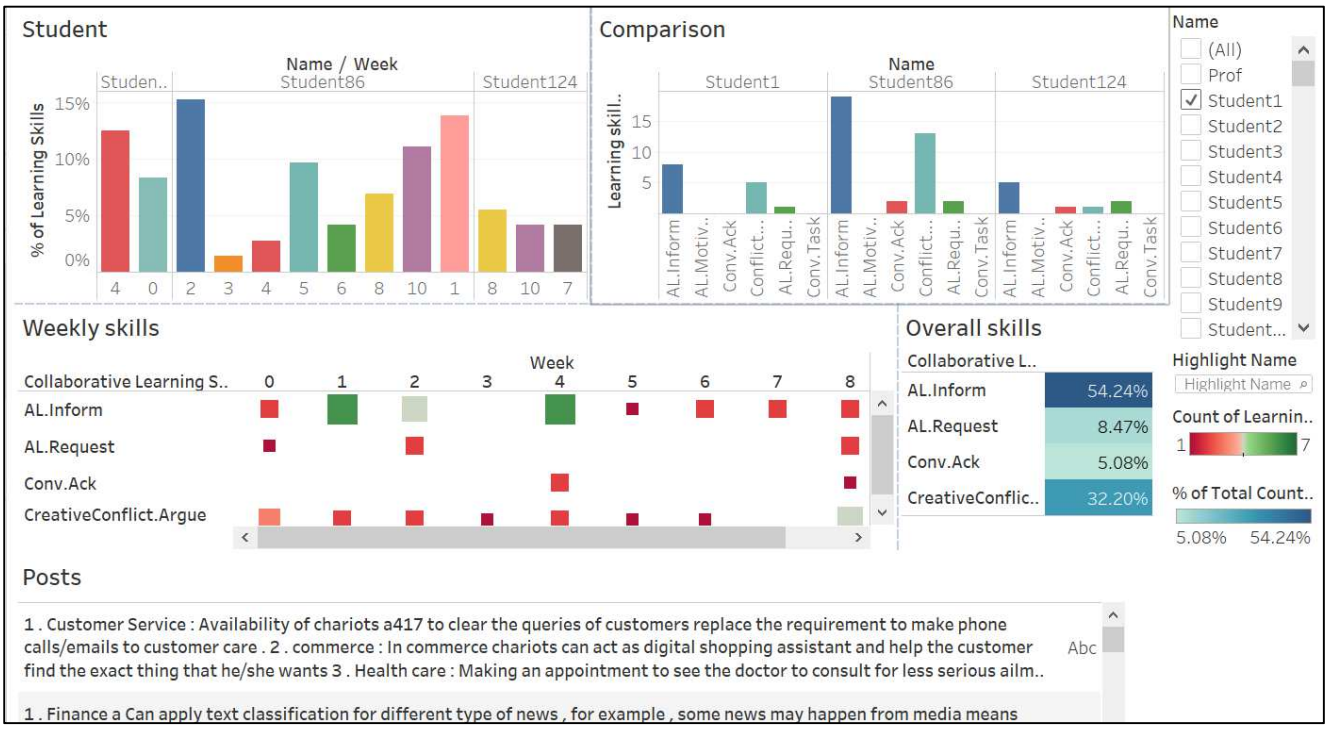


Fig. 3. Dashboard for Collaborative Learning Insights

mere repetition of the course content. Therefore, we came up with a design where questions were asked that promoted the student to conduct research and then participate in the forum. Moreover, we also observed that the questions that related to topics beyond the class were found to be more interesting and motivated the students to participate in forum discussions. Table V shows the discussion forum settings for the course. Note that the questions are a mix of business and technical aspects which align with the course objectives.

TABLE V. DISCUSSION FORUM SETTINGS FOR CASE STUDY

Name	***Forum***	***Week-Topic***	***Title***	***Body***
Anonymized identification of student who posted the thread	Forum in which thread originated from	Categorised by the weekly content covered in class	Subject of the thread	Content of the thread

3) *Participants and Posts*

LMS Discussion Forum on Text Mining and Language Processing is designed with weekly threads. This dataset consists of approximately 800 posts, and it includes the anonymous student identifier, title, and topic of the discussion as well as the content of the post. We had 126 students and one professor. The topic consists of #week, and the topics are separated by "-" which can be split with regular expression.

B. *Experiments and Evaluations*

To answer the research questions from a more objective standpoint, we conducted an exploratory analysis to study the general statistics of the student learning skills as well as individual behaviors. To answer RQ1, we conducted manual evaluations on the outputs generated by the tool for the mapping and scoring stage.

Evaluation of Tool: To evaluate the solution model, we selected sample posts and manually tagged the posts to the various discourse relations. We then compared it with the output generated by the tool. The results of this process are shown in Fig. 4.

Student	Week	Manual	Algo	Accuracy
126	1	Tempoal.Synchrony (1) Active_Learning.Inform (2) Contingency.Cause (10) Temporal.Asynchronous (4)	Tempoal.Synchrony (5) Active_Learning.Inform (22) Contingency.Cause (10) Temporal.Asynchronous (4) Contingency.Cause (10) Motivate.Encourage (2) CreativeConflict.Argue (1)	100.00%
	2	Active_Learning.Inform(6) Contingency.Cause (1) Temporal.Synchrony (2)	Active_Learning.Inform(6) Contingency.Cause (1) Temporal.Synchrony (2)	100.00%
	4	Active_Learning.Inform(4) Active_Learning.Motivate(1) Temporal.Synchrony (2)	Active_Learning.Inform(5) Temporal.Synchrony (2)	85.71%
	5	Temporal.Synchrony (4) Active_Learning.Inform (19) Contingency.Cause (5) Temporal.Asynchronous (5) Contingency.Condition (5) Conversation.Acknowledge (1) CreativeConflict.Argue (1)	Temporal.Synchrony (7) Active_Learning.Inform (19) Contingency.Cause (5) Temporal.Asynchronous (2) Contingency.Condition (5) Conversation.Acknowledge (1) CreativeConflict.Argue (1)	92.50%

Fig. 4: Evaluation process of the solution model

To calculate accuracy, we use the difference in the count between manual tagging and system tagging. As shown in Fig. 4, the accuracy is the percentage of system-generated tags that are very similar to the manually assigned tags. The overall accuracy of the solution model is 94.72%.

C. *Dashboard*

To answer RQ2, we created a dashboard and conducted an exploratory visual analysis as shown in Fig. 5, 6 and 7.

1) *Participation Analysis*

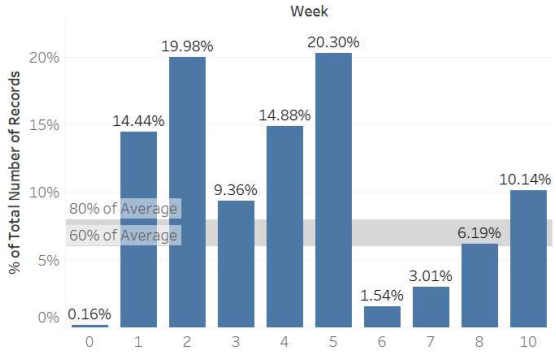


Fig. 5. Weekly participation details (x-axis indicates the lesson week).

2) Learning conversation and Group processing skills

a) Overall analysis

Collaborative Learning Skills	
Active_Learning.Inform	67.26%
Active_Learning.Motivate	0.36%
Active_Learning.Request	1.52%
Conversation.Acknowledge	1.30%
Conversation.Task	0.07%
CreativeConflict.Argue	29.48%

Fig. 6. Overall conversation skills statistics

b) Comparison analysis

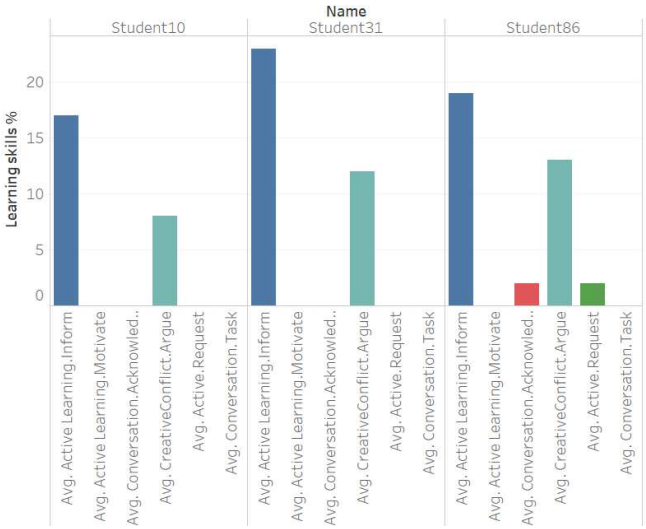
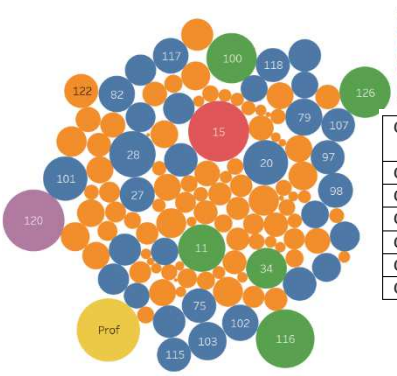


Fig. 7. Comparison analysis

D. Discussions

a) Conversation Skills - Grades Analysis

In this section, we further analyze the correlation between students' grades and their conversation skills profiles. To provide insights to the instructor, a cluster analysis approach is used where the students are clustered into groups based on the conversation skills in the discussion forum. Subsequently, the average grade of the cluster provides the correlation between the students' conversation profiles and the grades. As shown in Fig. 8-a), the students in Cluster 1 have a higher average grade than the students in Cluster 2. Correspondingly, in Fig. 8-b), we can observe that the students in Cluster 1 demonstrate higher collaborative learning skills in the discussion forum.



Cluster	Avg Grade	No of Users
C1	79.51613	31
C2	78.5814	88
C3	79	1
C4	83.2	5
C5	0	1
C6	86	1

(a) Six clusters of students in the course and average grades

Cluster	C1	C2
Active Learning.Inform	**83.71**	29.727
Active Learning.Motivate	**14.097**	1.8523
Active.Request	**58.161**	4.5114
Conversation.Acknowledge	**49.516**	1.1705
Conversation.Task	0	**0.52273**
CreativeConflict.Argue	**83.71**	29.352

(b) Cluster analysis in terms of clusters

Fig. 8. Correlations of graded with conversation skills using cluster analysis

b) Benefits of the IDALS Framework

Ability to track user behavior: Based on the results generated from the discourse analysis, we gathered each student's Soller's skills probability scores based on their participation in the discussion forum. Using the data gathered, we can track the student's weekly discussion behaviors from week 1 to week 10, compare the differences in scores, and identify the changing patterns of behaviors over the span of 10 weeks. It satisfies the objective of the instructors wanting to gain a better understanding of the students' behavior and profile.

Provide a gauge for the instructor: With the tracking ability, the instructors are provided with an overview of the class and individual students' profiles. With these insights, instructors can design course materials that cater to individual student's needs and create a personalized learning experience for the students through online learning. For example, the instructor can create personalized discussion threads for each student. Thus, we believe that the discourse analysis did well in the aspect of providing a measure for instructors to prepare their materials catering to student's needs and profiles.

Grouping students based on the profile: Upon generating the Soller's probability scores for each student, we investigated other ways to make full use of the data and decided to use clustering for grouping students of the similar Soller's profile

type together. Using K-Means clustering, we identified 6 clusters that can be used by the instructor in the future for pre-assigning groups. As the current method of grouping is usually randomized or based on grouping students of different grades together, students can end up in an unsuitable group due to the incompatibility in their working styles. By using the generated data as an estimator, groups can be formed by matching compatible students' Soller's types.

c) Limitations of the IDALS Framework

NLP technologies: Developing a system to analyze student communication is not a trivial task since even the latest natural language understanding technologies today combined with CSCL tools are still limited in their ability to understand and interpret student communication. For example, we use the raw statistics of discourse markers, and this may benefit the long posts and in future this should be addressed with normalization models. Another challenge is the limitation of dictionary models where the synonym words, spelling errors, paraphrased sentences, or ungrammatical sentences may reduce the accuracy of the solution model. To address these challenges, an AI layer that focuses on normalizing the language features of the posts should be integrated into the framework.

Design of discussion forums: An effective design and implementation of the discussion board are crucial in enabling better participation of the students [13]. At the same time, the instructors' role in facilitating the discussion also has a significant impact on the participation of students. Encouraging active participation also increases the likelihood that all group members will learn the subject matter and decreases the likelihood that only a few students understand the material and thereby leaving the others behind. One of the future directions is to combine this framework with the design features of the discussion forum that can enable to measure the effectiveness of learning the course material together with the social skills profile management.

Learner's profile and collaborative skills framework: An intelligent system not only generates learners' profiles based on the collaborative skills but also provides recommendations to the faculty or students on the gaps in the discussion forums. To achieve this, the learner's pre and post skills should be captured and integrated into the system. The faculty should also give some basic training to the students to help them understand the learning outcomes of the discussion forum in terms of social skills development. Soller's framework is the best place to start with for such awareness. However, this framework only provides the profiling at the micro-level. Integrating such micro profile to principles of personalized learning frameworks [30] will increase student engagement and achievement.

VII. CONCLUSION

In today's digital world, online discussion forums are widely used by students as a learning strategy. Discussion forums allow for constructive interactions both in-class and outside-of-class, thereby creating a round-the-clock learning environment. Further, vibrant and constructive discussions by students generate a wealth of knowledge which can be re-used in the future. Discussion forums contain discussion "threads" related to key course topics. In this paper, we presented an Integrated Discourse Analysis and Collaborative Learning Skills (IDALS) framework based on in-class discussions. We proposed the application of discourse analysis and collaborative learning frameworks to discussion forums to gain further insights into the student's learning in a classroom. Our solution leverages Soller's framework and Penn Discourse Treebank 2.0 to understand interactions at the discourse and semantic level. It uses unsupervised automated techniques to diagnose interactions in textual data. We applied IDALS framework and evaluated our solution model on discussion threads from a graduate-level text mining course. Firstly, IDALS framework supports discourse analysis by generating students' Soller's skills probability scores based on their participation in the discussion forum, and the instructor can track how the scores change over the course of a semester. Secondly, IDALS framework provides useful insights to the instructor about students' profiles to help the instructor design personalized learning experiences for the students. Lastly, the framework leverages K-Means clustering to group students based on their Soller's types. This approach can lead to improved group formation where students end up with suitable group members in terms of working styles.

REFERENCES

[1] A. Soller, A. Lesgold, F. Linton, and B. Goodman, "What makes peer interaction effective? Modeling effective communication in an intelligent CSCL," 1999.

[2] A. Soller, "Supporting social interaction in an intelligent collaborative learning system," International Journal of Artificial Intelligence in Education, vol. 12, pp. 40–62, 2001.

[3] R. Killen, "Effective teaching strategies: Lessons from research and practice," 2003.

[4] S. Gottipati, V. Shankararaman and M. N. Gokarn, "Automated Discussion Analysis - Framework for Knowledge Analysis from Class Discussions," 2020 IEEE Frontiers in Education Conference (FIE), 2020, pp. 1-8, doi: 10.1109/FIE44824.2020.9273819.

[5] N. Kemp and R. Grieve, "Face-to-face or face-to-screen? undergraduates' opinions and test performance in class-room vs. online learning," Frontiers in Psychology, vol. 5, 2014.

[6] J. Gee, J. "Discourse analysis: What makes it critical?" In R. Rogers (Ed.), An Introduction to critical discourse analysis in education (2nd Ed.), pp. 23-45, New York: Routledge, 2011.

[7] L. Rex and L. Schiller, "Using discourse analysis to improve classroom interaction," Using Discourse Analysisto Improve Classroom Interaction, pp. 1–163, 2009.

[8] E. Miltsakaki, R. Prasad, A. Joshi, and B. Webber, "The penn discourse treebank," in LREC, 2004.

[9] R. Prasad, N. Dinesh, A. Lee, E. Miltsakaki, L. Robaldo, A. Joshi, and B. Webber, "The penn discourse treebank2.0." in LREC, 2008.

[10] S. Walsh, Exploring classroom discourse: Language in action. Abingdon, Oxon: Routledge, 2011.

[11] K. Hyland, Metadiscourse: Exploring interaction in writing. London: Continuum, 2005.

[12] K. Aijmer, English Discourse Particles. Amsterdam: John Benjamins, 2002.

[13] E. Burge, "Learning in a computer conferenced contexts: The learners' perspective," 1994.

[14] D. Warriner, K. T. Anderson, "Discourse Analysis in Educational Research," In K. King, Y. J. Lai, S. May (eds) Research Methods in Language and Education. Encyclopedia of Language and Education (3rd ed.). Springer, Cham. 2017.

[15] R. Rogers, (Ed.). An introduction to critical discourse in education. New York: Routledge, 2011.

[16] B. Rymes, Classroom discourse analysis: A tool for critical reflection. New York: Hampton Press, 2009.

[17] N. Markee, The Handbook of Classroom Discourse and Interaction . Malden, MA: Wiley Blackwell, 2015.

[18] R. Wodak and P. Chilton, A New Agenda in (Critical) Discourse Analysis. Amsterdam: John Benjamins, 2005.

[19] M. N. Gokarn, S. Gottipati and V. Shankararaman, "Cognitive and Social Interaction Analysis in Graduate Discussion Forums," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-8, doi: 10.1109/FIE43999.2019.9028485..

[20] B. Kumaravadivelu, "Critical Classroom Discourse Analysis." TESOL Quarterly 33 (3): 453–484, 1999.

[21] J. Park and C. Cardie, "Improving implicit discourserelation recognition through feature set optimization," in SIGDIAL Conference, 2012.

[22] Y. Kido and A. Aizawa, "Discourse relation sense classi-fication with two-step classifiers," inCoNLL Shared Task,2016.

[23] E. Stepanov and G. Riccardi, "Sentiment polarity classification with low-level discourse-based features," Proceedings of the Second Italian Conference on ComputationalLinguistics CLiC-it 2015, pp. 269–273, 2015.

[24] M. K. Chandrasekaran, C. D. Epp, M.-Y. Kan, and D. Litman, "Using discourse signals for robust instructorintervention prediction," in AAAI, 2017.

[25] M. G. Sindoni, "Online Conversations: A Sociolinguistic Investigation into Young Adults' Use of Videochats." Classroom Discourse 2 (2): 219–235, 2011.

[26] E. Piurko. Discourse markers: their function and distribution in the media and legal discourse. M.A. Thesis. Lithuanian University, 2015.

[27] Y. Zhou and N. Xue, "PDTB-style discourse annotation ofchinese text," in ACL, 2012.

[28] B. Hurst, R. Wallace, and S. B. Nixon, "The Impact of Social Interaction on Student Learning," Reading Horizons: A Journal of Literacy and Language Arts, 52 (4), 2013.

[29] S. Bird, "NLTK: The natural language toolkit," ArXiv, vol. cs.CL/0205028, 2006.

[30] S. Patrick, K. Kennedy, and A. Powell, "Mean What You Say: Defining and Integrating Personalized," Blended and Competency Education, 2013.

[31] S. Gottipati, V. Shankararaman and R. Ramesh, "TopicSummary: A Tool for Analyzing Class Discussion Forums using Topic Based Summarizations," 2019 IEEE Frontiers in Education Conference (FIE), 2019, pp. 1-9, doi: 10.1109/FIE43999.2019.9028526.