Adopting Learning Analytics to Promote Collaboration in Online Discussions Written in Portuguese

[α]**Máverick André,** [α]**Anderson Pinheiro,** [α]**Vitor Rolim,** [β]**Rafael Dueire Lins,**
[β]**Taciana Pontual Falcão,** [β]**Rafael Ferreira Mello**

[α]Universidade Federal de Pernambuco, Centro de Informática
[β]Universidade Federal Rural de Pernambuco, Departamento de Computação

Corresponding author: rafael.mello@ufrpe.br

Abstract. Learning Management Systems (LSM) have been adopted to provide interaction between instructors and students in online learning courses. In these environments, large volumes of messages are exchanged, making it difficult for the participants to follow all the discussions.
When messages are ignored, students' opportunities to effectively construct knowledge through online discussions are reduced. In this context, this paper proposes a method for mining the content of messages written in Brazilian Portuguese to enhance students and instructors' interactions in online discussion forums. Four different automatic methods were developed, one for each of the following four dimensions of the originality and collaboration processes: (i) Semantic textual similarity to measure originality; (ii) Expressing appreciation towards other participants; (iii) Recognizing group presence; (iv) Sharing information and resources; (v) Soliciting feedback/Answering questions. The use of the methods for analyzing online discussions forums was validated in a case study conducted in a real course. A discussion forum tool was developed to provide feedback for students and instructors about the originality and collaboration level during online discussions. Results showed the potential of the proposal in promoting students' collaboration, enhancing interaction and reducing plagiarism in discussion forums.

André, M., Pinheiro, A., Rolim, V., Lins, R.D., Falcão, T.P., Mello, R.F. (2021) Adopting Learning Analytics to Promote Collaboration in Online Discussions Written in Portuguese. Computer-Based Learning in Context, 3 (2), 1-15.

Keywords: Learning Analytics, Online Discussion, Natural Language Processing, Collaboration, Originality Detection

1 Introduction

Learning Analytics (LA) can offer practical solutions to analyze the massive amount of data generated in Learning Management Systems (LMS), for example by applying automatic methods to extract information from the LMS generated data (Lang et al., 2017). It can be used to improve student engagement and performance (Coffrin et al., 2014), measuring the time elapsed by students in performing tasks (Kovanović et al., 2017a), and to model students' "behaviour" and participation (KickmeierRust et al., 2016). In its early days, much research in LA focused on extracting information from log data about the interaction with technology. However, recent research has focused on analyzing the textual contents of educational resources, such as analyzing essays, the production of academic texts, answers to open-ended questions, and online discussion forums (Lárusson and White, 2012; Simsek et al., 2015; Dascalu et al., 2015; Kovanović et al., 2015).

Among the resources available in a LMS, educational forums are widely used to encourage student participation, to answer questions, and to share resources (Hew and Cheung, 2008), being an asynchronous tool that can enhance collaboration in online learning courses (Kovanović et al., 2017b). Although the adoption of discussion forums can bring many benefits, a large volume of posts makes it difficult for the instructor and students to keep up with all interactions and leverage opportunities for social knowledge construction (Marbouti and Wise, 2016). Thus, a relevant issue arises: "*Can large volumes of online discussion data be automatically analyzed to provide meaningful insights for the students and instructors?*" There are different approaches to deal with this problem. One line of research proposes methods that analyze interactions within online educational discussion using shallow features, which basically uses statistics about the interactions and simple text analysis. For instance, methods based on direct statistics (e.g., number of posts and replies) and word cloud (Moreno-Marcos et al., 2018; Hu et al., 2018). In this case, the methods do not consider deeper text analysis such as text semantics.

Another possibility is to adopt natural language processing methods, such as sentiment analysis (Wen et al., 2014), text classification (Kovanović et al., 2016), and text summarization (Bhatia et al., 2014). Although the works cited produced significant results, they were primarily designed for texts in English only. The mapping of such methods to another language, such as Brazilian Portuguese, is not always easy or possible due to the absence of text mining tools and resources for languages other than English, e.g. LIWC (Linguistic Inquiry and Word Count) (Tausczik and Pennebaker, 2010) and Coh-Metrix (Graesser et al., 2004; Barbosa et al., 2021).

The present work aimed to provide methods for analyzing large volumes of collaborative discourse among students in educational forums written in Brazilian Portuguese, thus contributing towards a broader support for learning analytics in different languages. Particularly, this work focused on the analysis of collaborative indicators within an online discussion. Moreover, we also analyzed the messages posted by the students in terms of the originality of the contributions. In order to accomplish this, we proposed:

i. Using machine learning and rule-based approaches to automatically identify the four analytics for the detection of collaborative indicators in discussion transcripts in Brazilian Portuguese namely, expressing appreciation towards other participants, recognizing group presence, sharing information and resources and asking for feedback/answering questions.
ii. Adapting a measure for semantic textual similarity originally developed in English to Brazilian Portuguese. Such measure allowed assessing the level of novelty of a post by comparing it with the preceding messages in the forum.

2 Background

2.1 **Discussion forums.** An educational discussion forum is an asynchronous communication resource often used in LMS to promote discussions and interactions amongst participants about some educational topic (Pendry and Salvatore, 2015). It is a communication channel where students and instructors can share questions, opinions, comments and answers. Educational forums are usually mediated by instructors, who assist and supervise students in their learning process (De Wever et al., 2006; Garrison et al., 2010; Ferreira-Mello et al., 2019). Instructors can adopt the forum for:

- Encouraging the creation of bonds between students (Joksimovic et al., 2014)
- Developing the capacity for critical discussion about a theme or subject (Yoo and Kim, 2014)
- Answering questions and comments (Lin et al., 2009); and
- Assessing students (Rubio and Villalon, 2016).

Despite the advantages brought by adopting forums as communication resources in educational environments, it may also lead to the problem of information overload (Wulf et al., 2014), hindering students from effectively constructing knowledge (Marbouti and Wise, 2016), and instructors from being able to facilitate the online discussions (Garrison et al., 2010). Following up a large number of posts may become a heavy workload for the instructor, which, if not properly handled, may lower the motivation level of the students, leading to a decrease in the quantity and quality of their posts. These problems may be addressed by adopting the methods proposed in this paper to extract learning analytics related to collaboration and originality.

2.2 **Collaboration in educational forums.** Several researchers, over the last few years, have been working on strengthening the collaborative aspects of discussion forums in online learning. Murphy (2004) lists general aspects of collaboration that are relevant to educational forums, including: social presence; articulating individual perspectives; accommodating or reflecting the perspectives of others; co-constructing shared perspectives and meanings; building shared goals and purposes; and producing shared artefacts. Usually, the works to improve collaboration in educational forums address some of these aspects.

Garrison et al. (1999) highlight the importance of social interaction in online discussions with the concept of social presence, which is “the ability of participants in a community of inquiry to project themselves socially and emotionally, as ‘real’ people (i.e., their full personality), through the medium of communication being used” (p. 94). Social presence includes three categories with several indicators, which form a road-map to interpret the concept of interactions.

All indicators proposed by Murphy (2004) and Garrison et al. (1999) are relevant to analyze social interactions. However, four of them appear in both works and are listed as important psychological factors to improve social participation in online discussions (Amichai-Hamburger et al., 2016; Castellanos-Reyes, 2020):

- Expressing appreciation towards other participants: This binary indicator relates to students' expression of positive sentiments towards previous posts or forum participants.
- Recognizing group presence: This indicator relates to students' acknowledgement of the presence of peers.
- Sharing information and resources: This indicator relates to the identification of links or resources shared in the educational forum.
- Asking for feedback/Answering questions: This indicator is used to identify if students are answering the questions proposed by the instructors or raised by peers, or asking for help to understand the questions.

Therefore, the method we proposed aimed to automatically analyze those four different collaboration features to provide useful analytics to students and instructors. Moreover, we provided a learning analytics visualization tool to promote self-regulation among students.

2.3 **Approaches to online discussion analysis.** The literature offers several approaches to using learning analytics in the context of online discussions (Joshi and Rosé, 2007; Yen, 2013; Kim et al., 2016; Wise et al., 2014; Ferreira et al., 2020), with some emerging work on methods for automated analysis of online discussions (Mu et al., 2012; Kovanović et al., 2016). For example, text classification can be used for assessing online participation of students and monitoring of the learning progress through collaboration, among others (Lui et al., 2007). Moreover, Rosé et al. (2017) summarized several applications and the importance of collaboration within online discussion in terms of student assessment and to support instructors' decision-making.

Profiling students' interactions in threaded discussions is one of the most relevant applications of text analysis in learning analytics. Ravi and Kim (2007) proposed the classification of students' posts individually as "discourse acts" which include the following categories: (i) complement (complement of a previous message); (ii) information (information, command or announcement); (iii) correction (correction or objection to an earlier message); (iv) elaboration (elaboration of an earlier message or description, including the elaboration of questions and answers); (v) question (a question about a problem, including questions about previous messages); and (vi) answer (response to an earlier question or suggestion). They used a set of n-gram (n varying from 1 to 4) and a Support Vector Machine (SVM) classifier to categorize the posts. In the best case, the method reached an accuracy of 92.70%.

Recognition of the kind of message (genre) in online discussion posts is another important topic that has been addressed in the literature. Lin et al. (2009) proposed a system that classifies the genre of a post using a bag of words (using all words, and only verbs, adverb, or contextual words) as features and the decision tree algorithm for classification. The genres that this system sought to identify were: (i) announcements (clarifies doubts); (ii) questions (proposes a question); (iii) interpretation (interpretation based on facts and ideas exposed); (iv) conflict (conflicting opinion); (v) affirmation (maintaining and defending the idea disagreed by others); and (vi) others (varied messages that are difficult to categorize). Better results were achieved using the combination of verbs with contextual words and adverbs with contextual words.

In addition to the classification of posts, the extraction of indicators from forums that may aid instructors in following up on the discussion is another topic considered in the literature. McLaren et al. (2007) proposed the use of Awareness Indicators to provide an interface that allowed the instructor to supervise all of the online discussion. This indicator classified the posts as positive or negative, which meant whether the student needed help or not. Although they used machine learning, the features (e.g., text length, shape type, number of in-links, number of out-links, and number of indirect links) were provided manually. Different decision tree algorithms were evaluated, reaching an accuracy of 85.00%.

Most of the existing research has focused on the analysis of online discussions in English. Recently, Neto et al. (2018) proposed the adoption of LIWC and Coh-Metrix, which are text mining tools to measure text readability and coherence, to classify higher-level classes as the cognitive presence, in Portuguese. Moreover, Barbosa et al. (2020) and Barbosa et al. (2021) used the same tool to proposed approaches to evaluate texts in English and Portuguese according to cognitive and social presences. While Barbosa et al. (2020) presented Multilanguage classifier, Barbosa et al. (2021) applied text translation to create a model for both languages. Although both works provided interesting results, they offer classifiers for high-level categories and not for specific indicators. Moreover, there has been little attention devoted to developing systems that can transform the information extracted by text analysis into insights for the instructors and students.

This paper aimed to address those gaps by providing both: (i) a method for the automated analysis of student-generated content in online discussions in Brazilian Portuguese in detailed categories; and (ii) student and instructor dashboards with insights obtained from the automated analysis of the texts.

3 Methodology

This section presents the indicators extracted from online discussions to create the proposed dashboards. The unit of analysis for both levels was the message posted in an online discussion. Each post can contain more than one indicator of the analyzed dimensions. The following sections present further details on each of the proposed text analysis methods.

3.1 **Expressing appreciation towards other participants.** Sentiment Analysis (SA) techniques were used to analyze the degree of explicit friendliness in the students' messages in the forum. The SA algorithm proposed by (Cambria, 2016) was used here to identify if a text message was either positive, negative or neutral compared with the previous messages posted by other students in the forum. A traditional approach to SA adopts semantic resources (e.g., lexical dictionaries) and machine learning methods. Following this idea, a Brazilian Lexical Dictionary called SentiLex was used (Silva et al., 2012). SentiLex encompasses more than 7,000 words divided into different grammatical functions (verb, noun, adverb, adjective) and polarities (positive, negative, neutral) as one feature for a machine learning algorithm.

As the problem was to identify the degree of appreciation between the participants, this work focused on extracting the sentiment related to other students in the discussion. Thus, besides the SentiLex, other four features were used:

- ***Proper name***: As the goal was to identify the appreciation to the post of another participant, it was important to identify the proper names listed in the post. Thus, this feature stored the number of proper names found. To accomplish this, a Named Entity Recognizer (NER) system was used (Sarmento, 2006);
- ***Distance Name/Complement***: This feature was used to determine if an instance of appreciation is related to a proper name or not, by measuring the distance between the proper name and the positive word extract from SentiLex.
- ***Information Sharing***: This feature used the information sharing analytics (presented in "Sharing information and resources" section) to eliminate the appreciation related to a resource, by looking at compliments made by a participant to another;
- ***Post Size***: This feature captured the size of a post, in terms of the the number of characters. The idea was that longer posts tend to have more information, thus increasing the chance of being allusive to a previous post or another participant in the discussion.

The above features were applied to an SVM classifier to categorize posts according to their appreciation towards other participants. Such an SVM classifier was chosen because it reached good results in a wide range of applications (Fernández-Delgado et al., 2014).

3.2 **Recognizing group presence.** We proposed a rule-based approach to recognize Group Presence (Murphy, 2004). If a post adhered to any of the four rules below, the system classified it as group presence:

i. The first rule identified the existence of words tagged as an interjection in the analyzed post's first three positions. This strategy was adopted because interjections such as "*hi*" and "*hello*" customarily appear at the beginning of a conversation, indicating group presence. The implementation of this rule made use of the part-of-speech class algorithm available at Spacy[1].
ii. The second rule followed a very similar idea, but it identified the greetings at the beginning of a post instead of the interjection. For instance, expressions such as "good morning", "good afternoon" and "good evening".
iii. The third rule searched for plural personal pronouns (e.g., us and you) and treatment pronouns (e.g., Mr., Mrs., Prof.) in the message.
iv. Finally, the last rule extracted terms representing collectives in educational settings, such as colleagues, friends, and companions, that are classified as nouns. If one or more expressions were found, then the analyzed post was said to have the group presence indicator.

3.2 **Sharing information and resources.** Students often share information and resources by posting websites in online discussions. Thus, this feature was based on a link identification approach. Moreover, an automatic approach was used to check if the links retrieved were related to the forum's topics.

[1] https://spacy.io/usage/linguistic-features

Initially, we used a regular expression to verify if the post contained any URL in its body. Then, the link was validated to check if it was still active. Next, based on the content of the web page, a TFISF (Term Frequency-Inverse Sentence Frequency) was performed to find the main concepts from the retrieved content (Neto et al., 2000). The computation of the TFISF followed the same idea of TFIDF presented in Section Originality, but here instead of calculating the TFIDF of a document, we used a sentence. The same TFISF analysis was made with the nouns in the previous posts in the forum. If the intersection of both outputs was non-empty, we considered that the information shared was related to the proposed content for discussion in the forum. Otherwise, the instructor received an alert to check the content of the posts. It is important to mention that this feature only covered resources written in Portuguese.

3.3 **Asking for feedback/Answering questions.** This module aimed to classify the posts in discussion forums in: asking for feedback (questions), answering questions (answer), and neutral comment. To accomplish that, different features were used to train an SVM classifier (Joachims, 2002). A Genetic Algorithm (GA) (Mitchell, 1998) was used in finding the best set of parameters for the SVM classifier.

In the feature extraction step, the information required for the classification process was retrieved from the posts. The feature extraction used CoGrOO (Silva and Finger, 2013) to perform a part of speech tagging in each post. Next, word vectors of each class (questions, answer and neutral) were created. To avoid vectors with high dimension, only verbs, nouns, pronouns, adjectives and adverbs were considered.

The word vectors were ranked decreasingly concerning the representativeness (measured by TFIDF), for each grammatical class, at the beginning of the vectors. Here, TF was the number of times a word appeared performing the same grammatical function within the universe of postings of a specific class (questions, neutral or answer); and IDF measured the relationship between the number of times a word appeared in a specific class divided by the total number of posts that had this word. Finally, for each post, a vector was generated using 15 features, encompassing the pair $< class, grammatical\ class >$, where the classes were question, answer and neutral; and the grammatical classes were verb, noun, pronoun, adjective and adverb. Each position of the vector received the number of words with the same grammatical function, which appeared in the post and in one of the word vector classes.

After the feature extraction, the postings' classification was performed through the application of an SVM classifier. To improve the results of the classifiers, the parameters were optimized using a GA.

The SVM received the feature vectors and performed the training using each set of parameters generated by the GA until a termination condition was met. The GA had its initial population represented by solutions (individuals) whose genetic material consisted of eight genes referred to the SVM parameters: $SVMType$, $probability$, $kernelType$, $gamma$, nu, $cacheSize$, $cost$, and $epsilon$.

The execution cycle of the GA was performed with the following steps:

i. ***Initial population generation***: initialization of the algorithm with individuals (solutions) with random parameters (genes)
ii. ***Evaluation of individuals***: running an SVM algorithm on the set of parameters of each individual of the population.

Each GA solution's objective was to parameterize the SVM to classify the posts. For each execution, the F-measure was calculated and used as the function to be optimized; in GA, the F-measure was the "fitness" of each individual.

To create the new generations of individuals, the following methods were used (Melanie, 1999):

i. ***Crossover***: aimed to mix the genetic material of the selected individuals (parents) with the purpose of generating new solutions (children) capable of exploring new spaces in the search field, using uniform crossing;
ii. ***Mutation***: inserted diversity into the population using the change (mutation) of some genes, given a probability defined a priori.

The last step was to select the surviving individuals. In this step, the steady-state method (Melanie, 1999) was used, which consists of selecting the best solutions within a universe constituted by the previous population and the population acts.

The final output of this module used an optimized SVM to classify the posts into question, answer and neutral. It was used to measure the degree of collaboration in asking for feedback and answers to questions.

3.4 **Originality.** The semantic textual similarity measure for Brazilian Portuguese proposed in (Cavalcanti et al., 2017) was adapted to educational settings to detect the level of originality of the post in a forum. It measures the degree of similarity between posts in the entire forum. To accomplish that, four features were extracted: TFIDF similarity, Word2Vec similarity, binary similarity, and the size of sentences. Details about each of the features are presented in the following subsections.

- **TFIDF.** The TFIDF (Term Frequency-Inverse Document Frequency) is a statistical measure that stands for the importance of a word in a set of documents (i.e., a set of forum messages in this case) (Ricardo, 1999), widely used in natural language processing. Two text processing steps were used: (i) each word was expanded using two synonyms from TeP (Thesaurus for Brazilian Portuguese) (Aluísio et al., 2008). As there is typically a number of words in a forum message, the values of TFIDF are typically too small. (ii) application of a stemming algorithm to reduce the sparsity of the data (Hartmann, 2016). The final similarity between two posts was calculated by applying the Cosine distance between their TFIDF vectors.
- **Word2vec.** The Word2vec model makes use of a neural network to build "classes of equivalence" of a given word. Through training, Word2vec translates texts into a numerical Kdimensional vector space. Each word in a text is represented as a vector, allowing to measure the degree of similarity between words as the distance between two vectors (Mikolov et al., 2013). The Word2vec model used in this paper was built using the original implementation of Word2vec[2] based on Wikipedia[3] and news texts obtained from the G1[4] portal from September 15, 2016, to December 5, 2016. The default parameters for word2vec training were adopted in the current study. At this step, the stopwords were removed from each of the posts, which underwent lemmatization before the similarity calculation. Then, the matrix similarity method presented in (Ferreira et al., 2016) was applied.
- **The Binary Similarity Matrix.** The method detailed in (Ferreira et al., 2016) of the similarity matrix between sentences was used here, but the similarity values between words were set to 1 if the words were equal and to 0, otherwise.
- **Size of Posts.** Based on the idea that two posts with a different number of words potentially convey different information, an additional feature representing the size of the post was added (Zhao et al., 2014). To obtain a value that represented the size of the post (Post Size), the number of words of the shortest post (shortPost) was divided by the number of words of the longest post (longPost), as shown in Equation 1. Before applying this method, the stopwords were removed.

$$PostSize = \frac{shortPost}{longPost} \tag{1}$$

- **Linear Regression.** A multiple linear regression model was applied to the four similarity measures outlined in previous sections to obtain the final similarity value between the posts. The linear regression step aimed to verify a functional relationship between a dependent variable and one or more independent variables (Seber and Lee, 2012).

In the educational scenario proposed, the final similarity measure was used to identify the degree of similarity among posts from an online discussion forum. In such a case, the originality of a message is inversely proportional to the degree of similarity between a message and all the other posts in a forum (Sánchez-Martí et al., 2018). Thus, a higher degree of similarity means a lower degree of originality.

4 Evaluation

4.1 **Context and Data.** The proposed methods were evaluated on a database containing 600 posts from a "*Programming Language*" module of a Computer Science undergraduate online course held at a Federal University in Brazil in 2017. The forum was designed to account for 20% of the students' final mark. A total of 35 students posted messages over four weeks of discussions. The course was specifically redesigned to promote online discussions with an emphasis on the originality of discussions and collaboration. The pedagogical goals of the forum were to be a Question & Answering (Q&A) virtual space, and also to promote discussions among the students. In this context, the

[2] http://code.google.com/p/word2vec

[3] https://dumps.wikimedia.org/ptwiki/20160920/

[4] http://g1.com.br/

student had to post at least two messages a week (one question and one reply) in order to increase the interactivity in the discussion. In the end, only 10% of the messages were posted by the instructor.

Two expert coders manually analyzed each post from the database, and a third person acted as a referee to solve occasional divergences. The posts were categorized according to the presence or absence of each indicator considered: Expressing appreciation towards other participants; Recognizing group presence; and Sharing information and resources. Table 1 shows the distribution of the category of the posts and the Cohen's Kappa coefficient (κ) (Cohen, 1960) of agreement between the evaluators.

Table 1. Post distribution in the evaluation dataset

Analytic	Presence	Absence	κ
Appreciation	78	522	0.86
Group presence	19	581	0.75
Sharing information	17	583	0.93

In addition to the experiment, a case study was performed in which the learning analytics extracted was applied in a real-time application to assess students' participation in the online discussion. An educational forum, called iFórum, was developed to present the extracted learning analytics for instructors and students. This case study was applied in a class of "*Advanced topics in artificial intelligence*" with 12 students in the last year of a Computer Science course at the Federal University in Brazil, during the first semester of 2017. During this course, the students discussed scientific papers sent to the forum by the instructor. The course guidelines recommended that the students presented their views on the paper. The students had one week to make posts for each paper assigned. The online discussion accounted for 20% of the final mark.

4.2 **Experiments.** The following evaluation measures were used to assess the proposal: (i) precision: the number of correctly predicted classes divided by the number of all predicted classes; (ii) recall: the number of correctly predicted classes divided by the number of all instances that should have been classified; and (iii) F-measure: the uniform harmonic mean of the precision and the recall (Forman, 2003).

As presented in the Method section, the proposed approach used an SVM classifier to identify the appreciation and a rulebased method to extract the group presence and the sharing of information. Table 2 shows the results obtained with the proposed approach.

Table 2. Proposed Methods Results

Analytic	Precision	Recall	F-Measure
Appreciation	94.50	94.70	94.60
Group presence	80.70	76.60	78.50
Sharing information	84.60	81.10	82.80

All features obtained a performance above 75% of F-measure, reaching up to 94.60% for appreciation, 78.50% for Group Presence and 82.80% for sharing information. It is important to note that the database was unbalanced due to the nature of the forum, in which most of the posts were questions or answers. Despite that, the proposed approach reached good results.

Moreover, as presented in the Method section, the problem of identifying "request for feedback" and "answering to questions" was translated into a classification problem regarding questions, answers and neutral comments. The same dataset was used to evaluate such an indicator. The 600 posts were divided into 60 posts for questions, 96 neutral comments and 444 answers. The coding was performed using the same methodology previously described, and the agreement reached $\kappa = 0.89$. Two scenarios were developed to evaluate the proposed classification: Scenario 1 (S1) - classification of posts using the features proposed; Scenario 2 (S2) - classification of posts using a TDIDF vector as features. For each scenario, the results were evaluated using a simple SVM classifier with default parameters and adopting the Genetic Algorithm (GA) to optimize the SVM parameters. The evaluation was performed using a 10-fold cross-validation approach.

Table 3 presents the results of the F-measure for each scenario. As the classes “question" and “answer" were equally important, the table shows the F-measure for each class.

Table 3. Question/Answer Evaluation: F-Measure

Method	Questions	Neutral	Answers
S1 (SVM)	87.40	97.90	98.70
S1 (SVM+GA)	97.80	99.70	98.80
S2 (SVM)	28.30	99.00	95.20
S2 (SVM+GA)	54.90	99.40	96.40

The results show that the adoption of the GA enhanced the F-measure for all the classes. We highlight the improvement in the class “question", raised from 87.40% to 97.80% and from 28.30% to 54.90% for S1 and S2, respectively. Besides, S1 outperformed S2 in all cases. For example, S1 improved the result of the identification of “question" in 77.04%, showing the proposed method’s efficacy. The combination S1 (SVM+GA) reached better results than the others.

Figure 1. Proposed educational forum.

4.2 **Case Study.** This section presents the tool created to support the collaboration in online discussion and the results of the proposed case study. Figure 1 presents an overview of the proposed tool. It shows the forum subject, the first post in the forum, a navigation menu and the “thermometer", which allowed to visualize the results of the learning analytics of the extracted indicators. This interface indicates the potential practical value of the developed algorithms in a real course. In terms of visualization for students, the novelty proposed in this forum was the inclusion of the two different “thermometer" bars (in the top-left part of figure 1) showing:

i. the student’s collaboration level, which encompassed the four collaboration features proposed. The final value of this level was measured by dividing the number of posts containing each of the collaboration features by the total number of posts;
ii. the level of originality, which is inversely proportional to the level of similarity between the students’ posts.

Those “thermometers” bars were visible to all participants (students and instructors) and intended to stimulate self-regulation among the students, especially for enhancing collaborative discussion and the quality of posts, avoiding similar messages. Moreover, an increase in the number of messages was also expected.

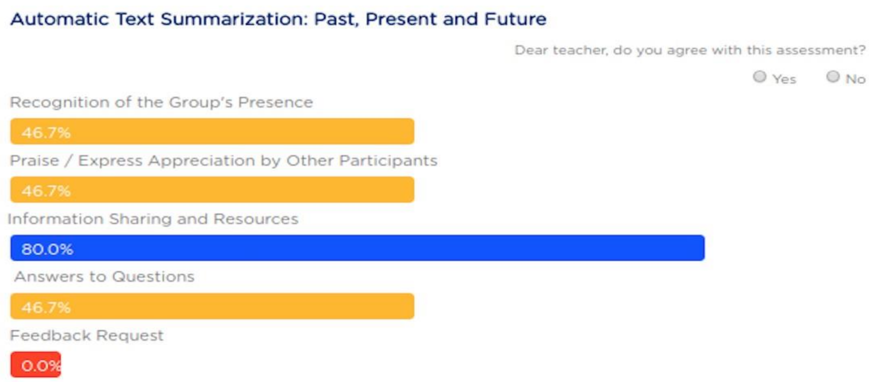


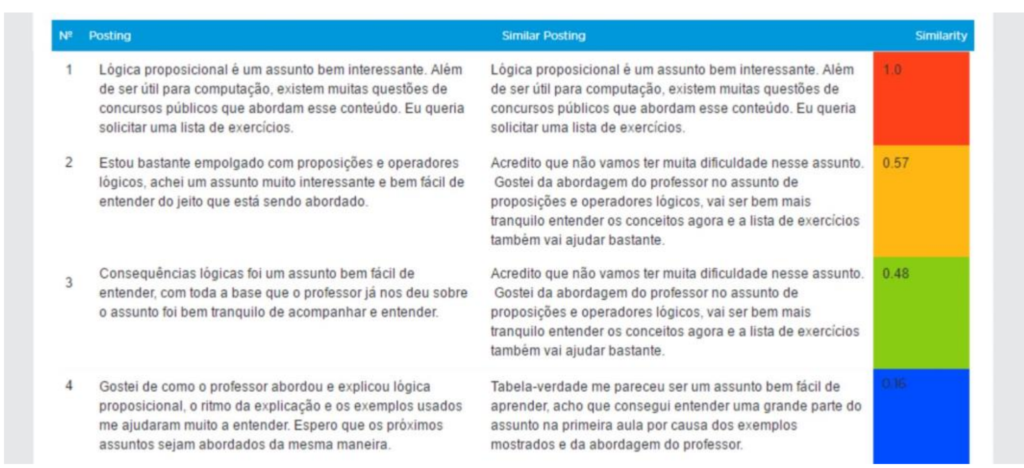
Figure 2. Collaboration per forum.

The proposed tool also contained learning analytics that could be accessed only by instructors, with details on the collaboration analytics per forum and per student and the similarity among posts. Figure 2 presents the level of collaboration analytics from a specific forum. It was also possible to visualize the same information for each student. Such information could aid the formative evaluation and to measure student participation in the online discussion.

Figure 3 presents the analytics about the similarity in the forum, which could be accessed only by instructors. For each post, it listed the most similar post and the similarity value. The similarities were split into 4 groups: blue (0-0.3), green (0.310.5), yellow (0.51-0.7) and red (above 0.7). Those groups were defined by the instructor responsible for this case study, but they can be configured according to the course's design. The colors blue and green meant that the posts did not bear much resemblance to the previous posts in the forum, i.e. they were potentially original posts. The yellow and red colors meant that the posts had /higher levels of similarity to previous posts.

In the first week of discussions (week1), a traditional forum using the Moodle platform was adopted to discuss the articles. At the end of week1, only ten posts were recorded. Two students did not participate in the discussions, and all others posted only once. It is important to mention that the posts entered were simply a summary of the paper proposed for discussion.

The second paper assignment was sent in the proposed tool using the learning analytics extracted. Again, the students had one week (week2) to interact in the forum. As a result, the number of posts on week2 was 30. Besides that, all the students enrolled in the course participated in the forum, and there was a more significant number of interactions between the students, counting not only posts that summarized the article but also texts that included all four categories of collaboration identified by the proposed approach. Each student posted at least twice.



Nº	Posting	Similar Posting	Similarity
1	Lógica proposicional é um assunto bem interessante. Além de ser útil para computação, existem muitas questões de concursos públicos que abordam esse conteúdo. Eu queria solicitar uma lista de exercícios.	Lógica proposicional é um assunto bem interessante. Além de ser útil para computação, existem muitas questões de concursos públicos que abordam esse conteúdo. Eu queria solicitar uma lista de exercícios.	1.0
2	Estou bastante empolgado com proposições e operadores lógicos, achei um assunto muito interessante e bem fácil de entender do jeito que está sendo abordado.	Acredito que não vamos ter muita dificuldade nesse assunto. Gostei da abordagem do professor no assunto de proposições e operadores lógicos, vai ser bem mais tranquilo entender os conceitos agora e a lista de exercícios também vai ajudar bastante.	0.57
3	Consequências lógicas foi um assunto bem fácil de entender, com toda a base que o professor já nos deu sobre o assunto foi bem tranquilo de acompanhar e entender.	Acredito que não vamos ter muita dificuldade nesse assunto. Gostei da abordagem do professor no assunto de proposições e operadores lógicos, vai ser bem mais tranquilo entender os conceitos agora e a lista de exercícios também vai ajudar bastante.	0.48
4	Gostei de como o professor abordou e explicou lógica proposicional, o ritmo da explicação e os exemplos usados me ajudaram muito a entender. Espero que os próximos assuntos sejam abordados da mesma maneira.	Tabela-verdade me pareceu ser um assunto bem fácil de aprender, acho que consegui entender uma grande parte do assunto na primeira aula por causa dos exemplos mostrados e da abordagem do professor.	0.16

Figure 3. Similarity among posts from forum.

Table 4. Question/Answer Evaluation: F-Measure

	Analytic	Number of Posts	Percentage of Posts
Week1	Appreciation	00	00.00%
	Group presence	03	30.00%
	Sharing information	00	00.00%
Week 2	Appreciation	06	20.00%
	Group presence	13	43.33%
	Sharing information	13	43.33%

Table 4 presents the increase in the number and percentage of collaborative posts from week1 to week2. It is important to observe that week1 and week2 had 10 and 30 posts, respectively. Moreover, Figure 4 shows students' individual behavior on week1 and week2 regarding the collaborative indicators. It is possible to see a change of behavior in the students' interactions after adopting the iForum tool. Every student enhanced the quality of their posts, in relation to the features analyses, in week2 (except student 2). This shows that the proposed visualization helped students not only to write more messages but also to incorporate the social skills analyzed.

Regarding the degree of originality of posts in the first week, the average similarity between posts was 0.28, with a standard deviation of 0.13; while in the second week (forum tool with the proposed measure), the mean was 0.27 with standard deviation 0.09. Although there was a reduction in the degree of originality, it did not represent a statistic difference. However, it is important to highlight that the level of plagiarism was lower with the use of the proposed tool, even though it had three times as many posts and almost ten times as many comparisons in the forum (the first and second week had 45 and 435 similarity comparisons, respectively). Moreover, in the second week, the posts had a higher degree of similarity with the instructor's first message.

As an automatic approach to improve originality, the tool sent a private message to the student who had written a post classified in the red group before its publication. It suggested that the student could re-write their post, referring to the previous messages (with a high degree of similarity).

Finally, the instructor used the learning analytics proposed to indicate the assignment of course grades of the students participating in the forum. The participation in this forum represented 10% of the final marks of the course. This 10% was determined using the automatically generated analytics and not the content of the post itself.

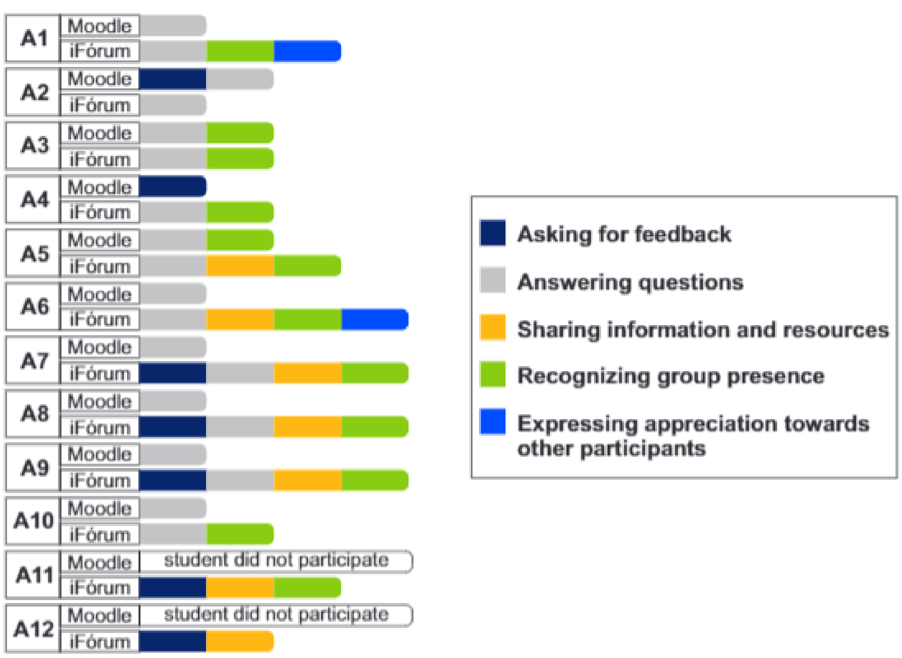


Figure 4. Behavior of students in the first week (Moodle) and second week (iFórum).

5 Discussion

It is important to notice that there is no other work in the literature that adopts automatic methods to identify: (i) expressing appreciation towards other participants, (ii) recognizing group presence and (iii) sharing information and resources for texts written in Portuguese. There are studies that proposed manual approaches to reach similar goals. For instance, a study that sought to identify collaboration in forums and chats was reported by (de Melo Ferreira et al., 2013), where two pairs of linguistics experts in Portuguese

were responsible for creating the rules based on a syntactic analysis. One of the pairs of experts analyzed the discussion forum and then proposed the rules. The second pair of linguists proposed rules without viewing the forum, creating more general rules. The results showed that the rules explicitly created for the forum obtained a better result with 93.95% accuracy, while the generic rules resulted only in 58.90% accuracy. The generic rules approach's low performance when creating the rules independently of the forum is the limitation of (de Melo Ferreira et al., 2013). The rule-based system was created to automatically generate learning analytics that allows the students to evaluate their performance and the instructors to measure the participants' performance in the discussion forums.

Gomes (2012) advocates a teaching-learning process centred on student interactions. He presents a system based on natural language processing techniques to analyze interactions in discussion forums automatically. Using the TFIDF approach, the postings were classified into the following classes: greeting, discussion, motivation, social, information, confirmation, negation, task, clarification, inquiry, and thanks. The highest accuracy obtained was 40% with the Bayesian classifier.

Recently, Ferreira et al. (2020) proposed a classifier based on structural features to develop a classifier for social presence in English. Although the authors reported an accuracy value up to 0.88, it would not be possible to reproduce the results for Portuguese directly due to the lack of resources are available to support text mining applications in Portuguese.

As an attempt to address the problem of limited resources for automatically analyze Portuguese Language, Barbosa et al. (2021) proposed the use of text translation techniques in combination with classification methods. The authors found that it was possible to increase the classification of online discussion messages in Portuguese by up to 55%, translating it to English before extracting the features.

There are two main differences between the approach proposed in the current paper and the previous works Barbosa et al. (2021); Ferreira et al. (2020): (i) both of the presented work relies somehow on English resources and data, which is not always available; (ii) the previous studies focused on high-level social presence indicators, while we proposed the analysis of low-level aspects.

Additionally, several works focus on the automatic cognitive presence for Portuguese (Barbosa et al., 2020; Neto et al., 2018). However, none of them had been reproduced to the analysis of low-level indicators of social presence.
Finally, the case study presented here showed that adopting the proposed learning analytics was associated with an increase in collaboration in educational forums in this scenario. However, further analysis is required for a generalizable conclusion. The number of collaborative posts increased whenever the students interacted with the indicators of originality and collaboration proposed. Moreover, the case study also presented details about the change of behavior for each student, as they started to write more messages considering the social elements evaluated in this study.

6 Conclusion

This work has two main contributions. First, text mining techniques were used to analyze the posts in a forum written in Brazilian Portuguese extracting four different features: (i) Expressing appreciation towards other participants; (ii) Recognizing group presence; (iii) Sharing information and resources; (iv) Asking for feedback/Answering questions. To the best of our knowledge, this paper is the first to address these problems in texts written in Portuguese. Previous works in the literature (Barbosa et al., 2020; Neto et al., 2018) presented attempts to classify the cognitive and social presences in Portuguese; however, they focused on higher-level classes rather than on specific indicators as proposed here.

Second, a case study in a real classroom was performed. In this case study, the instructor proposed a discussion activity about a scientific paper related to the course topic. In the first week, they used the traditional forum, and in the second week, a forum that shows the originality and collaboration learning analytics to the students. In terms of content, the difference between the two weeks was the paper used as the theme of discussion.

The number of posts increased from 10 (in the first week) to 30 in the second week, and the number of posts with collaborative indicators also rose. Besides the positive impact of the proposed analytics, other aspects may have influenced the number of posts in the second week, such as the number of activities of the other modules, the level of interest on the paper focus of the discussion, and the motivation of the students. Besides, the learning analytics information was used by the instructor to assess the interactions among students.

This work's main limitation is the small size of the dataset for the quantitative experiment and the cohort of students (12) in the case study, and the case study's duration (two weeks). The following points may be pursued as lines for further work:

- Replicating the case study in other courses and with a larger number of students;
- Using data from other online discussions to increase the number of posts used in the training step, as some machine learning algorithms (such as deep learning algorithms) perform better on larger corpora. This would also improve the reliability of the results in the quantitative evaluation;
- Adapting the proposed methods to other languages, such as English and Spanish. As the methods for natural language processing are usually language-dependent, the proposed methods cannot be directly adopted;
- Increasing the number of collaboration indicators (Murphy, 2004) such as: (i) Summarizing or reporting on content, (ii) Directly disagreeing with/challenging statements made by another participant, and (iii) Introducing new perspectives;
- Conducting a focus group to understand the students' perception of the proposed tool;
- Performing an empirical study using a control group to evaluate the efficacy of the proposed method.

Acknowledgements

This research was partially supported by the fellowship 310060/2018-4 of the National Council for Scientific and Technological Development – CNPq.

References

Aluísio, S.M., Specia, L., Pardo, T.A., Maziero, E.G., Fortes, R.P., 2008. Towards brazilian portuguese automatic text simplification systems, in: *Proceedings of the Eighth ACM Symposium on Document Engineering*, ACM, New York, NY, USA. pp. 240–248. doi:10.1145/1410140.1410191.

Amichai-Hamburger, Y., Gazit, T., Bar-Ilan, J., Perez, O., Aharony, N., Bronstein, J., Dyne, T.S., 2016. Psychological factors behind the lack of participation in online discussions. *Computers in Human Behavior* 55, 268–277.

Barbosa, A., Ferreira, M., Mello, R.F., Lins, R.D., Gaševic, D., 2021. The impact of automatic text translation on´ classification of online discussions for social and cognitive presences, in: *Proceedings of the International Conference on Learning Analytics & Knowledge*, pp. 605–614. Doi: 10.1145/3448139.3448147.

Barbosa, G., Camelo, R., Cavalcanti, A.P., Miranda, P., Mello, R.F., Kovanović, V., Gaševi´ c, D., 2020. Towards automatic´ cross-language classification of cognitive presence in online discussions, in: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pp. 605–614.

Bhatia, S., Biyani, P., Mitra, P., 2014. Summarizing online forum discussions-can dialog acts of individual messages help? in: *EMNLP*, pp. 2127–2131.

Cambria, E., 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems* 31, 102–107.

Castellanos-Reyes, D., 2020. 20 years of the community of inquiry framework. *TechTrends* , 1–4.

Cavalcanti, A.P., de Mello, R.F.L., Ferreira, M.A.D., Rolim, V.B., Tenório, J.V.S., 2017. Statistical and semantic features to measure sentence similarity in portuguese, in: *2017 Brazilian Conference on Intelligent Systems* (BRACIS), IEEE. pp. 342–347.

Coffrin, C., Corrin, L., de Barba, P., Kennedy, G., 2014. Visualizing patterns of student engagement and performance in moocs, in: *Proceedings of the fourth international conference on learning analytics and knowledge*, ACM. pp. 83–92.

Cohen, J., 1960. A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 37–46.

Dascalu, M., Trausan-Matu, S., Dessus, P., McNamara, D.S., 2015. Discourse cohesion: A signature of collaboration, in: *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, ACM. pp. 350–354.

De Wever, B., Schellens, T., Valcke, M., Van Keer, H., 2006. Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education* 46, 6–28.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., 2014. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research* 15, 3133–3181. URL: http://jmlr.org/papers/ v15/delgado14a.html.

Ferreira, M., Rolim, V., Mello, R.F., Lins, R.D., Chen, G., Gaševic, D., 2020. Towards automatic content analysis of´ social presence in transcripts of online discussions, in: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pp. 141–150.

Ferreira, R., Lins, R.D., Simske, S.J., Freitas, F., Riss, M., 2016. Assessing sentence similarity through lexical, syntactic and semantic analysis. *Computer Speech & Language* 39, 1–28.

Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., Romero, C., 2019. Text mining in education. Wiley Interdisciplinary Reviews: *Data Mining and Knowledge Discovery* 9, e1332.

Forman, G., 2003. An extensive empirical study of feature selection metrics for text classification. *Journal of machine learning research* 3, 1289–1305.

Garrison, D.R., Anderson, T., Archer, W., 1999. Critical Inquiry in a Text-Based Environment: *Computer Conferencing in Higher Education. The Internet and Higher Education* 2, 87–105.

Garrison, D.R., Anderson, T., Archer, W., 2010. The first decade of the community of inquiry framework: A retrospective. *The internet and higher education* 13, 5–9.

Gomes, G.A.F., 2012. Eu-tu: O emprego da classificação automática de mensagens em fóruns eletrônicos de discussões para análise do processo de ensino e aprendizagem centrado em interações. Rio de Janeiro, *PPGI/IM/iNCE/UFRJ*.

Graesser, A.C., McNamara, D.S., Louwerse, M.M., Cai, Z., 2004. Coh-metrix: Analysis of text on cohesion and language. *Behavior research methods, instruments, & computers* 36, 193–202. Doi: 10.3758/BF03195564.

Hartmann, N.S., 2016. Solo queue at assin: Combinando abordagens tradicionais e emergentes. *Linguamática* 8, 59–64.

Hew, K.F., Cheung, W.S., 2008. Attracting student participation in asynchronous online discussions: A case study of peer facilitation. *Computers & Education* 51, 1111–1124.

Hu, Q., Huang, Y., Deng, L., 2018. A method for analysis of online discussion forum in moodle, in: *2018 13th International Conference on Computer Science & Education (ICCSE), IEEE*. pp. 1–4.

Joachims, T., 2002. Learning to classify text using support vector machines: *Methods, theory and algorithms. Kluwer Academic Publishers.*

Joksimovic, S., Gasevic, D., Kovanović, V., Adesope, O., Hatala, M., 2014. Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The internet and higher education* 22, 1– 10.

Joshi, M., Rosé, C.P., 2007. Using transactivity in conversation for summarization of educational dialogue, in: *Workshop on Speech and Language Technology in Education*.

Kickmeier-Rust, M.D., Bedek, M., Albert, D., 2016. Theorybased learning analytics: Using formal concept analysis for intelligent student modelling, in: *Proceedings on the International Conference on Artificial Intelligence* (ICAI), The Steering Committee of the World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp). p. 97.

Kim, D., Park, Y., Yoon, M., Jo, I.H., 2016. Toward evidencebased learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education* 30, 30–43.

Kovanović, V., Gaševi´ c, D., Hatala, M., Siemens, G., 2017a.´ A novel model of cognitive presence assessment using automated learning analytics methods, in: *Measurement in Digital Environments White Paper Series*. SRI International.

Kovanović, V., Joksimovi´ c, S., Gaševi´ c, D., Hatala, M., ´ Siemens, G., 2015. Content analytics: the definition, scope, and an overview of published research. *Handbook of Learning Analyitcs* .

Kovanović, V., Joksimovi´ c, S., Katerinopoulos, P., Michail, C., ´ Siemens, G., Gaševic, D., 2017b. Developing a mooc ex-´ perimentation platform: insights from a user study, in: *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, ACM. pp. 1–5.

Kovanović, V., Joksimovi´ c, S., Waters, Z., Gaševi´ c, D., Kitto, ´ K., Hatala, M., Siemens, G., 2016. Towards automated content analysis of discussion transcripts: A cognitive presence case, in: *Proceedings of the sixth international conference on learning analytics & knowledge, ACM*. pp. 15–24.

Lang, C., Siemens, G., Wise, A., Gaševic, D., 2017. *Handbook´ of learning analytics*.

Lárusson, J.A., White, B., 2012. Monitoring student progress through their written point of originality, in: Proceedings of the *2nd International Conference on Learning Analytics and Knowledge*, ACM. pp. 212–221.

Lin, F.R., Hsieh, L.S., Chuang, F.T., 2009. Discovering genres of online discussion threads via text mining. *Computers & Education* 52, 481–495.

Lui, A.K.F., Li, S.C., Choy, S.O., 2007. An evaluation of automatic text categorization in online discussion analysis, in: Advanced Learning Technologies, 2007. ICALT 2007. *Seventh IEEE International Conference on, IEEE*. pp. 205–209.

Marbouti, F., Wise, A.F., 2016. Starburst: a new graphical interface to support purposeful attention to others' posts in online discussions. *Educational Technology Research and Development* 64, 87–113.

McLaren, B.M., Scheuer, O., De Laat, M., Hever, R., De Groot, R., Rosé, C.P., 2007. Using machine learning techniques to analyze and support mediation of student e-discussions.*Frontiers in Artificial Intelligence and Applications* 158, 331.

Melanie, M., 1999. An introduction to genetic algorithms. Cambridge, Massachusetts London, England, Fifth printing 3, 62–75.

de Melo Ferreira, F.J., Miranda, S.K.O., de Barros Costa, E., da Costa, F.P.D., Rocha, H.J.B., 2013. Um modelo de fórum de discussão com suporte às interações entre aprendizes utilizando mapas conceituais, in: *Brazilian Symposium on Computers in Education*, p. 416.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J., 2013. Distributed representations of words and phrases and their compositionality, in: *Advances in neural information processing systems*, pp. 3111–3119.

Mitchell, M., 1998. An introduction to genetic algorithms. *MIT press*.

Moreno-Marcos, P.M., Alario-Hoyos, C., Muñoz-Merino, P.J., Estévez-Ayres, I., Kloos, C.D., 2018. A learning analytics methodology for understanding social interactions in moocs. *IEEE Transactions on Learning Technologies* 12, 442–455.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C., Fischer, F., 2012. The acodea framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International journal of computer-supported collaborative learning* 7, 285–305.

Murphy, E., 2004. Recognising and promoting collaboration in an online asynchronous discussion. *British Journal of Educational Technology* 35, 421–431.

Neto, J.L., Santos, A.D., Kaestner, C.A., Freitas, A.A., 2000. Generating text summaries through the relative importance of topics. *Lecture Notes in Computer Science*, 300–309.

Neto, V., Rolim, V., Ferreira, R., Kovanović, V., Gaševi´ c, D., ´ Lins, R.D., Lins, R., 2018. Automated analysis of cognitive presence in online discussions written in portuguese, in: *European conference on technology enhanced learning,* Springer. pp. 245–261.

Pendry, L.F., Salvatore, J., 2015. Individual and social benefits of online discussion forums. *Computers in Human Behavior* 50, 211–220.

Ravi, S., Kim, J., 2007. Profiling student interactions in threaded discussions with speech act classifiers. *Frontiers in Artificial Intelligence and Applications* 158, 357.

Ricardo, B.Y., 1999. Modern information retrieval. *Pearson Education India*.

Rosé, C.P., Howley, I., Wen, M., Yang, D., Ferschke, O., 2017. Assessment of discussion in learning contexts, in: *Innovative Assessment of Collaboration. Springer*, pp. 81–94.

Rubio, D., Villalon, J., 2016. A latent semantic analysis method to measure participation quality online forums, in: Advanced Learning Technologies (ICALT), *2016 IEEE 16th International Conference on*, *IEEE*. pp. 18–19.

Sánchez-Martí, A., Puig, M.S., Ruiz-Bueno, A., Regós, R.A., 2018. Implementation and assessment of an experiment in reflective thinking to enrich higher education students' learning through mediated narratives. *Thinking Skills and Creativity* 29, 12–22.

Sarmento, L., 2006. Siemês–a named-entity recognizer for portuguese relying on similarity rules. *Computational Processing of the Portuguese Language*, 90–99.

Seber, G.A., Lee, A.J., 2012. Linear regression analysis. volume 936. John Wiley & Sons.

Silva, M., Carvalho, P., Sarmento, L., 2012. Building a sentiment lexicon for social judgement mining. *Computational Processing of the Portuguese Language*, 218–228.

Silva, W.D.C.M., Finger, M., 2013. Improving cogroo: the brazilian portuguese grammar checker, in: Proceedings of the *9th Brazilian Symposium in Information and Human Language Technology*.

Simsek, D., Sandor, A., Shum, S.B., Ferguson, R., De Liddo, A., Whitelock, D., 2015. Correlations between automated rhetorical analysis and tutors' grades on student essays, in: *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge*, ACM. pp. 355–359.

Tausczik, Y.R., Pennebaker, J.W., 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology* 29, 24– 54. Doi: 10.1177/0261927X09351676.

Wen, M., Yang, D., Rose, C., 2014. Sentiment analysis in mooc discussion forums: What does it tell us? in: *Educational data mining* 2014.

Wise, A., Zhao, Y., Hausknecht, S., 2014. Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics* 1, 48–71.

Wulf, J., Blohm, I., Leimeister, J.M., Brenner, W., 2014. Massive open online courses. *Business & Information Systems Engineering (BISE)* 6, 111–114.

Yen, C.H., 2013. A framework of e-learning analytics for asynchronous discussion forums, in: Advanced Learning Technologies (ICALT), *2013 IEEE 13th International Conference on, IEEE.* pp. 26–28.

Yoo, J., Kim, J., 2014. Can online discussion participation predict group project performance? investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education* 24, 8–32.

Zhao, J., Zhu, T., Lan, M., 2014. Ecnu: One stone two birds: Ensemble of heterogenous measures for semantic relatedness and textual entailment., in: *SemEval@ COLING*, pp. 271– 277.