

Received January 15, 2022, accepted January 29, 2022, date of publication February 7, 2022, date of current version February 11, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3149271

Improving MOOCs Using Information From Discussion Forums: An Opinion Summarization and Suggestion Mining Approach

OMAIMA ALMATAIFI¹ AND ADITYA JOHRI²

¹Department of Information Systems, King Abdulaziz University, Jeddah 21589, Saudi Arabia

²Department of Information Sciences and Technology, George Mason University, Fairfax, VA 22030, USA

Corresponding author: Omaima Almatrafi (oalmatrafi@kau.edu.sa)

This work was supported in part by the U.S. National Science Foundation under Award 1712129 and Award 1707837.

ABSTRACT Discussion forums are integral to MOOCs and a useful resource for collaborative learning. In this paper we examine if the posts made on forums by participants can also provide meaningful information to assess and improve the effectiveness of MOOCs. We present an empirical approach that uses posts in MOOC forums to summarize participants' opinion towards aspects of a course and identify suggestions for improving a course. Specifically, our approach: (1) detects participants' attitude towards aspects of a course (e.g. professor, lecture, assignment) at a context or sentence level, (2) extracts suggestions for course improvement, which is a novel space in MOOC learning analytics, and (3) aggregates and displays the results visually. The study used lexicon and rule-based approach and was able to identify aspect-based sentiments and suggestions related to course design elements with a good overall score (0.41 kappa score). By summarizing opinions from a vast amount of textual data on forums our approach allows instructors to improve their course and thereby student engagement and learning.

INDEX TERMS Discussion forums, MOOC, sentiment analysis, suggestion mining, visualization.

I. INTRODUCTION

Massive Open Online Courses (MOOCs) attract and enroll a high number of students. This ability to scale-up instruction is one of the core values provided by MOOCs. Yet, most MOOCs are still taught by an individual instructor or a relatively small team of instructors [1]. In a traditional educational setting, especially in small classes, it is easier for instructors to monitor student progress and to collect feedback from students while the course is underway. To improve future offerings, instructors typically distribute a course evaluation form (i.e. questionnaire) at the end of the course to gather students' feedback and this is often sufficient to understand how the course can be improved. Due to the high student-to-teacher ratio in MOOCs, traditional methods for feedback are not efficient [2], [3]. Consequently, an innovative approach is needed to manage the course, especially to monitor student progress and analyze student feedback. The ability to detect and understand student feedback is especially

The associate editor coordinating the review of this manuscript and approving it for publication was Camelia Delcea .

pertinent for MOOCs given the high rate of attrition that has been reported. Real-time feedback and the ability to make suitable changes accordingly can be a useful mechanism to curb disengagement. Given the low completion levels for MOOCs, participants who complete the final evaluation will most likely be those who completed the course, which means a) there will be fewer total participants and the survey will not be representative, and b) the voice of students who dropped out may not be heard. Also, phrasing of any questionnaire can be biased [4] therefore collecting student opinions through natural discourse or interaction, without any kind of prompts, is more effective.

One aspect of a MOOC that has shown promise as a way to understand course dynamics and monitor progress is the discussion forum [5]. In MOOCs, the discussion forum is the main channel for social learning and interaction among participants. Participants use it to seek or share knowledge, discuss ideas, express opinions, report technical issues, and build social connections, all of which make it a natural source for course feedback [6], [7]. Discussion forums though have a limitation. The high number of unstructured posts hampers

instructors' ability to keep up with the rapid flow of postings, and more importantly, it prevents them from using information shared on the forums to make informative decisions that could improve learners' retention and both current and future course offerings. An alternative and more efficient way to gather feedback would be to use computational models to analyze and summarize participants' opinions and suggestions regarding the course and its aspects from the content of their contributions in discussion forums and provide on-going evaluation of various course-related aspects. Researchers in [8] and [9] emphasize the importance of analyzing students' reviews using sentiment analysis to inform MOOC instructors and designers. This paper presents a study that had three goals: 1) to mine learners' opinions regarding the course at a finer granularity (aspect level) using sentiment analysis; 2) to mine learners' suggestions and feedback for the course; and 3) to provide a summary of learners' opinions on course elements. The contribution of this work will help instructors monitor learners' attitudes about the course from a natural source and highlight participants' needs and expectations for the course. The information can be used by instructors to make informed interventions based on learners' preferences, what they liked and/or disliked about the course, and their suggestions for improving the course. These actions can be taken at two levels: 1) individual-level intervention - respond to learner's post directly, or 2) course-level adjustment - adjust the course design or pedagogy to fit participants' needs and improve the current and future iterations of the course. As a result of a well-designed course, student engagement and learning can be increased [10]. Note: We use the terms *course-related aspects* or *course elements* interchangeably to refer to different attributes of a course including instructional staff, course content, and assessments.

II. RELATED WORK

In this section, we present related work on sentiment analysis in MOOC and distinguish between sentiment analysis and aspect-based sentiment analysis. Next, we introduce the suggestion mining task and end with a discussion of summarization for making sense of information. Given the significant volume of research on MOOCs and on forums in MOOCs within the past couple of years, we limit the review of prior work to empirical papers most relevant to our problem statement.

A. MOOC DISCUSSION FORUMS

Discussion forums within MOOCs have been studied extensively given the useful and interesting data they generate through user interaction and dialogue. Studies have found that within the large data corpus of discussion forums, both positive and negative comments are present. Although negativity may be an outcome of discussion or comments by a comparatively small number of participants, it can have a disproportional impact on learners and instructors [11]. On the other hand, positive interactions can promote student participation. Adamopoulos in [12] collected learners'

attitudes from course evaluation reviews towards MOOC aspects such as the course materials, assignments, professors, and discussion forums. The study found that learners' positive attitude towards professor(s), assignments and course material have a positive effect on learners' retention. Ramesh *et al.* in [13] found that learners who participated in the discussion forums and dropped out had expressed negative sentiment regarding course logistics. At the same time, the study found that when these posts were responded to and resolved, it changed the attitude of the learners toward the course and increased their likelihood of completing the course. This calls for more efficient ways to process the posts [14] and mine learners' sentiment, so that individual support intervention can be provided promptly.

B. MOOC SENTIMENT ANALYSIS

Sentiment analysis is a text analysis task that uses machine learning and natural language processing techniques to identify writers' attitude and opinion. An opinion is classified based on its sentiment polarity to positive, negative or neutral. Aspect-based sentiment analysis is more complex; it identifies both the sentiment and the aspect associated to it (opinion target) and thus it gives a detailed information.

The majority of MOOC literature focuses on sentiment analysis and has been applied for a diverse range of functions and at four different levels. At the course level, one research study explored the collective sentiment of all participants and explored the trending opinion toward the course [15]. The study reported a correlation between the ratio of sentiment expressed in the discussion forums and the dropout rates each day. In addition, when researchers also explored the evolution of learners' sentiment at the course level, they found that learners start enthusiastically, express peak negative sentiments before deadlines, and then positive sentiment increases significantly again at the end of the course [16], [17]. At a user-level, the goal is to assess the effect of participants' sentiment on their behavior. Studies showed that participants who demonstrate motivation and express enthusiastic posts are more likely to stay in the course [18]. Similarly, participants who express or are exposed to negative emotions (confusion) are more likely to drop out [19]. Ramesh *et al.* used sentiment analysis to predict participant engagement and performance [20]. However, negative sentiment was not a strong predictor of engagement. In fact, engaged and disengaged learners were equally likely to post negative comments in the forums. At a post level, the goal is to detect if the post holds an opinion and if so, determine the sentiment orientation of that post [21]. Studies have developed models using supervised and deep learning methods to classify messages into positive and negative posts across domains [22], [23]. Seven different approaches (supervised and unsupervised machine learning algorithms) have been examined to perform sentiment analysis on MOOC data on a post-level, and found that Random forest and Dictionaries approaches produced good results considering that sentiment analysis and opinion mining is not an easy task due to its subjective nature [16].

There is a scarcity of studies that applies sentiment analysis in MOOC at the aspect level. Recently, a study used weakly supervised technique for aspect-based sentiment analysis to evaluate students' reviews collected from Coursera and other traditional classroom settings [24]. This research is the closest to the first aim in our research, yet the means of data is different. In our case, the data were extracted from discussion forums in MOOC, which is not purposely designed for course evaluation.

C. SUGGESTION MINING

Studies on user generated content in MOOC discussion forums show that participants not only share their course experiences but also propose suggestions for course improvement. Therefore, to complement the sentiment analysis in our study we integrate a "suggestions for improvement" [25] detection model to extract participant suggestions regarding course-related aspects from user generated content in the discussion forums. Suggestion mining as a problem has been investigated mainly in reviews and Twitter data for commercial purposes [26]. The goal is to extract participants' wishes and suggestions from user generated content in order to help brand owners improve the next iteration of the product and to help fellow customers make better decisions when buying a product. This business analytics task can also be transferred to a learning analytics task to assist instructors and course designers in improving their course offerings, and assist learners and policy makers in making better decisions such as taking or promoting the course. Negi *et al.* [27] studied suggestion mining, in particular, explicit suggestions, and defined it as "the text which directly proposes, recommends, or advises an action or an entity" (p. 120). In that study, forum posts (product support forum) was first introduced for suggestion mining. The result of their analysis reported that deep neural network algorithms seem to provide the best results compared to SVM and rule association for in-domain and cross domains evaluation. To the best of the author's knowledge, this paper is the first paper to introduce suggestion detection from forum posts in MOOC which can be considered as a learning analytics task for course improvement.

D. OPINION SUMMARIZATION

The summarization and aggregation of the opinions is what is called opinion summarization, which is a subtask in sentiment analysis. It aims to postulate a concise summary from large number of unstructured data on a topic. There are two approaches for opinion summarization: a visual-based approach and a text-based approach [29]. A visual-based approach provides an attractive and easy to understand representation, while a text-based approach offers more details about the aspects summary (i.e. displays statistics and displays frequent or important phrases). Although it is important to provide summarization for the aspects opinion and provide the results in a manner that is easy to interpret, only a small number of studies in sentiment analysis and opinion mining literature included this task in their research [30].

In the educational context, it will enable both instructors and learners to easily understand the positive and negative aspects of a course. According to Dave and colleagues [28], the ideal opinion-mining tool is supposed to "process a set of search results for a given item, generating a list of product attributes (quality, features, etc.) and aggregating opinions about each of them (poor, mixed, good)" (pg. 519).

In this paper, opinion summarization and suggestion mining will be carried out at the sentence and aspect level. Every aspect will be classified into one of the following values: positive, negative, neutral, or suggestion. This work is distinct from the previous work in the following ways. First, it is mining learners' attitude towards course-related aspects at a finer grain level (sentence or aspect level), especially since sample data showed that posts include two course-related aspects on average. Second, it introduces the task of suggestion detection regarding course-related aspects from MOOC forums. Third, it provides an opinion summary by aggregating the sentiment polarity regarding course elements.

Table 1 presents examples of the various labels an aspect can receive given a post piece (context).

III. METHOD

A. DATA

The data in this study was sampled from MOOCPosts dataset [31]. MOOCPosts contains about 30,000 discussion forum posts that are hand labeled on six dimensions: a) binary dimensions: question, answer, and opinion; and b) ordinal dimensions (scored 1-7): confusion, urgency, and sentiment.

Although the data is labeled, the labels do not pertain to the purpose of this study for the following reasons. First, opinionated posts in many cases occur in a knowledge sharing context (e.g. "*It is sad to see girls in the video denied their rights for education.*") Second, the sentiment score is an overall score about the sentiment polarity of the post, while the goal of this study is to identify the aspect of the course and its opinion. Some posts may contain more than one aspect and the overall sentiment will not help (e.g. "*The book is difficult to read but the professor makes it easy to follow*"). Therefore, for the purposes of this study, a sample of the dataset was selected and hand-coded to identify sentiment polarity associated with the course element.

The sample was selected from the Humanities and Science domain since it has the largest number of courses; the courses are Economy, Statistics, Environmental Physiology, and Women's Health. Only posts that contain one of the keywords that explicitly refer to one of the course components were included. The method for selecting the keywords (aspects) will be described in the method section. A stratified sampling approach was carried out to select 1000 instances (Table 2) based on the overall sentiment category distribution: considering sentiment score in the range [0-3]: negative, [3-5]: neither, and [5-7]: positive. The sample showed that a post has two course-related aspects on average. The distribution of the sentiment categories in the

TABLE 1. Examples of context, aspects, and labels.

Post (context)	Aspect	Label
While I thought all the modules were interesting to an extent, I really loved the Stress module as there were so many applications to day-to-day life.	Module	Positive
I am not concerned with the grading system and have not questioned previous sloppy grading mistakes.	Grading	Negative
The answer to the quiz says that Cp is a valid metric?	Quiz	Neither
Having the CC at the bottom of the video would be easier i think	Video	Suggestion

TABLE 2. Data overview.

Data	Size(#posts)
Original (MOOCPosts dataset)	29,584
Humanities posts (H)	9,711
Humanities+Aspect (H+A)	3011
Sample of (H+A)	528 posts (1000 aspects)

sample resembles the original dataset with about (2.7% negative posts, 19% positive posts, and 78% neutral posts).

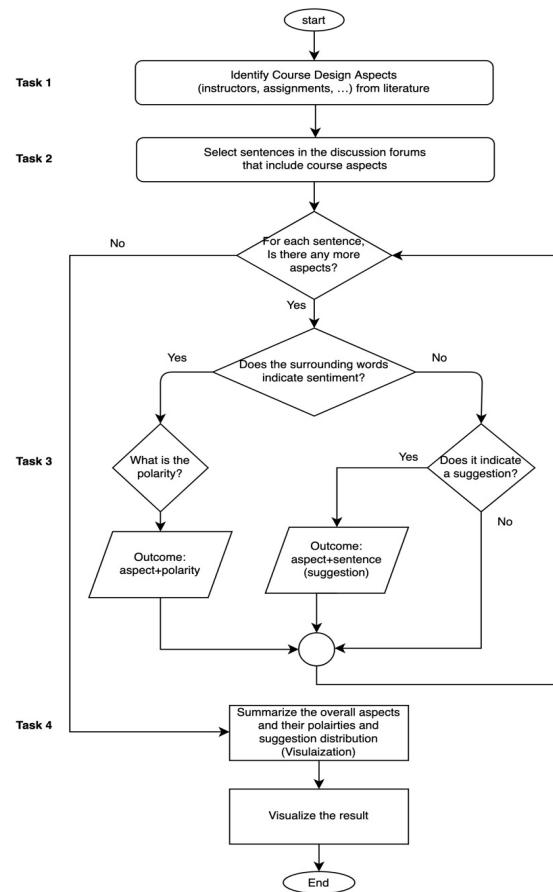
The sample was coded by at least three human coders via Figure Eight (Human-in-the-Loop) to be used for evaluation. To ensure annotation quality, coders needed to maintain an accuracy of 70% to have their labels included in the final result. In the evaluation step, only instances that had high confidence among the coders were used, which yielded up to 906 instances for evaluation.

B. IMPLEMENTATION

The process of course-related aspects opinion summarization and suggestion mining can be divided into four sub-tasks: aspect identification, post-preprocessing, suggestion detection or sentiment polarity detection, lastly, aspect sentiment summarization. Figure 1 summarizes the steps and tasks to accomplish the aim of this paper.

C. TASK 1: ASPECT IDENTIFICATION

The goal of this task is to identify course-related aspects that are of interest to understand participants' opinions related to them. An initial list of course-related aspects was gathered from the literature. Adamopoulos in [12], identified four aspects for course evaluation: assignments, professor, discussion forum, and course material. Ramesh *et al.* in [13] presented frequent words in logistical posts, and Comer *et al.* in [11] surveyed course-related aspects that received negativity, which include a) course/course content which include platform/instructional design, grading, readings, assignments, lectures; b) instructor; and c) peer interaction/feedback. Furthermore, Wen and colleagues studied the sentiment for the following course-related aspects: course, lecture, assignment, and peer-assessment [15]. This yielded a list of the following potential aspects: platform, course, class, material, reading, lecture, quiz, exam, assignment, homework (or HW), peer, grading, instructor, professor (or prof.), and teaching assistant (or TA). Furthermore, natural language processing was applied to extract the most frequent nouns that appear in the posts. From among the 100 most frequent nouns found, the following nouns - video, discussion, section, article, and book- were added to have a wider coverage of

**FIGURE 1.** Flowchart of the steps carried out to present course-aspects opinion and suggestions summarization.

course-related aspects. The final list of aspects is presented in Table 3. The aspects were grouped into categories.

A limitation of this technique is that it will constrain the sentiment analysis and suggestion extraction to posts that contain an aspect explicitly. On the other hand, when the aspect list is determined in advance, it can be used as keywords to search the dataset. This will allow instructors to monitor any new course design aspect by just adding the words into the list of the aspects they want to monitor.

D. TASK 2: PREPROCESSING

For preprocessing, Natural Language Toolkit (NLTK) was used which contains libraries for text processing such as

TABLE 3. List of aspects grouped into categories and the number of their occurrences in the sample.

Course Category	Aspects	Count (#occurrences)
Assessments	Homework, HW, Assignment, Quiz, Exam, Grading	159
General(Course)	Course, class	649
Instructional staff	Professor, prof., TA, Teaching Assistant, Instructor	74
Material	Lecture, Video, Module, Material, Reading, Discussion, Section, Article, Book, Chapter, Topic	1267
Others	Platform, Peer	27

tokenization, parsing, stemming, and tagging. In this study it was used to break down the post into sentences (sentence tokenizer). Then if the sentence has one of the words in the aspect list, it will be qualified for the analysis. Otherwise, the sentence will be skipped. In addition, WordNet lemmatizer and part-of-speech tagger were used for aspect identification (find nouns), and suggestion mining (linguistic rules).

E. TASK 3

1) SENTIMENT POLARITY DETECTION

The method used to find the sentiment polarity of a sentence is Valence Aware Dictionary and sEntiment Reasoner (VADER). According to sentiBench, VADER has the best performance in 3-class experiments compared with other algorithms, and it has been consistently among the top performing methods on various datasets [32]. VADER uses syntactical rules and lexicons to compute the sentiment polarity of a text [33]. It considers negation and contraction (e.g., “not good”, “wasn’t very good”), intensity from punctuation (e.g., “Good!!!”), word capitalization (“GOOD!!!”), degree modifiers (e.g., “very, kind of”), acronym and slang (e.g., “lol”), and emoticons and utf-8 encoded emojis. The VADER lexicon list contains 7500 words that were based on existing sentiment lexicon. VADER lexicon is unique in the sense that all lexicons have been hand scored by 10 individuals (-4 to 4) to convey the intensity score of the word in addition to its polarity (valence-based lexicon list). Although the lexicons were mostly domain independent, they were attuned to social media sentiment. Thus, some common words in the MOOC domain was added, which include: accessible, available, learning, working, understand, incorrect, irrelevant, informative, looking forward. In addition, words that were content specific were removed (e.g. violence, poverty, slavery, stress). Furthermore, the contrastive conjunction “however” was added to VADER implementation to adjust the computation for sentences that show two different ideas similar to the case of “but”.

VADER will return a set of scores: positive, negative, neutral, and compound. The compound score was used to determine the sentiment polarity of the text. The labels in this study followed the below criteria:

- Positive sentiment: compound score ≥ 0.45
- Negative sentiment: compound score ≤ -0.20
- Neutral sentiment: >-0.20 and <0.45

2) SUGGESTION MINING

Suggestion sentences in many cases exhibit certain linguistic patterns regardless of the suggestion domain [27]. In particular, most of the wishes and suggestions contain phrases that involve modal verbs such as ‘could’, ‘should’, and ‘would’ [26]. In this paper, the rule-based classification approach proposed in [26] was utilized with some modifications (highlighted in bold) to fit the characteristics of MOOC forums and the purpose of identifying specific type of suggestion which is related to the course improvement. At the beginning, during the process of analyzing the words, the contraction (‘d) was replaced by would, although linguistically it could mean had or would. This would allow the rules not to miss a possible suggestion.

The rules (Figure 2):

1) Rule#1: Modal verb (**past tense**) + Auxiliary verb + <window of size 3> Adjective – **Subject (I or you)**

This rule will detect all statements and questions that exhibit the following pattern: a modal verb such as might, could, would and should, followed by an auxiliary verb (be, have been), followed by an adjective within three words window after the auxiliary verb. e.g., “*It would have been better to add titles to the videos*”.

Unlike suggestions in review forums, the text in educational forums may contain modal verbs to explain point of views and discuss subject matters. Thus, the past tense constraint is added to the rule. In addition, the constraint, subject is not I or you, is added to dismiss suggestion/wish to the self or fellow participant. (e.g., “*I would like to upgrade my knowledge and I am pleased to take this course.*”). These two constraints will reduce the number of false positives detected (i.e. statement identified as suggestion while it is not).

2) Rule#2: Modal verb (**past tense**) + <window of size 3> Preference verb **or ‘request’**

This rule will detect all statements that exhibit the following pattern model verb (past tense), followed by a preference verb such as ‘love’, ‘like’, ‘prefer’ within three words window. ‘Request’ is added since the context in MOOC is more interactive, and participants can directly post suggestions to the course staff (e.g., “*I would request course staff to post details regarding deadlines for quizzes*”). It indicates a suggestion that can improve the course.

- e.g., “*I’d love to have more examples about this topic*”.
- 3) Rule#3: Has one of the wishful words + **Not followed by ‘T’**
- This rule will classify all statements that include one of these wishful words as a suggestion [‘wish’, ‘hope’, ‘recommend’, ‘suggest’] e.g., “*I hope this course will remain open*”. The rule has been adjusted to consider only suggestion related to course improvement by dismissing statements that has wishful word when the wishful word is followed by I because then it indicates a personal wish (e.g., “*I hope I can catch up and keep up with the course*”). In MOOC forums and specifically at the beginning of the course many participants will post and share their personal wishes out of the course.
- 4) Rule#4: Rule of “needs to” or “urge”
- Statements that directly use the word needs to. The authors also find that participants in MOOC forums use the word ‘urge’. Thus, it has been added to the rule taken from [26].
- e.g., “*I urge you to include a discussion of these concerns in future*”.

```
# Check if the sentence satisfies one of the suggestion for improvement rules
def isSuggestion(sentence):
    tagged_sentence = pos_tag(sentence.split())
    if (has_model_be_adj(tagged_sentence) |
        has_model_preference(tagged_sentence) |
        has_suggestion_word(tagged_sentence)):
        return TRUE
    return FALSE
```

FIGURE 2. High level suggestion mining function.

F. TASK 4: OPINION SUMMARIZATION

For each aspect, aggregate the number of positive and negative occurrences and the number of suggestion phrases. Neutral posts are not of interest in the summarization process, so they are excluded in the summary generation. At the end, each course design component and its sub-categories (aspects) will be presented in simple charts to communicate the statistics regarding the number and percentage of occurrences on the three dimensions (positive, negative, and suggestions). Visualization type was chosen carefully to allow experts and non-experts to get sense of the course in terms of its aspects at glance. In particular, pie and bar charts were used with attributes such as color and tooltip to add more information. Tableau software was used for this task.

IV. RESULT AND DISCUSSION

Human labels showed that aspect identified in post pieces referred to an aspect of the current course 95% of the time. This means that the predefined list of aspects was useful in targeting pieces of interest. In some cases, the retrieved posts may enclose these words but refer to another meaning (e.g. classification, the word class referring to social class, or the term “of course”), which do not refer to course attributes. This section will present the evaluation

of the sentiment and suggestion classification (task 3) and present the sentiment summarization for course-related aspects (task 4).

Since the data is qualitative in nature and the classes are categorical and imbalanced, the appropriate metric to use is Cohen Kappa. This metric compares the expected accuracy and the observed accuracy by measuring inter-coder agreement by taking agreement by chance into account. The expected values are the human labeled annotations. Only instances attained more than 0.5 confidence were employed in the evaluation process, which resulted in 906 instances.

The overall model achieved Cohen Kappa score of 0.41 in identifying both the sentiment and the suggestion for course-related aspects at a context level. This kappa score exemplified fair to good reliability score. In other words, there is a moderate strength of agreement between the model results and the human annotations according to kappa’s guide interpretation [34].

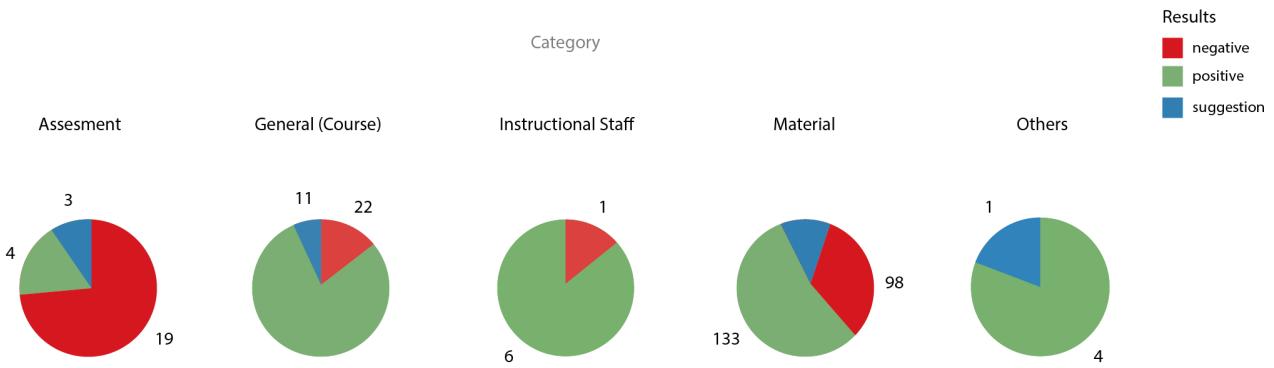
Furthermore, precision, recall, and F1 score were used to report the model performance since it is an imbalanced classification problem (Table 4). The number of instances in the table reflects the original distribution of MOOC posts; most of the posts are neutral, does not hold a sentiment or a suggestion. VADER (with some modification) was able to recall about 65% of the sentiment classes with the highest precision for the neutral class and lowest precision for the negative class. Some words in the negative class were hard to distinguish such as ‘problem’ and ‘error’. For example, the word ‘error’ can refer to an error in a quiz question or it can refer to a discourse about statistical error (e.g., how to compute error rate?) in statistics courses which in this case is not indicative of a negative sentiment. It occurs because the approach relies heavily on the lexicons.

TABLE 4. The evaluation of the classification (precision, recall, f1 score).

	P	R	F1	Support (# instances)
Positive	0.59	0.68	0.64	231
Neutral	0.81	0.66	0.73	577
Negative	0.37	0.63	0.47	82
Suggestion	0.22	0.50	0.31	16

A. SENTIMENT POLARITY DETECTION DISCUSSION

There are few studies in the surveyed MOOC literature that reported the evaluation of opinion or sentiment analysis. The sentiment polarity classification achieved an overall F1 score of 93.3% using LSTM with GloVe word embeddings when applied on course evaluation data [24]. Another study that utilized data from similar medium, MOOC discussion forums, reported F1 score in the sentiment classification as the following (positive = 0.19, negative = 0.62, and neutral = 0.43) [35]. In absolute numbers, the results as shown in table 4 is higher than in [35] when classifying

**FIGURE 3.** Course component opinion and suggestion summary.**TABLE 5.** Suggestion results compared to related work.

	Using General Rules [26]	Suggestion for Improvement Rules
Precision	0.08	0.22
Recall	0.44	0.50
F1 score	0.14	0.31

positive and neutral sentiments. However, since the dataset and the assumptions made in the current study are different, making a direct comparison of performance is difficult.

B. SUGGESTION MINING DISCUSSION

In general, opinionated text holds not only sentiment but also suggestions. Although suggestion is rare in reviews and comments [26](about 1.6% of our sample data conveyed suggestion for improvement), they are very helpful for course improvement because they communicate direct wishes to the course designers. Table 5 shows the result of identifying the suggestion using the general rule-based mining for suggestions (first column), and the result after adding some rules and constraints for the narrowed purpose of identifying suggestions for MOOC improvement (second column). The modified version showed an improvement in the ability to detect suggestion for improvement from post pieces. The F1 score has doubled for the positive class. In addition, the modified rules were able to recall 50% of the posts that have suggestions for improvement.

A qualitative look into the misclassified suggestions indicate that the extracted context were indeed a) a suggestion but are not targeted to the course improvement (e.g. *I took course x and I recommend it, where x course is not the current course*); b) in the form of suggestion but related to the course content (e.g. *The three non-medical interventions that I think might be effective in treating HIV/AIDS are: ...*); c) suggestions for fellow participants (e.g. *I highly recommend week 1 of his lectures*). On the other hand, the missed suggestion (i.e. the ones that fail to match the rules) were

a) mainly, suggestions not related to course design structured as a listing of suggestions (e.g. *incorporating into the curriculum a sex education class*); b) in the negation form (e.g. *not that I didn't like the other modules, but a change every once in a while is always positive*). c) in rare cases, the constraints affected the extraction of the suggestion (e.g. *if I could express a wish, I would like a follow up book*)

C. OPINION SUMMARIZATION

Communicating the information about learners' attitudes and suggestions towards the course to the instructors and course designers has to be done in an easily interpretable manner. Figure 3 illustrates the results of the classification on the sample data, which means it doesn't represent any specific course. The headings of the figure show the course design categories and the chart shows the percentages of the different classifications except for neutral class (omitted, not important). The numbers represent the count of class (positive, negative, suggestion) occurrences.

For example, the instructor can infer from Figure 3 that participants have a positive opinion regarding the instructional staff (mainly green- positive) while assessments received many negative comments (represented by red color). This kind of visualization helps course designers identify areas of course design improvements.

Figure 4 displays the breakdown of the aspects of one of the categories. Content/Course material is picked for illustration since it has the largest number of posts. A tooltip mark is used to show the context in which the post mentioned the aspect. For example, there are 17 posts that showed a negative attitude towards video. For example, the instructor can also see the suggestions by clicking on the squares in the suggestion column, e.g., “It would be great if the file naming convention you use fit with the naming convention in the Courseware section.” This suggestion has been identified in the Course/content category, particularly identified as related to ‘section’ aspect. This type of representation is simple and easy to use.

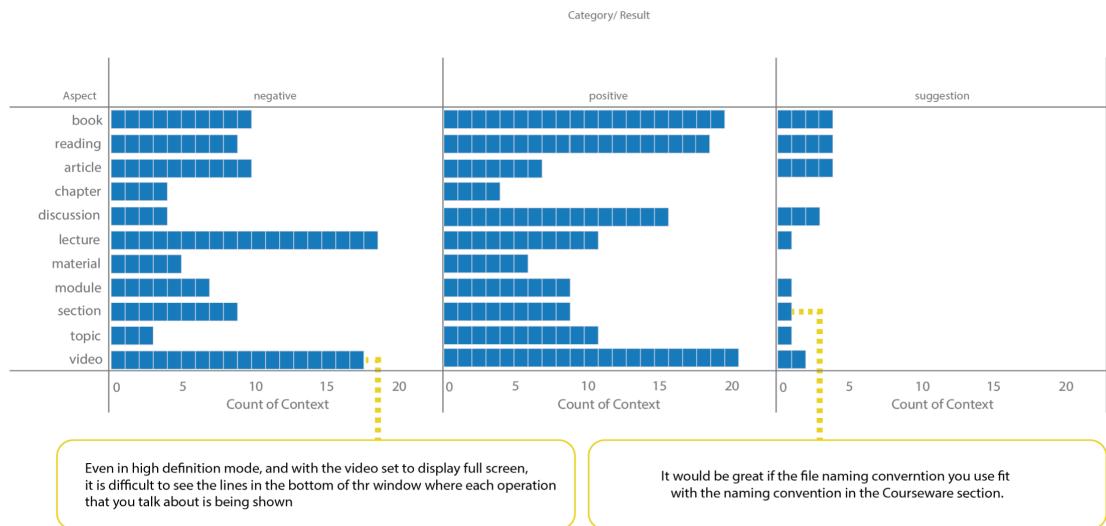


FIGURE 4. Course content/material aspects summary (positive, negative and suggestion).

V. LIMITATIONS AND FUTURE WORK

In this paper, the data was collected from different courses in the humanities and sciences domain. Future work may include a variety of domains to extend the model across domains. The classification task succeeded in achieving a good reliability in this study nonetheless there is a room for improvement. First, we plan to investigate supervised and deep learning approaches in a comparative study to identify the best performing one for aspect-based sentiment analysis in MOOC. Second, in the suggestion detection, future work may investigate ways to distinguish course related suggestion and suggestion specific for course improvement. Another promising direction for the suggestion mining would be to apply the automatic suggestion detection in forums that encourage participants to offer suggestions. Moreover, a potential improvement for the visualization is the integration of the time dimension so instructors can filter and monitor the course-related aspects as the course proceeds. Finally, courses in the future can have a discussion forum dedicated to providing suggestions by participants and this analysis technique can be used to quickly classify them into their sentiments and take action.

VI. DISCUSSION AND CONCLUSION

In this paper we present a study of using feedback from discussion forums in MOOCs to improve a course. Our approach summarizes MOOC participants' attitudes and suggestions in regard to course-related aspects from posts in the forums. Specifically, the approach entails detecting participants attitude towards course-related aspects, extracting suggestion for course improvement, and then aggregating the result and displaying the findings in a visual representation. The aspect sentiment analysis at the sentence/context level has achieved a good performance using VADER with few modifications.

Regarding suggestion for improvement mining, the problem has not been studied in MOOC literature. The extracted suggestions are useful for the instructors if they are tuned to it and take necessary actions, which can improve participants' learning experience. The method used (rule based approach) was able to recall half of the suggestions for course improvement yet there is a room for improvement. Lastly, the summarization using visual aid will help instructors intuitively interpret the effectiveness of course elements from learner generated text in the discussion forums (i.e. what worked well and what needs improvement). This work if applied in real-time can assist instructors better manage the course through the detection of learners' attitudes and suggestions. The immediate feedback about the course can be used to help instructors assess, adjust and improve the current and future iterations of the course.

ACKNOWLEDGMENT

This work is based on the first author's dissertation completed at George Mason University. The authors would like to acknowledge the authors of Stanford MOOC posts dataset. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

REFERENCES

- [1] S. Chaturvedi, D. Goldwasser, and H. Daumé, III, "Predicting instructor's intervention in MOOC forums," in *Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics*, 2014, pp. 1501–1511.
- [2] P. Topali, A. Ortega-Arranz, A. Martínez-Monés, and S. L. Villagrá-Sobrino, "'Houston, we have a problem': Revealing MOOC practitioners' experiences regarding feedback provision to learners facing difficulties," *Comput. Appl. Eng. Educ.*, vol. 29, no. 4, pp. 769–785, Nov. 2020.
- [3] R. F. Kizilcec, C. Piech, and E. Schneider, "Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses," in *Proc. 3rd Int. Conf. Learn. Anal. Knowl.*, 2013, pp. 170–179.

- [4] W. Pena and A. Melgar, "Ontology-based information extraction from Spanish Forum," in *Computational Collective Intelligence*. Cham, Switzerland: Springer, 2015, pp. 351–360.
- [5] K. Stephens-Martinez, M. A. Hearst, and A. Fox, "Monitoring MOOCs: Which information sources do instructors value?" in *Proc. 1st ACM Conf. Learn. Scale Conf. (L@S)*, Mar. 2014, pp. 79–88.
- [6] S. Mak, R. Williams, and J. Mackness, "Blogs and forums as communication and learning tools in a MOOC," in *Proc. 7th Int. Conf. Netw. Learn.*, 2010, pp. 275–285.
- [7] W. Zou, X. Hu, Z. Pan, C. Li, Y. Cai, and M. Liu, "Exploring the relationship between social presence and learners' prestige in MOOC discussion forums using automated content analysis and social network analysis," *Comput. Hum. Behav.*, vol. 115, Feb. 2021, Art. no. 106582.
- [8] M. Misuraca, G. Scipi, and M. Spano, "Using opinion mining as an educational analytic: An integrated strategy for the analysis of students' feedback," *Stud. Educ. Eval.*, vol. 68, Mar. 2021, Art. no. 100979.
- [9] K. Lundqvist, T. Liyanagunawardena, and L. Starkey, "Evaluation of student feedback within a MOOC using sentiment analysis and target groups," *Int. Rev. Res. Open Distrib. Learn.*, vol. 21, no. 3, pp. 140–156, May 2020.
- [10] L. D. Fink, "The power of course design to increase student engagement and learning," *Peer Rev.*, vol. 9, no. 1, pp. 13–17, 2007.
- [11] D. Comer, R. Baker, and Y. Wang, "Negativity in massive online open courses: Impacts on learning and teaching and how instructional teams may be able to address it," *InSight, J. Scholarly Teach.*, vol. 10, no. 1, pp. 92–113, Aug. 2015.
- [12] P. Adamopoulos, "What makes a great MOOC? An interdisciplinary analysis of student retention in online courses," in *Proc. ICIS*, 2013.
- [13] A. Ramesh, D. Goldwasser, B. Huang, H. Daume, and L. Getoor, "Understanding MOOC discussion forums using seeded LDA," in *Proc. 9th Workshop Innov. Use NLP Building Educ. Appl.*, 2014, pp. 28–33.
- [14] M. S. Boroujeni, T. Hecking, H. U. Hoppe, and P. Dillenbourg, "Dynamics of MOOC discussion forums," in *Proc. 7th Int. Learn. Anal. Knowl. Conf. (LAK)*, Mar. 2017, pp. 128–137.
- [15] M. Wen, D. Yang, and C. Rosé, "Sentiment analysis in MOOC discussion forums: What does it tell us?" *Proc. 7th Int. Conf. Educ. Data Mining*, 2014, pp. 130–137.
- [16] P. M. Moreno-Marcos, C. Alario-Hoyos, P. J. Muñoz-Merino, I. Estévez-Ayres, and C. D. Kloos, "Sentiment analysis in MOOCs: A case study," in *Proc. IEEE Global Eng. Educ. Conf. (EDUCON)*, Apr. 2018, pp. 1489–1496.
- [17] J.-S. Wong, B. Purse, A. Divinsky, and B. J. Jansen, "Analyzing MOOC discussion forum messages to identify cognitive learning information exchanges," *Proc. Assoc. Inf. Sci. Technol.*, vol. 52, no. 1, pp. 1–10, 2015.
- [18] M. Wen, D. Yang, and C. Rosé, "Linguistic reflections of student engagement in massive open online courses," in *Proc. 8th Int. Conf. Weblogs Social Media (ICWSM)*, 2014, pp. 525–534.
- [19] D. Yang, M. Wen, I. Howley, R. Kraut, and C. Rose, "Exploring the effect of confusion in discussion forums of massive open online courses," in *Proc. 2nd ACM Conf. Learn. Scale (L@S)*, Mar. 2015, pp. 121–130.
- [20] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé, III, and L. Getoor, "Modeling learner engagement in MOOCs using probabilistic soft logic," in *Proc. NIPS Workshop Data Driven Educ.*, 2013, pp. 1–7.
- [21] A. Agrawal, J. Venkatraman, S. Leonard, and A. Paepcke, "YouEDU: Addressing confusion in MOOC discussion forums by recommending instructional video clips," in *Proc. 8th Int. Conf. Educ. Data Mining*, 2015, pp. 297–304.
- [22] A. Bakhtaria, "Towards cross-domain MOOC forum post classification," in *Proc. 3rd ACM Conf. Learn. Scale (L@S)*, Apr. 2016, pp. 253–256.
- [23] X. Wei, H. Lin, L. Yang, and Y. Yu, "A convolution-LSTM-based deep neural network for cross-domain MOOC forum post classification," *Information*, vol. 8, no. 3, p. 92, Jul. 2017.
- [24] Z. Kastrati, A. S. Imran, and A. Kurti, "Weakly supervised framework for aspect-based sentiment analysis on students' reviews of MOOCs," *IEEE Access*, vol. 8, pp. 106799–106810, 2020.
- [25] C. Brun and C. Hagege, "Suggestion mining: Detecting suggestions for improvement in users' comments," *Res. Comput. Sci.*, vol. 70, no. 79, pp. 171–181, 2013.
- [26] J. Ramanand, K. Bhavsar, and N. Pedanekar, "Wishful thinking: Finding suggestions and 'buy' wishes from product reviews," in *Proc. NAACL HLT Workshop Comput. Approaches Anal. Gener. Emotion Text (CAAGET)*, Jun. 2010, pp. 54–61.
- [27] S. Negi, K. Asooja, S. Mehrotra, and P. Buitelaar, "A study of suggestions in opinionated texts and their automatic detection," in *Proc. 5th Joint Conf. Lexical Comput. Semantics*, 2016, pp. 170–178.
- [28] K. Dave, S. Lawrence, and D. M. Pennock, "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews," in *Proc. 12th Int. Conf. World Wide Web*, 2003, pp. 519–528.
- [29] W. Maharani, D. Widayantoro, and M. Khodra, "Aspect-based opinion summarization: A survey," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 2, pp. 448–456, 2018.
- [30] R. Piriyani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000–2015," *Inf. Process. Manage.*, vol. 53, no. 1, pp. 122–150, Jan. 2017.
- [31] A. Agrawal and A. Paepcke. (2014). *The Stanford MOOCPosts Data Set*. [Online]. Available: <http://datastage.stanford.edu/StanfordMoocPosts/>
- [32] F. N. Ribeiro, M. Araújo, P. Gonçalves, M. A. Gonçalves, and F. Benevenuto, "SentiBench—A benchmark comparison of state-of-the-practice sentiment analysis methods," *EPJ Data Sci.*, vol. 5, no. 1, pp. 1–29, Dec. 2016.
- [33] C. Hutto and E. Gilbert, "Vader: A parsimonious rule-based model for sentiment analysis of social media text," in *Proc. 8th Int. AAAI Conf. Web Social Media*, 2014, pp. 216–225.
- [34] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, p. 159, Mar. 1977.
- [35] A. Ramesh, S. H. Kumar, J. Foulds, and L. Getoor, "Weakly supervised models of aspect-sentiment for online course discussion forums," in *Proc. 53rd Annu. Meeting Assoc. Comput. Linguistics 7th Int. Joint Conf. Natural Lang. Process.*, vol. 1, 2015, pp. 74–83.

OMAIMA ALMATRAFI received the B.S. degree in computer science from King Abdulaziz University (KAU), Jeddah, Saudi Arabia, in 2008, and the M.S. degree in information systems and the Ph.D. degree in information technology from George Mason University, USA, in 2013 and 2018, respectively. She is currently an Assistant Professor with the Department of Information Systems, KAU. Her research has been published in several conferences and academic journals. Her research interests include computer-supported collaborative learning, learning analytics, and educational data mining.



ADITYA JOHRI is a Professor with the Department of Information Sciences and Technology, Volgenau School of Engineering, George Mason University, Fairfax, VA, USA. He studies the use of information and communication technologies (ICT) for learning and knowledge sharing, with a focus on cognition in informal environments. He is the Co-Editor of the *Cambridge Handbook of Engineering Education Research* (CHEER; Cambridge University Press, 2014).