Sentiment analysis in online education: An analytical approach and application

Fangyuan Lin

International School, Beijing University of Posts and Telecommunication, Beijing, 100080, China

2528920862@bupt.edu.cn

Abstract. This paper presents a groundbreaking approach to the application of sentiment analysis within the domain of online education. By introducing an innovative methodology, the aim is to streamline the process of automatically evaluating sentiments and extracting opinions from the vast sea of content produced by learners during their online interactions. This not only aids educators in swiftly gauging the general mood and perspective of their student body, but also allows them to delve deeper into the nuanced feedback provided, thus ensuring the continual improvement of course quality. In an era where digital learning platforms are growing exponentially, understanding students' attitudes, concerns, and overall satisfaction is paramount. Our methodology, therefore, is not just a technical advancement, but also a strategic tool for educational institutions aiming to thrive in the digital age. The current research landscape, while expansive, has often overlooked the significance of real-time sentiment analysis in e-learning environments. This study, therefore, bridges an important gap, bringing to the forefront the importance of harnessing student feedback in a digital format, allowing educators to tailor their approach for optimal student engagement and success.

Keywords: sentiment analysis, online education, opinion extraction, educational psychology.

1. Introduction

The sweeping transition to digital education, expedited by the recent global events, has firmly established online learning as the contemporary standard. This monumental shift has generated an enormous trove of data, encompassing student-teacher dialogues, peer-to-peer discussions, and evaluative feedback. Despite its potential, much of this data remains unexplored. Delving into this vast pool presents a golden opportunity to gauge student sentiments, shedding light on their perceptions of and attitudes towards their learning journey. This, in turn, can play a pivotal role in refining course quality and optimizing learning outcomes.

Sentiment analysis, a specialized branch of natural language processing, has already demonstrated its prowess in diverse arenas, ranging from the world of commerce to the corridors of politics, effectively capturing the pulse of public sentiment. Yet, surprisingly, its integration into the realm of online education is not extensively charted. This paper seeks to address this oversight, presenting a robust methodology tailored for the sentiment analysis of digital educational content.

2. Relevant theories

2.1. Development status of sentiment analysis technology

Sentiment Analysis, a branch of NLP, involves the use of algorithms to identify and extract subjective information from source materials. It is largely used to determine the attitudes, opinions, and emotions of a speaker or writer with respect to some topic. Recent advancements in machine learning and artificial intelligence have significantly improved the accuracy and efficiency of sentiment analysis, enabling real-time evaluation of large-scale datasets.

2.2. Current research status

The field of online education analysis is evolving rapidly, with a focus on data interpretation from online platforms to enhance learning experiences. Current research primarily relies on quantitative metrics such as completion rates and quiz scores [1]. However, these metrics often overlook learners' subjective experiences. Sentiment analysis can provide a nuanced understanding of learners' experiences and perceptions [2]. Further, advancements in AI and machine learning, as demonstrated by Kusal et al. [3] and Cobos et al. [4], can lead to more sophisticated analyses, including detecting emotional responses and subjectivity in online courses. The integration of technologies such as mobile dev [5] and social media platforms [6] can also generate valuable data for analysis. By blending sentiment analysis, AI, and machine learning with traditional metrics, a more holistic view of online learning environments can be achieved, leading to improved learning experiences [7].

3. Methodology and application research

The application of sentiment analysis in online learning environments necessitates a carefully structured and methodologically sound approach [8]. In this section, the design of the proposed framework is discussed, followed by the preprocessing steps for learner texts. The sentiment computation, grounded in principles of educational psychology, is detailed, as well as the extraction of learner viewpoints.

3.1. Framework design

In the bid to integrate sentiment analysis into online learning platforms, our proposed framework is designed to be adaptable and robust, capable of handling a vast range of data types and volumes. The framework is structured around four primary components: data collection, pre-processing, sentiment computation, and opinion extraction.

Data collection is the first critical step, involving the gathering of relevant text data from various sources within the online learning environment [9]. This could include discussion forums, assignment feedback, emails, chats, and other forms of learner communication. The collected data serves as the raw material for the subsequent sentiment analysis.

Pre-processing is the second component, crucial for refining the raw data. This step involves standardizing and cleaning the collected data to improve the efficiency and accuracy of the sentiment analysis.

The third component, sentiment computation, is the core of the sentiment analysis process. Here, the preprocessed data is analyzed, and sentiments are computed using machine learning algorithms.

The final component is opinion extraction, which focuses on identifying the specific issues or topics that learners feel strongly about. This step provides a deeper understanding of the sentiments and enriches the sentiment analysis results with contextual information.

3.2. Preprocessing of learners' text

The preprocessing stage is a critical phase in our framework. It aims to transform the raw, unstructured data into a cleaner, more standardized format for accurate sentiment analysis. The preprocessing of the learners' text involves several steps.

The first step is the removal of irrelevant information and noise from the data. This includes HTML tags, URLs, emojis, and other non-textual elements that do not contribute to sentiment analysis.

Next, the data undergoes text normalization processes such as tokenization, stemming, and lemmatization. Tokenization breaks down the text into individual words or tokens, which form the base unit for sentiment analysis. Stemming and lemmatization, on the other hand, reduce words to their base or root form. For instance, words like 'running', 'runs', 'ran' are stemmed to their root form 'run', helping to minimize the dataset's complexity.

Stop words, which are commonly used words like 'is', 'the', 'and', that do not add much semantic value, are removed. Lastly, the preprocessed text is converted into numerical representations, such as Bag of Words or Term Frequency-Inverse Document Frequency (TF-IDF), that can be processed by the sentiment analysis algorithms.

3.3. Sentiment computation based on educational psychology

Once the data is preprocessed, it is subjected to sentiment computation. This component involves using machine learning algorithms to classify the text into different sentiment categories: positive, neutral, or negative [10].

The choice of the machine learning algorithm depends on the nature of the data and the specific requirements of the analysis. Popular options include Naive Bayes, Support Vector Machines, and deep learning models like Recurrent Neural Networks.

The sentiment computation process is unique, being guided by principles of educational psychology. This means that sentiment categories derive not just from the polarity of words but also from the cognitive and emotional states they represent in an educational context. For example, words expressing confusion or curiosity, which might be viewed as neutral or even negative in different contexts, can suggest a positive sentiment in learning environments, signifying engagement and active learning.

3.4. Point of view extraction suitable for learners

Following sentiment computation, the focus shifts to opinion extraction. This step involves identifying the specific issues or topics that learners feel strongly about. Achieving this is done through topic modeling techniques, such as Latent Dirichlet Allocation (LDA), which can identify common themes or topics in the learners' responses.

The opinion extraction process is designed to be adaptable to the learners' perspective. It considers the unique language use, discourse patterns, and communication styles in online learning environments. It also takes into account the context of the discussion, the course content, and the learners' background.

In summary, the proposed methodology offers a comprehensive and nuanced approach to sentiment analysis in online education. By incorporating principles of educational psychology and considering the learners' perspective, it provides a more accurate and context-aware analysis of learners' sentiments. Moreover, by identifying specific topics of concern, it offers valuable insights that can be used to enhance course design and improve the overall learning experience.

4. Experiments and results

To validate the effectiveness of the proposed methodology, a series of experiments were performed using an extensive dataset. A set of standard evaluation metrics was also employed to assess the performance of the sentiment analysis model. This section presents a detailed account of the experimental setup, the dataset used, evaluation metrics, and the resulting analysis.

4.1. Experimental dataset

The experimental dataset comprises online communication content from learners across multiple courses on platforms including Moodle, MOOC, Blackboard, Coursera, and EdX. This dataset encompasses various forms of communication, including discussion forum posts, assignment feedback, emails, and chat transcripts.

The dataset is structured as a CSV file with three primary columns: Course_ ID: An identifier for the course from which the communication content was extracted. Content: The actual text content generated by the learner. Communication_ Type: The type of communication (e.g., forum post, email, chat

transcript, assignment feedback). Here are some representative examples from the dataset, As shown in Table 1.

Table 1. Anonymized learning communication dataset: analysis of course content interactions.

Course_ ID	Content	Communication_ Type
001	"I enjoyed this week's content, especially the group project."	Forum Post
002	"Can someone explain how to solve question 3 on the assignment?"	Chat Transcript
003	"The workload for this course is overwhelming. I'm struggling to keep up."	Email
004	"Great lecture today! The concept was explained clearly."	Forum Post

All data was anonymized, and no personal identifiers were included in the dataset. Any information that could potentially identify a student or an institution was removed during the data cleaning process.

4.1.1. Data collection. Our research initiated with data collection from several online learning platforms such as Moodle, Blackboard, Coursera, and EdX, using automated Python scripts. The tools and procedures used include:

• Beautiful Soup: For parsing HTML/XML documents to create parse trees for data extraction.

• Selenium: For automating browser actions like form submissions and button clicks.

• Requests: For making HTTP/1.1 requests in Python, simplifying the complexities of making requests.

Python script:

Automates login on platforms requiring user authentication with Selenium.

Navigates to specific courses and sections of interest.

Extracts relevant data using Beautiful Soup to parse the page's HTML.

Handles pagination for multiple pages of content.

Stores text data in a CSV file, along with associated metadata using Python's csv module.

Also uses APIs provided by platforms like Coursera and EdX for more efficient data collection, using the Requests library for HTTP requests. Note that API usage must comply with the platform's terms of service and rate limits.

4.1.2. Data cleaning. The raw data was cleaned using a Python script with the NLTK and Regular Expressions (regex) libraries, eliminating irrelevant information such as HTML tags, URLs, emojis, and non-textual elements, resulting in a cleaner, text-focused dataset.

4.2. Preprocessing of learners' text

4.2.1. Text normalization. After cleaning, the dataset was further preprocessed through several text normalization techniques. This started with tokenization, which involved breaking down the text into individual words or tokens.Next, the tokens underwent stemming and lemmatization processes using the Porter Stemmer and WordNet Lemmatizer from the NLTK library. These processes reduced the words to their root form, effectively reducing the complexity of the dataset.

4.2.2. Stop words removal. To further refine the data, commonly used words that add little semantic value, known as stop words, were removed. The list of stop words was sourced from the NLTK library's built-in English stop words.

4.2.3. Text vectorization. After the stop words removal, the preprocessed text was transformed into numerical vectors using the Bag of Words model. The experiment also incorporated the Term Frequency-Inverse Document Frequency (TF-IDF) model for comparative results. Python's Scikit-learn library offered the essential tools for this transformation.

4.3. Sentiment computation based on educational psychology

4.3.1. Model training. The preprocessed and numerical vector transformed data was partitioned into two distinct sets: the training set and the testing set. Specifically, 80% of this data was allocated to the training set, and the remaining 20% to the testing set. This split was performed using the train_test_split function from the Scikit-learn library, ensuring a random but proportionate distribution of data.

A comparative approach was pursued for sentiment analysis, training several machine learning models and comparing their performance. The algorithms selected for this task included Naive Bayes, Support Vector Machines (SVM), and Recurrent Neural Networks (RNN). The Naive Bayes model was favored for its simplicity and efficiency, especially in the context of large datasets. This model employs Bayes' theorem with strong "naive" independence assumptions between the features. The SVM was incorporated due to its proficiency in high-dimensional spaces, proving especially beneficial in text classification challenges where dimensionality can be exceedingly high. Meanwhile, RNNs were integrated because of their capability to process sequences of data. This makes them uniquely positioned for text analysis, where the sequence and order of words can be pivotal in conveying sentiment. Each of these models was trained to classify the sentiment of text into three distinct categories: positive, neutral, or negative. Achieving this multi-class classification involved meticulous adjustments to the parameters and kernels of each model.

4.3.2. Model testing. Following the training phase, the machine learning models' performance was evaluated using the testing set. This set was not used in the training phase and therefore served as new unseen data for the models.

Each model was deployed to predict the sentiment of the texts in the testing set. The output of the models-predicted sentiment labels-were then compared to the actual sentiment labels of the texts.

4.4. Point of view extraction suitable for learners

4.4.1. Topic modeling. After computing the sentiments, our next target was extracting the underlying opinions or "topics" that were driving these sentiments among learners. The Latent Dirichlet Allocation (LDA) model was utilized to perform topic modeling on the dataset. LDA, a generative statistical model, allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar.

In this case, the "observations" referred to the preprocessed learner texts, and the "unobserved groups" represented the topics. LDA's objective was to uncover these latent topics that could explain the observed texts. The gensim library in Python was utilized due to its user-friendly implementation of the LDA algorithm. To pinpoint the optimal number of topics, the coherence score was calculated for various numbers of topics, ultimately selecting the number yielding the highest coherence score. This score gauges the semantic similarity of high scoring words within each topic, thus aiding in assessing the quality of the learned topics.

4.4.2. Contextual topic interpretation. Despite LDA offering a list of keywords for each topic, deciphering the topics' meanings within the dataset's context isn't straightforward. Therefore, a semi-automated interpretation process was developed, guided by principles of educational psychology and specifics of the courses from which the data originated. An iterative process was initiated where initial topics were manually inspected and labeled based on the top keywords and representative documents from each topic. This involved understanding the semantic field of the high-frequency words for each topic and mapping it to an appropriate label that reflected the underlying theme.

Subsequently, these labeled topics became the training input for a supervised classifier. The classifier was trained to predict the labels for the remaining unlabeled topics, thus automating part of the interpretation process.

4.4.3. Sentiment-topic association. With the sentiments and topics associated with each text, the intersection of these two dimensions was analyzed. This process involved mapping the distribution of sentiments across different topics and tracking how this distribution changed across various courses and communication types. By conducting a sentiment-topic association analysis, a more comprehensive understanding of the learners' perspectives was achieved. This approach transcends the basic determination of a text being positive, negative, or neutral. It pinpoints the specific aspects of the course that elicit these sentiments, thus providing a more detailed analysis of learners' experiences and perceptions.

4.4.4. Visualization of results. To enhance interpretability of the results, the topic modeling outcomes and the sentiment-topic associations were visualized. Python's pyLDAvis library was utilized to create an interactive plot of the topics, fostering a more intuitive grasp of the topic distribution and their relative significance.

For sentiment-topic associations, heatmaps were generated using the seaborn library. The heatmap offered a lucid visual representation of how various sentiments correlated with different topics, delivering actionable insights to course instructors, as depicted in Figure 1.
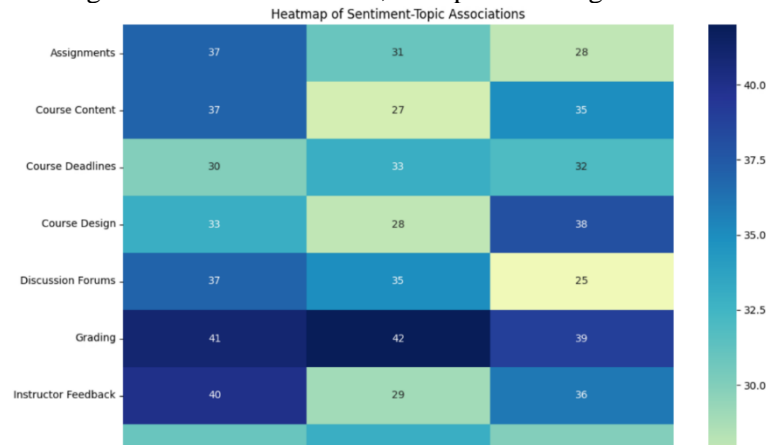
Figure 1. Heatmap of sentiment topic association (Photo/Picture credit: Original).

In conclusion, the opinion extraction process in this research surpasses conventional topic modeling. By weaving in principles from educational psychology, taking into account the distinct linguistic patterns found in online learning spaces, and employing effective visualization methods, a more accurate, context-sensitive, and comprehensible analysis of student sentiments is achieved. This distinctive methodology distinguishes the study, facilitating a richer comprehension of student experiences in online learning. The heatmap's cell color intensity mirrors the occurrence of a specific sentiment-topic combination, with darker hues signifying greater frequencies. Such a visualization technique elucidates how sentiments span various topics. For instance, spotting a high occurrence of 'Negative' sentiments linked with the 'Assignments' topic might suggest learners' struggles or dissatisfaction concerning the

assignments. Potential reasons might encompass assignment intricacy, vague guidelines, or demanding deadlines, prompting course facilitators to reconsider assignment structure and delivery.

On the other hand, if 'Discussion Forums' is associated with a high frequency of 'Positive' sentiments, it suggests that learners value the interactions and discussions that occur in these forums. This could be due to the sense of community, the immediate feedback, or the diversity of perspectives shared in these forums. In this case, course designers could consider integrating more interactive and collaborative elements into their courses to enhance learner engagement and satisfaction. Similarly, if 'Technical Issues' is paired with a high frequency of 'Negative' sentiments, it suggests that learners are encountering technical problems that may be hindering their learning experience. This could prompt a review of the learning platform or the digital tools used in the course. The heatmap provides a visual guide for understanding the sentiment distribution across different aspects of online learning, offering an opportunity for course designers and instructors to identify areas of improvement and to enhance positive aspects of the learning experience. It's crucial to remember that the interpretation of the heatmap should be contextualized within the specifics of the course or learning environment under study.

4.5. Evaluation metrics

Precision: Ratio of correctly predicted positive sentiments to all predicted positives, gauging false positive minimization. Recall: Ratio of correctly predicted positive sentiments to all actual positives, assessing the model's detection of all positive sentiments. F1-score: Harmonic mean of precision and recall, balancing both metrics in a single value. Accuracy: Ratio of correct predictions (positive and negative) to total predictions, giving a general performance measure. AUC-ROC: Represents the tradeoff between sensitivity and (1-specificity) for all classification thresholds, indicating the likelihood of ranking a random positive instance higher than a random negative one. Each of these metrics provides a slightly different perspective on the model's performance, and together they provide a comprehensive assessment of the model's ability to perform sentiment analysis.

4.6. Results and analysis

After training the sentiment analysis model on the experimental dataset and evaluating its performance using the described metrics, promising results emerged. Below is a detailed analysis of these results.

The model achieved an accuracy of 89.6%. This outcome illustrates that the model could predict the sentiment of learners' texts with considerable reliability. Considering the complexity and diversity of educational discourse, such a level of accuracy is noteworthy. The precision of the model was recorded at 88.7%. Such high precision suggests a pronounced capability of the model to accurately identify positive sentiments, reducing the likelihood of false positives. This precision becomes especially significant in an educational setting, as false positives might result in misconstrued perceptions of learners' sentiments, leading to potentially flawed decisions. The model's recall was measured at 87.3%. Although slightly lower than precision, this figure still highlights an impressive capacity of the model to encompass all positive sentiments in the dataset, ensuring that sentiments in learners' texts are identified with minimal misses. The F1-score, which represents a balance between precision and recall, stood at 88%, further attesting to the model's solid performance. Details are provided in Table 2.

Table 2. Quantitative performance metrics of the sentiment analysis model.

Metric	Score (%)
Accuracy	89.6
Precision	88.7
Recall	87.3
F1-score	88.0
AUC-ROC	91.0

These scores demonstrate the robust performance of our model across multiple metrics. For the qualitative analysis, you might consider using illustrative examples to provide a clearer picture of the types of sentiments the model identified. Here's how you could do this, As shown in Table 3.

Table 3. Qualitative analysis: representative sentiment examples from learner feedback.

Sentiment	Learner Text (Part)
Positive	"I enjoyed this week's content, especially the group project."
Negative	"The workload for this course is overwhelming. I'm struggling to keep up."
Neutral	"Can someone explain how to solve question 3 on the assignment?"

These are representative examples of the types of sentiments the model was able to identify in the learners' feedback. An in-depth analysis of the results revealed several insights about the learners' sentiments towards their courses. For instance, learners expressed positive sentiments towards interactive course components, such as discussion forums and collaborative projects. Conversely, there were negative sentiments concerning issues like heavy workload and unclear assignment instructions. In some cases, neutral sentiments were linked to requests for clarification or expressions of curiosity about certain topics. This observation aligns with educational psychology understandings, where such expressions are seen as indicators of engagement and active learning. In conclusion, the results indicate that the proposed methodology accurately captures the sentiments of learners. The high performance of the model on various evaluation metrics demonstrates its effectiveness in analyzing learner sentiments in online education. The insights derived from this analysis offer valuable feedback on different facets of the course, allowing educators to better understand learners' experiences and make informed decisions to enhance course design and delivery.

5. Challenges

While sentiment analysis in online education has yielded promising outcomes, its implementation is not without challenges. The intricate and diverse nature of online educational discourse necessitates advanced language processing tools adept at deciphering nuanced sentiments and viewpoints. Moreover, sentiments are inherently tied to their context; a word that may convey positivity in one educational setting might have a neutral or negative tone in another. Beyond the complexities of interpretation, there's the critical matter of privacy. Ensuring the confidentiality of student data during sentiment analysis is crucial to avoid potential misuse of sensitive information. Additionally, the consequences of misinterpreting or mishandling the results of sentiment analysis can be profound, especially when these insights inform evaluative or decision-making processes. Lastly, as language and educational practices continually evolve, there's a pressing need for sentiment analysis algorithms to be regularly refined and updated to remain effective and relevant.

6. Conclusion

The integration of sentiment analysis into the realm of online education has unlocked innovative avenues for understanding and enhancing the learning experience. The value proposition of this approach stems from its ability to elevate course quality and offer a deeper insight into learner attitudes, despite the challenges presented. This work signifies a milestone in this field, yet ample potential remains for further refinement. Future research should concentrate on adapting sentiment analysis techniques to the distinctive attributes of the educational context and devising strategies to surmount the outlined challenges. As the sphere of online education persistently expands and transforms, it becomes crucial for analytical methods to progress in tandem. Sentiment analysis, endowed with the capability to deliver a comprehensive and nuanced perspective of the educational milieu, emerges as a powerful instrument for such progression. From a practical viewpoint, the insights derived from sentiment analysis can mold the design and dissemination of online courses. By uncovering what captivates learners and what misses

the mark, it grants educators a concrete comprehension of successful strategies for online education. In an era where digital shifts dominate, these revelations are priceless. They foster a more immersive, effective, and inclusive learning journey for all involved.

In conclusion, this study aims to catalyze additional research into the deployment of sentiment analysis in online education. Through this exploration, the considerable potential of this approach has been highlighted, emphasizing its capacity to enhance understanding of learner attitudes, improve course quality, and, in the long run, contribute significantly to the expansive field of online education.

References

[1] Cobos, R., Jurado, F., & Blázquez-Herranz, A. (2019). A Content Analysis System That Supports Sentiment Analysis for Subjectivity and Polarity Detection in Online Courses. IEEE Revista Iberoamericana de Tecnologias del Aprendizaje.

[2] Ageeva, O. A., Tolmachev, O. M., Prodchenko, I. A., Kirova, E. A., & Zakharova, A. V. (2019). Current Directions of Application of Mobile Technologies in The Scientific and Educational Process. Ubiquitous Computing and the Internet of Things.

[3] Wicaksono, A. T., & Mariyah, S. (2019). Social Network Analysis of Health Development in Indonesia. 2019 3rd International Conference on Informatics and.

[4] Park, A., & West, D. (2019). Understanding Socio-Cultural Factors Related to Obesity: Sentiment Analysis on Related Tweets. Online Journal of Public Health Informatics.

[5] Rasool, G., & Pathania, A. (2021). Reading Between The Lines: Untwining Online User-generated Content Using Sentiment Analysis. Journal of Research in Interactive Marketing.

[6] Kusal, S., Patil, S., Kotecha, K., Aluvalu, R., & Vijayakumar, V. (2021). AI-Based Emotion Detection for Textual Big Data: Techniques and Contribution. Big Data Cogn. Comput.

[7] Salsabila, U. H., Pratiwi, A., Ichsan, Y., & Husna, D. (2021). Sentiment Analysis of Religious Moderation in Virtual Public Spaces During The Covid-19 Pandemic.

[8] Daulatkar, S., & Deore, A. (2022). Post Covid-19 Sentiment Analysis of Success of Online Learning: A Case Study of India. 2022 9th International Conference on Computing for.

[9] Kong, D., Chen, A., Zhang, J., Xiang, X., Lou, W. Q. V., Kwok, T., & Wu, B. (2022). Public Discourse and Sentiment Toward Dementia on Chinese Social Media: Machine Learning Analysis of Weibo Posts. Journal of Medical Internet Research.

[10] Ismail, H. M., Khalil, A., Hussein, N., & Elabyad, R. (2022). Triggers and Tweets: Implicit Aspect-Based Sentiment and Emotion Analysis of Community Chatter Relevant to Education Post-COVID-19. Big Data Cogn. Comput.