Opinion Mining Models for Learner Feedback on Massive Open Online Courses

Christy Jackson J
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
christyjackson.j@vit.ac.in

Narendra G O
School of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
narendra.go2019@vitstudent.ac.in

Suriyakrishnan Sathish
School of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
suriyakrishnan.s2019@vitstudent.ac.in

Maheysh V
School of Computer Science & Engineering
Vellore Institute of Technology
Chennai, India
maheysh.v2019@vitstudent.ac.in

Aravinda Boovaraghavan
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
aravinda.b2019@vitstudent.ac.in

Makesh Srinivasan
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
makesh.srinivasan2019@vitstudent.ac.in

Hashwanth S
School of Computer Science and Engineering
Vellore Institute of Technology
Chennai, India
hashwanth.s2019@vitstudent.ac.in

***Abstract*—Feedback and rating in MOOC's (Massive Open Online Courses) are critical components in the creation of courses as well as a successful learning process. These data can be evaluated to discover any probable patterns and, depending on the input, the course is likely to be improved. Recent research in sentiment analysis has shown that it may be used to identify emotions. Positive and negative feedback classification can be used as an initial parameter to determine the learner's degree of satisfaction. The major goal of our work includes investigating various machine learning methods for sentiment analysis utilizing existing MOOC course feedback as a case study. The emotions shown in the feedback might give information about the learner's degree of pleasure. Logistic Regression, KNN, Naïve Bayes, Decision Tree and Random Forest were the methods adopted in the research.**

***Keywords*—MOOC, Sentiment Analysis, Classification, Natural Language Processing, Bag of Words**

I. Introduction

The advent of technology has profoundly transformed the educational industry. The various methods in which technology could improve learning are an emerging field of interest for researchers. Massive Open Online Courses (MOOCs) [1] ignited focus into the education sector. Widely popular and proven MOOC learning sites include Coursera [2], Udacity, edX [3], Australia's Open2Study, and UK Futurelearn. MOOC focuses on building classroom environments that are freely available by unrestricted involvement and online programs. The capabilities entail open access, customizability, flexible content verification, accessible infrastructure, and academic objectives. MOOCs give learners help from virtual communities and resources for online contact with the course teachers [4].

Based on opinions taken from reviews of students, earlier research shows that the most significant consideration for students is who is delivering the course [5].But understanding what can be improved in the course for a given MOOC to be taught by the same team of instructors is more consequential. Recent studies on the use of social media have demonstrated that several psychological intuitions can be identified and analysed by emotion analysis. An early step in evaluating MOOC's emotions includes determining if the response is positive or negative. It should offer an overview of how learners feel about the course. This leads to better decisions to improve learner experience and satisfaction, which is critical in maintaining the MOOC's performance [6]. Additionally, there is no straightforward solution in MOOCs to address sentiment analysis, nor a study of available techniques.

2022 International Conference on Electronic Systems and Intelligent Computing (ICESIC) | 978-1-6654-8385-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICESIC53714.2022.9783508

978-1-6654-8385-8/22/$31.00

This research provides a detailed comparison of various machine learning methods that are applied to the feedback messages for polarity identification. Therefore, as the overall goal is to identify patterns based on the research, a dataset from an existing MOOC is taken and studied to diagnose habits that enrolled users exhibit based on their emotions

II. Related Work

Analysis of emotions is used to describe the disparate types of behavior. In specific, various methods will classify certain states or patterns, like process mining [7] or conversation analysis. Discussion research takes advantage of the prediction system used in this research to infer positive emotions. Apart from forecasting affective states in MOOCs, related methods are used for forecasting dropouts, determining whether the learner can complete the course (or whether the learner can obtain a certificate), or the rating that the student will achieve. Several independent variables were used in some situations, mainly related to system utilisation (e.g. inactivity times [8]), group participation, video-observation activities [9], and previous task performance. In [10], the authors have predicted average assignment marks and observed the greatest association between the number of prior quizzes attempted and the ranking. Sinha and Cassel [11] have projected grades and graded them into multiple categories: low, medium, moderate, and very moderate.

Mackness et al. [12] questioned how to build a MOOC that would provide meaningful outcomes for the participants. Much of the previous work associated with this topic included collecting data and interviews [13], [14]. In comparison, predictive text interpretation, image extraction, and data mining techniques were used in some previous e-learning work to extract views from user-generated material, such as comments, forums, or blogs [15]–[17]. Ideology is necessary to optimise as it has been shown that learners with a good attitude are more inspired in E-learning settings [18] Accordingly, earlier research shows that frustration is associated with impaired performance and behaviour. Conversely, dissatisfaction was less related to worse learning [19]. Depending on user-generated textual comments published during the classes, Adamopoulos [20] conducted sentiment analysis to gather the opinions of students on MOOC features such as the course features and institution features. The aim in that study was to decide which variables influence the success of the course, and it is, therefore, necessary to answer the similar but separate issue of what can be changed when the course is presented again.

Piryani, Madhavi, and Singh [21] researched Sentiment Mining and Sentiment Analysis studies, and their findings showed that machine learning (supervised) methods prevailed (67.2%) over (unsupervised) lexicon techniques. Moreover, comments (e.g. film comments) were the most common data types, followed by tweets and news stories. All controlled instances and unsupervised objects shall be identified as points of reference. Authors in [22] evaluated three (supervised) ML algorithms namely Naïve Bayes, Maximum Entropy Classification, and Support Vector Machines (SVM). The reason for choosing these algorithms is that they had different concepts were employed in earlier classification studies. We considered SVM to be the highest accuracy in that analysis but acknowledged the fact that the problem of classifying feelings has higher complexity than few other classification techniques since this message might sometimes be communicated in a sarcastic language, and the preprocessing stage has some subjectivity. Likewise, Boiri and Moens [23] utilised unigrams to be using the same techniques as binary functions (the occurrence or non-occurrence of the feature). Turney [24] has described, in an unsupervised technique, comments using the difference between words ”excellent” and ”poor,” with accuracies ranging from 0.66 for movie reviews to 0.84 of automobile reviews. However, Hu and Liu [25] made a genius intervention in calculating the directionality of the phrases by remembering the words of opinion (adjectives) and scanning for them in positive and negative terms dictionary definitions, taking into consideration the possible effect of negative terms.

While there are many methods to execute sentiment analysis but there are only a handful of submissions in MOOCs that provide an assessment of the method. This article will pave the way for a novel innovation in the field and will include a summary of the various methods of machine learning for the classification of MOO Courses input.

III. Dataset

While the datasets contains reviews of courses from popular online MOOC websites such as Coursera. The attributes are described in Table I.

TABLE I
Attribute Description

S.No	*Attribute Name*	*Attribute Data Type*	*Attribute Description*
1	CourseId	Continuous	Unique identifier for a review.
2	Review	Discrete	Assessment of the course given by the learner.
3	Label	Discrete [1-5]	Rating of the course given by the learner in the range of 1-5 (very negative-very positive)

IV. Approach

The sentiment analysis carried out in this study focuses on assessing the polarity of the sentiments of the learners in feedback, i.e. whether these are positive or negative feedback. The approach followed in this study involves Natural Language Processing (NLP) and various classification models.

A. Data Preprocessing

1) Non-Alphanumeric Characters: Alphanumeric characters are retained whereas non-alphanumeric characters are removed.

2) Stemming and Lemmatization: Reducing the word to its root stem brings uniformity to the words used in the text. However, stemming may reduce a word to meaningless form and lemmatization is adhered to, for production of uniform meaningful root words.

3) Stopwords Removal: Removal of unwanted words like 'a', 'the' etc. from the text, as they do not impact the classification.

4) Lowercasing: Lowering the case of all the words in the text reduces the dimensions of the matrix representation of the bag of words (BoW) model.

B. Bag of Words

Algorithms can be applied only to numbers and hence the preprocessed text and its tokens is converted to the number format using bag of words model, represented by count vectorization, counting the number of occurrences of each word.

C. Classification Models

The bag of words model produces vectors with binary values, which are fed into different classification models, as an experiment to achieve better and meaningful accuracy. Models included in the research are:

1) Logistic Regression: Logistic regression is a binary classifier, which takes binary values like 1 or 0. Sigmoid function (1) is used for classifying the output. Each class is considered individually for multi-class classification, using One versus All principle.

$$y = \frac{e^{\beta_0 + \ \beta_1 x}}{1 + e^{\beta_0 + \ \beta_1 x}} \tag{1}$$

where y is the independent variable to be classified, x is the dependent variable, β_0 is the bias term and β_1 is the coefficient term for every single input x.
The sigmoid function takes values from 0 to 1 and if y is more than 0.5, the output class is 1 else 0, concordant with probability rules.

2) K-Nearest Neighbors Algorithm: K-NN is a lazy learner algorithm, which stores the dataset, followed by taking action on the dataset. It assumes similarity between the new data points and available data, using various distance measures like Euclidean distance, Manhattan distance etc.

3) Naïve Bayes Algorithm: It is a probabilistic algorithm based on Bayes' theorem with an assumption of independence among independent variables.

$$P\,(C|x) = \frac{P\,(x|C)\,P(C)}{P(x)} \tag{2}$$

where $P\,(C|x)$ is the probability of the output to be class C, given x, $P\,(x|C)$ is the probability of the output to be x, given class C. P(C) and P(x) are the prior probabilities of the class C and input x.
The class to which the given input will be classified as per the posterior probability of the classes. A given point gets assigned to the class with the highest posterior probability.

4) Decision Tree Algorithm: This is a tree based graph used to obtain every possible solution, based on the constraints provided.

5) Random Forest Algorithm: Random forest algorithm is an extensive model of decision tree where multiple decision trees are used.

V. Results and Discussion

Classification models were evaluated using metrics like precision, recall and F1-score.

TABLE II
MODEL ACCURACY

S.No.	Algorithm	Accuracy (%)
1	Logistic Regression	75.4
2	Random Forest	74.9
3	K-NN	73.6
4	Decision Tree	69.1
5	Naïve Bayes	27.9

Fig.1 of the distribution of ratings by linear regression shows that 80% of the reviews were very positive. 14% were positive reviews while neutral, negative and very negative account to a minimal percentage. This indicates that the course they took was highly useful and satisfactory for the users.

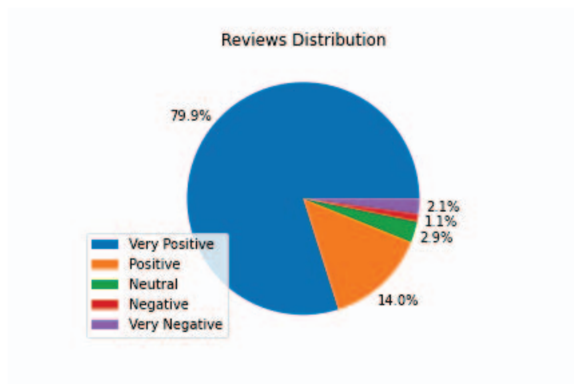

Fig. 1. Distribution of reviews (Logistic Regression).

Fig.2 compares the actual review and reviews predicted by the logistic regression model.
Fig.3 indicates that the computer science courses were inherently popular over the other courses during the most popular courses with its framework courses-Python and R among the top 10. Beyond the top 10 courses, an equilateral trend of technical and non-technical courses were observed.
The word cloud in Fig.4 shows the most popular terms used in Python-based courses, as predicted by logistic regression. User sentiment is largely positive with words such as 'good', 'great' and 'learn' taking part in the reviews. However, certain words, nominal in numbers, like 'difficult' and 'challenging' indicate negative and neutral sentiments.
Algorithm accuracy can be improved, with improvement in the pre-processing NLP techniques followed by stochastic algorithms like logistic regression, support vector machines, multi-layer perceptrons, with words encoded as in

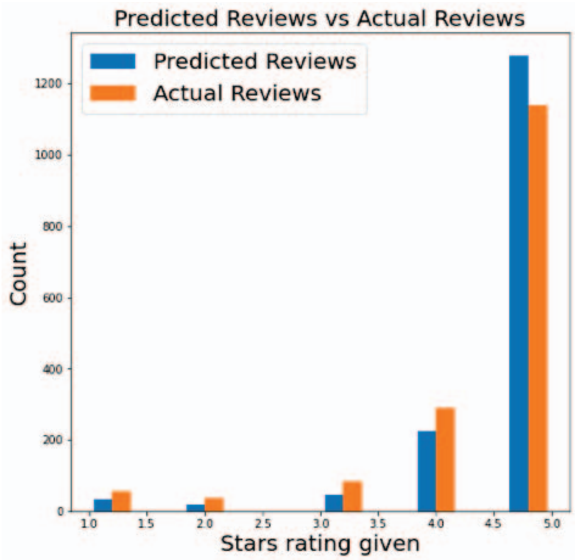

Fig. 2. Predicted vs Actual Reviews (Logistic Regression).

Fig. 4. Word Cloud for Python-based courses (Logistic Regression).

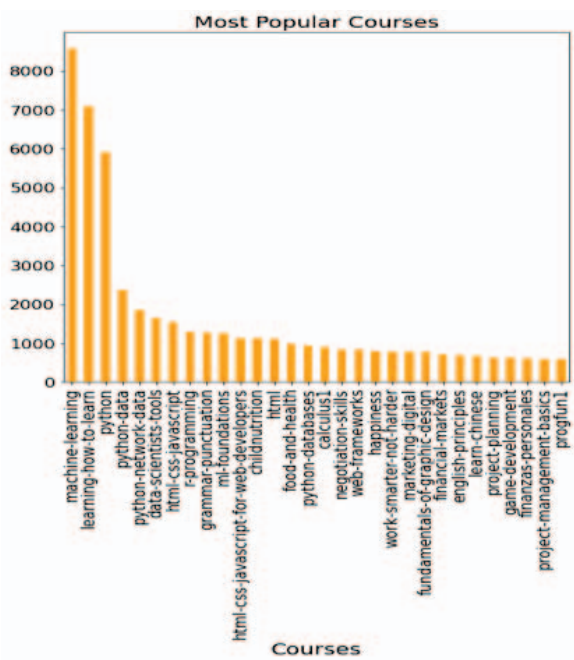

Fig. 3. Popularity of Courses (Logistic Regression).

the proposed system, using n-gram or TF-IDF models 3. Classification thresholds for individual categories can also be experimented with. Values above the threshold can be classified as YES (1) and below the threshold can be classified as NO (0).

Since, the ratio of number of samples to the number of words per sample (W) $<$ 1500, small multi-layer perceptrons with n-gram input can perform better or equivalent to sequence models, with equivalent computation time. If W $>$1500, fine-tuned pre-trained embedding with sepCNN model provides expected results. Hyperparameter tuning, if applicable, can also be experimented with, changing the number of layers, number of units per layer, dropout rate, learning rate, kernel size, embedding dimensions etc.

$$w = \frac{NumberofSamples}{NumberofWordsperSample} \tag{3}$$

VI. Conclusion

In this study, various opinion mining models were used to identify the polarity of the course reviews. Patterns with respect to reviews and ratings were obtained, which can help improvement of the courses offered to the learners. Logistic regression achieved the highest comparable accuracy, indicating majorly positive sentiment or 5-star ratings. Our findings have given an insight into the domain of online learning, and effectiveness of MOOC courses among the general population. This is crucial, especially now because MOOC courses have become a primary source of knowledge and in some cases even forms of formal education. Learning about the impacts, success and effectiveness can help us better structure our programs, and provide better recommendations to learners.

References

[1] J. Kay, P. Reimann, E. Diebold, and B. Kummerfeld, "Moocs: So many learners, so much potential..." *IEEE Intelligent systems*, vol. 28, no. 3, pp. 70–77, 2013.

[2] C. Severance, "Teaching the world: Daphne koller and coursera," *Computer*, vol. 45, no. 8, pp. 8–9, 2012.

[3] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton, "Studying learning in the worldwide classroom research into edx's first mooc." *Research & Practice in Assessment*, vol. 8, pp. 13–25, 2013.

[4] J. Kim, P. J. Guo, D. T. Seaton, P. Mitros, K. Z. Gajos, and R. C. Miller, "Understanding in-video dropouts and interaction peaks inonline lecture videos," in *Proceedings of the first ACM conference on Learning@ scale conference*, 2014, pp. 31–40.

[5] C. Shi, S. Fu, Q. Chen, and H. Qu, "Vismooc: Visualizing video clickstream data from massive open online courses," in *2015 IEEE Pacific visualization symposium (PacificVis)*. IEEE, 2015, pp. 159–166.

[6] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining mooc clickstreams: Video-watching behavior vs. in-video quiz performance," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3677–3692, 2016.

[7] S. Mak, R. Williams, and J. Mackness, "Blogs and forums as communication and learning tools in a mooc," in *Proceedings of the 7th International Conference on Networked Learning 2010*. University of Lancaster, 2010, pp. 275–285.

[8] T. Phan, S. G. McNeil, and B. R. Robin, "Students' patterns of engagement and course performance in a massive open online course," *Computers & Education*, vol. 95, pp. 36–44, 2016.

[9] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.

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

[11] D. Leony, P. J. M. Merino, J. A. R. Valiente, A. Pardo, and C. D. Kloos, "Detection and evaluation of emotions in massive open online courses." *J. UCS*, no. 5, pp. 638–655, 2015.

[12] S. Halawa, D. Greene, and J. Mitchell, "Dropout prediction in moocs using learner activity features," *Proceedings of the second European MOOC stakeholder summit*, vol. 37, no. 1, pp. 58–65, 2014.

[13] C. Ye and G. Biswas, "Early prediction of student dropout and performance in moocs using higher granularity temporal information," *Journal of Learning Analytics*, vol. 1, no. 3, pp. 169–172, 2014.

[14] Z. Ren, H. Rangwala, and A. Johri, "Predicting performance on mooc assessments using multi-regression models," *arXiv preprint arXiv:1605.02269*, 2016.

[15] T. Sinha and J. Cassell, "Connecting the dots: Predicting student grade sequences from bursty mooc interactions over time," in *Proceedings of the second (2015) ACM conference on learning@ scale*, 2015, pp. 249–252.

[16] J. Mackness, S. Mak, and R. Williams, "The ideals and reality of participating in a mooc," in *Proceedings of the 7th international conference on networked learning 2010*. University of Lancaster, 2010, pp. 266–275.

[17] C. O. Rodriguez, "Moocs and the ai-stanford like courses: Two successful and distinct course formats for massive open online courses." *European Journal of Open, Distance and E-Learning*, 2012.

[18] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, H. A. Parada G, M. Muñoz-Organero, and A. Rodríguez-De-Las-Heras, "Analysing the impact of built-in and external social tools in a mooc on educational technologies," in *European Conference on Technology Enhanced Learning*. Springer, 2013, pp. 5–18.

[19] D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in *2007 IFIP international conference on network and parallel computing workshops (NPC 2007)*. IEEE, 2007, pp. 788–792.

[20] H. H. Binali, C. Wu, and V. Potdar, "A new significant area: Emotion detection in e-learning using opinion mining techniques," in *2009 3rd IEEE international conference on digital ecosystems and technologies*. IEEE, 2009, pp. 259–264.

[21] A. El-Halees, "Mining opinions in user-generated contents to improve course evaluation," in *International conference on software engineering and computer systems*. Springer, 2011, pp. 107–115.

[22] J. Moshinskie, "How to keep e-learners from e-scaping," *Journal of Interactive Instruction Development*, vol. 14, no. 1, pp. 8–11, 2001.

[23] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments," *International Journal of Human-Computer Studies*, vol. 68, no. 4, pp. 223–241, 2010.

[24] P. Adamopoulos, "What makes a great mooc? an interdisciplinary analysis of student retention in online courses," 2013.

[25] R. Piryani, D. Madhavi, and V. K. Singh, "Analytical mapping of opinion mining and sentiment analysis research during 2000–2015," *Information Processing & Management*, vol. 53, no. 1, pp. 122–150, 2017.