Student Evaluations in Teaching - Emotion Classification using Neural Networks

Abstract—Student evaluation of teaching effectiveness plays an important role in Higher Education. Evaluations serve as Formative (identify areas of improvement in the process) and Summative (assess the end goal) measurements of teaching. Educational institutions collect these evaluations in both qualitative and quantitative forms. Qualitative evaluations serve as a bridge for students to express their feelings about the teaching methodology used, instructor efficiency, classroom environment, learning resources and others. Identifying student emotions help instructors to have good intellectual insight about the actual impact of teaching. Teaching models include traditional models, modern flipped class-room models, and active learning approaches. Light-weight team is an active learning approach, in which team members have little direct impact on each other's final grades, with significant long-term socialization. We propose and extend previous method for assessing the effectiveness of the Light-weight team teaching model, through automatic detection of emotions in student feedback in computer science course by using Neural Network model. Neural Networks have been widely used and shown high performance in variety of tasks including but not limited to Text Classification and Image Classification. It is highly deemed to work great with huge volume of data. In this study we discuss how sequential model can be used with smaller data sets and it performs well, compared to the baseline models such as Support Vector Machines and Naive Bayes.

Keywords—Classification, Educational Data Mining, Neural Networks, Student Evaluations, Teaching Methods.

I. Introduction

Quality of education is one of the primary factors which requires constant attention and improvement. Student evaluations of teaching serve as both formative and summative measure in the process of quality education. Literature dates back to 1920's [1] with the works of Remmers to assess the student evaluation agreements with alumni and peers [2], [3]. Educational institutions collect student evaluations in both quantitative and qualitative forms. The quantitative feedbacks include a likert-type scale in which responses are scored along a range, to capture the level of agreement and disagreement. Qualitative feedbacks serve as a bridge for students to express their feelings about the teaching method used, instructor efficiency, classroom environment, learning resources and others.

One of the emerging approach in the field of teaching is Active Learning approach. Light-Weight teams [4] is an Active Learning approach, in which team members have no direct impact on each other's final grades, yet there is a significant component of peer teaching, peer learning and long-term socialization. This innovative pedagogical approach has been studied in Computer Science undergraduate courses and has been reported to have high levels of student engagement [4], [5].

Emotion Mining is the process of detecting and analyzing human feelings about events, issues and or services. Qualitative feedbacks aids in the process of identifying student emotions. Authors Tzacheva et al. [6], [7], study the effectiveness of teaching model and their impact on student learning styles and experience in classroom and identify factors that help in performance and positive attitude of students towards Computer science course. They propose a novel method for assessing pedagogical innovation thorugh detection of emotions in text, produced by student participants, in computer science courses. The results show that implementation of Active Learning methods increase positive emotions among students and improve their learning experience.

Educational Data Mining is a new field which involves identifying patterns of student behaviours and learning by use of Machine Learning and Data Mining technologies. Neural Networks in Data Mining is a mathematical model which has its roots in biological neural network. Neural networks have achieved impressive results in several classification tasks [8], [9], [10], [11]. It is widely perceived that Neural Networks performs well with huge volume of data. Since student evaluations of teaching has limited data availability considering the number of students registering for a course, very limited works have used Neural networks in the education data mining field. Researchers use the classical machine learning models like Naive Bayes, Support Vector Machine for sentiment classification of student evaluations data. In this work we use sequential learning model on the student feedbacks for emotion classification and compare with the traditional models.

The reminder of the paper is organised as follows, section II focusses on related work; followed by method, experiments and results in section III and IV.

II. Related Works

A. Clasification - Traditional Machine Learning Models

Authors Altrabsheh et al. [12], collect real time student feedback and label the data into three sentiment class 'positive', 'neutral', and 'negative' with help of three experts. The learning performance was investigated with the following machine learning techniques: Naive Bayes, Complement Naive Bayes, Maximum Entropy, and Support Vector Machine. They achieve good results with Support Vector Machine and Complement Naive Bayes. In a similar way authors Leong et al. [13] use prompt feedback and propose the use of short message service (SMS) for student evaluation and explore the application of text mining in particular Sentiment Analysis ('positive' and 'negative')on SMS texts. They show the positive and negative aspects of lecture in terms of the conceptual words extracted and text link analysis visualization.

Authors Dhanalakshmi et al. [14], classify student's feedback into 'positive' or 'negative' and suggest that Naive Bayes performs better with good recall. Authors Jagtap &

Dhotre [15] classify student feedback data into 'positive' and 'negative' categories by using of hybrid approach combining Hidden Markov Model (HMM) and Support Vector Machine (SVM). Though they have concluded that applying advance feature selection method combined with hybrid approach work well for complex data, their works did not show the results of classification model for validation.

Authors Rajput et al. [16] apply text analytics methods on student's feedback data and obtain insights about teacher's performance with the help of tag clouds, and sentiment score. In this work the authors use sentiment dictionary Multi-Perspective Question Answering (MPQA) Stoyanov et al. [17] to find words with positive and negative polarity. By combing the word frequency and word attitude the overall sentiment score for each feedback is calculated. Finally they have compared the sentiment score with Likert scale based teacher evaluation and conclude that Sentiment score with word cloud provide better insights than Likert scale results.

B. Classification - Neural Networks

Neural Networks is widely used in several classification tasks and proven to achieve best results. But it is still in the infancy stage with Educational Data. Most of the works in literature focus on predicting student performance using Artificial Neural Networks. For instance, Guo et al. [18] use multiple level representations with unsupervised learning and fine tune neural network layers through back propagation. They use High school data with different kinds of information including background and demographic data, past study data, school assessment data, study data, and personal data. Compared to the traditional methods like Support Vector Machines and Naive Bayes, their model achieve better performance. Authors Musso et al. [19], also use student background information along with cognitive and non-cognitive measures to predict student academic performance using Artificial Neural Networks achieve greater accuracy compared to discriminant analyses method.

While the above methods use non-text data for classification, the following researchers use text data. Online discussion forum is a popular tool for student communication and collaboration in web-based courses. Authors Wei et al. [20] use Stanford MOOC posts dataset [21] to identify 'confusion', or 'urgency' and sentiment of the posts. They propose a transfer learning framework based on convolutional neural network and long short-term memory model. Student Evaluation of Teaching Effectiveness (SETE) serves as an important aspect in validating the teaching models, resourses and effectiveness of teaching and learning outcomes. Authors Galbraith et al. [22] use Neural Networks to measure student learning outcomes from SETE's.

There is not much work in applying Neural networks for sentiment classification from student evaluation of teaching. In this work we use sequential model with 1D convolution and word embedding for automatic classification of emotions from student evaluations.

TABLE I
SAMPLE - STUDENT COMMENTS

Easily available to communicate with if needed.
The course has a lot of valuable information.
Get rid of the group project.
There was no enthusiasm in the class.
The instructor should make the class more lively and interactive.
Best professor

III. METHODOLOGY

We use Web-Based course evaluation system to collect Data for the study. This system is administered by a third-party Campus Labs. In assistance with Center for Teaching and Learning, Campus Labs collect the student feedback for course evaluations. The student feedbacks for an instructor is collected for the terms of 2013 to 2017 including Fall, Spring and Summer sections of courses handled by the instructor. After the data collection from Campus Labs, jsoup [23] a Java library is used to process the html files and extract the comments. The data contains 1070 instances. Sample student comments shown in "Table. I".

A. Pre-Processing

We use python Natural Language Toolkit (NLTK) [24] to process the Qualitative student feedbacks and make it suitable for emotion labeling and classification. The steps include removing certain special characters like punctuation, splitting the sentence into pieces of words called tokens, case-folding, stop-words removal. The pre-processed dataset contains close to 800 records in the dataset

B. Emotion Labeling

Labeling the data is the most significant task for any supervised machine learning algorithm. In this work we use the National Research Council - NRC Lexicon [25], [26] for this purpose. NRC Emotion lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, disgust, surprise, trust, joy, and sadness) and two sentiments (positive and negative). The Annotations in the lexicon are at WORD-SENSE level. Each line has the format: <Term> <AffectCategory> <AssociationFlag>.

Student comments is processed and if a match to word is found then the score is incremented accordingly based on the Flag value in lexicon, here if a word is present twice then automatically based on the frequency score for that particular emotion will be incremented. After the entire comment is processed the Emotion which has the highest score is assigned as the final Emotion with respect to that student comment.

C. Classification

Classification is the process of predicting the class labels of given data points, and it belongs to the category of Supervised Learning. The learning algorithms for classification are broadly divided into two types as lazy learning (memory-based learning system) and eager learning (optimized learning

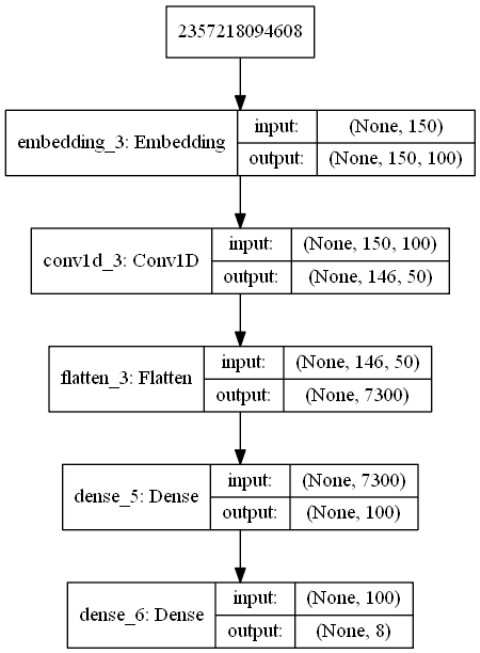

Fig. 1. Neural Networks Model Summary

system) algorithms. Lazy learning algorithms store all the training data and defe the process until it recieves a query or test set to process. Whereas the eager learnng algorithm learns the classifier structure with the training data and use the learning to predict the test instances. The former takes less time learning and more time classifying while the later is the opposite.

Some examples of lazy learning algorithms include K-Nearest Neighbor, Case-Based reasoning; while Naive Bayes, Neural networks, Decion Tree are examples of Eager learning. In this paper we use Keras [27] a high-level neural network API in python for automatic classification of emotion from student evaluation data. The classification model is based on Keras sequential model, which is a linear stack of layers. We use the 1D convolutional kernal with dense (fully connected) layer compiled with Adaptive Moment Estimation (Adam) optimizer and categorical crossentropy as loss function. Finally the model is trained using Epochs = 5 and Batch size = 2.

IV. Experiments and Results

In this section we describe our experiments and results. The data for this study is collected from Campus Labs website. The data extracted consists 1070 records. The pre-processed dataset contains close to 820 records in the dataset. For labeling the data - student feedback comments with different types of Emotion, we use the National Research Council - NRC Lexicon [25], [26].

There are several classification algorithms that have been applied to text classification problems. In this work we use traditional Naive Bayes and Support Vector Classification methods as a baseline to compare the neural networks implementation.

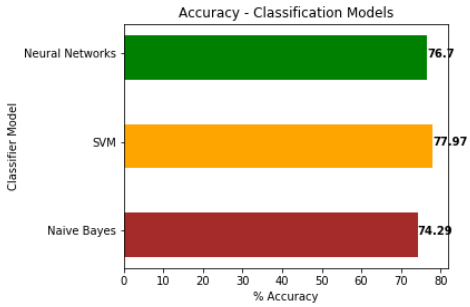

Fig. 2. Classifier - Accuracy.

A. Naive Bayes Classifier and Support Vector Machine Classifier

One of the popular use of text pre-processing in the traditional methods is use of TF-IDF (Term Frequency - Inverse Document Frequency) which is a popular weighting scheme used in information retrieval and text mining applications. It is a statistical measure to evaluate the importance of words in the document or corpus. TF-IDF is mainly composed of two terms: Term Frequency "(1)" and Inverse Document Frequency "(2)".

$$TF(t) = \frac{\text{Number of times term t appears in a document}}{\text{Total number of terms in the document}} \quad (1)$$

$$IDF(t) = \frac{\text{Total number of documents}}{\text{Number of documents with term t in it}} \quad (2)$$

The student evaluations dataset is processed with TF-IDF and given as input to the Naive Bayes and Support Vector classification. We achieve accuracy of approximately 74.79% with Naive Bayes and 77.97% with Support Vector Machine.

B. Neural Networks Classifier

In order for the text input to be understood by the neural network algorithm, it is required to process the text before passing to the classifier model to be trained. For this purpose words are replaced with unique numbers and combined with embedding vector to make it semantically meaningful. We achieve an accuracy of approximately 76.7% which is very much in close approximation with the traditional models

V. Conclusion

In this paper we apply neural networks classifier for emotion detection in student evaluation of teaching. We use Keras Deep Learning API. Using appropriate number of epochs for training on the source domain results in better performance. We also compare the neural networks model with the traditional text classification models like Naive Bayes and Support Vector Machine. We notice that neural networks yields (76.7%) similar performance to traditional text classification models like Naive Bayes (74.79%) and Support Vector Machine (77.97%), though the size of the dataset is not big. which is considered to be a drawback when using neural networks for classification.

In future we plan to extend this work by collecting student survey to identify actionable patterns, that helps improve the teaching model, learning environment to a better state.

REFERENCES

[1] H. K. Wachtel, "Student evaluation of college teaching effectiveness: A brief review," *Assessment & Evaluation in Higher Education*, vol. 23, no. 2, pp. 191–212, 1998.

[2] H. Remmers, "The relationship between students' marks and student attitude toward instructors.," *School & Society*, 1928.

[3] H. Remmers, "To what extent do grades influence student ratings of instructors?," *The Journal of Educational Research*, 1930.

[4] C. Latulipe, N. B. Long, and C. E. Seminario, "Structuring flipped classes with lightweight teams and gamification," in *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*, pp. 392–397, ACM, 2015.

[5] S. MacNeil, C. Latulipe, B. Long, and A. Yadav, "Exploring lightweight teams in a distributed learning environment," in *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pp. 193–198, ACM, 2016.

[6] A. Tzacheva and J. Ranganathan, "Emotion mining from student comments a lexicon based approach for pedagogical innovation assessment," *European Journal of Education and Applied Psychology*, no. 3, pp. 3–13, 2018.

[7] A. Tzacheva, J. Ranganathan, and R. Jadi, "Multi-label emotion mining from student comments," in *Proceedings of the 2019 The 4th International Conference on Information and Education Innovations*, pp. 1–5, ACM, 2019.

[8] M. Aono and S. Himeno, "Kde-affect at semeval-2018 task 1: Estimation of affects in tweet by using convolutional neural network for n-gram," in *Proceedings of The 12th International Workshop on Semantic Evaluation*, pp. 156–161, 2018.

[9] Y. Kim, H. Lee, and K. Jung, "Attnconvnet at semeval-2018 task 1: attention-based convolutional neural networks for multi-label emotion classification," *arXiv preprint arXiv:1804.00831*, 2018.

[10] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification," in *Twenty-ninth AAAI conference on artificial intelligence*, 2015.

[11] A. Severyn and A. Moschitti, "Twitter sentiment analysis with deep convolutional neural networks," in *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 959–962, ACM, 2015.

[12] N. Altrabsheh, M. Cocea, and S. Fallahkhair, "Learning sentiment from students' feedback for real-time interventions in classrooms," in *Adaptive and Intelligent Systems* (A. Bouchachia, ed.), (Cham), pp. 40–49, Springer International Publishing, 2014.

[13] C. K. Leong, Y. H. Lee, and W. K. Mak, "Mining sentiments in sms texts for teaching evaluation," *Expert Syst. Appl.*, vol. 39, pp. 2584–2589, Feb. 2012.

[14] V. Dhanalakshmi, D. Bino, and A. M. Saravanan, "Opinion mining from student feedback data using supervised learning algorithms," in *2016 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1–5, March 2016.

[15] B. Jagtap and V. Dhotre, "Svm and hmm based hybrid approach of sentiment analysis for teacher feedback assessment," *International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)*, vol. 3, no. 3, pp. 229–232, 2014.

[16] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-based sentiment analysis of teachers' evaluation," *Applied Computational Intelligence and Soft Computing*, vol. 2016, p. 1, 2016.

[17] V. Stoyanov, C. Cardie, and J. Wiebe, "Multi-perspective question answering using the opqa corpus," in *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pp. 923–930, Association for Computational Linguistics, 2005.

[18] B. Guo, R. Zhang, G. Xu, C. Shi, and L. Yang, "Predicting students performance in educational data mining," in *2015 International Symposium on Educational Technology (ISET)*, pp. 125–128, IEEE, 2015.

[19] M. F. Musso, E. Kyndt, E. C. Cascallar, and F. Dochy, "Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks.," *Frontline Learning Research*, vol. 1, no. 1, pp. 42–71, 2013.

[20] 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, 2017.

[21] A. Agrawal, J. Venkatraman, S. Leonard, and A. Paepcke, "Youedu: addressing confusion in mooc discussion forums by recommending instructional video clips," 2015.

[22] C. S. Galbraith, G. B. Merrill, and D. M. Kline, "Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? a neural network and bayesian analyses," *Research in Higher Education*, vol. 53, no. 3, pp. 353–374, 2012.

[23] J. Hedley, "jsoup: Java html parser," 2009.

[24] J. Perkins, *Python text processing with NLTK 2.0 cookbook*. Packt Publishing Ltd, 2010.

[25] S. M. Mohammad and P. D. Turney, "Crowdsourcing a word–emotion association lexicon," *Computational Intelligence*, vol. 29, no. 3, pp. 436–465, 2013.

[26] S. M. Mohammad and P. D. Turney, "Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon," in *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pp. 26–34, Association for Computational Linguistics, 2010.

[27] A. Gulli and S. Pal, *Deep Learning with Keras*. Packt Publishing Ltd, 2017.