# Predicting students' emotions using machine learning techniques

Nabeela Altrabsheh, Mihaela Cocea, and Sanaz Fallahkhair

School of Computing, University of Portsmouth
Lion Terrace, Portsmouth, United Kingdom
{nabeela.altrabsheh,mihaela.cocea,sanaz.fallahkhair}@port.ac.uk

**Abstract.** Detecting students' real-time emotions has numerous benefits, such as helping lecturers understand their students' learning behaviour and to address problems like confusion and boredom, which undermine students' engagement. One way to detect students' emotions is through their feedback about a lecture. Detecting students' emotions from their feedback, however, is both demanding and time-consuming. For this purpose, we looked at several models that could be used for detecting emotions from students' feedback by training seven different machine learning techniques using real students' feedback. The models with a single emotion performed better than those with multiple emotions. Overall, the best three models were obtained with the CNB classifier for three emotions: amused, bored and excitement.

## 1 Introduction

Emotions play an important role in the learning process, and, thus, their detection can improve our understanding of the role they play [2]. For example, positive emotions can increase students' interest in learning, increase engagement in the classroom and motivate students [2], and happy students are generally more motivated to accomplish their goals.

Research on the prediction of specific emotions from text is in its early days, with very few studies reported in this area [1], end even fewer with focus on education [6]. Moreover, from these approaches, only some use machine learning in their approach for emotion prediction from text, e.g. [1, 6].

In this paper we focus on the prediction with machine learning of emotions relevant for learning from students' textual feedback in a classroom context, which to the best of our knowledge, has not yet been investigated. To establish which emotions are relevant for learning, research evidence is used from previous studies. To investigate the prediction of the identified emotions from text, we experimented with several preprocessing and machine learning techniques.

There are four main steps to create prediction models from text with machine learning: preprocessing the data, selecting the features, applying the machine learning techniques and evaluating the results. *Preprocessing* is the process of cleaning the data from unwanted elements and has been applied in specific emotion prediction, e.g. removal of stop words and stemming [4]. *Feature selection* allows a more accurate analysis of the sentiments and detailed summarization of the results. The most common feature is unigrams which is found in many

research works, e.g. [4]. A variety of *machine learning techniques* have been used for polarity and emotions prediction from text. In our experiments we used classifiers previously shown to work well [6, 4]: Naive Bayes (NB), Multinomial Naive Bayes (MNB), Complement Naive Bayes (CNB), Support Vector Machines (SVM), Maximum Entropy (ME), Sequential Minimal Optimization (SMO), and Random Forest (RF). The models are *evaluated* using accuracy, precision, recall, F-score, Area under curve (AUC), kappa statistic and error rate.

Previous research on emotions related to learning indicates a variety of emotions. Some studies grouped similar emotions, which may be beneficial as using too many emotions is difficult to classify and may cause conflicts [5]. This, however, could be misleading for the real emotion, such as in the case of grouping disgust with boredom or frustration with sadness [7].

We chose the following emotions in our research due to their importance in learning and their common use in previous research [5, 7]: Amused, Anxiety, Appreciation, Awkward, Bored, Confusion, Disappointed, Embarrassed, Engagement, Enthusiasm, Excitement, Frustration, Happy, Motivated, Proud, Relief, Satisfaction, Shame, and Uninterested.

## 2   Emotion Prediction from Students' Feedback

The data was collected from lectures taught in English in Jordanian universities. The students submitted feedback, opinions, and feelings about the lecture through Twitter. For each tweet, they were asked to choose one emotion from a set of emotions provided (the 19 emotions from the previous section).

The total amount of data collected was 1522 tweets with the corresponding emotion label, one tweet from each student. Some of the emotions appeared more frequently than others.The most frequent emotions were: Bored (336), Amused (216), Frustration (213), Excitement (178), Enthusiasm (176), Anxiety (130), Confusion (73), and Engagement (67).

We experimented with all the emotions combined and then subtracted, in turn, the emotion with the lowest number of data. In total, we experimented with 15 models: all emotions (8 classes); 7 emotions (All except engagement) +other (8 classes); 6 emotions (7 emotions except confused) + other (7 classes); 5 emotions (6 emotions except anxiety) + other (6 classes); 4 emotions (5 emotions except enthusiasm) + other (5 classes); 3 emotions (4 emotions except excitement) + other (4 classes); 2 Emotions (Amused, Bored) + other (3 classes); and each emotion + other (2 classes).

We experimented with two levels of preprocessing: (a) high preprocessing, which includes: tokenization, covert text to lower case, remove punctuation, remove numbers, remove stop words, remove hashtags, remove URLs, remove retweets, remove user mentions in tweets, and remove Twitter special characters; (b) low processing, which includes: tokenization, covert text to lower case, and remove stop words. The high preprocessing was only used for all the emotions combined model, i.e. the 8 classes model. Due to the low results observed for this model, only low preprocessing was used for the other models.

All the models were tested using 10-fold cross-validation; the accuracy, error rate and kappa statistics were used to assess the overall performance of the classifiers, while precision, recall, F-score and Area under curve (AUC) were used to assess the ability of the classifier to correctly identify the specific emotions. For the multi-class models, the metrics for the emotion classes were averaged. The best results for each model are represented in Table 1. These were chosen based on the highest precision, recall, F-score, and AUC.

**Table 1.** Best models

|  | Method | Accuracy | Error | Precision | Recall | F-score | kappa | AUC |
|---|---|---|---|---|---|---|---|---|
| ALL Preprocessed | ME | 0.28 | 0.72 | 0.28 | 0.27 | 0.25 | 0.14 | 0.64 |
| ALL W/O Preprocessing | ME | 0.31 | 0.69 | 0.30 | 0.31 | 0.27 | 0.17 | 0.67 |
| 7 Emotions+ other | MNB | 0.26 | 0.75 | 0.26 | 0.24 | 0.24 | 0.14 | 0.62 |
| 6 Emotions+ other | MNB | 0.27 | 0.73 | 0.27 | 0.26 | 0.27 | 0.14 | 0.62 |
| 5 Emotions+ other | MNB | 0.28 | 0.72 | 0.28 | 0.30 | 0.29 | 0.14 | 0.63 |
| 4 Emotions+ other | MNB | 0.33 | 0.67 | 0.30 | 0.37 | 0.33 | 0.16 | 0.65 |
| 3 Emotions + other | ME | 0.43 | 0.57 | 0.32 | 0.33 | 0.30 | 0.18 | 0.62 |
| *2 Emotions+ other* | *CNB* | *0.53* | *0.47* | *0.35* | *0.48* | *0.41* | *0.23* | *0.63* |
| **Amused+ other** | **CNB** | **0.65** | **0.35** | **0.23** | **0.56** | **0.33** | **0.14** | **0.61** |
| Anxiety+ other | CNB | 0.66 | 0.34 | 0.12 | 0.40 | 0.18 | 0.05 | 0.54 |
| **Bored+ other** | **CNB** | **0.66** | **0.34** | **0.38** | **0.63** | **0.47** | **0.24** | **0.65** |
| Confused+ other | CNB | 0.65 | 0.35 | 0.06 | 0.39 | 0.11 | 0.02 | 0.53 |
| Engagement+ other | CNB | 0.64 | 0.36 | 0.04 | 0.27 | 0.07 | 0.02 | 0.47 |
| Enthusiasm + other | CNB | 0.62 | 0.38 | 0.15 | 0.41 | 0.22 | 0.04 | 0.53 |
| **Excitement+ other** | **CNB** | **0.67** | **0.33** | **0.21** | **0.60** | **0.32** | **0.16** | **0.64** |
| Frustration+ other | CNB | 0.68 | 0.32 | 0.24 | 0.48 | 0.48 | 0.14 | 0.6 |

The results indicate that in terms of accuracy the models with a single emotion performed better than the multi emotion models. Accuracy alone, however, does not indicate how well a classifier can predict specific emotions. The precision, recall, and F-score for the emotion classes indicate how well the models perform in terms of detecting emotion. Since precision, recall, and F-score for the emotion class(es) are relatively low, the accuracy is due to the correct identification of the other class (which is also the majority class in most cases).

AUC is the probability that a classifier will rank a randomly chosen positive instance higher than a randomly chosen negative one. It is usually used in machine learning to test how well the models perform, however, it can be noisy as a classification measure [3]. Our AUC results are all relatively low, indicating that our models are not very good at identifying specific emotions. Similarly, the kappa statistic for all model are low, indicating that the models perform better than chance, but only to a limited level. This statistic is comparable across models, regardless of their number of classes, thus indicating that the "2 emotions + other" model is performing better than some of the 2-class models.

When looking at the machine learning techniques that lead to the best identification of emotions (i.e. the recall for the emotion class), Multinomial Naive Bayes (MNB) and Maximum Entropy (ME) led to the best results for the multi-

emotion models. However for the single emotion models, Complement Naive Bayes (CNB) performs best, leading to the highest recall for all eight models.

The results indicate that some emotions can be more easily detected than others. When looking at the overall picture and the balance of the evaluation metrics considered, three 2-class models (in bold) and a 3-class model (in italics) stand out in Table 1. The models are: Amused + other, Bored + other, Excitement + other and Amused + Bored + other.

## 3 Conclusions and Future Work

In this paper, we investigated different models to detect specific emotion from students' real-time feedback. We found that the most frequent learning emotions in students' feedback were: Amused, Anxiety, Bored, Confusion, Engagement, Enthusiasm, Excitement, and Frustration.

We found that some emotions are more easily detectable than others, i.e. Amused, Anxiety and Bored; however, all models were relatively low in performance. The detection of emotion from text is a difficult process due to the different interpretations of words, as well as the limited data availability; hence further investigation is needed to improve the models.

Future work includes experimenting with other n-grams such as bigrams and trigrams, using learning-related emotion lexicons, and investigating the relation between learning emotions and sentiment polarity.

## References

1. Danisman, T., Alpkocak, A.: Feeler: Emotion classification of text using vector space model. In: Convention Communication, Interaction and Social Intelligence, vol. 1, pp. 53–59 (2008)
2. D'Mello, S., Jackson, T., Craig, S., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R., et al.: Autotutor detects and responds to learners affective and cognitive states. In: Workshop on Emotional and Cognitive Issues at the International Conference on Intelligent Tutoring Systems (2008)
3. Hanczar, B., Hua, J., Sima, C., Weinstein, J., Bittner, M., Dougherty, E.R.: Small-sample precision of roc-related estimates. Bioinformatics **26**(6), 822–830 (2010)
4. Kaur, J., R Saini, J.: Emotion detection and sentiment analysis in text corpus: A differential study with informal and formal writing styles. International Journal of Computer Applications **101**(9), 1–9 (2014)
5. Kort, B., Reilly, R., Picard, R.W.: An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In: Advanced Learning Technologies, pp. 43–436. IEEE Computer Society (2001)
6. Tian, F., Gao, P., Li, L., Zhang, W., Liang, H., Qian, Y., Zhao, R.: Recognizing and regulating e-learners emotions based on interactive chinese texts in e-learning systems. Knowledge-Based Systems **55**, 148–164 (2014)
7. Tian, F., Zheng, Q., Zheng, D.: Mining patterns of e-learner emotion communication in turn level of chinese interactive texts: Experiments and findings. In: Computer Supported Cooperative Work in Design (CSCWD), pp. 664–670. IEEE (2010)