Opinion Mining for Predicting Peer Affective Feedback Helpfulness

Mouna Selmi[1], Hicham Hage[2] and Esma Aïmeur[1]
[1]Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Canada
[2]Computer Science Department, Faculty of Natural & Applied Sciences, Notre Dame University, Zouk Mosbeh, Lebanon

Keywords: E-Learning, Peers' Interaction, Peer Affective Feedback, Classification, Machine Learning, Natural Language Processing, Sentiment Analysis, Opinion Mining.

Abstract: Peer feedback has become increasingly popular since the advent of social networks, which has significantly changed the process of learning. Some of today's e-learning systems enable students to communicate with peers (or co-learners) and ask or provide feedback. However, the highly variable nature of peer feedback makes it difficult for a learner who asked for help to notice and benefit from helpful feedback provided by his peers, especially if he is in emotional distress. Helpful feedback in affective context means positive, motivating and encouraging feedback while an unhelpful feedback is negative, bullying and demeaning feedback. In this paper, we propose an approach to predict the helpfulness of a given affective feedback for a learner based on the feedback content and the learner's affective state. The proposed approach uses natural language processing techniques and machine learning algorithms to classify and predict the helpfulness of peers' feedback in the context of an English learning forum. In order to seek the best accuracy possible, we have used several machine learning algorithms. Our results show that Naïve-Bayes provides the best performance with a prediction accuracy of 87.19%.

1 INTRODUCTION

Affective and psychological factors seem to affect the learner's motivation and performance in a learning context (Robinson et *al.*, 2009). This is mainly useful for distant learning systems where learners lack face to face interactions with the tutor and their co-learners. To meet the learner's affective needs, these systems can adapt the learning activities to the learner's affective and psychological state and/or elicit affective feedback from co-learners to help him overcome his emotional distress. This feedback type became popular in learning environment over the last few years, especially with the advent of social networks (Ortigosa et *al.*, 2014). Although, this feedback type may not achieve the quality of tutor feedback, its advantage is that it can often be given in a more frequent and voluminous manner (Lu and Law, 2012). However not all peer feedback is helpful for the feedback requester (Walker et *al.*, 2012). As an illustration, let us take the example of a learner, *Bob*, enrolled in an online English course. He gets frustrated whenever he has to speak in English. To overcome his frustration, he decided to post a message on his class forum and ask his co-learners (peers) to help him with some advices

Bob wrote: *"I am originally from China. Whenever i speak to native (English speaker), I feel very frustrated and i'll start to stammer. The phrasing, sentence structure & grammar of my sentences become all in a mess."*

In response to his request, Bob receives a lot of feedback from his peers. Nonetheless, not all the feedback he received was positive (e.g. advising, motivating). For instance, one of his co-learners started ridiculing him because of his origin rather than helping him with some advices. If Bob faces negative feedback (e.g. ridiculing) first, while he is experiencing already negative emotions (frustration), this may worsen his affective state, prevent him from noticing positive ones and even push him **to give up his learning**. However, if positive and effective feedback is presented to Bob first, this will help him to be more confident and encourage him to pursue his goal of learning. Hence, it is important to filter peers' feedback and protect learners from negative ones. In (Selmi et *al.*, 2013), a privacy framework has been proposed to protect the feedback requester (in emotional distress), in the context of peer affective

Selmi M., Hage H. and Aïmeur E..
Opinion Mining for Predicting Peer Affective Feedback Helpfulness.
DOI: 10.5220/0005158704190425
In *Proceedings of the International Conference on Knowledge Management and Information Sharing* (KMIS-2014), pages 419-425
ISBN: 978-989-758-050-5

feedback, from abusive peers. Nonetheless, this is not enough, since it does not protect the learner against a negative feedback provided unintentionally by a good-wiling peer. As countermeasure, it is strongly needed to propose an approach that evaluates the quality of peers' affective feedback.

Previous research has considered the assessment and quality evaluation of peers review in collaborative e-learning environment (Nicol et *al.*, 2006). However, peer reviews and assessment are cognitive feedback. They are context independent, and target the content by specifying and evaluating aspects of the work. Whereas affective feedback is context dependent and uses affective language to bestow praise and criticism, or to give encouragement and support in order to improve the individual performance. To the best of our knowledge, no prior work in the educational literature has attempted to evaluate automatically the quality of peer affective feedback. Therefore, in this paper, we propose an approach to classify and predict the quality of peer affective feedback in a learning context. For this purpose, we use machine learning techniques and natural language processing. Furthermore, we consider in this classification contextual information such as the affective and psychological state of the feedback requester. This evaluation will help the learner noticing and finding the relevant feedback without being confronted to negative feedback that may worsen his affective state and negatively affects his learning.

The paper is organized as follows: an overview of some of the related work in regard to peer feedback evaluation, sentiment analysis (also known as opinion mining) for text classification is provided in the next section. This is followed by our methodology of peer affective feedback classification in section 3. The data collection, experiments setup and findings are presented in section 4 together with a discussion of our results. Section 5 concludes the paper and provides an overview of future works.

2 RELATED WORK

Our approach is novel in its consideration of peer affective feedback. Nonetheless, it is related to many previous works in peer feedback evaluation and classification techniques.

2.1 Peer Feedback Quality Evaluation

In the literature, there are several perspectives on peer feedback quality evaluation. A first perspective originates from peer reviews. In this context, Nandi et *al.*, (2012) proposed a framework with a set of criteria for feedback evaluation including social cues and feedback consistency. Nonetheless, they focused in their evaluation framework on feedback type rather than the feedback content. Similarly, Rabbany et *al.*, (2014), analyzed both the content and the structure of learners' feedback using social analysis techniques including community mining. Their purpose behind analyzing learners' feedback is to collect data and statistics about discussed topics, so they used a set of different criteria to evaluate feedback quality.

Although these approaches are efficient at evaluating the content of the feedback, they do not consider the affective aspects in feedback and are not tailored for the context of peer affective feedback.

2.2 Opinion Mining for Quality Prediction

Opinion mining, also called sentiment analysis, focuses on the polarization of opinion: positive, negative or neutral, and is generally used in product reviews (Siering and Muntermann, 2013). In this context, Lu et *al.*, (2010) incorporated social context features to predict reviews quality by investigating the reviewer's identity and his social connections or relationships. Similarly, Lu and Law, (2010) used a set of human observed features to distinguish helpful reviews from unhelpful ones from online consumers' reviews of different products. In educational context, Xiong and Litman, (2011) focused on peer review helpfulness in writing and claimed that the combination of different types of features was useful for helpfulness prediction for product reviews as well as peer reviews.

Even so several properties distinguish peer affective feedback from peer reviews and other types of reviews; we will draw upon these studies on peer reviews to tailor their utility on peer affective feedback quality evaluation and helpfulness prediction.

3 PREDICTING PEER AFFECTIVE FEEDBACK HELPFULNESS

The affective feedback that learners seek can refer to their mastery goal such as their performance of a new language or their self-improvement goal. When the learner asks for peer affective feedback, he is required to self-report his affective or psychological state (frustrated, demotivated, bored, anxious, etc.) to help peers give as useful feedback as possible. However, in response to his request, he may receive various feedback, which can be positive (if advising, encouraging and expressing concern or empathy) or negative (if ridiculing or bullying him)—see Table 1.

Because our goal of classification is to help learners who experience negative emotions to notice and benefit from positive feedback before having to confront negatives ones, we focus only on a set of negative emotions that require providing affective support, from a learning perspective (Fishbach et *al.*, 2010). This set contains the most common negative emotions in a learning context, such as *boredom*, *frustration*, *anxiety* and *demotivation*.

It is important to note that according to several motivation theories affective comments that provoke positive feelings help boost student interest, motivation, and self- efficacy, even when they are not task-focused or informative (Fishbach et *al*., 2010).

Furthermore, novices are concerned with evaluating their engagement and they are more likely to adhere to a goal after receiving positive (versus negative) feedback. Based on their findings we assume that it is simpler to classify peer affective feedback as either "positive" or "negative" when dealing with feedbacks given to novices only. This helps us to select our dataset collection in order to evaluate our approach whose details are described in the next section.

To attain our classification goal, we need to extract the polarity of peers messages and their opinions from the feedback they provide. To do that, we have relied on machine learning algorithms to classify peers' affective feedback. Hence, we first consider the representation of a given peer affective feedback as input to the machine learning algorithm. We use natural language processing techniques to automatically represent each peer feedback as a vector of text attribute values.

Before converting the text feedback to data, there are many *preprocessing steps* that should be applied. The first step is the *tokenization,* which serves to break up the feedback into tokens that correspond to words in our analysis. Then, stop words are removed to reduce the *feature dimensionality*. The next step is called *stemming* which refers to reducing words to their stem or root. However, in the context of affective feedback, we specifically seek adjectives that describe the affective and psychological state of the learner to automatically extract this state using tagging tools.

The next step is to compute the frequencies of the different words of the feedback and use the result vector as a representation of the feedback that refers to using *bag of words* as *linguistic model*. In addition to that, we will focus on *bi-grams* extracted from the peer feedback that appears frequently together. The idea behind this choice is that bi-grams, in our context, may be more indicative than separate words.

Apart from considering the message written by the learners, contextual features, such as the affective or psychological state that initiated the feedback request, must be considered. Indeed, since in the context of affective feedback, no peer comment can be classified as always positive regardless of the learner affective and psychological state, its consideration is indispensable for the sentiment analysis task.

After feedback preprocessing steps, the word-feedback matrix is established. In our data context, the lines are peers' feedback and the columns are words and bi-grams (called also features) together

Table 1: Example of peer feedback.

Feedback request	Peer feedback	
	Positive	Negative
"Whenever i speak to native (English speaker), I feel very frustrated and i'll start to stammer. The phrasing, sentence structure & grammar of my sentences become all in a mess."	Peer 1: Find an international community of people who all have English as their second language, and who don't have the same native language as you. Then, among people who also do a lot of mistakes you'll not feel frustrated you will be more confident, and will start to concentrate on ideas that you want to express, not on mistakes that you do	Peer 2: Chinese stammering…I am sorry for you interlocutor Peer 3: you make me laugh… Peer 4: English is important, but it is not essential

with the learner negative affective or psychological state that initiates the feedback request.

As for classification model, machine learning algorithms, both supervised and unsupervised, could be applied for this task such as Naïve Bayes (NB), Support Vector Machine (SVM), *k*-Nearest Neighbors (*k*-NN), Decision Trees (C4.5), Association Rules, etc.

The mining of peers' affective feedback using natural language processing techniques and machine learning algorithms poses generally several challenges. To face these challenges, we evaluate different options and choices throughout the data, the features and the classification algorithms while testing our approach.

4 TESTING AND VALIDATION

When seeking possible resources of peers' feedbacks regarding learner emotions, a good source would be online discussion forums where users express themselves frequently and spontaneously to get feedbacks from others users. It is very common for second language speakers to experience negative affective or psychological states, such as frustration, anxiety and demotivation, when it comes to using what they have learned or evaluating their learning level.

With this in mind, we focused on discussion forums for English learning. In fact, getting over anxiety, frustration and demotivation caused by the language learning has been the subject of many discussions among peers. The main steps of the process of feedback classification can be categorized into four main stages: *data collection, preprocessing*, *feature selection* and *learning*.

To create and put to the test our model, we used Rapid Miner (Prekopcsák et *al.*, 2011) which allows us to experiment numerous families of machine learning classifiers.

Data collection

A collection of 300 feedback requests with affective and psychological context was gathered from different English learning forums. The focus was on 4 most reported affective and psychological states in English forum discussion: *frustrated*, *demotivated*, *anxious* and *bored.*

Additionally, 30 graduate students (15 females and 15 males) whose first language is not English were recruited to fill out a survey. The survey took place from December 2013 to January 2014 and included two sections: the first one was *providing feedback* to ten different requests posted by learners on English forums (as illustrated in Table 1). The second section was *labelling of the feedback* given by others peers as positive or negative. In this context, a neutral feedback carrying no explicit negative words was considered as positive. The agreement between raters is moderate (FleissKappa 0.58). This labelling serves to the classifier training phase and test.

Data Preprocessing

Before applying a data mining algorithm, the data have to be preprocessed. Feedbacks shorter than 3 words are removed. We use natural language processing techniques to automatically represent each peer feedback as a vector of text attribute values. The first step in the data preprocessing is parsing and removing stop words. A set of frequently occurring words (also called tokens) are then collected from each feedback. This process is called tokenization which is based on punctuation and spaces to separate tokens. Each token is then converted to its morphological format to reduce the space of words or features.

The statistics obtained from the computational linguistic process are then used to build the model for the sentiment analysis task.

For example, the text processing of the message feedback given in the Table 1 gives the word vector as follow:

<*Speak, native, English, speaker, frustrated, start, stammer, phrasing, sentence (2) structure, grammar, become, mess*. >

Even though this may increase the feature dimensionality, the use of a feature selection method would mitigate this problem.

Feature Selection

In a text, there may be sets of words that always go together. Going back to our example (see Table 1), the pair *feel* and *frustrated* appear together two times in the same posting. Identifying such pairs or bi-grams allows us to reduce the features dimensionality. This feature selection technique allows us also to parse each feedback message and capture the sets of significant words in our context of peer affective feedback.

Learning

The last step in the classification process is to apply the desired machine-learning algorithms to obtain a classifier. Several algorithms were used and compared: J48 implementation of C4.5 decision-trees, Naïve-Bayes, Association rules with Naïve-Bayes, and K-Nearest Neighbors. We have chosen Decision trees, Naïve Bayes and K-NN because they are widely used in classification task especially in sentiment analysis. As for Association rules, we have chosen to experiment this algorithm because the discovery of interesting association relationships among words containing the feedbacks can help us classifying peer affective feedbacks.

Testing

In order to effectively use our limited data, we used *k*-fold cross validation for all experiments with different values of *k* to evaluate the performance (*k*=3, *k*=5). We have tried different value of *k*-fold cross validation to examine the results for each configuration. In *k*-fold cross validation, the training set are randomly divided into *k* samples where a single sample is retained for testing the model and the remaining *k-1* samples serve for the model training. The validation process is then repeated *k* times where the *k* samples are used only once for validation. The final validation result is obtained by averaging the results of the *k* folds. We have chosen this validation method because all examples are used for training and validation where each example is used only once for validation. This helps avoid making decisions that give good results on training data but do not generalize well.

In order to minimize the number of misclassification on the training dataset, we ran a series of experiments with different classification configurations. The findings of these experiments are reported in the next subsection.

Results

We first considered the bag of words representation of a given peer feedback as input to the mining algorithm. It consists in simply computing the frequencies of the different words in a given feedback and uses the result vector as input to feed the mining algorithm. Here we also focus on bi-grams extracted from the peer feedback. The idea behind this choice is that bi-grams in our context are more indicative than separate words.

The evaluation of single label classifiers is generally conducted using classic metrics, such as *Precision*, *Recall* and *F-measure* (Pang et *al.*, 2002). *Prediction accuracy* is the selected metric for evaluating the performance of our model since our goal is to obtain a classifier that generalizes well. The results obtained for different classifiers are shown in Table 2. The accuracies thus obtained differ from one classifier to another and depend on the classification setting, such as whether the linguistic model uses bags of words or bi-grams, etc.

Based on the results highlighted in Table 2, Naïve-Bayes provides the best accuracy of 87.19% when using a bi-grams model. We can say that we have taken a best choice by focusing on bi-grams in peer affective feedback classification.

Table 2: Final results applying machine learning algorithms.

Algorithm	Settings	Accuracy (%)	
		Bags	Bi-grams
Naïve-Bayes		86.11	87.19
k-NN	*k* = 3	67	60.51
k-NN	*k* = 5	55.51	49.49
C4.5	confidence = 0.25	78.50	79.51
Association Rules	support = 0.1 confidence = 0.8	55.51	65.76

The obtained accuracy is a good result with respect to the sentiment analysis literature that have found results between 80-87% when classifying movie reviews (Pang et *al.*, 2002) and (Martínez-Cámara et *al.*, 2011) with an accuracy of 82.90% and 86.84, respectively, using SVM. Naïve-Bayes classifies 87.19% of examples correctly at the cost of a loss of 0.66% of good corrections. The confusion matrix of this algorithm, illustrated in Table 3, shows the classification details.

Table 3: Confusion matrix of Naïve Bayes.

Predicted class	Actual class		
	Positive	Negative	Class precision (%)
Positive	130	17	88.44
Negative	19	115	85.82

We believe that the accuracy we have found is good and promising considering the context of peer affective feedback and the training data collected from discussion forums. These platforms are generally considered as very noisy because messages exchanged between peers are generally informal and contain many mistakes, as well as *emoticons* and symbols. This characteristic of these environments makes the data preprocessing and the sentiment analysis particularly challenging.

On other hand, our work is different from existing works which focused on predicting the helpfulness of peer reviews because it takes into consideration especially the learner affective and psychological state in the sentiment analysis task. In fact, unlike the others works, we believe that it is not sufficient to consider only the feedback message when dealing with affective and psychological factors which affect the learning process. In our work, we do not only classify a feedback as positive or negative we also predict the helpfulness of a peer feedback given the emotional or psychological state of the learner who asked for it. The obtained high accuracy shows that it is possible to successfully predict if a peer feedback is helpful or unhelpful for a given student. This finding will allow us to adapt the learner's interactions to his affective and psychological state in order to promote his learning, which is the ultimate goal of e-learning systems.

5 CONCLUSION

In this paper, we propose an approach to predict the helpfulness of a given feedback for a learner based on the feedback content and the learner's affective state. To do this, we use natural language processing techniques and machine learning algorithms by combining linguistic and contextual features such as the learner's affective state. In our experiment, we show that Naïve-Bayes performs well using bi-grams and classified correctly 87.19% of examples. In addition, we show that the accuracy of different machine learning approaches experimented depends upon classification features such as the linguistic model. In this context, we have provided a proof of concept using only 300 peers' feedback as training data, which is *insufficient* compared to what is needed for the task of opinion mining. Nonetheless, the findings of our approach remain valid and could be improved in future works with the collection of more data and feedback evaluation from the learners.

In this work, we use most of the words that appear in peer affective feedback to prove that the classification and quality prediction may help the learners notice and benefit from positive feedback while avoiding negative ones. Further experiments to study this dependence relationship will be conducted in future works.

Other factors will also be studied in future directions such as peers' expertise as it may help predict the feedback quality. Finally, we will investigate further the impact of classifying and helping learners notice relevant feedback and whether or not this affects positively their learning.

REFERENCES

Robinson, J., McQuiggan, S. and Lester, J., 2009. Evaluating the consequences of affective feedback in intelligent tutoring systems. *in Affective Computing and Intelligent Interaction and Workshops*, 2009. ACII.

Ortigosa, A., José, M. M., Carro R.M., 2014. Sentiment analysis in Facebook and its application to e learning. *Computers in Human Behavior* 31: pp.527-541.

Lu, J., Law, N. 2012. Online peer assessment: effects of cognitive and affective feedback. *Instructional Science*, 40(2): pp. 257-275.

Walker, E., Rummel, N., Walker, S., Koedinger, K.R., 2012. Noticing relevant feedback improves learning in an intelligent tutoring system for peer tutoring. in *Intelligent Tutoring Systems*. Springer.

Selmi, M., Hage, H., Aïmeur, E., 2013. Privacy framework for peer affective feedback, *in Proceeding of the 9th International conference on Signal Image technology & Internet based Systems* (SITIS 2013), pp. 1049-1056. IEEE.

Nicol, D.J., Macfarlane-Dick, D., 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2): pp. 199-218.

Nandi, D., Hamilton, M., Harland, J., 2012. Evaluating the quality of interaction in asynchronous discussion forums in fully online courses. *Distance Education*, 33(1): pp. 5-30.

Rabbany, R., Elatia, S., Takaffoli, M., Zaiane, A.R., 2014. Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective, *in Educational Data Mining*. Springer. pp. 441-466.

Siering, M., Muntermann, J., 2013. What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions. *in Wirtschaftsinformatik*.

Lu, Y., Tsaparas, P., Ntoulas, A., Polanyi, L. 2010., Exploiting social context for review quality prediction. *in Proceedings of the 19th international conference on World wide web*. ACM.

Xiong, W., Litman, D., 2011. Understanding differences in perceived peer-review helpfulness using natural language processing. *in Proceedings of the 6th Workshop on Innovative Use of NLP for Building Educational Applications*. Association for Computational Linguistics.

Fishbach, A., Zhang, Y., Trope, Y., 2010. Counteractive evaluation: Asymmetric shifts in the implicit value of conflicting motivations. *Journal of Experimental Social Psychology*, 46(1): pp. 29-38.

Prekopcsák, Z., Makrani, G., Henk, T., Gaspar, C., 2011. Radoop: Analyzing big data with rapidminer and hadoop. *in Proceedings of the 2nd RapidMiner Community Meeting and Conference* (RCOMM 2011).

Pang, B., Lee, L., Vaithyanathan, S.: Thumbs up? 2002. sentiment classification using machine learning techniques. *in Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, 10. Association for Computational Linguistics.

Martínez-Cámara, E., Martín-Valdivia, M.T., Ureña-López, L.A., 2011. Opinion classification techniques applied to a spanish corpus, *in Natural Language Processing and Information Systems*. Springer. pp. 169-176.