Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning

Christos Troussas, Maria Virvou
Department of Informatics
University of Piraeus
Piraeus, Greece
{ctrouss, mvirvou}@unipi.gr

Kurt Junshean Espinosa, Kevin Llaguno, Jaime Caro
Department of Computer Science
University of the Philippines
Cebu city, Philippines
kpespinosa@up.edu.ph,
kevinllaguno.cmsc110@gmail.com,
jdlcaro@dcs.upd.edu.ph

***Abstract*—The growing expansion of contents, placed on the Web, provides a huge collection of textual resources. People share their experiences, opinions or simply talk just about whatever concerns them online. The large amount of available data attracts system developers, studying on automatic mining and analysis. In this paper, the primary and underlying idea is that the fact of knowing how people feel about certain topics can be considered as a classification task. People's feelings can be positive, negative or neutral. A sentiment is often represented in subtle or complex ways in a text. An online user can use a diverse range of other techniques to express his or her emotions. Apart from that, s/he may mix objective and subjective information about a certain topic. On top of that, data gathered from the World Wide Web often contain a lot of noise. Indeed, the task of automatic sentiment recognition in online text becomes more difficult for all the aforementioned reasons. Hence, we present how sentiment analysis can assist language learning, by stimulating the educational process and experimental results on the Naive Bayes Classifier.**

Keywords— affective interaction; facebook; naive bayes classifier; perceptron classifier; rocchio classifier; sentiment analysis

I. INTRODUCTION

Sentiment analysis is becoming a popular study these days, mainly because of the fact that social networking sites include online users who are free to express their thoughts, feelings and impressions concerning a specific topic. In fact, nowadays, any kind of marketing business is currently immersing to the new trends of businesses. Apart from written surveys, the companies also extend their customer satisfaction analysis through the web, in order to gather a large amount of data.

Few studies on sentiment analysis have already been presented. These studies are targeted to Twitter, for tweet updates about a specific topic, mostly on brands of products [1]. These systems collect raw data from twitter, using hash tags, like #exampletopic, and use the data as a corpus to be feed upon implementing the classifying method. However, gathering and analyzing data from the social networking site like Twitter has one known downside. Every twitter update is restricted to 140 characters in length [1]. For this reason, Twitter users tend to use heavy abbreviations and fragmented expressions.

The social networking site Facebook will be the targeted website for this paper. This is because Facebook, unlike Twitter, has 5000 characters for every status update [14]. For this reason, a clearer sentence construction would be more expectable. Moreover, the number of Facebook users is absolutely abundant, namely it is a good sample for creating a corpus. Concerning that sample, the determination of the polarity of the people's opinion would be quite interesting.

Sentiment analysis is involved in the study of opinion mining. Furthermore, sentiment analysis is commonly used by advertisers, movie creators and other organizations that wish to acquire their customers' reaction on a specific topic. Although the simplest way to gather opinions is in the form of surveys, there are few drawbacks, which consist of great handicaps of the marketing research. The problems emerging of this approach are the conduct of a survey for each product or feature, the format, the distribution and timing of the survey, and the reliance on the good will of people to take the survey. All the aforementioned problems need high maintenance for the marketing research group's view.

Opinion mining in sentiment analysis also faces few challenges for the system developers' perspective. In the case of opinions, not all words used in the sentence have significance. Some words are classified as noise because they are of no use in the process of classifying the polarity of the opinion. Also, there are words like "not" and whenever they are added to a positive word, they will attach a negative meaning to the existent opinion. Aside from words, symbols like sad face "☹" or a happy face "☺" present significance in natural language processing. Hence, there are not only the words, which take place in the observation procedure. Another challenge of this study is the way of gathering/collecting data for corpus. Corpus, by definition, is the collection of writings or recorded remarks used for linguistic analysis, like sentiment analysis. One way is to use existing corpus, but the common existing corpus is a movie corpus.

In view of the above, we have developed a system that is able to classify an opinion using sentence-level classification whether it entails positive, negative or neutral emotions. Opinions are in the form of status updates in the social networking site Facebook. The specific objectives of our study are:

- To develop our own corpus through a Facebook application.
- To properly train the system to accept inputs in the form of status updates from the corpus, disregarding updates that do not contain words or face emoticons.

- The ability of the system to classify the polarity of an opinion per status update basis, during the testing phase,

Our Facebook application, which is used to gather data for the corpus, can be used by Facebook users worldwide. Corpus will be limited in quantity, 5000 for positive sentiments and the other 5000 for the negative sentiments.

- The study will focus on the users' Facebook status updates.
- The study will not include Facebook posts like photo stories, application stories or other similar stories.
- Facebook is used in our study, as the main user interface.

The rest of the paper is organized as follows. In section 2, we present the related scientific literature concerning the sentiment analysis in Social Networking Services. In section 3, we describe the architecture and the methodology of our system. In section 4, we present the sentiment analysis for language learning in Facebook. In section 5, we present our experimental results and a discussion about them. Finally, in section 6, we present the conclusions and our future work.

II. Related Work

Sentiment analysis has been handled as a Natural Language Processing task at many levels of granularity. Starting from being a document level classification task, it has been handled at the sentence level and more recently at the phrase level. There are many established methods for sentiment analysis at the sentence and paragraph level.

In [8], the authors discussed the application of support vector machines in sentiment analysis with diverse information source.

In [11], the authors applied minimum cuts in graphs to extract the subjective portion of texts they were studying and used machine learning methods to perform sentiment analysis on those snippets of texts only.

In [13], the authors discussed categorizing texts into polar and neutral first before determining whether a positive or negative sentiment is expressed through the text. However, in [7], the authors operate on the premise that little neutrality exists in online texts.

In [7], the authors developed techniques that algorithmically identify large number (hundreds) of adjectives, each with an assigned score of polarity, from around a dozen of seed adjectives. Their methods expand two clusters of adjectives (positive and negative word groups) by recursively querying the synonyms and antonyms from WordNet. Since recursive search quickly connects words from the two clusters, they implemented several precaution measures such as assigning weights which decrease exponentially as the number of hops increases. This confirms that the algorithm-generated adjectives are highly accurate by comparing them to the results of manually picked word lists. It is worth pointing out that this work uses Lydia as the backbone to process large amount of news and blogs.

In [3], the authors provided a good survey of various techniques developed in online sentiment analysis. It covers concept of emotion in written text (appraisal theory), various methodologies which can be broadly divided into two groups: (i) symbolic techniques that focuses on the force and direction of individual words (the so-called "bag-of words" approach), and (ii) machine learning techniques that characterizes vocabularies in context. Based on the survey, the authors found that symbolic techniques achieves accuracy lower than 80% and are generally poorer than machine learning methods on movie review sentiment analysis. Among the machine learning methods, they considered three supervised approaches: Support Vector Machine (SVM), Naive Bayes Multinomial (NBM), and maximum Entropy (Maxent). They found that all of them deliver comparable results on various feature extraction (unigrams, bigrams, etc) with high accuracy at 80%~87%.

Another significant effort for sentiment classification on Twitter data is conducted by [2]. The authors use polarity predictions from three websites as noisy labels to train a model and use 1000 manually labeled tweets for tuning and another 1000 manually labeled tweets for testing. They however do not mention how they collect their test data. They propose the use of syntax features of tweets like retweet, hash tags, link, punctuation and exclamation marks in conjunction with features like prior polarity of words and POS of words.

In [5], the authors perform sentiment analysis on feedback data from Global Support Services survey. One aim of their study is to analyze the role of linguistic features like POS tags. They perform extensive feature analysis and feature selection and demonstrate that abstract linguistic analysis features contributes to the classifier accuracy.

In [6], the authors use distant learning to acquire sentiment data. They use tweets ending in positive emoticons like ":)" ":-)" as positive and negative emoticons like ":(" ":-(" as negative. They build models using Naive Bayes, MaxEnt and Support Vector Machines (SVM), and they report SVM outperforms other classifiers. In terms of feature space, they try a Unigram, Bigram model in conjunction with parts-of-speech (POS) features. They note that the unigram model outperforms all other models. Specifically, bigrams and POS features do not help.

In [10], the authors take a naive approach to collect and classify 300000 tweets into three categories: (i) tweets queried with emoticon queries such as ":-)", ":)", "=)" indicate happiness and positive emotion (ii) tweets with ":-(", ":(", "=(", ";(" implies dislike or negative opinions, and (iii) tweets posted by newspaper accounts such as "New York Times" are considered objective or neutral. This serves as the training set for Naive Bayes Multinomial (NBM), which they found to be superior to Support Vector Machine (SVM) and Conditional Random Field (CRF) as the classifier to unigrams, bigrams, and trigrams. The result indicates that bigrams provides the best accuracy.

However, after a thorough investigation in the related scientific literature, we came up with the result that there is not any sentiment analysis and recognition using the Naive Bayes classifier for language learning in Facebook. Most of the aforementioned approaches, however, are primarily based on ngram models. Moreover, the data they use for training and testing is collected by search queries and is therefore biased. In contrast, we present features that achieve a significant gain over a unigram baseline. In addition we explore a different method of data representation and report significant improvement over the unigram models. Our data are a random sample of streaming Facebook statuses unlike data collected by using specific queries. The size of our hand-labeled data allows us to perform cross validation experiments and check for the variance in performance of the classifier across folds.

III. Architecture And Methodology

The main methodology for Sentiment Analysis is the Classifier method specifically the Naive Bayes Classifier Method where in a status update is being classified as positive or negative.

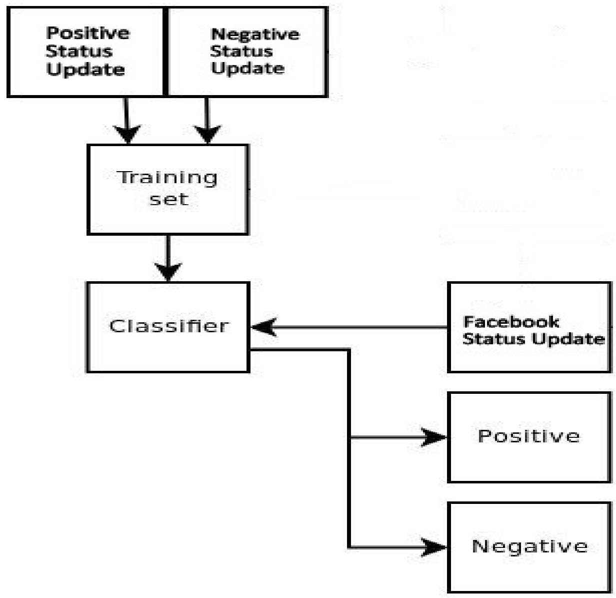

Figure 1. Main methodology of Naive Bayes Classifier

Fig. 1 shows the overview of sentiment analysis using Naive Bayes Classifier.

A. Naive Bayes Classifier

Bayesian classifiers are based around the Bayes rule, a way of looking at conditional probabilities that allows you to flip the condition around in a convenient way. A conditional probably is a probably that event X will occur, given the evidence Y. That is normally written P(X | Y). The Bayes rule allows us to determine this probability when all we have is the probability of the opposite result and of the two components individually: P(X | Y) = P(X) P(Y | X) / P(Y). This restatement can be very helpful when we are trying to estimate the probability of something based on examples of it occurring.

In this case, we are trying to estimate the probability that a document is positive or negative, given its contents. We can restate that, so that is in terms of the probability of that document occurring if it has been predetermined to be positive or negative. This is convenient, because we have examples of positive and negative opinions from our data set above.

The point of view that renders this process a "naive" Bayesian one is that we make a large assumption about how we can calculate the probability of the document occurring; it is equal to the product of the probabilities of each word within its occurrence. This implies that there is no link between one word and another word. Independence assumption it is called.

We can estimate the probability of a word occurring, given a positive or negative sentiment by looking through a series of examples of positive and negative sentiments and counting how often it occurs in each class. This is what makes this supervised learning, the requirement for pre-classified examples to train on.

B. Corpus Creation

Corpus by definition is a collection of writings or recorded remarks used for linguistic analysis. In this application, recorded remarks are classified into two groups, the negative and positive feeling in Facebook users' status updates. Range of 5000 – 10000 status updates will be the targeted number for corpus. It will be evenly divided for two classes, negative and positive. It is recommended that corpus is large in number and for this reason the number of 5000 data appears to provide very satisfactory results.

The Facebook Query Language (FQL) object enables running FQL queries using the Graph API. FQL enables people to use a SQL-style interface to query the data exposed by the Graph API. It provides some advanced features not available in the Graph API, including batching multiple queries into a single call. Using this developer features, data will be collected from Facebook users based on the records in Facebook. The system will be trained on the sentiments of the people, to whom our Facebook language learning application, presented in [12], is addressed. The collected data will be manually identified whether it is positive or negative. Positive status updates will then be stored in a class, just as the negative ones.

C. Classification

A conditional probability is a probability that event X will occur, given the evidence. So, our initial formula looks like this:

P(sentiment | sentence) = P(sentiment) P(sentence | sentiment) / P(sentence)

We can drop the dividing P(line), as it's the same for both classes, and we just want to rank them rather than calculate a precise probability. We can use the independence assumption to let us treat P(sentence | sentiment) as the product of P(token | sentiment) across all the tokens in the sentence. So, we estimate P(token | sentiment) as:

count(this token in class) + 1 / count(all tokens in class) + count(all tokens)

The extra 1 and count of all tokens is called "add one" or Laplace smoothing and stops a 0 finding its way into the multiplications. If there was not any sentence with an unseen token in, it would score zero.

The classify function starts by calculating the prior probability (the chance of it being one or the other before any tokens are looked at) based on the number of positive and negative examples; in our example, that will always be 0.5, as for each observation (positive/negative status update), there are the same amount of data. We then tokenize the incoming document and for each class multiply together the likelihood of each word being seen in that class. We sort the final result and return the highest scoring class.

Our study classifies the polarity of the context/status update in a sentence level. Sentence level, in most cases, is more accurate than the phrase level because every status update has its own style in addressing users' sentiment.

IV. SENTIMENT ANALYSIS IN LANGUAGE LEARNING

Emotions are complex states of mind and body, consisting of physiological, behavioral, and cognitive reactions to situations that can be managed and directed. Cognitively, individuals interpret an event as one that may be sad, dangerous or happy. Behaviorally, a student may seek comfort when s/he is sad or run and seek help when s/he faces danger. It is critical to recognize the crucial link between emotions, thought, and action. Our emotional state has the potential to influence our thinking [4]. For example, students learn and perform more successfully when they feel secure, happy and excited about the subject matter [9]. Although emotions have the potential to energize students' thinking, emotional states also have the potential to interfere with learning. If students are overly excited or enthusiastic, they might work carelessly or quickly rather than working methodically or carefully [4]. Moreover, sentiments such as anger, anxiety and sadness have the potential to distract students' learning efforts by interfering with their ability to involve in the educational process successfully. Sentiments can interfere with students' learning in several ways, including limiting the capacity to balance emotional issues with tutoring. Some students may have difficulty in learning because their minds are cluttered with distracting thoughts and memories. For example, a student who is distressed might be thinking so much about a bad memory that little mental ability is left to think about other matters. If students are working to cope with emotions, they might not have sufficient resources available to engage in learning. In these situations, students may need extra prompts to be helped with learning [4]. Some students might need a reminder to help them remain focused and to redirect their attention to the Facebook educational application. Some students might need one-on-one time with their peers, which can be achieved by instant or asynchronous text messaging in Facebook, in order to help the process of their feelings or the resolution of a problem.

Sentiment analysis and recognition may be incorporated into sophisticated educational applications in Social Networking Services, such as Facebook, by providing adaptive interaction based on the user's emotional state. Regardless of the various emotional paradigms, neurologists and psychologists have made progress in demonstrating that sentiment is very significant in the educational process, as students are not susceptible of learning throughout the whole time of being tutored. Hence, the way that people feel may play an important role in their cognitive processes as well. The major challenge in sentiment analysis is the effort to improve the accuracy of recognizing people's emotions. Ideally, evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' sentiments. It is hoped that the multimodal approach may provide not only better performance, but also more robustness.

In view of the above, we focused on measuring the emotional level of each user and then according to this level and his/her personal user model, we provide him/her with advice concerning the ability to start or proceed with the language learning application. Fig. 2 illustrates how sentiment analysis can be involved in the educational process.

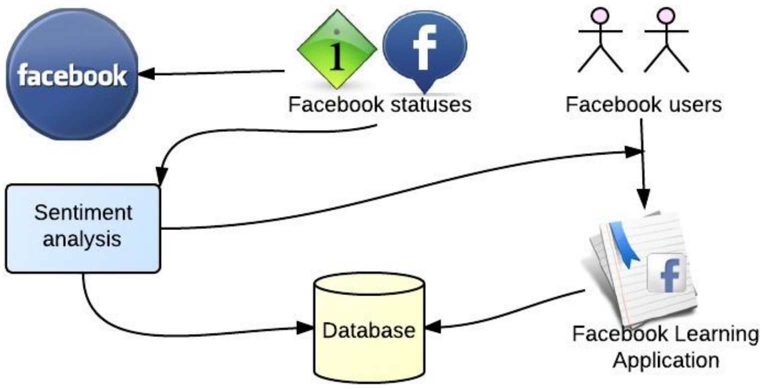

Figure 2. General Architesture

V. EXPERIMENTAL RESULTS AND DISCUSSION

In this study, we used three classifiers, namely Naive Bayes, Rocchio and Perceptron, in order to compare their performance in predicting whether a Facebook status update is positive or negative. We collected around 7000 status updates from 90 users. The status updates were then manually labeled as positive or negative. The following table contains sample of status updates in each class.

TABLE I. SAMPLE OF STATUS UPDATES

Sample of negative status updates:	Sample of positive status updates:
Freaking full of doubt.	Just finished making pancakes for breakfast.. oh and the yummiest part, it comes with a free strawberry syrup!
i really don't like to shave my hair but i have to. :/ frustrated :(	inspired by you! <3
can't sleep...	11 days to go before

	Christmas :)

Since there were a lot fewer negative samples, we based the distribution of the final dataset from it. We used the following data distribution for training and testing set (50%-50%):

TABLE II. DATA DISTRIBUTION

	Training	Testing
Positive	1131	1131
Negative	1131	1131

The dataset for each partition was randomly selected. The same training and testing set were used for each classifier. The classifiers were compared in terms of these three metrics: precision, recall and F-score performance using the computations shown below.

$$\mathrm{Precision} = \frac{tp}{tp + fp} \quad (1)$$

$$\mathrm{Recall} = \frac{tp}{tp + fn} \quad (2)$$

$$F - \mathrm{score} = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}} \quad (3)$$

The following tables summarize the results.

TABLE III. NAIVE BAYES PRECISION AND RECALL PERFORMANCE

Naive Bayes Classifier	Actual Positive	Actual Negative
Predicted Positive	0.76	0.23
Predicted Negative	0.35	0.65

TABLE IV. ROCCHIO PRECISION AND RECALL PERFORMANCE

Rocchio Classifier	Actual Positive	Actual Negative
Predicted Positive	0.73	0.24
Predicted Negative	0.27	0.76

TABLE V. PERCEPTRON PRECISION AND RECALL PERFORMANCE

Perceptron Classifier	Actual Positive	Actual Negative
Predicted Positive	0.65	0.35
Predicted Negative	0.52	0.48

The following table compares the precision, recall, and the F-scores of each classifier.

TABLE VI. PRECISION, RECALL AND F-SCORE COMPARISON OF THE THREE CLASSIFIERS

	Naive Bayes Classifier	*Rocchio Classifier*	*Perceptron Classifier*
Precision	0.77	0.75	0.65
Recall	0.68	0.73	0.56
F-score	0.72	0.74	0.60

Based on the F-score, Rocchio classifier has the best performance and Perceptron the least. Naive Bayes performed almost as well as Rocchio, but has a significantly lower recall than the latter.

VI. CONCLUSIONS AND FUTURE WORK

In this paper, we present the sentiment analysis for language learning using Naive Bayes Classifier. Furthermore, we present important features that achieve a significant gain over a unigram baseline. In addition, we explore a different method of data representation and report significant improvements over the unigram models. Our data is a random sample of streaming Facebook statuses and were not collected by using specific queries. The size of our hand-labeled data allows us to perform cross validation experiments and check for the variance in performance of the classifier across folds. Finally, we present our experimental results, which show that the accuracy in analyzing the sentimental state of Facebook users, using the Naive Bayes Classifier, is really high.

It is in our future plans to perform deeper study on the sentiment analysis and recognition of the Facebook statuses in order to further promote the language learning procedure. Hence, the emotional state of a Facebook user will be examined in depth and will consist of a crucial factor of his/her learning ability.

REFERENCES

[1] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment analysis of Twitter data", LSM '11 Proceedings of the Workshop on Languages in Social Media, Association for Computational Linguistics, pp. 30-38, 2011.

[2] L. Barbosa, J. Feng, "Robust sentiment detection on twitter from biased and noisy data", Proceedings of the 23rd International Conference on Computational Linguistics, pp. 36–44, 2010

[3] E. Boiy, P. Hens, K. Deschacht, M. - F. Moens, "Automatic Sentiment Analysis in On-line Text", Proceedings of the 11th International Conference on Electronic Publishing, pp. 349-360, 2007.

[4] L. Darling-Hammond, S. Orcutt, K. Strobel, E. Kirsch, I. Lit, D. Martin, "Feelings Count: Emotions and Learning", Stanford University School of Education, pp. 89-104.

[5] M. Gamon, "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis", Proceedings of the 20th international conference on Computational Linguistics, pp. 841–847, 2004.

[6] A. Go, R. Bhayani, L. Huang, "Twitter sentiment classification using distant supervision", Technical report, Stanford, 2009.

[7] N. Godbole, M. Srinivasaiah, and S. Skiena, "Large-scale sentiment analysis for news and blogs", Proceedings of International Conference on Weblogs and Social Media, 2007.

[8] T. Mullen, N. Collier, "Sentiment analysis using support vector machines with diverse information sources", Proceedings of Conference on Empirical Methods in Natural Language Processing, pp. 412-418, 2004.

[9] K. Oatley, S. Nundy, "Rethinking the role of emotions in education", D. R. Olson & N. Torrance (Eds.), The handbook of education and human development: New models of learning, teaching and schooling, pp. 257-274, 1996.

[10] A. Pak, P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining”, Proceedings of the Seventh International Conference on Language Resources and Evaluation, 2010.

[11] B. Pang, L. Lee, “A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts”, ACL '04 Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics, pp. 271-278, 2004.

[12] M. Virvou, C. Troussas, J. Caro, K. J. Espinosa, “User modeling for language learning in Facebook”, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 15th International Conference on Text, Speech and Dialogue, Volume 7499 LNAI, pp. 345-352, 2012.

[13] T. Wilson, J. Wiebe, P. Hoffman, “Recognizing contextual polarity in phrase level sentiment analysis”, Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 347-354, 2005.

[14] Inside Facebook: http://www.insidefacebook.com/2011/09/21/5000-character-limit-floating-navigation-bar/