Analysis of sentiment change over time using user status updates from social networks

Milica Ćirić*, Aleksandar Stanimirović*, Leonid Stoimenov*
* Faculty of Electronic Enginneering, University of Niš, Niš, Serbia
milica.ciric@elfak.ni.ac.rs, aleksandar.stanimirovic@elfak.ni.ac.rs, leonid.stoimenov@elfak.ni.ac.rs

***Abstract*— Social networks' users usually post status updates of their opinions on and reactions to events almost instantly. Because of this, pre-announced events, especially ones that divide the public, lead to generation of a large quantity of status updates in the days just before and after the event. Status updates published before the event usually contain speculations about what will happen, and those published after represent the reaction to what actually happened. In this paper we analyze the change of sentiment in tweets collected in following few days after a specific pre-announced event. We also discuss can this analysis be used for prediction, as an indicator of to which extent will this event fulfill its designated long term goal.**

I. Introduction

Text sentiment classification deals with identifying the sentiment expressed in a text of certain length. Usually, the goal is to assign one of two labels: positive or negative, although there are researches than focus on subjective and objective classes [1, 2], or have multiple degrees of positivity/negativity [3, 4, 5]. As such, it can be applied in many fields, some of which include: classifying product customer feedback [3], preprocessing input data for recommender systems [6] or determining the citizens' position on government policies and decisions [7].

Besides just determining sentiment, it is also interesting to analyze sentiment change over time. The effect of a certain event to public opinion can be explored by tracking sentiment change on social networks.

Social networks are especially convenient for tracking sentiment change due to the fact that their users post status updates very often, even up to several times an hour. Also, the option for sharing other's statuses enables fast spreading of its content. One of the most popular social networking platforms is Twitter [8]. It allows its users to post messages of maximal length of 140 characters, also known as "tweets". Since the launching in July 2006, its popularity has rapidly grown and it now has over 600 million active users [9].

Tweets can contain all kinds of information, from daily events in the life of the user to live news from political events. A common form of a tweet is a reference to a person, event, product and a personal opinion about the referenced entity. The # symbol, called a hashtag, is used to mark keywords or topics in a tweet. It was created organically by Twitter users as a way to categorize messages. By analyzing tweets one could observe which are the popular entities and also the "public opinion" about them.

During the research presented in this paper we collected around 100000 tweets from a 4 day period following the release of Lady Gaga's single Applause and classified them by their sentiment. For classifying we used Naïve Bayes classifier which is one of the mathematically simplest classifiers, but gives results in the rank with much more sophisticated classifiers. Rather that developing our own implementation of Naïve Bayes classifier, we used Mallet open source library and extended it with some new features. According to the findings from a previous research [10] we used a combination of predefined and developed features that produced the best results for Twitter sentiment classification.

II. Related work

The most popular domain for sentiment classification is certainly classification of movie reviews [4, 11, 12] and product reviews in general [13, 14]. The length of reviews can range from a couple of sentences to several paragraphs, but the text is usually well formed. Another research direction concerning product reviews sentiment classification is domain adaptation [11, 14], i.e. managing the difference in domains of training and testing data.

Twitter sentiment classification is rather different than movie and product reviews, mostly due to the language style and the shortness of the tweets, but is an increasingly popular topic [1, 5, 15, 16, 17]. The upside of the informal and slang-like language is the presence of emoticons or smileys that can be used as an indicator of the sentiment the author wants to express [5, 11, 15]. Automatically assigning sentiment based on emoticons greatly facilitates the creation of training and testing corpuses. In [18] instead of classifying the sentiment of a tweet, the authors determine the general sentiment about a specified hashtag.

Even texts of small length can express sentiments about multiple entities and classifying the overall sentiment can give a false result. Therefore, some researchers focus on identifying the sentiment pointed to a specific entity, so called target-dependent sentiment classification [1].

There have been multiple researches that analyze the connection between sentiment in tweets addressing election candidates during the campaign and the outcome of the elections [19, 20, 21]. Some of them have discovered that the prediction using sentiment data would have given good results [20], while in other cases the prediction wouldn't have been accurate [21]. In [22] the authors track the behavior streams from groups of users with similar biases and present time-dependent collective measures of their opinions. They tested the effectiveness of their system by tracking groups' Twitter responses to a common stimulus set: the 2012 U.S. presidential election debates.

Authors of [23] explore how public mood patterns, as evidenced from a sentiment analysis of Twitter posts published during a 5 month period in 2008, relate to fluctuations in macroscopic social and economic indicators in the same time period. They attempt to identify a quantifiable relationship between overall public mood and social, economic and other major events in the media and popular culture. Research presented in [24] utilized a stream of millions of tweets on Twitter to explore how people feel about Thanksgiving and Christmas through real-time sentiment analysis. With help of Twitter Streaming API, the author discovered the patterns of sentiment changes by hour before and after the two holidays.

III. DATA CORPUS AND CLASSIFIER

Since Twitter is currently one of the most popular social networking platforms and we already had experience with Twitter sentiment classification [10, 25] we decided to use tweets as our data corpus. Tweets can be obtained from Twitter with the use of Search API [26] or Streaming API[27]. Search API performs a search according to the specified query and returns a collections of relevant Tweets. Search API is not meant to be an exhaustive source of Tweets and not all Tweets are indexed or made available via the search interface. Streaming API, on the other hand, gives low latency access to Twitter's global stream of Tweet data. We decided to use the Search API and perform queries in regular intervals in order to collect tweets.

In order to monitor change of sentiment over time, we needed an event that was announced and expected with anticipation. Choosing Lady Gaga and the release of her single Applause, which was heavily announced and promoted on Twitter, enabled us to get a lot of tweets during a 4 day period. Additionally, the fact that opinions about Lady Gaga and her actions are often very divided, implied that we will probably get data with strong sentiments, making them easier to classify. The idea was to collect tweets in the days before and after the release of her single Applause. However, due to a leak incident, the single was released a week earlier than announced [28], and we were only able to start the collection of tweets after the release. This is also one of the reasons we decided to use Search API, since it enables querying past data.

Search API offers a wide range of parameters that can be defined for a query. For our purpose, we searched for tweets containing key words "lady gaga", those that were directed to her (to:ladygaga) and those that were mentioning her (@ladygaga). Overall, we acquired 100944 tweets during a 4 day period following the release of the Applause single. These tweets were then classified by their sentiment.

Classification was performed with the use of Naive Bayes classifying algorithms. Naive Bayes classifier (NB) [29] is one of the simplest classifier. It is a probabilistic classifier and it is based on the Bayes theorem. The premise is that each feature of an instance is completely unrelated any other feature of the same instance, i.e. probability that an instance belongs to a certain category is influenced by each feature independently.

The probability that instance *i* belongs to category *c* can be expressed with (1).

$$P(c \mid i) \propto P(c) \prod_{1 \le k \le n_i} P(f_k \mid c) \qquad (1)$$

In (1) $P(c)$ represents the prior probability that an instance belongs to the category c, $P(f_k|t)$ represents the probability of feature f_k occurring in an instance that belongs to the class c and n_i is the number of features in instance i.

The assumption of feature independence is not very realistic for the real world and it is the reason this classifier is called naïve. However, as simple and inefficient as Naïve Bayes classifier may seem in theory, in practice it usually gives good results, often as good as much more complex classifiers. For example, in [10] in comparison of Naïve Bayes, MaxEnt and Balanced Winnow, NB has slightly worse results than MaxEnt, but much better than Balanced Winnow.

In our research we used Mallet [30], an open source Java library for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text. This gave us the opportunity to use built-in functionalities of Mallet's Naïve Bayes implementation and also extend the library to fulfill our specific needs or try a new approach.

The classifiers used were trained during the research published in a previous paper, dealing with context and sentiment classification [10]. For sentiment classification we collected tweets with positive, negative and objective sentiment. However, due to the small number of objective tweets, compared to the number of positive and negative ones, we decided to proceed with only these two categories. Collected tweets were selected according to the emoticons they contained. In order to decrease the influence of emoticons in classifying and increase the influence of the words used, we subsequently removed emoticons from about 90% of the collected tweets.

For training the classifiers we started with adding a feature for each word in the tweet and then added various data processing pipes for filtering features. These filtering steps can help in extracting features that carry more information and removing others that can be considered as noise. They are widely used in various researches [1, 5, 15, 17]. Some of the data processing pipes we used were built in Mallet library, some we were able to get by customizing existing Mallet data processing pipes and some we had to implement ourselves. Analyzing results gotten with different combinations showed that the best results for sentiment classification were produced when using the following pipes:

- Lowercase normalization
- Adding some application specific features (e.g. sport results, music ranking, …)

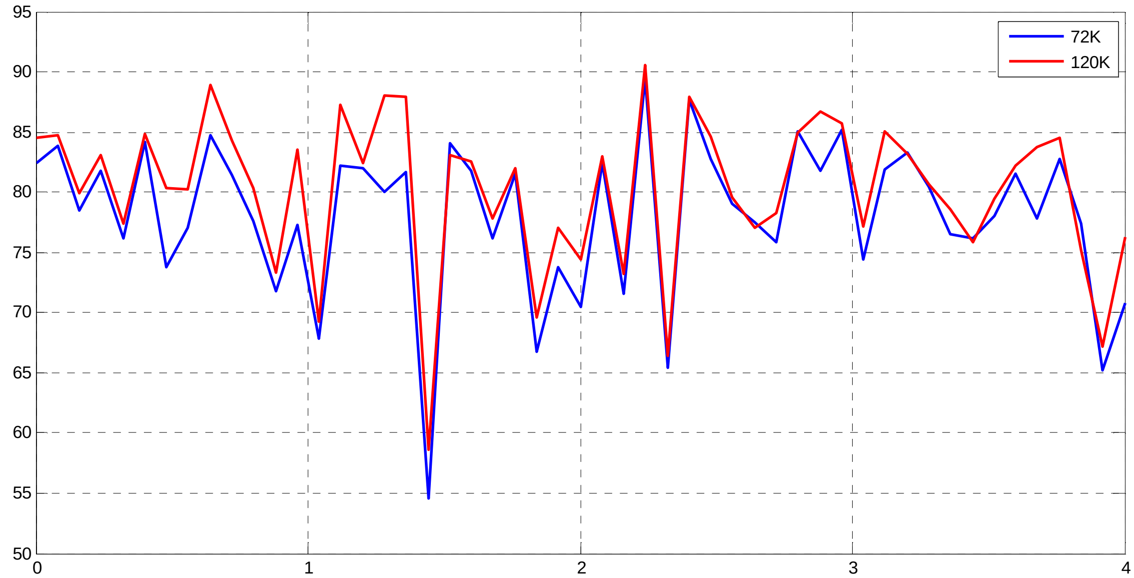

Figure 1. Change of percent of positive tweets during the four day period

- Adding negation features
- Removing punctuation
- Some primitive stemming
- Adding bigram and n-gram features

One other conclusion based on classification results was that for sentiment classification the mere presence of the feature is a better indicator of the class, than the number of appearances of the said feature. In other words, the multivariant Bernoulli model is more suited for sentiment classification than the multinomial model.

As a result of this research we trained two classifiers with two training sets of different sizes (72 thousand and 120 thousand training samples), using feature processing pipes mentioned above, which we will refer to as 72K and 120K classifier from now on.

IV. RESULTS

Twitter Search API returns tweet objects in JSON format and, besides the actual text of the tweet, they also contain various other properties. Some of them include: the author of the tweet, if the tweet is a retweet, time of posting, location if it is specified, etc. We preprocessed Search API query results to extract the text and posting time for each tweet. During classifying we calculated the number of positive and negative tweets posted during each hour of the monitored period. Fig. 1 shows the change of the percent of positive tweets during a 4 day period following the release of the Applause single.

It can be seen that positive tweets make up the majority of collected tweets during the whole period. It is reasonable to assume that Lady Gaga's fans are the most invested in tweeting about her new single, which is why these results are not surprising. It would probably be

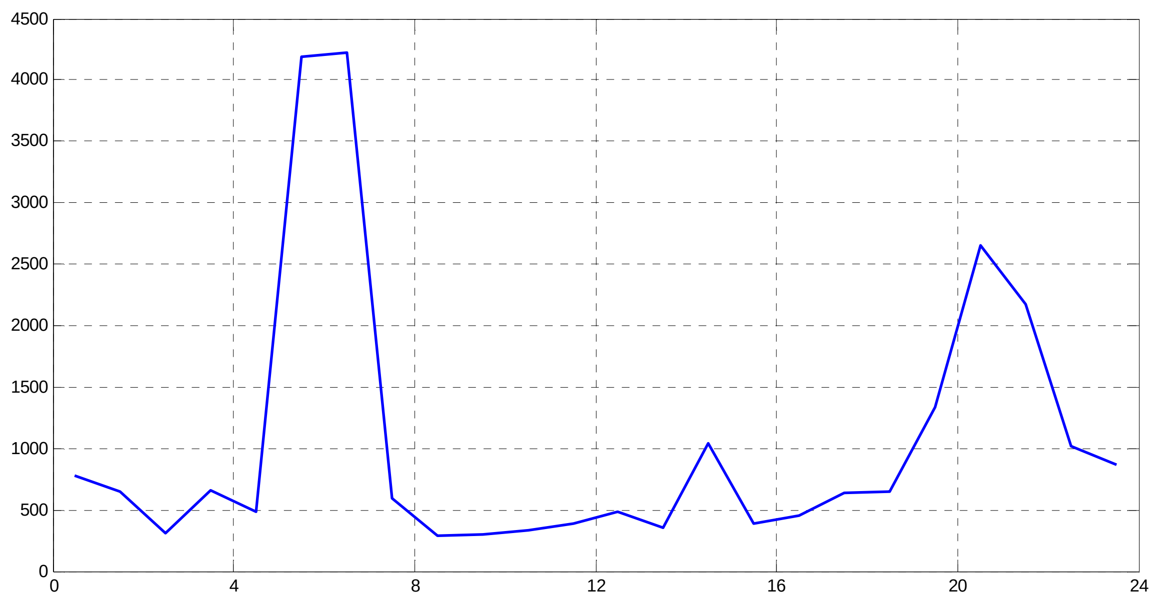

Figure 2. Change of the number of total tweets over a 24h period of day

interesting to examine the exact content of tweets that were posted in during periods that correspond to graph extreme values. The presented extracted data most likely could have been used as a one of the parameters for future sales estimation. Another possible application for such data would be estimation of the chart position that the single will peak at and the duration of time it will hold a position on the chart.

Data shown on Fig. 2 represents the total number of tweets collected for each hour in the day. Since Twitter API returns the time of posting in UTC time, the data from the graph is presented according to that. There are three peaks that stand out. The global maximum corresponds to the time period between 6h and 7h UTC, i.e. around midnight in North America time zones. There are also two smaller peaks around15h and 21h UTC. These data could be useful for determining the time patterns of Twitter use. Comparing these data to other determined time patterns of use could show the specificities of this particular case, for further analysis.

Off course, the classifiers are not perfect and there are misclassified tweets. Some examples that were misclassified as negative include:

- *Of course we saw it, and we died beacuse it's to much perfect.*
- *sorry it looks like I cut your hand and neck hhh.. url*
- *Bow down.. Queen is back bitches haha!*

It can be noticed that these examples include words with originally negative meanings. However, they are used in positive context which the classifier doesn't detect. Coordinating the perceived meaning of the word/phrase with the context of use is a possible research topic to be considered in the future.

V. Conclusion

Social networks represent a good source for data that can be used to model public opinion. In that context, one of the most analyzed social networks is Twitter. Twitter status messages - tweets, have their own characteristics that distinguish them from other domain data, which is why the classifier should also be trained with Twitter data. In Twitter sentiment classification, preprocessing the features in order to improve the accuracy usually gives good results.

Tracking sentiment change over time can give a good insight into public change. Also, it can probably be used as a parameter for prediction of long term goals, such as record sales or election results, and that is a research direction worth exploring. Additionally, it is possible to track sentiment over a period of time for multiple singles of the same artist, and compare their placements on charts. Comparing these data for multiple artists is also a possibility.

Acknowledgment

Research presented in this paper was funded by the Ministry of Education, Science and Technological Development of the Republic of Serbia, within the project "Technology Enhanced Learning", No. III 47003.

References

[1] L. Jiang, M. Yu, M. Zhou, X. Liu, T. Zhao, ”Target-dependent Twitter Sentiment Classification”, Proceedings of the ACL HLT, 2011, pp. 151-160

[2] B. Pang, L. Lee, "A Sentimental Education Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts", *Proceedings of the ACL*, 2004, pp. 271-278

[3] 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*, 2004, pp. 841—847

[4] A. Kennedy, D. Inkpen, "Sentiment Classification Of Movie Reviews Using Contextual Valence Shifters", *Computational Intelligence*, vol. 22, no. 2, 2006, pp. 110-125

[5] A. Pak, P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining”, Proceedings of the 7th LREC, 2010, pp. 1320–1326.

[6] A. Stanimirović, M. Ćirić, B. Džonić, L. Stoimenov, N. Petrović, "Sistem za davanje preporuka baziran na tehnologijama semantičkog web-a", *Zbornik radova YU INFO 2012*, 2012, pp. 147-152

[7] G Stylios, D Christodoulakis. J Besharat, MA Vonitsanou, I Kotrotsos, A Koumpouri, S Stamou, "Deciphering the Public's Stance towards Governmental Decisions", *Proceedings of the 10th European Conference on E-Government*, 2010, pp. 376-381

[8] Twitter, https://twitter.com

[9] Twitter statistics, http://www.statisticbrain.com/twitter-statistics/

[10] M. Ćirić, A. Stanimirović, N. Petrović, L. Stoimenov, "Naive Bayes Twitter classification", *Proceedings of SAUM 2012*, 2012, pp. 248-251

[11] J. Read, "Using Emoticons to reduce Dependency in Machine Learning Techniques for Sentiment Classification", *Proceedings of the ACL Student Research Workshop*, 2005, pp. 43-48

[12] B. Pang, L. Lee, S. Vaithyuanathan, "Thumbs up? Sentiment Classification using Machine Learning Techniques", *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, vol. 10, 2002, pp. 79-86

[13] H. Cui, V. Mittal, M. Datar, "Comparative Experiments on Sentiment Classification for Online Product Reviews", *Proceedings of the 21st national conference on Artificial intelligence*, vol. 2, 2006, pp. 1265-1270

[14] J. Blitzer, M. Dredze, F. Pereira, "Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification", In ACL, 2007, pp. 187-205

[15] A. Go, R. Bhayani, L. Huang, “Twitter Sentiment Classification using Distant Supervision”, CS224N Project Report, Stanford, 2009.

[16] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, R. Passonneau, "Sentiment Analysis of Twitter Data", *Proceedings of the ACL 2011 Workshop on Languages in Social Media*, 2011, pp. 30-38

[17] A. Davies, Z. Ghahramani, "Language-independent Bayesian sentiment mining of Twitter", *Proceedings of the 5th SNAKDD Workshop on Social Network Mining and Analysis*, 2011, pp. 99-106

[18] X. Wang, F. Wei, X. Liu, M. Zhou, M. Zhang, "Topic Sentiment Analysis in Twitter: A Graph-based Hashtag Sentiment Classification Approach", *Proceedings of the CIKM '11*, 2011, pp. 1031-1040

[19] H. Wang, D. Can, A. Kazemzadeh, F. Bar, S. Narayanan, "A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle", ACL '12 Proceedings of the ACL 2012 System Demonstrations, 2012, pp. 115-120

[20] A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, "Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment", *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media,* 2010, pp. 178-185

[21] M. Choy, L. F. M. Cheong, N. L. Ma, P. S. Koo, A sentiment analysis of Singapore Presidential Election 2011 using Twitter data with census correction, *Proceedings of CoRR. 2011*, 2011

[22] Y. R. Lin, D. Margolin, B. Keegan, D. Lazer, Voices of Victory: A Computational Focus Group Framework for Tracking Opinion Shift in Real Time, 22nd International World Web Conference, 2013

[23] J. Bollen, H. Mao, A. Pepe, Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena, *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 2011, pp. 450–453

[24] W. Hu, Real-Time Twitter Sentiment toward Thanksgiving and Christmas Holidays, *Social Networking*, 2013

[25] M. Ćirić, A. Stanimirović, N. Petrović, L. Stoimenov, "Comparison of different algorithms for sentiment classification ", *Proceedings of TELSIKS 2013*, 2013, pp. 567-570

[26] The Huffington Post, http://www.huffingtonpost.com/2013/08/12/lady-gagas-applause-early-release-single-leaks_n_3744177.html

[27] Twitter Search API, https://dev.twitter.com/docs/api/1.1/get/search/tweets

[28] Twitter Streaming API, https://dev.twitter.com/docs/api/streaming

[29] C. D. Manning, P. Raghavan, H. Schütze, *Introduction to Information Retrieval*, Cambridge University Press, New York, 2008.

[30] K. McCallum, "MALLET: A Machine Learning for Language Toolkit.", http://mallet.cs.umass.edu, 2002.