Sentiment detection for predicting altruistic behaviors in Social Web: a case study

Jacek Filipczuk
and Emanuele Pesce
Dipartimento di Informatica
Universitá degli Studi di Salerno
84084 Fisciano (SA) - Italy

Sabrina Senatore
Dipartimento di Ingegneria dell'Informazione
ed Elettrica e Matematica Applicata
Universitá degli Studi di Salerno
84084 Fisciano (SA) - Italy

***Abstract*—With the advent of Social Web, the user has become an active consumer which shares information and participates in social networks, online communities, blogs, wikis, feeds and chats. The volunteer person-power is a valuable resource which creates innovative content and helps other users to make right decisions with his own opinions, suggestions, advice. Opinions and suggestions have an amazing impact on the online user community: they may unexpectedly influence decision-making activities starting from simply buying or not a smartphone until to social events, political actions, and even marketing strategies. This paper aims at studying the role played by the sentiments in influencing the user actions. Particularly, it analyzes how sentiments expressed in the text can move the reader to do altruistic actions. The idea is from RAOP community where users write posts asking or offering a free pizza.**

Our work achieves a comparative analysis of machine learning methods on a ROAP dataset, that collects original posts where users asked for a free pizza. The goal is to extract the sentiments expressed in natural language, in the textual requests, in order to predict which user request will be satisfied (getting a free pizza). Finally, a posteriori "affective" analysis shows the predominant emotions expressed in the satisfied requests, that move the readers to have an altruistic behavior.

I. Introduction

Social computing [1] is a recent computing paradigm which provides an environment for human social interaction, collaboration and information sharing. It has become a hot topic attracting broad interest from not only research communities but also technological, software fields as well as business strategies. Recent trends in Social Computing evidence the real-time application development and future direction of science and technology, in the light of the user behavior perspective. Sentiments and emotions that motivate people to collaborate, sharing and taking actions represent indeed, a side of the same coin. Studies [2], [3] have demonstrated that the emotions and hedonic factors affect the user behaviors in such activities: in [3] for instance, a study on business persons indicate that altruism and job autonomy have a strong interaction effect on the willingness to help others through information systems (where electronic bulletin boards, web communities, and knowledge management systems are the conceivable information systems).

This paper introduces the sentiment analysis for the social Web: the idea is from ROAP (Random Act of Pizza), an online community devoted to giving away free pizza to strangers that ask for one (and they do not offer anything tangible in return). In [4], a study on ROAP competition is presented; the authors analyze this type of altruistic actions and evidence that the success of the requests is related to two different factors: social factors (who is asking and which relationship exists between the recipient and the donor) and linguistic factors (how they are asking and which textual form promotes successful requests). Our work instead, aims at investigating the sentimental factors that move a reader to satisfy a request from a stranger. To this purpose, the original ROAP dataset was analyzed and extended with additional sentiment-based features extracted by the textual requests; then, a comparative study of machine learning methods was carried out in order to predict the succeed requests (i.e., the requests that are satisfied, getting a free pizza). In addition, a posteriori analysis on the requests that succeed shows what emotions characterize these requests. The paper is organized as follows. Section II sketches a literature review of some related works. Section III briefly describes the ROAP competition, evidencing the main user's interaction. Section IV gives a logical overview of the process and then, in the following sections (Sections V-VII) the design and implementation details are provided as well. Particularly, a comparative study across the main machine learning methods is provided in Section VI. Section VII covers a-posteriori "affective" analysis which will show the main emotions hidden in the requests that move the reader to behave altruistically. Finally, conclusion and future works close the paper.

II. Related work

In recent years, there has been a growing interest in automatic extraction of sentiments as well as emotions occurring in the written language. A lot of applications spreading to several domains which range from consumer product reviews, health care and financial services to social events, political elections and market analysis [5] are interested in monitoring the moods that drive users in the their Web activities.

978-1-5090-1897-0/16/$31.00 ©2016 IEEE

This wave of application domains especially breaks in Social Web, where the growth of social applications and tools is a fertile ground for retrieving sentiment and emotional data expressed in the user-generated content. Sentiment and emotion detection is studied also to the user behaviour analysis: the users feeling is often condensed in a few textual sentences, but it produces a sane reaction moving the readers in some affective state that can generate an empathetic reply (such as "likes") or disapprove that sentiment. Altruistic traits are at the basis of many online communities whose existence is motivated by the information sharing and human interaction. Although helping others may not contribute to improving the helper's own performance (in some cases, might even detract from it), altruistic behaviors are often rewarding from human being perspective and assume a crucial value from an organizational viewpoint [3]. Studies regarding altruistic behaviour are mainly based on altruistic personality traits [6] that are not motivated by situational pressures or constraints [7]; altruistic behaviour are often activated when the perception of the need for helping is clear, and when there is the freedom to choose to help or not [8]. Understanding the factors that move the user to be altruist has implications for questions in social psychology and linguistic pragmatics: it is important to know what is being requested; what the giver receives in return; how the request has been formulated. In particular in [4], social and linguistic factors that make a successful request are studied, especially emphasizing the reasons that motivate people to give when they do not receive anything in return. Our approach mainly focuses on capturing sentiments expressed in the requests and predict if a request succeeds or not, by the analysis of the sentiments expressed in the text of the request.

The automatic extraction of sentiments as well as emotions in texts is becoming a large and growing area of interest [9], [10]. Sentiment Analysis (SA) especially, has been extensively studied in recent years. Most sentiment detection techniques aim at capturing the overall sentiment expressed in a whole text. This approach is called document-level SA, in contrast to phrase-level SA, which identifies sentiments expressed in sentences, outlining that the identification of the local sentiment is more reliable than the global document sentiment [11]. Existing work mainly concentrates on the use of three types of features: lexicon features, POS features, and microblogging features for sentiment analysis and opinion mining [12]. Some approaches combine these three types of features, giving major emphasis to POS tags with or without word prior polarity involved [13]. Close to our approach, there are the works which exploit machine learning techniques to sentiment analysis. In [14], a machine-learning method applies text-categorization techniques to some subjective parts of the document, by means of techniques for finding minimum cuts in graphs. In [15] a classification in positive and negative reviews is accomplished by summarizing sentiments. Also in [16], a machine learning approach identifies sentiments in blog, review, forum from Web, with the aim to discover in the feelings that people express with regard to certain consumption products. Several sentiment analysis techniques have been devised and evaluated on many domains [17], [18], [19] focusing on the categorization of overall sentiment [14], or recognizing the sentiments at the feature level [10], [20], [17], [18].

III. A ROAP COMPETITION

Random Acts Of Pizza or, briefly RAOP[1] is a community which hosts a competition based on a simple idea: people post messages on the site, some of those messages are from people wanting to give pizza away and others are from people looking to get a pizza. In the ROAP community, users write posts asking or offering a free pizza; generally a user makes a post where he describes why he would have a pizza. The reasons may be very different. For example, a user wrote that he is alone, without money and feels quite sad, so a pizza would cheer him up, another one wrote a joke asking in return a free pizza. ROAP is a sub-community of Reddit.com, a known website which gathers various communities in forums called subreddits. Every user can create a post in a subreddit. Users interact in many ways: they comment posts, or click on buttons to upvote (show that they like the post or comment) or downvote (show that they dislike the post or comment). A post is usually formed by a text, but it can include links to other web sites, photos or even videos. The content of a post changes based on the subreddit where it is posted. For example, if you visit a subreddit for cars photos, then most of the posts will be with photos showing cars.

IV. Logical Overview

Figure 1 shows the logical overview of the global process. The input dataset is given to the *Preprocessing* phase; a selection of original features, considered "relevant" in our analysis is accomplished (*Feature selection*); then a further analysis is carried out, to extract additional features and extend the given dataset with sentiment-based information (*Feature derivation*). The data matrix generated by these features is used in the *Machine Learning* phase, across some well-known machine learning algorithms. Results are compared to evaluate the prediction performance of a request success. Finally, a *Posteriori analysis* phase executes a further textual processing of the posts, detecting the emotions enclosed in success posts in order to assess which emotions drive altruistic actions.

V. Preprocessing

The preprocessing phase is crucial in this approach. It is strictly related to the extraction of relevant features. The input dataset was gained from the Kaggle[2] website. It is formed by a collection of textual requests (for a free pizza), with a set of metadata and further auxiliary information. For instance, each request has information like the time when it was written, an identifier number, the number of upvotes, the number of downvotes, etc. Almost all data are numerical, except for the title and the description of a request that are textual. Some

[1] http://www.randomactsofpizza.com/info.html
[2] https://www.kaggle.com/

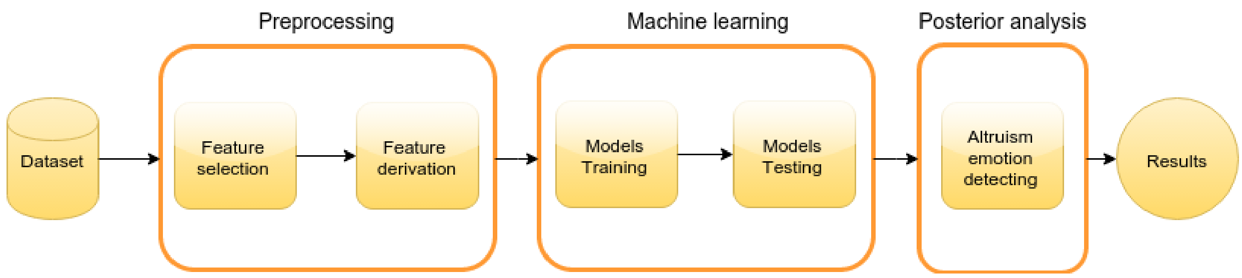


Fig. 1: Logical overview of the whole process

features from the original dataset are discarded because they are not relevant for the purpose of our approach; instead new features targeted at the sentiment analysis are taken in consideration in the machine learning approaches.

A. Selected Features

Table I shows the self-describing features selected from the original dataset, grouped in four main classes and listed as follows:

- **Text-based features**: are two textual features of the original dataset: *request_title* and *request_text*, respectively related to the title and the description of the request.
- **Post-based features**: data values related to how the request has been accepted; for instance, they are the number of upvotes and downvotes associated with the request, at the time it was collected; number of comments, information related to the timestamp of the request.
- **User-related features**: data related to the requester actions and the reaction of the other users to the request: data values related to the number of days since the first post of the requester, number of comments, posts, subreddits at time of request, retrieval in ROAP and Reddit; values related to upvotes, downvotes at time of request and retrieval, etc.
- **Outcome feature**: a boolean value called *requester_received_pizza*, used to train machine learning algorithm; it indicates the success of the request, i.e., whether the requester received pizza.

B. Derived Features

Two new feature types have been generated from the textual analysis of the posts and added to the original selected features. They are as follows:

- **Sentiment-based feature**: provides a numeric value that represents the polarity of the text enclosed in the features $request\_title$ and $request\_text$. The polarity calculation is based on a Python library for Sentiment Analysis, called *NLTK* API[3]. Given a text, NLTK returns three numerical values which represent the probability that the text will be negative, neutral and positive. The values relative to the negativity and the positivity will add up to 1, while the neutrality is standalone. The final result is

[3]http://www.nltk.org/api/nltk.html

TABLE I: Features Classes

Post-based features
number_of_downvotes_of_request_at_retrieval
number_of_upvotes_of_request_at_retrieval
request_number_of_comments_at_retrieval
unix_timestamp_of_request_utc
User-related features
requester_days_since_first_post_on_raop_at_request
requester_number_of_comments_at_request
requester_number_of_comments_at_retrieval
requester_number_of_comments_in_raop_at_request
requester_number_of_comments_in_raop_at_retrieval
requester_number_of_posts_on_raop_at_request
requester_number_of_posts_on_raop_at_retrieval
requester_number_of_posts_at_request
requester_number_of_posts_at_retrieval
requester_number_of_subreddits_at_request
requester_subreddits_at_request
requester_upvotes_minus_downvotes_at_request
requester_upvotes_minus_downvotes_at_retrieval
requester_upvotes_plus_downvotes_at_request
requester_upvotes_plus_downvotes_at_retrieval
requester_account_age_in_days_at_request
requester_account_age_in_days_at_retrieval
Text-based Features
request_title
request_text
Outcome feature
requester_received_pizza

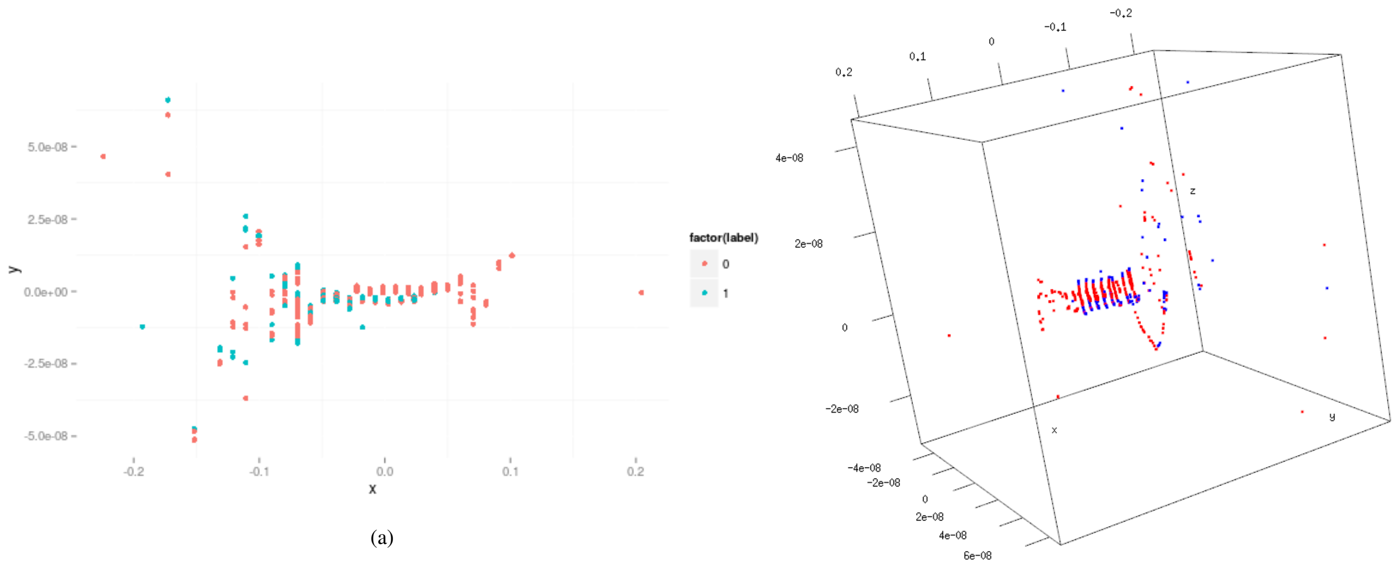
a text label which will be *neutral* if the value associated with the probability is greater than 0.5; otherwise, the label will be *positive* or *negative*. Rather than exploiting the three values, our feature compresses them in one digest value, whose calculation is given as follows. Given a text txt, let $pPos$ and $pNeu$ be values (computed by the *NLTK* API) associated respectively with the **positive** and **neutral** sentiments of txt, the Sentiment Compression (SC) label of txt, is given as:

$$SC(txt) = pPos \cdot sign(0.5 - pNeu) \quad (1)$$

where *sign* function is so defined:

$$sign(x) = \begin{cases} -1 & \text{if } x \leq 0, \\ +1 & \text{if } x > 0. \end{cases}$$

Just to give an clarifying example, if $SC(txt) = +0.3$, means that the text txt is not neutral (as the *sign* function value is positive); then, since the positive and negative polarity values sum up to 1, the predominant polarity for txt is negative, with value $1 - 0.3 = 0.7$. If



(a) (b)

Fig. 2: Multidimensional scaling applied on the dataset, in 2D (a) and 3D (b). Data are mainly concentrated in a central zone of the space. Labels corresponds to the success (1) or the unsuccess (0) of the requests.

$SC(txt) = -0.4$, then txt is labeled as neutral (recall that 0.4 represents the probability that txt is labeled as positive).

- **Success frequency feature**: collects the word frequency in the textual features, taking into account only the succeed requests. For each word w in a text txt, this feature keeps track about how many times w appears in txt and also in the all the succeed requests. The feature *success frequency rate* $SFreq(txt)$ of a given text txt is calculated as follows:

$$SFreq(txt) = \frac{\sum_{w \in txt} f_{txt}(w) \cdot f_S(w)}{length(txt)} \qquad (2)$$

where $f_{txt}(w)$ is the frequency of the word w in the given text txt, $f_S(w)$ is the frequency of that word w wrt. all the succeed requests, and $length(txt)$ is the number of words enclosed in the text txt.

VI. Machine Learning

The Machine Learning (ML) phase accomplishes the training of the different machine learning models and tests them in order to compare and find the best results.

A. Dataset description

The input dataset is formed by a collection of 5671 requests. A $[5671 \times 25]$ matrix has been defined, where the rows are the samples (requests) and the columns are the features. Precisely the 25 features are so summarized: 4 are post-based features (see Table I), 17 are user-related features, 2 features substitute the original text-based features: they are now SC($request\_title$) and SC($request\_text$) and finally, two additional features are SFreq($request\_title$) and SFreq($request\_text$). The data matrix has been standardized and some outliers have been removed. The size of the final matrix was $[5548 \times 25]$. Before training the models, the data set has been studied, in order to understand how the data are distributed in the space and select the most appropriate ML algorithms for the training. Figure 2 shows the multidimensional scaling in two and three dimensions (respectively, Figure 2a and Figure 2b), applied to our data. Labels in the figures corresponds to the success (1) or the unsuccess (0) of requests. The data visualization evidences that the data are very close: most of the requests are concentrated in a central zone of the space, whereas a few requests are arranged on the marginal area.

B. Training & Testing Models

The dataset has been split with random sampling without repetition as follows: training set: $[3884 \times 25] \approx 70\%$ and testing set: $[1664 \times 25] \approx 30\%$. An accurate setting of the parameters has been done, in order to train the models in the best possible way. Some models work better that others giving interesting results. In both cases there was the necessity to train the models with different parameters to find the ones that give good results. The classification models used in the experimentation are as follows:

- Support Vector Machine [21], with
  - Linea Kernel
  - Gaussian Kernel
  - Polynomial Kernel
  - Spline Kernel
- Random Forest [22]
- k-Nearest Neighbor [23], with
  - k = 1, 5, 15, 25, 51
- Naive Bayes [24]

After the training step, the classification models have been tested on the remaining data. The results on the testing data are compared with the feature $requester\_received\_pizza$. Table II shows the results of the classification phase: for each classifier, values of *accuracy*, *precision* and *recall* are shown. Similarly, Figure 3 shows the graphical representation of these results. The performance (in terms of Accuracy, Precision and Recall) of each classifier is shown and compared to the other models. In general, all the models evidence good performance in terms of accuracy; the ML model work like binary classifiers: the results confirm a good discriminative power, especially for the the SVM Linear Kernel and the Random Forest. Also the results in terms of the precision are interesting: the percentage of the returned success requests is very high with respect to all the returned requests that are considered a success. The recall instead shows lower values: in the best case, the results evidence that the returned success requests are just over half of the effective success requests. Anyway, since the models behave as binary classifiers, the overall performance can be considered good and accurate (especially in terms of accuracy), considering also that in the data visualization of Figure 2, the points distribution was very condensed.

TABLE II: *Accuracy*, *Precision* and *Recall*, for each classifier.

Classifier	**Accuracy**	**Precision**	**Recall**
SVM (Linear Kernel)	0.8522	0.7410	0.5281
SVM (Gaussian Kernel)	0.8089	0.6392	0.4975
SVM (Polynomial Kernel)	0.8474	0.7826	0.4924
SVM (Spline Kernel)	0.7620	0.5216	0.3554
Random Forest	0.8630	0.8319	0.4974
K-NN (k = 1)	0.7157	0.3905	0.3045
K-NN (k = 5)	0.7873	0.5986	0.2225
K-NN (k = 15)	0.7530	0.6071	0.0803
K-NN (k = 25)	0.7686	0.9091	0.0496
K-NN (k = 51)	0.7458	0.9000	0.0209
Naive Bayes	0.7662	0.5260	0.4236

The results emphasize the predictive power of the overall framework: the additional features based on the sentiment analysis contribute to characterize the nature of the requests and support the classification models: given a request, the framework will predict if it will have or not success; in other words, given a free pizza request, the overall model says if it can inspire or not altruism.

VII. A Posteriori Analysis

This phase starts from the results of our experimentation. Its goal is to analyze the emotions enclosed in the textual requests to figure out what are the most common emotions in the success requests; i.e., which emotions characterize a success request. So far, the work concerned to predict if a text can inspire altruism or not, by using ML models. This step instead focuses on trying to understand what characterizes altruism in a social context; precisely, which sentiments and emotions (more than others) inspire altruistic actions.

To this purpose, the text concerning requests which have been labelled as *altruistic* (i.e., success in the a priori knowledge of the dataset, see $requester\_received\_pizza$ in Table I) have been taken into account. The text analysis has been accomplished by the O2MC Api[4] which allows the extraction of a set of four opposite pair of emotions: *Joy- Sadness, Trust-Disgust, Fear-Anger,* and *Surprise-Anticipation.* For each individual emotion, a value in a range (0,1) is given: 0 means that the emotion is really weak while 1 means that the emotion is really strong. The emotions extracted from a text, with higher values are called *dominant*. So, the frequencies of the dominant emotions in all the textual requests of our dataset have been estimated. Table III shows the percentage of succeeded requested, associated with each emotion.

TABLE III: Emotions expressed in the satisfied requests (in percentage).

Joy	**Trust**	**Fear**	**Surprise**
87.77	4.11	2.26	2.05
Sadness	**Disgust**	**Anger**	**Anticipation**
1.51	1.78	1.85	1.64

It is easy to see that on 1258 safistied requests, the dominant emotion is *Joy* (87.77%); in other words, a free pizza was offered when the textual request expressed the emotion joy. Some minimal influence is given also by the emotions like *Trust*, whose percentage of occurrence is higher than the remaining emotions. The analysis of the results from the emotional viewpoint highlights that in general, a text which expresses *Joy* has a good chance to inspire altruistic actions. Then, by observing the pairs of opposite emotions, shown in Table III, let us notice that the emotions *Joy, Trust, Fear, Surprise* assume higher values than their own opposite emotions.

VIII. Conclusions

This work presents a machine learning approach to evaluate how the sentiments inspired by a written text can influence the user actions. The approach extends the original dataset, adding new features that give more weight to the textual content besides to the original features, in order to obtain more decisional power. The experimental results confirm the effectiveness of our approach, in terms of accuracy and precision and outline the predictive power: given a request, the overall model can say if it will inspire or not altruistic actions. A further a posteriori analysis allows us to improve knowledge about altruism, inspecting the emotions that imply altruistic behaviors. As stated, positive emotions, such as joy inspire selfless actions target to satisfy or support the user requests. The results seem promising, especially if some improvements will be taken into account in the future development. For instance,

[4]http://www.o2mc.io/portfolio-posts/text-analysis-restful-api-language-polarity-and-emotion/

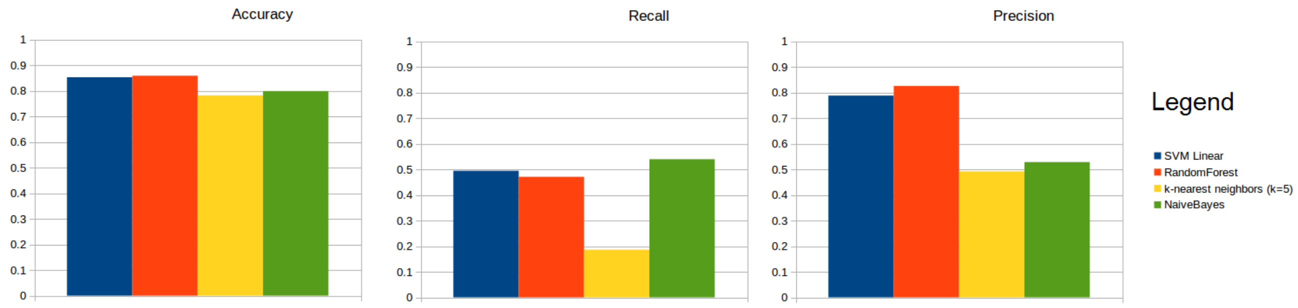


Fig. 3: A comparative histogram-based view: Accuracy, Precision and Recall for each classifier (see Table II)

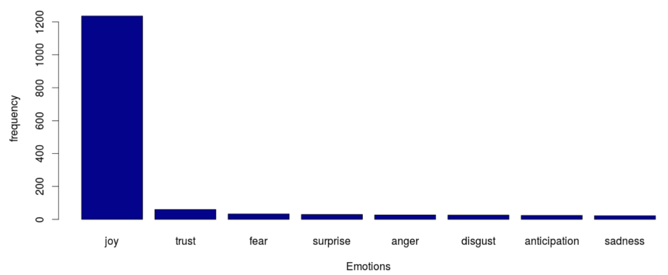


Fig. 4: Emotions expressed as the number of satisfied requests.

adding emotion-driven features could improve the predictive performance of ML methods (especially in terms of recall). As a further future investigation, contextual information could be introduced in the text analysis, to improve the significance of the detected emotions.

References

[1] C.-Y. Wang, S.-C. Chou, and H.-C. Chang, "Emotion and motivation: Understanding user behavior of web 2.0 application," in *Information Technology: New Generations, 2009. ITNG '09. Sixth International Conference on*, April 2009, pp. 1341–1346.

[2] B. Furneaux and D. Nevo, "Beyond cognitions: A call for greater consideration of emotion in information systems decision theories," in *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, 2008, pp. 451–.

[3] G. Lee and W. J. Lee, "Altruistic traits and organizational conditions in helping online," *Comput. Hum. Behav.*, vol. 26, no. 6, pp. 1574–1580, Nov. 2010.

[4] T. Althoff, C. Danescu-Niculescu-Mizil, and D. Jurafsky, "How to ask for a favor: A case study on the success of altruistic requests," in *Proceedings of ICWSM*, 2014.

[5] D. Ghazi, D. Inkpen, and S. Szpakowicz, "Prior and contextual emotion of words in sentential context," *Comput. Speech Lang.*, vol. 28, no. 1, pp. 76–92, Jan. 2014. [Online]. Available: http://dx.doi.org/10.1016/j.csl.2013.04.009

[6] J. Rushton, "The altruistic personality," in *Development and Maintenance of Prosocial Behavior*, ser. Critical Issues in Social Justice, E. Staub, D. Bar-Tal, J. Karylowski, and J. Reykowski, Eds. Springer US, 1984, vol. 31, pp. 271–290.

[7] E. B. King, J. M. George, and M. R. Hebl, "Linking personality to helping behaviors at work: An interactional perspective," *Journal of Personality*, vol. 73, no. 3, pp. 585–608, 2005.

[8] B. A. Fritzsche, M. A. Finkelstein, and L. A. Penner, "To help or not to help: Capturing individuals'decision policies," *Social Behavior and Personality: an international journal*, vol. 28, no. 6, pp. 561–578, 2000.

[9] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," *IEEE Intelligent Systems*, vol. 28, no. 2, pp. 15–21, 2013.

[10] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures on Human Language Technologies*, vol. 5, no. 1, pp. 1–167, 2012.

[11] T. Nasukawa and J. Yi, "Sentiment analysis: capturing favorability using natural language processing," in *Proceedings of the 2nd international conference on Knowledge capture*, ser. K-CAP '03. New York, NY, USA: ACM, 2003, pp. 70–77. [Online]. Available: http://doi.acm.org/10.1145/945645.945658

[12] H. Saif, Y. He, and H. Alani, "Semantic sentiment analysis of twitter," in *Proceedings of the 11th international conference on The Semantic Web - Volume Part I*, ser. ISWC'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 508–524.

[13] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of twitter data," in *Proceedings of the Workshop on Languages in Social Media*, ser. LSM '11. Stroudsburg, PA, USA: Association for Computational Linguistics, 2011, pp. 30–38.

[14] B. Pang and L. Lee, "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts," in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, ser. ACL '04, Stroudsburg, PA, USA, 2004.

[15] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing - Volume 10*, ser. EMNLP '02, Stroudsburg, PA, USA, 2002, pp. 79–86. [Online]. Available: http://dx.doi.org/10.3115/1118693.1118704

[16] E. Boiy and M.-F. Moens, "A machine learning approach to sentiment analysis in multilingual web texts," *Information Retrieval*, vol. 12, no. 5, pp. 526–558, 2009. [Online]. Available: http://dx.doi.org/10.1007/s10791-008-9070-z

[17] I. Titov and R. T. McDonald, "Modeling online reviews with multi-grain topic models," in *WWW*, J. Huai, R. Chen, H.-W. Hon, Y. Liu, W.-Y. Ma, A. Tomkins, and X. Z. 0001, Eds. ACM, 2008, pp. 111–120.

[18] M. Cataldi, A. Ballatore, I. Tiddi, and M.-A. Aufaure, "Good location, terrible food: detecting feature sentiment in user-generated reviews," *Social Network Analysis and Mining*, pp. 1–15, 2013. [Online]. Available: http://dx.doi.org/10.1007/s13278-013-0119-7

[19] V. Loia and S. Senatore, "A fuzzy-oriented sentic analysis to capture the human emotion in web-based content," *Know.-Based Syst.*, vol. 58, pp. 75–85, Mar. 2014.

[20] H. M. Liu B. and J. Cheng, "Opinion observer: Analyzing and comparing opinions on the web," in *Proceedings of International World Wide Web Conference*. New York, USA: ACM, 2005, pp. 342–351.

[21] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab – an S4 package for kernel methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004. [Online]. Available: http://www.jstatsoft.org/v11/i09/

[22] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002. [Online]. Available: http://CRAN.R-project.org/doc/Rnews/

[23] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, 4th ed. New York: Springer, 2002, iSBN 0-387-95457-0. [Online]. Available: http://www.stats.ox.ac.uk/pub/MASS4

[24] C. Weihs, U. Ligges, K. Luebke, and N. Raabe, "klar analyzing german business cycles," in *Data Analysis and Decision Support*, D. Baier, R. Decker, and L. Schmidt-Thieme, Eds. Berlin: Springer-Verlag, 2005, pp. 335–343.