The Social Mood of News:
Self-reported Annotations to Design Automatic Mood Detection Systems

Firoj Alam, Fabio Celli, Evgeny A. Stepanov, Arindam Ghosh, and Giuseppe Riccardi
Department of Information Engineering and Computer Science,
University of Trento, Italy
{firoj.alam,fabio.celli,evgeny.stepanov,arindam.ghosh,giuseppe.riccardi}@unitn.it

Abstract

In this paper, we address the issue of automatic prediction of readers' mood from newspaper articles and comments. As online newspapers are becoming more and more similar to social media platforms, users can provide *affective* feedback, such as mood and emotion. We exploited the self-reported annotation of mood categories obtained from the metadata of the Italian online newspaper *corriere.it* to design and evaluate a system for predicting five different mood categories from news articles and comments: indignation, disappointment, worry, satisfaction and amusement. The outcome of our experiments shows that overall, bag-of-word-ngrams performs better compared to all other feature sets, however, stylometric features perform better for the mood score prediction of articles. Our study shows that such self-reported annotations can be used to design automatic systems.

1 Introduction and Background

Participating in social media has become a mainstream part of our daily life - we read articles, comments, other people's statuses and provide feedback in terms of emotions through written content. Currently, newspapers are also being designed as social media platforms to facilitate users to provide their opinion along with emotional feedback. Since, currently, our social participation is mostly done through social media platforms, therefore, the online content, including social media and newspapers' content, is growing so rapidly. Turner reported in (Turner et al., 2014) that by 2020 it might reach 44 trillion gigabytes including news articles and user generated content such as likes, dislikes, emotions, tastes, identities, and data collected by sensors (Liu, 2007).

Such increasing amount of digital data creates an unprecedented amount of opportunities for business and individuals as well as it poses new challenges to process and generate concrete summaries out of it. For example, everyday journalists need to deal with large quantity of information whenever they need to prepare a historical/follow-up report or a summary from a large collection of documents. For example, journalist might want to know how particular *topics* of a news are associated with *users' mood*. The importance of such studies and their use-case has also been reported in (Riccardi et al., 2015). The challenges include automatic processing of such semi-structured or unstructured data in different dimensions such as linguistic style, interaction, sentiment, mood and other social signals. Finding the collective information of such signals requires automatic processing, which will be useful for various professionals, specifically psychologists and social and behavioral scientists. Among other affective dimensions, the mood and sentiment are particularly important to analyze the consumer behavior towards brands and products (Pang and Lee, 2008; Stieglitz and Dang-Xuan, 2013).

In the past few decades, the affective dimension of text has been mainly analyzed in terms of positive and negative polarity (Pak and Paroubek, 2010a; Kouloumpis et al., 2011; Cambria et al., 2016a), although more detailed dimensions are proven to be very useful. In particular, moods such as tension, depression, anger, vigor, fatigue, and confusion in tweets have been found to be good predictors of stock market exchanges (Bollen et al., 2011). It has also been demonstrated that it is possible to predict anger, sadness, and joy from Livejournal blogs with performances up to 78% accuracy (Nguyen et al., 2010). Moreover, it is also possible to distinguish Twitter users who are likely to share contents generating joy or amusement from the ones who are likely to share contents generating sadness, anger or disappointment with an accuracy of around 61% (Celli et al., 2016). An increasing number of studies focused on analyzing sentiment in terms of positive and negative polarity from a short text (microblog) (Akkaya et al., 2009; Paltoglou and Thelwall, 2010). For the machine classification perspective, a research application SentiStrength utilizes a different source of information towards assigning a sentiment score in a short text (Thelwall et al., 2011; Stieglitz and Dang-Xuan, 2013). To design automatic detection and classification systems a typical approach of getting reference annotation is to use either lexicon, automatic system, such as SentiStrength, (Bollen et al., 2011; Stieglitz and Dang-Xuan, 2013; Ferrara and Yang, 2015; Kim and Salehan, 2015), manual annotation by the expert annotators, *or* users' self-reported annotation (Nguyen et al., 2014; Mishne and others, 2005). Recent advances in knowledge based NLP for sentiment analysis can be found in (Cambria et al., 2016b). In another study (Cambria, 2016), author presented a framework for sentiment analysis, which include knowledge-based system and machine learning module.

Self-reported mood annotation by the users of the blog posts has also been studied previously as can be seen in (Go et al., 2009; Pak and Paroubek, 2010b; Pak and Paroubek, 2010b). In citedavidov2010enhanced, Davidov et al. used twitter hastags as labels for designing the automatic classification system. Similar study has also been conducted by Kunneman et al. in (Kunneman et al., 2014). There are still many challenges in designing an automatic system using self-reported annotations because the annotations are not done in a consistent manner. Users annotate them based on their *self-perception* and social media platform are not designed by following any psychological instruments or instructions. The obvious advantages of such annotations are that 1) they are cost-effective, and 2) we are getting user's natural affective expressions.

In this work, our goal was to investigate whether such annotations can be useful for designing an automatic system. We investigated two different approaches: 1) a system that can predict the mood score for each mood category from the articles and comments, 2) a classification model to classify either positive or negative mood. For the experiments, we compared the performance of different sets of features, which included word-ngrams, character-ngrams, stylometric, psycholinguistic and ngrams of part-of-speech. Our study is inline with the study presented by Nguyen et al (Nguyen et al., 2014), where the authors investigated a different set of features along with different machine learning algorithms for the feature selection and classification problem. Our work differs in a way that we focused on the prediction of mood in a scale of $[0...1]$ and the utilization of different sets of features, and used both articles and comments. Our experiments with comments are very complex due to the noisy structure of the data. The reason for focusing on prediction is to have a fine-grained view in a form of 'emotional sphere' of a comment or an article because text may contain a blend of emotional manifestations in a separate part of the text. The motivation of utilizing stylometric and psycholinguistic features is that mood may be expressed with certain kinds of idiosyncratic vocabulary and style, which these features might capture.

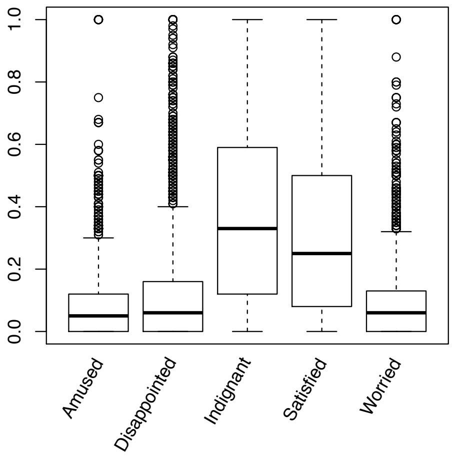
The structure of the paper is as follows. In Section 2 we present the details of the corpus. Then, we report the experimental procedures in Section 3 and report the results of the experiments in Section 3.2. Finally, discussions and conclusions are appear in Section 4 and 5, respectively.

2 Corpus

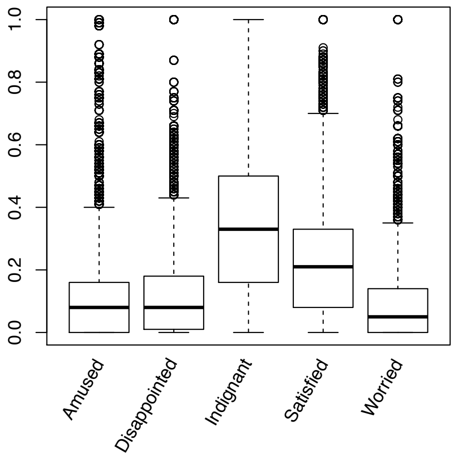
We collected the data from *Corriere della Sera*. It is one of the most popular Italian daily newspapers, in which the online platform is structured as a social media platform (Boyd et al., 2010). In particular, the platform of the Corriere provides: 1) a semi-public profile[1] for each registered user, 2) articulates a list of users connected by a relationship of interest and 3) allows to view their list of connections to other registered users 4) it also include mood metadata reported by the readers as their 'self-perception'.

The annotations of moods are available at the article and at the author levels. Therefore, the mood scores for each article was directly obtained from *Corriere* article's metadata, which is basically average scores of the users' reported mood for that article. Whereas, the mood scores for the comments were

[1]By semi-public we refer that Corriere provides user's average mood scores, number comments and votes made, interest and number of following people but not any demographic information.



(a) Articles



(b) Comments

Figure 1: Box-plots for the reference mood scores of each mood category

obtained from the mood scores of the user who commented. Mood scores of the user are part of the users' personal profile and describe all the moods they declared after reading the articles. A small part of this dataset has also been studied in (Celli et al., 2014; Celli et al., 2016) for studying mood and finding the association between mood, personality traits and interaction style.

For this study, we collected $\sim$ 2200 articles and the associated $\sim$ $300K$ comments. As a part of preprocessing, we filtered some data to remove outliers, for each mood category, for articles and comments, respectively. Outliers are computed based on the mood scores that appeared independently in each mood category of the articles and comments. As can be seen in the Figure 1 for some articles we have outlier scores for amused, disappointed and worried, and for comments we also have outliers for satisfied too. For example, outliers for commnets in amused category are score above 0.4, which are basically scores above the upper outer fence in the boxplot. Subsequently, data was partitioned into the train, development and test set with 60%, 20%, and 20% respectively, for the training, validation and evaluation of the prediction system. For the sake of replicability of the experiments, we will publicly release the data information (articles' links with training, development and test split) on github[2]. As a part of preprocessing, we removed URLs from the text even though the URL itself represents some information.

In Figure 1, we present box-plots of the mood score distribution for the articles and comments, respectively. From the figures, we observed that there are similar distributions in the mood categories for both articles and comments. For example, for indignation and satisfaction the scores of the data points vary between 0.1 to 0.6. From the data, we also observed that in many cases users tend to annotate articles when the content of the articles represent the emotional feelings of indignation or satisfaction.

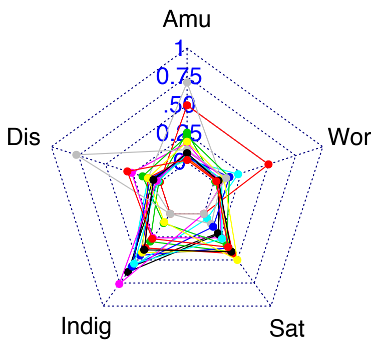


Figure 2: Spider plot of the reference mood scores from the selected comments. Amusement (Amu), Disappointment (Dis), Indignation (Indig), Satisfaction (Sat), Worry (Wor).

[2]https://github.com/nlpresources/Corriere-mood-data

A lexical analysis has been performed on articles and comments to understand the complexity of the task. We observed that for articles the average number of the tokens is 550 per article with maximum 3188 and minimum 44. Whereas for comments, the average is 44 with a maximum of 285 and minimum of 1 token. A closer look at the comments with higher number of tokens reveals that people usually talk about national issues such as economy, taxes, and environmental causes. There is a difference between article and comments in terms of language style. For example, the written style of the articles is more formal, whereas the text in comments is more noisy and informal, containing repetitions, emoticons, jargon, abbreviations, non-standard grammar, and urls. The noisy structure is very common in any social media conversation as also reported in (Nguyen et al., 2014; Alam et al., 2013). In Figure 2, we present a spider-plot with reference mood scores from the selected comments, which range from 0 to 1. As can be seen in the figure, the mood scores of indignation and satisfaction are higher than other categories.

In order to get a clear understanding of the labels such as mood scores and category for comments and articles we present an example of an user's comment in Figure 3. As can be seen in the Figure, the comment is labeled with the five mood scores for five mood categories using the user's self-report. These mood scores were then turned into a class label score (see Section 3.2.2) to define class label such as positive and negative. For this example it is negative (Neg)

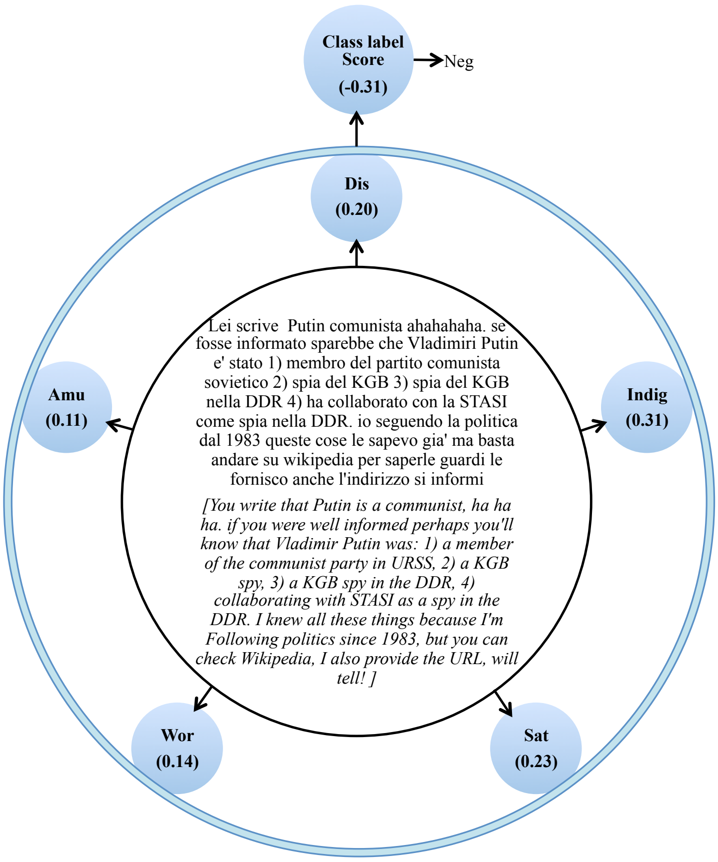


Figure 3: An example of self-reported annotation of a comment with mood scores and category. English translation is provided with Italic form. Mood category is negative for this example.

3 Methodology

We experimented with a different set of feature for the prediction of the mood and designing the classification system using the articles as well as comments. Such feature set includes bag-of-word-ngrams and bag-of-character-ngrams, psycholinguistic, stylometric, and ngrams of part-of-speech. In addition to investigating individual feature set, we also experimented them with feature level combination. However, the performances are poor, which we do not report in this paper.

For the mood score prediction task we used the Random Forest algorithm, whereas for the classification system we used Support Vector Machines (SVMs). This is because, from our initial experiments using these two algorithms with two different tasks, we found that these two are suited for the tasks we investigated. For example, Random Forest outperformed SVMs for the prediction task, reported in (Celli et al., 2016).

3.1 Features

3.1.1 Bag-of-word-ngram

We investigated the bag-of-word-ngrams, with $3 >= n >= 1$, and their logarithmic term frequencies (tf) multiplied with inverse document frequencies (idf) – tf-idf. Although the bag-of-words model has many drawbacks such as data sparsity and high dimensionality, it is the simplest and has been known to work well in most text-based classification tasks. As the bag-of n-grams approach results in a large dictionary which increases computational cost, we selected 5K most frequent n-grams.

3.1.2 Bag-of-character-ngram

Similar to the bag-of-word-ngrams, we also extracted and evaluated bag-of-character-ngrams, with $6 >= n >= 2$ and tf-idf transformation. The motivation of experimenting with this feature set is due to its success in sentiment classification task as reported in (Abbasi et al., 2008).

3.1.3 Part-of-Speech features (POS):

To extract POS features we used TextPro (Pianta et al., 2008) and designed the feature vector using bag-of-ngram approach, with $3 >= n >= 1$ and tf-idf transformation.

3.1.4 Stylometric Features

The use of stylometric features has its root in the domain of authorship identification (Yule, 1939; Abbasi and Chen, 2008; Bergsma et al., 2012; Cristani et al., 2012). Its use has also been reported for text categorization and discourse classification problems (Koppel et al., 2002; Celli et al.,). In authorship identification task, stylometric feature is defined as different groups such as lexical, syntactic, structural, content specific, idiosyncratic and complexity-based (Koppel et al., 2002; Abbasi and Chen, 2008; Cristani et al., 2012). In this work, we use the term *stylometric* to refer to the complexity-based[3] features reported in (Tanaka-Ishii and Aihara, 2015; Tweedie and Baayen, 1998). The extracted feature set is presented in Table 1.

In addition to the features listed in Table 1, we also extracted some word and character based low-level features and then projected them onto statistical functionals. They include count of word-ngram (n=2 to 3) and character ngram (n=2 to 4). The statistical functions include mean, median and standard deviation. The number of the resultant feature set is 97.

3.1.5 Psycholinguistic Features

To extract the psycholinguistic features from the articles and comments we utilized the Linguistic Inquiry Word Count (LIWC) (Pennebaker et al., 2001), which is a knowledge-based system developed by Pennebaker et al. over the past few decades. The uses of these features have been studied in different research areas of psychology and sociology, and mostly used to study gender, age, personality, honesty, dominance, deception, and health to estimate the correlation between these attributes and word use (Mairesse et al., 2007; Tausczik and Pennebaker, 2010). The success of these features has also been reported in literature (Nguyen et al., 2014; Alam and Riccardi, 2014; Danieli et al., 2015). The types of LIWC features include the following:

- General: word count, average number of words per sentence, a percentage of words found in the dictionary and percentage of words longer than six letters and numerals.
- Linguistic: pronouns and articles.
- Psychological: affect, cognition, and biological phenomena.
- Paralinguistic: accents, fillers, and disfluencies.

[3]Also uses the term constancy measure or lexical richness in literatures.

Table 1: Stylometric features

General
• word count = N • dictionary size = V
Length-based features:
• Average word length • Short word ratio (length = 1-3) to N
Frequency-based Ratios
• Ratio of Hapax Legomena to N • Ratio of Hapac Dislegomena to N
Lexical Richness using transformations of N and V:
• Mean Word Frequency = N/V • Type-Token Ratio = V/N • Guiraud's $R = V/sqrt(N)$ • Herdan's $C = log(V)/log(N)$ • Rubet's $K = log(V)/log(log(N))$ • Maas $A = (log(N) - log(V))/log^{2(N)} = a^2$ • Dugast's $U = log^{2(N)}/(log(N) - log(V))$ • Lukjanenkov and Neistoj's $LN = (1 - V^2)/(V^2 * log(N))$ • Brunet's $W = N^{(V^{(-a)})}, a = 0.172$
Lexical Richness using Frequency Spectrum:
• Honore's $H = b(log(N)/a - (V(1,N)/V)), b = 100, a = 1$ • Sichel's $S = V(2,N)/V$ • Michea's $M = V/V(2,N)$ • Herdan's $V = sqrt(sum(V(i,N) * (V(i,N)/N)^2) - 1/V)$ • Yule's $K = a(-1/N + sum(V(i,N) * (V(i,N)/N)^2)), a = 1$ • Simpson's $D = sum(V(i,N)(V(i,N)/N)(V(i,N) - 1)/(N - 2))$ • Entropy = $V(i,N)(-log((V(i,N)/N))^s * (V(i,N)/N)^t, s = t = 1$
• Length ratios 30 features

- Features about personal concern include work and home.
- Punctuation and spoken categories.

Since it is a knowledge based system, therefore it is packaged with dictionaries for different languages including Italian. For this work, we used the Italian version of the dictionary (Alparone et al., 2004), which contains 85 word categories. In addition, we also extracted 5 general descriptors and 12 punctuation categories, constituting a total of 102 features. The LIWC feature processing differs according to types of features, which includes counts and relative frequencies (see (Tausczik and Pennebaker, 2010)).

3.2 Experiments

3.2.1 Mood Score Prediction Experiments and Results

For the mood score prediction experiments, we utilized Random Forest as a learning algorithm (Breiman, 2001). It is a decision tree based algorithm where each decision tree is generated by randomly sampling instances and features, then the score of the forest is computed by averaging the scores from the trees. For this experiment, the number of the tree is set to 100. The number of trees can be optimized, which has not been done in this study and could be done in future. We used the development set to do some preliminary experiments. Then, to obtain the results on the test set, we trained the model by combining the training and development sets. For each task and feature set, we normalized each feature to have zero mean and unit variance.

We measured the performance of the mood score prediction system using Root-Mean-Square-Error (RMSE). Moreover, we computed the baseline results by randomly generating the scores using the gaussian distribution based on the prior mean and standard deviation, as presented in Table **??**.

Table 2: Results on the test set with different feature sets. RMSE lower is better. Baseine results are computed by randomly selected from gaussian distribution based on prior mean and standard deviation. Base: Baseline, W-ng: word ngram, C-ng: character ngram. Amusement (Amu), Disappointment (Dis), Indignation (Indig), Satisfaction (Sat), Worry (Wor).

	Article						**Comments**					
Class	**Base**	**W-ng**	**C-ng**	**POS**	**Style**	**LIWC**	**Base**	**W-ng**	**C-ng**	**POS**	**Style**	**LIWC**
Amu	0.130	0.100	0.100	0.102	0.120	0.102	0.170	0.118	0.118	0.119	0.119	0.120
Dis	0.150	0.108	0.112	0.116	0.128	0.120	0.180	0.126	0.127	0.127	0.128	0.128
Indig	0.380	0.266	0.274	0.280	0.247	0.278	0.350	0.245	0.244	0.246	0.246	0.247
Sat	0.370	0.267	0.276	0.271	0.166	0.275	0.230	0.165	0.164	0.165	0.165	0.166
Wor	0.130	0.095	0.096	0.099	0.118	0.099	0.170	0.118	0.117	0.118	0.118	0.118
Avg	0.230	0.167	0.172	0.174	**0.156**	0.175	0.220	**0.154**	**0.154**	0.155	0.155	0.156

In Table 2, we present the results with different feature sets. We obtained better results with stylometric features for the mood categories of the articles and second best results with word-ngram features. For the comments, we obtained better results with word- and character- ngram and almost similar results across other feature sets. Our results for the articles and comments are statistically significant ($p < 0.05$) compared to the random baseline, which we computed using t-test. For the comments, the results are similar across different feature sets within third decimal digits as we see in Table 2. The difference in performance is observed after 3rd decimal digits across different feature sets. It might be because POS, stylometric and LIWC feature sets are not able to capture better information from the users' comments due to the noisy structures. The other reason could be due to the higher variations in terms of a number of tokens in the comments, which create higher sparseness. In terms of the performance and number of features in stylometric feature set, we speculate that these features might be useful in cross-language/domain experiments.

3.2.2 Mood Classification Experiments and Results

For the classification task, we first transformed the mood scores into binary classes such as positive and negative. In order to do that, first we computed an overall mood $class\ label\ score$ by subtracting the sum of "Disappointment", "Worry" and "Indignation" scores from the sum of "Amusement" and"Satisfaction", as shown in Equation 1. Then, we turned this score into either of the two classes: positive and negative, as shown in Equation 2. We ignored the instances with score equals to zero. After the data-split into train, development and test set the class distribution for the articles remains 63% (negative) *vs* 37% (positive), and for comments it is 53% (negative) *vs* 47% (positive).

$$class\,label\,score = (amusement + satisfaction) - (disappointment + worry + indignation) \quad (1)$$

$$class\_label\_instance(i) = \begin{cases} pos & if\ score > 0 \\ neg & if\ score < 0 \end{cases} \quad (2)$$

We designed the classification model using Support Vector Machines (SVM) (Platt, 1998) and used linear kernel to tackle the problem of higher dimensions in some feature sets. We measured the performance of the system using macro-averaged precision, recall, F1 measure and accuracy. Baseline results are computed by randomly selecting the class labels, such as positive or negative, based on the prior class distribution of the training set as shown in Table 3.

In Table 3, we present the classification results for the articles and comments. For the articles, we obtained better results using word-ngrams and the second best is character-ngrams. For the comments, we obtained similar results with both word and character ngrams, however a minor improvement is observed with character ngrams. The performance of POS, LIWC and stylometric feature sets are lower comparatively. For the classification experiments, the results are also statistically significant ($p < 0.05$) compared to the random baseline. The statistical test has been performed using McNemar's test.

Table 3: Classification results on the test set using different feature sets

Exp	Articles				Comments			
	P	R	F1	Acc	P	R	F1	Acc
Baseline	47.89	47.93	47.90	53.51	49.93	49.92	49.93	50.04
Word-ngram	62.20	61.70	**61.89**	58.73	54.44	54.27	54.06	54.71
Char-ngram	55.95	56.22	55.76	56.69	55.33	55.16	**55.03**	55.71
POS	53.97	53.96	53.96	56.24	52.29	52.12	51.59	52.96
Style	54.43	52.76	50.56	59.64	52.37	52.26	51.96	52.93
LIWC	54.30	54.05	54.02	57.37	52.43	52.31	51.98	53.01

4 Discussion

For the prediction task, the overall results for comments are better compared to articles, whereas, for the classification task, the results are better for articles than comments. From our investigations of different feature sets, we observed that overall bag-of-word-ngrams perform better in both tasks. For the prediction task of the experiment with articles, we obtained better results using stylometric features. Hence, regarding this feature set, our observation is that it does not contain language specific information, therefore, we can explore it further for cross-domain/language study. Regarding the use of self-reported annotation, our findings suggest that more investigation is necessary for the annotation by observer/expert annotators to understand their reliability. One important issue is that in this self-reported annotations, users have not followed any instructions or have not maintained any psychological instruments while expressing their affective opinions.

5 Conclusion

In the paper, we present our work for the prediction and classification of mood from news article and comments in which we utilized self-reported mood annotations and investigated the different type of feature sets. For the mood score prediction task, we obtained better results using bag-of-word-ngrams and stylometric features for both articles and comments. For the classification task, we obtained better results with bag-of-word-ngrams. The prediction and classification tasks with comments became difficult due to the noisy structure of the data. Since the self-reported data are increasing over time, therefore, further investigation by the expert annotators of the annotations would be helpful to use this kind of data for designing automatic systems. The other important focus would be cross-corpus study to see how the trained models can be useful in a different corpus.

Acknowledgments

The research leading to these results has received funding from the European Union - Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 610916 - SENSEI - http://www.sensei-conversation.eu/.

References

Ahmed Abbasi and Hsinchun Chen. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS)*, 26(2):7.

Ahmed Abbasi, Hsinchun Chen, and Arab Salem. 2008. Sentiment analysis in multiple languages: Feature selection for opinion classification in web forums. *ACM Transactions on Information Systems (TOIS)*, 26(3):12.

Cem Akkaya, Janyce Wiebe, and Rada Mihalcea. 2009. Subjectivity word sense disambiguation. In *Proc. of the Conference on EMNLP*, pages 190–199. Association for Computational Linguistics.

Firoj Alam and Giuseppe Riccardi. 2014. Fusion of acoustic, linguistic and psycholinguistic features for speaker personality traits recognition. In *Proc. of ICASSP*, pages 955–959, May.

Firoj Alam, Evgeny A Stepanov, and Giuseppe Riccardi. 2013. Personality traits recognition on social network-facebook.

F. Alparone, S. Caso, A. Agosti, and A. Rellini. 2004. The italian liwc2001 dictionary. Technical report, LIWC.net, Austin, TX.

Shane Bergsma, Matt Post, and David Yarowsky. 2012. Stylometric analysis of scientific articles. In *Proc. of the NACL*, pages 327–337. ACL.

Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proc. of ICWSM*, pages 1–10.

Danah Boyd, Scott Golder, and Gilad Lotan. 2010. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, pages 1–10. IEEE.

Leo Breiman. 2001. Random forests. *Machine learning*, 45(1):5–32.

Erik Cambria, Soujanya Poria, Rajiv Bajpai, and Björn Schuller. 2016a. Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *the 26th International Conference on Computational Linguistics (COLING), Osaka*.

Erik Cambria, Björn Schuller, Yunqing Xia, and Bebo White. 2016b. New avenues in knowledge bases for natural language processing. *Knowledge-Based Systems*, 108(C):1–4.

Erik Cambria. 2016. Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31(2):102–107.

Fabio Celli, Evgeny A. Stepanov, and Giuseppe Riccardi. Tell me who you are, i'll tell whether you agree or disagree: Prediction of agreement/disagreement in news blog.

Fabio Celli, Giuseppe Riccardi, and Arindam Ghosh. 2014. Corea: Italian news corpus with emotions and agreement. In *Proceedings of the First Italian Conference on Computational Linguistics CLiC-it 2014 & and of the Fourth International Workshop EVALITA 2014*, pages 98–102. Pisa University Press.

Fabio Celli, Arindam Ghosh, Firoj Alam, and Giuseppe Riccardi. 2016. In the mood for sharing contents: Emotions, personality and interaction styles in the diffusion of news. *Information Processing & Management*, 52(1):93–98.

Marco Cristani, Giorgio Roffo, Cristina Segalin, Loris Bazzani, Alessandro Vinciarelli, and Vittorio Murino. 2012. Conversationally-inspired stylometric features for authorship attribution in instant messaging. In *Proc. 20th ACM Multimedia*, pages 1121–1124. ACM.

Morena Danieli, Giuseppe Riccardi, and Firoj Alam. 2015. Emotion unfolding and affective scenes: A case study in spoken conversations. In *Proc. of Emotion Representations and Modelling for Companion Systems (ERM4CT) 2015,*. ICMI.

Emilio Ferrara and Zeyao Yang. 2015. Quantifying the effect of sentiment on information diffusion in social media. *PeerJ Computer Science*, 1:e26.

Alec Go, Richa Bhayani, and Lei Huang. 2009. Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1:12.

Dan Kim and Mohammad Salehan. 2015. The effect of sentiment on information diffusion in social media.

Moshe Koppel, Shlomo Argamon, and Anat Rachel Shimoni. 2002. Automatically categorizing written texts by author gender. *Literary and Linguistic Computing*, 17(4):401–412.

Efthymios Kouloumpis, Theresa Wilson, and Johanna Moore. 2011. Twitter sentiment analysis: The good the bad and the omg! In *ICWSM*.

FA Kunneman, CC Liebrecht, and APJ van den Bosch. 2014. The (un) predictability of emotional hashtags in twitter.

Hugo Liu. 2007. Social network profiles as taste performances. *Journal of Computer-Mediated Communication*, 13(1):252–275.

François Mairesse, Marilyn A Walker, Matthias R Mehl, and Roger K Moore. 2007. Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res.(JAIR)*, 30:457–500.

Gilad Mishne et al. 2005. Experiments with mood classification in blog posts. In *Proceedings of ACM SIGIR 2005 workshop on stylistic analysis of text for information access*, volume 19, pages 321–327. Citeseer.

Thin Nguyen, Dinh Phung, Brett Adams, Truyen Tran, and Svetha Venkatesh. 2010. Classification and pattern discovery of mood in weblogs. In *Advances in knowledge discovery and data mining*, pages 283–290. Springer Berlin Heidelberg.

Thin Nguyen, Dinh Phung, Brett Adams, and Svetha Venkatesh. 2014. Mood sensing from social media texts and its applications. *Knowledge and information systems*, 39(3):667–702.

Alexander Pak and Patrick Paroubek. 2010a. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, pages 1320–1326.

Alexander Pak and Patrick Paroubek. 2010b. Twitter as a corpus for sentiment analysis and opinion mining. In *LREc*, volume 10, pages 1320–1326.

Georgios Paltoglou and Mike Thelwall. 2010. A study of information retrieval weighting schemes for sentiment analysis. In *Proc. of the 48th ACL*, pages 1386–1395. ACL.

Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2):1–135.

James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71.

Emanuele Pianta, Christian Girardi, and Roberto Zanoli. 2008. The textpro tool suite. In *LREC*. Citeseer.

John Platt. 1998. Sequential minimal optimization: A fast algorithm for training support vector machines. Technical report, Microsoft Research.

Giuseppe Riccardi, Frederic Bechet, Morena Danieli, Benoit Favre, Robert Gaizauskas, Udo Kruschwitz, and Massimo Poesio. 2015. The sensei project: Making sense of human conversations. In *International Workshop on Future and Emergent Trends in Language Technology*, pages 10–33. Springer.

Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, 29(4):217–248.

Kumiko Tanaka-Ishii and Shunsuke Aihara. 2015. Computational constancy measures of texts—yule's k and rényi's entropy. *Computational Linguistics*.

Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1):24–54.

Mike Thelwall, Kevan Buckley, and Georgios Paltoglou. 2011. Sentiment in twitter events. *JASIST*, 62(2):406–418.

Vernon Turner, John F Gantz, David Reinsel, and Stephen Minton. 2014. The digital universe of opportunities: Rich data and the increasing value of the internet of things. *IDC Analyze the Future*.

Fiona J Tweedie and R Harald Baayen. 1998. How variable may a constant be? measures of lexical richness in perspective. *Computers and the Humanities*, 32(5):323–352.

G Udny Yule. 1939. On sentence-length as a statistical characteristic of style in prose: With application to two cases of disputed authorship. *Biometrika*, 30(3/4):363–390.