User Sentiment Detection: A YouTube Use Case

Smitashree Choudhury[1], John G. Breslin[1, 2]

[1] DERI, National University of Ireland, Galway, Ireland
[2] School of Engineering and Informatics, National University of Ireland, Galway, Ireland
smitashree.choudhury@deri.org, john.breslin@nuigalway.ie

Abstract. In this paper we propose an unsupervised lexicon-based approach to detect the sentiment polarity of user comments in YouTube. Polarity detection in social media content is challenging not only because of the existing limitations in current sentiment dictionaries but also due to the informal linguistic styles used by users. Present dictionaries fail to capture the sentiments of community-created terms. To address the challenge we adopted a data-driven approach and prepared a social media specific list of terms and phrases expressing user sentiments and opinions. Experimental evaluation shows the combinatorial approach has greater potential. Finally, we discuss many research challenges involving social media sentiment analysis.

1. Introduction

The rapidly-increasing popularity of social media sites such as Facebook[1], Flickr and YouTube is primarily due to the ease of use and simplicity of these systems for the creation, collaboration and sharing of resources (images, video) even from non-technical users. For video sharing, YouTube is the most popular site on the Web. According to a recent [1] study, YouTube[2] accounts for 20% of Web traffic and 10% of total Internet traffic. YouTube provides many social mechanisms to judge user opinion and views about a video by means of voting, rating, favourites, sharing and negative comments, etc. This context information is useful in studying user and community behaviour and perspective. Analysis of user comments provides a useful data source for many applications such as comment filtering, personal recommendation, and user profiling to name a few. In this paper, we opted for an unsupervised lexicon-based feature interpretation to analyze the sentiment orientation of user comments and we used the publicly-available sentiment lexicon called SentiWordnet [2]. But the question remains whether the available lexicons are sufficient for a dynamic domain where language changes and evolves so fast. Can the existing lexicons capture these dynamics? How can we detect the sentiment polarity of colloquial phrases? We propose to extend existing resources with a social media specific

[1] http://www.facebook.com
[2] http://www.youtube.com

phrase list frequently used by users. Usage of the list is further validated using the community-created dictionary urbandictionary.com[3].

The application of comment analysis will enable the video creator to view constructive comments and filter out spam comments. This is much more useful in the domain of news items where there are thousands of user comments generated for each news item every day, and also has relevance for product reviews and educational videos. Our study focused on the top 4,000 videos across five major categories provided by YouTube classifications.

The rest of the paper is organized as follows. Section 2 describes related work in the field of sentiment analysis. Section 3 describes the data corpus and its statistics. Our approach for sentiment analysis is described in section 4. This is followed by a discussion of the results in section 5. Finally we conclude our paper with some future directions.

2. Related Work

We will describe some studies related to sentiment analysis and social media user studies. Early work in this area includes Turney [17] and Pang [18] who applied different methods for detecting the polarity of product reviews and movie reviews respectively. It is less clear how sentiment analysis techniques can be employed in the context of social website analysis where the language tends to be more freeform and informal. Blitzer et al. [6] extended a learning model for sentiment classification on a product review data set in cross-domain settings. Researchers in [3] reported high accuracy in English movie reviews and 72%-83% accuracy in non-English reviews. Tan et al. [8] present a domain-specific classifier to label the top *k* unlabeled instances and to learn a new model for a new domain.

Existing research based on SentiWordnet has focused on identifying opinionated words. The linguistic rule-based approach of Chaumartin [7] uses SentiWordNet in combination with WordNet Affect to detect emotion and valence values for words in headlines. His approach uses the frequency of sentiment-bearing words and types of WordNet relations instead of polarity scores. Fahrni and Klenner [9] focus on the target-specific polarity determination of adjectives. A prior-polarity lexicon of adjectives is derived from SentiWordNet. Devitt and Ahmad [10] have exploited this resource in combination with WordNet's semantic content for sentiment polarity detection in financial news. Zhang et al. [11] extract subjective adjectives from SentiWordNet in order to estimate the probability that a document contains opinion-bearing expressions. Bermingham et al. [12] studied online radicalisation among YouTube users by analysing user comments and their language usage.

In [13] the dependency of helpfulness of product reviews from Amazon users on the overall star rating of the product is examined and a possible explanation model is provided. "Helpfulness" in that context is defined by Amazon's notion of how many users

[3] http://www.urbandictionary.com

rated a review and how many of them found it helpful. In [14] the temporal development of product ratings and their helpfulness and dependencies on factors such as the number of reviews or the effort required (writing a review vs. just assigning a rating) are studied. Works on sentiment classification and opinion mining such as [15] deal with the problem of automatically assigning opinion values (e.g. "positive" vs. "negative" vs. "neutral") to documents or topics using various text-oriented and linguistic features. Recent work [5] in this area also makes use of SentiWordNet to improve classification performance. However not many studies in sentiment analysis and opinion mining has focused on the noisy data such as user comments in social media sites such as Youtube.

3. Data Corpus

We created a data corpus consisting of information from around 20,000 videos. We used the YouTube API to collect the most popular and most relevant videos across five major categories including 10 different sub-categories such as politics and news, science and technology, travel, music, movie, sports, gaming, people and blogs. For each category we collected the 2,000 most popular videos.

For each video we collected up to 50 comments if available along with the usernames, text of the comment, etc. The complete collection includes data about 19,743 videos and more than 500,000 associated comments both short and long.

Fig. 1 describes the comment distribution over various categories. Music commands the highest average comments followed by news, movie, people and travel.

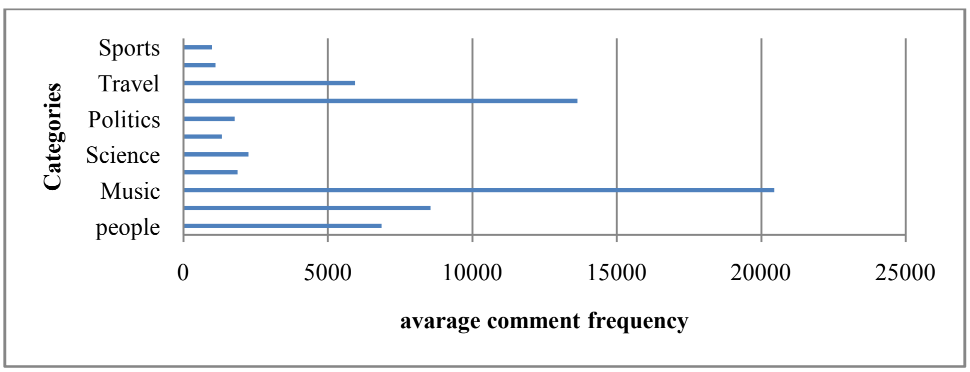

Figure 1. Comment distribution over categories.

3.1 Content Pre-Processing

The textual content of comments gives us a glimpse of a user's view on the video content. Therefore the choice of words helps in detecting the valence of the content. For efficient processing and gathering of language statistics, we indexed all our data with Lucene[4] The top 20 words from positively-classified comments included many positive adjectives (cool, awesome, amazing, great, nice, etc.); informal acronyms indicating a happy state (e.g. lol); and positive exclamations (e.g. wow). The top 20 words from negatively classified comments included negation terms (don't, never); negative adjectives (sad, fake, scary, hate etc.) and various profanities or insulting terms. These top terms were extracted from the comments in terms of their frequency. After collecting the video data and the corresponding comments we carried out some basic pre-processing of the text content. Comments were subjected to stop-word removal, and term stemming. We used Porter stemming to stem the terms and then used SentiWordnet for sentiment polarity detection. SentiWordnet is a publicly-available thesaurus annotated with the sentiment polarity of each synset from WordNet 2.0 [4].

4. Our Approach

We adopted a combination of lexicon-based approaches as background knowledge to detect the comment sentiments. Below we will describe the primary sentiment lexicon SentiWordnet and the manually-created social media specific phrase list.

4.1 SentiWordnet (SWN)

For analysing user sentiment as reflected in comments, we used SentiWordnet, a sentiment annotated lexicon built on top of WordNet. WordNet is a lexical dictionary describing terms and their semantic relationships (hypernyms, meronyms, hyponyms, etc.). Each term is represented as a synset in WordNet and each synset contains synonym terms. Each term in the synset is described with a triple of positive, negative and neutral/objective sentiment scores. For instance, the term "worst" carries the triple of {pos. = .25, neg. = .75, neutral = 0}. The triple score sums up to 1. SWN consists of around 207,000 word-sense pairs or 117,660 synsets. It provides entries for nouns (71%), verbs (12%), adjectives (14%) and adverbs (3%).

4.2 Negation and Social Media Aware Phrase List (SMAPL)

Negation detection in a sentence is crucial for identifying the term sentiment. Terms do not communicate the meaning in isolation but much more if identified in a window of context. The word "good" generally carries a positive sentiment, but when it appears in a phrase with the qualifiers "not so good", this will change the polarity of the phrase from positive to negative. It is imperative to identify such contextual situations for proper sentiment detection. We have created a list of negation expressions by observing comments and used the list to scan each sentence for their presence. In case an expression is found we reverse the term polarity given in the SentiWordnet dictionary. When

[4] http://lucene.apache.org/

considering negation expressions, words such as “no”, “never”, ”none”, “not”, “hardly”, “seldom” are all considered. Social media interactions such as comments do not confirm strictly to the conventional linguistic style as observed in blogs, news, and even product reviews. These interactions are informal, full of acronyms, slang words and are constantly changing. Any cross-domain model will result in poor performance if adapted without proper training. For this reason, we created an informal terms list normally used to express opinion and feelings in social media and social networking sites.

Table 1. A subset of SMAPL,

douchebag	hardly painful	not good
nope	me gusta	lmfao
omg	damnit	rulez
freakin	scumbag	Damn good

In order to incorporate social media specific terms and phrases, we prepared a list from our dataset. To prepare the domain-specific list which we called the “Social Media Aware Phrase List” we used two different sources: (1) our existing data and 2) flagged comments. For the existing comments, we followed these steps:

1. Indexed the comment dataset in a Lucene index
2. Extracted the high frequency terms
3. Filtered the terms not available in SentiWordnet as possible candidates

We also selected the flagged comment with the assumption that these comments must have sentiments with high intensity as they were given “thumbs down” as a mark of unacceptability by other users. An analysis of flagged comments revealed that there are four major categories of comments that are normally flagged: (1) self promotion: where people ask others to subscribe to or watch their video; (2) propaganda: contains messages expressing strong beliefs on topics like religion, socialism, and conspiracies; (3) abusive comments: contain extremely sexist and racist comments; and (4) other: these cannot be classified under the above three categories because the content looks absolutely normal, however perhaps due to opposing views or some other dislike of the commenter, they mark the message as “flagged”. These flagged comments can be used to build comment classifiers for rating and filtering purposes.

Our assumption proved correct when we noticed that flagged comments carried relatively strong and abusive words compared to the non-flagged comments.

4.3 Sentiment Analysis and Polarity Detection

Sentiment analysis is the task of identifying positive and negative opinions, emotions, and evaluations [17]. Our study evaluates the user comment sentiment by using SentiWordnet and our customised social media specific phrase list. Fig. 2 shows an abstract view of the system modules.

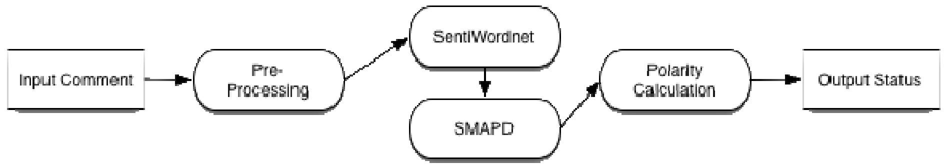

Figure 2. System architecture.

The system starts with a pre-processed user comment as input and detects the number of sentiment-bearing words that exist in the comment. For each word in the list, we calculate the sentiment score by retrieving the number of entries and the triple for each entry. In case of a single entry, the decision will be in favour of the class (positive, negative or neutral) with the highest score. In case of multiple entries of a term, such as different entries (senses from Wordnet), we averaged the scores of each class over each entry in order to achieve some normalisation.

The above step will give us a triple of three scores for a term. To determine the sentiment polarity of the entire comment, we aggregate the score of all terms in the comment and the frequency of each class is counted (the decision will again be in favour of the high-frequency class or the highest average of the three classes). Below is an example of a user comment and the sentiment scores of terms.

Table 2. First pass of a user comment.

Comment	*Sentiment polarity score (pos., neg., neutral)*
I think it would benefit religious people to see things like this, not just to learn about our home, the Universe, in a fun and easy way, but also to understand that non-religious explanations don't leave people hopeless and helpless as they think: they inspire people with awe, understanding and a thirst for exploration. Can you ask for more?	benefit: 0.0: 0.125: 0.875 fun: 0.0: 0.0: 1.0 easy: 0.0: 0.625: 0.375 understand: 0.375: 0.125: 0.5 leave: 0.0: 0.0: 1.0 hopeless: 0.0: 0.75: 0.25 helpless: 0.0: 0.875: 0.125 inspire: 0.0: 0.125: 0.875 awe: 0.5: 0.125: 0.375 understanding: 0.0: 0.0: 1.0 thirst: 0.25: 0.0: 0.75
Document class	Neutral

4.4 Negation Detection

The sentiment score is refined with negation detection in the sentence. After the first pass of term polarity detection, the comment is passed through the negation detection stage. The negation phrase lexicon consists of a list of high-frequency terms that have more probability of appearing in a negative context. Once the input is identified as carrying such phrases, the previously determined sentiment score is reversed from positive to negative or vice versa.

4.5 Social Media Aware Phrase Detection (SMAPD)

This phase consists of scanning the comments for any terms or adjectives typically observed in social media interactions but that have no entry in the SentiWordnet set. This list could be improved further and used as a comment classification feature. But for now, we only scan the text with a window of four terms for the presence and absence of the term or phrase in the list. If detected, we give a fixed polarity value of 0, 1 or -1.

5. Result of Sentiment Analysis

For evaluation, we extracted 100 comments from each category randomly and manually annotated with positive and negative sentiments as the ground truth. Evaluation was conducted both with SentiWordnet alone and in combination with the extended list.

Table 3. Results of sentiment analysis with SentiWordnet.

	Science & Technology		*Entertainment*		*Sports*		*News & Politics*	
	GT	Detected	GT	Detected	GT	Detected	GT	Detected
Positive	40	30 (75%)	45	26 (57%)	35	23 (65%)	23	16 (69%)
Negative	35	12 (35%)	39	11 (28%)	39	17 (43%)	54	17 (31%)

Table 3 above shows our positive and negative detection against the ground truth across all the categories. The result shows that most of them are misclassified as objective or neutral. The positives are comparatively better detected than the negatives. Table 4 below shows the detection performance once we add the list expressing the negation expressions and social media specific negative and positive terms. The trend is similar but the improvement in negative sentiment detection is better than the positive sentiments. This confirms our assumption that the present approach does well but needs more training in social media scenarios.

Table 4. Results of sentiment analysis including the second phase.

	Science & Technology		*Entertainment*		*Sports*		*News & Politics*	
Positive	GT	Detected	GT	Detected	GT	Detected	GT	Detected
	40	34 (85%)	45	32 (71%)	35	29 (82%)	23	17 (73%)
Negative	35	21 (60%)	39	23 (58%)	39	27 (69%)	54	28 (51%)

5.1 Discussion

The results in Fig. 3 show that negative detection (right) performed better than the positive (left) detection in the second stage. This confirms our hypothesis that an extension to the existing lexicons is crucial for an improved performance. It is useful to remember that the comments used for extracting phrases were not part of the evaluation data.

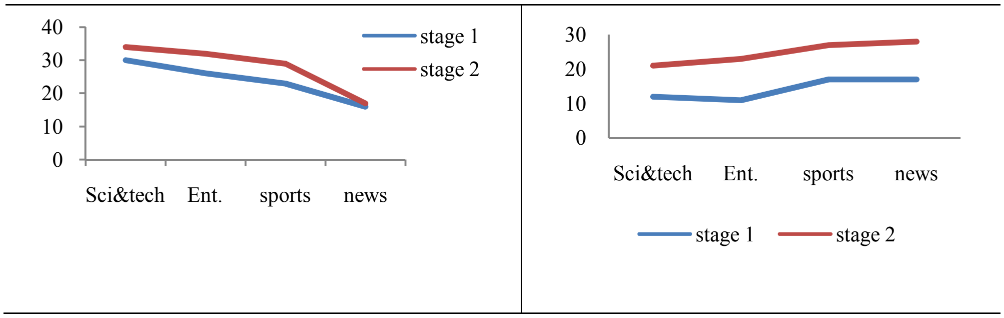

Figure 3. Positive (left) and negative (right) sentiment detection using both methods.

Compared to many other studies for product and movie reviews, the accuracy score is low for the social media comment use case. So far, no study has reported on their evaluation's performance using YouTube comments, and therefore we consider there is a need to carry out more training and to explore the specific characteristics related to these types of comments.

The application of SentiWordnet alone is not giving satisfactory results when compared to the state-of-art achieved in other content domains. Even in combination with the extra negation detection and SMAPD phase, the results show room for improvement.

We will now describe some of the limitations in sentiment analysis for informal comments as found on YouTube and the associated research challenges, and some of the main reasons for not achieving the expected results.

1. SentiWordnet and other lexicons are primarily suitable for news, blogs, product reviews and movie reviews where people follow certain linguistic syntaxes more when compared to many social media sites (especially YouTube), where the language is more informal and consists of many adopted slangs. Therefore, we suspect that cross domain applications need more careful feature selection and training.
2. The lack of context disambiguation is another known limitation of SentiWordnet; however some contextual information could be derived from YouTube videos to assist with disambiguation of terms.
3. In YouTube comment threads, there may be some benefit in analysing and taking into account the temporal impact of a series of comments (i.e. the impact of negative and positive comments on subsequent comments and actions).
4. For the degree of sentiment, a simple positive or negative will not be sufficient, but the intensity of the sentiment will help guide subsequent impact.

6. Conclusions and Future Work

In this work, we proposed a lexicon-based unsupervised sentiment detection of user comments for YouTube videos. Since the existing sentiment lexicons are not geared towards social media conversations and interaction patterns, we extended a social media specific lexicon expressing sentiments and opinions of the user. We showed that the combined use of the existing dictionary SentiWordnet and the extended list performs better in detecting the user sentiments from the comments. The result also showed that recall of negative sentiment is poorer compared to the positives, which may be due to the wide linguistic variation used in expressing frustration and dissatisfaction. Further studies need to be performed to validate the result. We described future work towards improving the social lexicon and statistically validating it so that it can be used across other domains.

Acknowledgements

This work was supported by Science Foundation Ireland under grant number SFI/08/CE/I1380 (Líon 2).

References

1. X. Cheng, C. Dale, and J. Liu. Understanding the characteristics of internet short video sharing: YouTube as a case study. In Technical Report arXiv:0707.3670v1 cs.NI, New York, NY, USA, 2007. Cornell University, arXiv e-prints.
2. Esuli. Automatic Generation of Lexical Resources for Opinion Mining: Models, algorithms and Applications. PhD in Information Engineering, PhD School “Leonardo da Vinci”, university of Pisa, 2008.
3. K. Denecke. Using SentiWordnet for multilingual sentiment analysis. In Data Engineering Workshop, 2008. ICDEW 2008, pages 507– 512, 2009.

4. C. Fellbaum, editor. WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA, 1998.
5. Siersdorfer, S., Chelaru, S., Nejdl, W., & San Pedro, J. (2010). How useful are your comments? *Proceedings of the 19th international conference on World Wide Web - WWW '10*, 891. New York, New York, USA: ACM Press. doi: 10.1145/1772690.1772781.
6. M. D. John Blitzer and F. Pereira. Biographies, Bollywood, boom-boxes and blenders: Domain adaptation for sentiment classification. Association of Computational Linguistics (ACL), 2007.
7. F. R. Chaumartin. Upar7: A knowledge-based system for headline sentiment tagging. In Proceedings of SemEval- 2007, pages 422–425, Prague, Czech Republic, June 2007. S.
8. Tan, G. Wu, H. Tang, and X. Cheng. A novel scheme for domain-transfer problem in the context of sentiment analysis. In CIKM '07, pages 979–982, New York, NY, USA, 2007. ACM.
9. Fahrni and M. Klenner. Old wine and warm beer: Target-specific sentiment analysis of adjectives. In AISB 2008, 2008.
10. Devitt and K. Ahmad. Sentiment polarity identification in financial news: A cohesion-based approach. In Proceedings of the 45th Annual Meeting of the ACL, pages 984–991, Prague, Czech Republic, 2007.
11. E. Zhang and Y. Zhang. UCSC on TREC 2006 blog opinion mining. In Text Retrieval Conference, 2006.
12. Bermingham, Adam and Conway, Maura and McInerney, Lisa and O'Hare, Neil and Smeaton, Alan F. (2009) Combining social network analysis and sentiment analysis to explore the potential for online radicalisation. In: ASONAM 2009 - Advances in Social Networks Analysis and Mining, 20-22 July, 2009, Athens, Greece.
13. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee. How opinions are received by online communities: a case study on Amazon.com helpfulness votes. In WWW '09: Proceedings of the 18th international conference on World Wide Web, 2009
14. F. Wu and B. A. Huberman. How public opinion forms. In Internet and Network Economics, 4th International Workshop, WINE 2008, Shanghai, China, pages 334–341, 2008.
15. M. Thomas, B. Pang, and L. Lee. Get out the vote: Determining support or opposition from Congressional floor-debate transcripts. In EMNLP '06: Proceedings of the ACL-02 conference on Empirical methods in natural language processing, pages 327–335, 2006
16. K. Toutanova and C. Manning. Enriching the knowledge sources used in a maximum entropy part-of-speech tagger. Proceedings of the 2000 Joint SIGDAT Conference EMNLP/VLC, 63-71, 2000, pages 63–70, October 2000.
17. Peter Turney (2002). "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews". *Proceedings of the Association for Computational Linguistics (ACL)*. pp. 417–424. http://arxiv.org/pdf/cs.LG/0212032.
18. Bo Pang; Lillian Lee and Shivakumar Vaithyanathan (2002). "Thumbs up? Sentiment Classificat*ion using Machine Learning Techniques". Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*. pp. 79–86. http://www.cs.cornell.edu/home/llee/papers/sentiment.home.html.