iFeel: A Web System that Compares and Combines Sentiment Analysis Methods

Matheus Araújo
UFMG
Belo Horizonte, Brazil
matheus.araujo@dcc.ufmg.br

Pollyanna Gonçalves
UFMG
Belo Horizonte, Brazil
pollyannaog@dcc.ufmg.br

Meeyoung Cha
KAIST
Daejeon, Korea
meeyoungcha@kaist.edu

Fabrício Benevenuto
UFMG
Belo Horizonte, Brazil
fabricio@dcc.ufmg.br

ABSTRACT

Sentiment analysis methods are used to detect polarity in thoughts and opinions of users in online social media. As businesses and companies are interested in knowing how social media users perceive their brands, sentiment analysis can help better evaluate their product and advertisement campaigns. In this paper, we present iFeel, a Web application that allows one to detect sentiments in any form of text including unstructured social media data. iFeel is free and gives access to seven existing sentiment analysis methods: SentiWordNet, Emoticons, PANAS-t, SASA, Happiness Index, SenticNet, and SentiStrength. With iFeel, users can also combine these methods and create a new Combined-Method that achieves high coverage and F-measure. iFeel provides a single platform to compare the strengths and weaknesses of various sentiment analysis methods with a user friendly interface such as file uploading, graphical visualizing, and weight tuning.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences; H.3.5 [**Online Information Services**]: Web-based services

General Terms

Human Factors, Measurement

Keywords

Sentiment analysis, Web applications, Comparison, Social media.

1. INTRODUCTION

The extreme popularity of Online Social Networks (OSNs) has led people to share nearly everything on the Web such as opinions about a prominent public event, URLs, news, and daily conversations. This means massive amount of data are generated everyday on online platforms. In Twitter alone users posts more than 400 million messages daily as of March 2013. Mining and utilizing such big social data hence can bring a number of new opportunities to businesses and markets.

Sentiment analysis is a popular method for mining OSN data and has many useful applications. It is common to find reviews or comments of products, services, events, and brand names on OSNs. From such unstructured data, sentiment analysis can reveal how people feel about a particular product or service, which is essential for businesses and companies. With the growing interest in both industry and academia, many tools and methods have been proposed for detecting sentiments embedded in text data including OpinionFinder [18], ANEW [2], PANAS-t [11], Emoticons, SentiWordNet [8], Happiness Index [7], SentiStrength [14], SASA [16], and SenticNet [1].

A number of strategies and techniques have been used for sentiment analysis such as machine learning [12], lexical dictionaries, natural language processing, and psychometric scales. The diversity of techniques results in differences in the types of sentiment categories each method reveals, assumptions made, and datasets used for validation. While each tool has its own strength and weakness, little effort has been paid on comparing them. In a previous work [10], we compared 8 existing sentiment tools across two scales: coverage (i.e., the fraction of data whose sentiment is revealed) and prediction performance (i.e., the fraction of data whose sentiment is correctly judged). No single method was a winner in terms of prediction performance, and each method had limited coverage. Based on these observations, we proposed a new approach, called Combined-Method, that is a mix of 7 license-free methods to yield high coverage and F-measure.

Extending our initial effort to compare and combine existing sentiment analysis methods, this paper presents a Web tool called iFeel (available at `http://www.ifeel.dcc.ufmg.br`). iFeel gives easy access to existing tools for anyone to use. Such feature is useful for those who would like to try sentiment analysis without much prior knowledge. iFeel can also help researchers evaluate a new method by providing a single platform to access existing tools without having to implement them. For example, a researcher may evaluate his new lexical dictionary that is built for a specific task such as testing sentiments related to automobiles, for instance, sentiments related to how people feel driving through a traffic jam in a city or a rural road.

Compared to other Web applications [9,15] that also provide sentiment analysis, iFeel is unique in that it allows access to not just one but *multiple* methods for comparison. In particular, by allowing users to explicitly determine weights for combining different

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2744-2/14/04.
http://dx.doi.org/10.1145/2567948.2577013.

methods for the Combined-Method, users can optimize the coverage and F-measure to meet their taste. For instance, one may increase the weight for SASA when handling data on politics, as SASA was initially tested on political debates. The iFeel system has an easy interface and allows users to either directly type in a message or upload a text file up to 10,000 lines of messages. The result of the sentiment analysis is presented both as numbers as well as graphical visualizations.

The remainder of this paper is organized as follows. We first briefly describe the sentiment methods implemented in the iFeel system. We then describe how we built the Web tool by laying out its architectural overview and the interface design. Next we discuss the performance of the iFeel system based on a wide range of test examples. Finally, we conclude and discuss future work.

2. SENTIMENT DETECTION METHODS

Existing sentiment analysis tools cover different mood categories such as anger, happiness, fatigue, etc. In order to conduct apple-to-apple comparison, iFeel groups these varying scales of sentiments into two representative categories: positive affect and negative affect. This section describes how this adaptation was done for each method implemented in the iFeel system.

2.1 PANAS-t

PANAS-t [11] is a psychometric scale for detecting mood fluctuations within Twitter and is based on another method, called Positive Affect Negative Affect Scale (PANAS) [17], which was built for analyzing structured text. PANAS-t covers eleven mood categories: joviality, assurance, serenity, surprise, fear, sadness, guilt, hostility, shyness, fatigue, and attentiveness, where the strength of the method is at tracking any increase or decrease in sentiment levels over time. We grouped joviality, assurance, serenity, and surprise as the positive affect; fear, sadness, guilt, hostility, shyness, and fatigue as the negative affect. We exclude attentiveness, as it is considered to have neutral affect.

2.2 Emoticons

This method extracts sentiments in texts using a large set of common emoticons like the smiley :) filtered from Web that express positive and negative sentiments. The complete list of emoticons we considered are listed in [10].

2.3 SentiWordNet

SentiWordNet [8] is a tool that is widely used in opinion mining, and is based on an English lexical dictionary called WordNet. This lexical dictionary groups adjectives, nouns, verbs, and other grammatical classes into synonym sets called synsets. To assign polarity, we considered the average scores of all associated synsets of a given text. We then take the relative strength of positive and negative affect in a given text.

2.4 Happiness Index

Happiness Index [7] is a sentiment scale based on the Affective Norms for English Words (ANEW) [2], which is a collection of 1,034 words commonly used associated with their affective dimensions of valence, arousal, and dominance. Happiness Index is scaled between 1 and 9, depending on the amount of happiness inferred from text. We consider the range between 1 to 4 as negative and the range of 5 to 9 as positive.

2.5 SentiStrength

SentiStrength [14] is a mix of supervised and unsupervised classification methods, including simple logistic regression, SVM, J48 classification tree, JRip rule-based classifier, SVM regression, and Naïve Bayes. This method extends the existing LIWC dictionary [13], which is made for structured text, to include a wide range of OSN contexts. LIWC is a text analysis tool that evaluates, among cognitive and structural components, the emotional (positive and negative affects) of a given text.

2.6 SASA

SailAil Sentiment Analyzer (SASA) [16] is based on machine learning techniques similar to SentiStrengh. The tool has been evaluated by the Amazon Mechanical Turk, where the turkers were recruited to label 17,000 Twitter messages related to the 2012 US Election as positive, negative, neutral, or undefined. In iFeel, we are using a model trained by authors of the method with a Naive Bayes classifier for detect positivity and negativity of a given text.

2.7 SenticNet

SenticNet [5] is based on artificial intelligence and semantic Web techniques. The tool uses Natural Language Processing (NLP) techniques to create a polarity for nearly 14,000 concepts and was evaluated in measuring the level of polarity in opinions of patients about the National Health Service in England [3]. SenticNet uses the affective categorization model Hourglass of Emotions [4] that provides an approach that classify messages as positive and negative of a given text.

3. THE IFEEL SYSTEM

3.1 System Design

The high-level workflow behind iFeel's components is depicted in Figure 1. The flow of this architecture follows the steps through which a user takes when she uploads a file for sentiment analysis. The process begins when the logged in user, who uploads a file in the system (Step 1). This file must be in the format accepted by the system, which consists of messages delimited by line breaks. The user can monitor the processing status of the file in *My Data* page, previous showed in Figure 4.

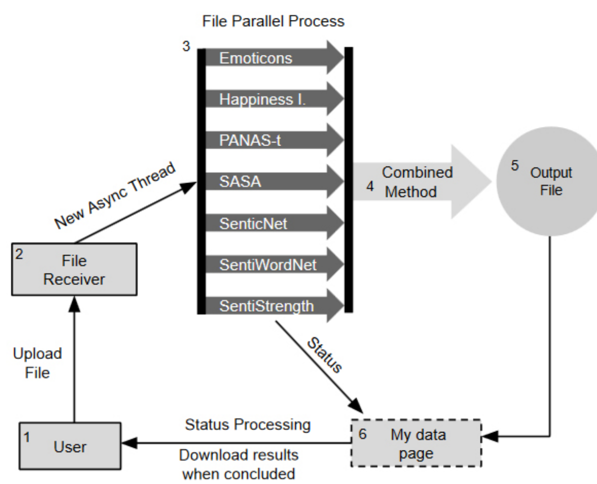

Figure 1: Overview of the iFeel architecture

Upon completion of the file upload (Step 2), iFeel creates an asynchronous thread for the seven methods that concurrently process the file, where each process is devoted to the first seven sentiment analysis tools described in the previous section. Methods are executed in parallel and the output of each process is stored in its own file (Step 3). Each process has its own rules in handling the data for the natural language processing of sentiment detection (e.g., parsing and cleaning texts, removing stop words, analyzing grammatical structures).

Because the existing methods cover different techniques such as from lexical base to machine learning, combining the results can achieve a highest coverage in sentiment analysis and relatively high prediction performance. This is what we call the Combined-Method in iFeel. Coverage in sentiment detection represents the fraction of data whose sentiment is revealed and accuracy represents the fraction of data whose sentiment is correctly judged. The combination is made in the two steps (Step 4). First, for each line of the input file, take the result of each method. Second, give a higher weight for the method that obtained best F-measure or exactly following the weights given by the user if there is any (see Figure 4). Finally, the results for existing sentiment methods and the Combined-Method are stored in the output file (Step 5), which is then made available to the user in the *My Data* page along with graphical visualization to help user interpret the result (Step 6).

3.2 System Performance

We measured the performance of iFeel in two ways. One is efficacy, which examines the gains from comparing and combining existing methods. Another is scalability, which tests how much time is needed to process big social data.

In terms of efficacy, we further consider two sub measures: accuracy and coverage, as described in the earlier section. The sentiment analysis methods in iFeel are each based on a different set of base techniques: some are based on lexical methods (e.g., PANAS-t, Emoticons, SentiWordNet and Happiness Index) and others are based on machine learning (e.g., SentiStrength and SASA). These differences naturally lead to varying degrees of coverage and accuracy. The relationship between accuracy and coverage across all methods are shown in Figure 2, indicating that no single method is a winner. Overall Combined-Method achieves the highest level of coverage as well as relatively high F-measures. This clearly demonstrates the need for combining and comparing various sentiment analysis methods.

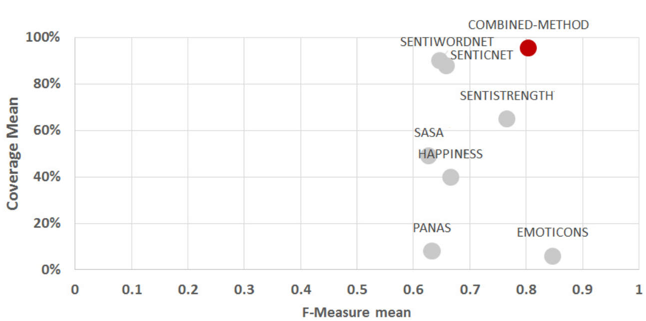

Figure 2: Trade off between the coverage vs F-measure for methods, including the Combined-method

Next, in order to test scalability of iFeel, we analyze the system capacity to handle a given size of input file. We used tweet data for test, which is popular data used for sentiment analysis. Figure 3 shows the CPU time for all methods across tweet files of increasing sizes, which are drawn from large tweet dataset in [6]. The tweet files have exponentially increasing number of tweets from 1,000 to 10 million. All tests were executed on a Dell Desktop, with Intel(R) Xeon(R) Processor (2.53GHz) with 24 Cores, and 94 Gigabytes of RAM, in a Ubuntu version 12.04.3.

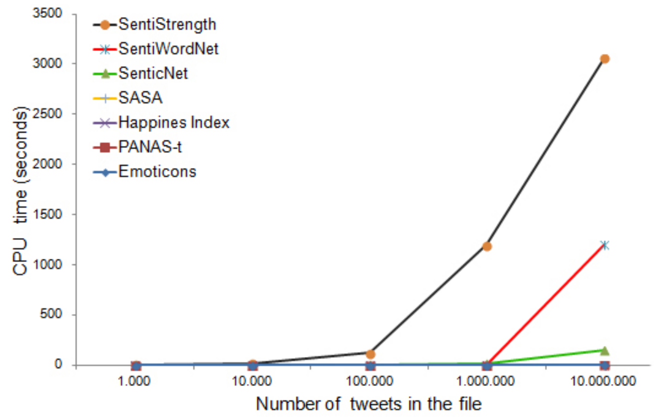

Figure 3: Performance of iFeel as a function of input file size

The figure presents the methods sorted from faster to the slower. SentiStrength, SentiWordNet, and SenticNet were the methods that took the most time to execute. This result is crucial for the efficacy of iFeel, because it can limit the size of input data. The execution time varied from one method to another. SentiStrength, for example, took more than 115 seconds of CPU time when treating an input file that contained more than 100,000 tweets. SentiWordNet and SenticNet scaled better to handling nearly one million tweets. As we can see in the figure, all methods executed faster in datasets with less than 10,000 tweets. Given that an average person tweets a few times a day (i.e., known average is 1.85 tweets per day) and has one hundred followers, 10,000 tweets could represent the amount of tweets a user reads in a given month. Overall, the test demonstrates that even when the size of the input file grows exponentially, the iFeel system does not require the same increasing amount of time for handling the input data.

4. SYSTEM DEMONSTRATION

We show a snapshot of the Web interface in Figure 4, where a user has entered the message "*It's not that I hate you, I just strongly disagree with you =/*". We can see the sentiment analysis result of this message on the bottom of the input interface. iFeel visualizes how polarity of the given message is judged differently by existing sentiment analysis methods. As we can see, in this example all methods detected sentiments in the message. SentiWordNet, PANAS-t and SASA detected positive sentiments, whereas the remaining methods detected negative sentiments. If we change our query to "*I never make the same mistake twice. Three...four times maybe, but never twice :)*", almost all methods tag the text as negative, Emoticons tagged it as positive and PANAS-t did not detected any feeling.

For the Combined-Method, users can fine tune the weights manually to optimize the coverage and F-measure depending on the input. For instance, users may increase the weight for SASA if they are handling politics-related data, because SASA was mainly trained for political debates. The demo video describes how a user can upload an input file and retrieve sentiment results with iFeel at `http://www.youtube.com/watch?v=7NsT1E-VWUY`. By publishing the iFeel tool free, we hope that future research and industry projects could easily utilize sentiment analysis.

5. CONCLUSION & FUTURE WORK

While the need for sentiment analysis is growing, there has been little effort in comparing the various sentiment analysis methods.

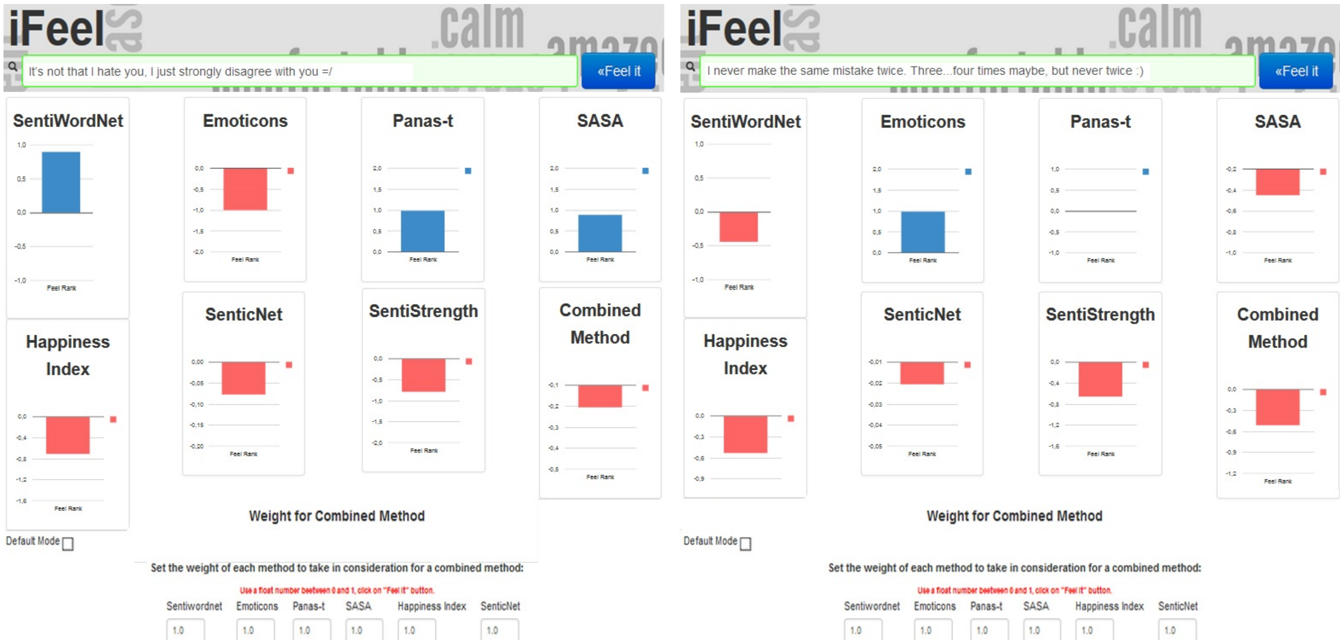

Figure 4: Screen snapshots of the iFeel system for two differen text input and user interface for fine tuning (bottom of the figures)

This paper presented iFeel, a Web tool for measuring the level of positive affect and negative affects that is based on consolidation of 8 existing sentiment analysis tools including PANAS-t, Emoticons, SentiWordNet, Happiness Index, SentiStrength, SASA, SenticNet, and Combined-Method. The iFeel system employs user friendly interface, by allowing text input to be typed directly or uploaded as a text file. We believe anyone from those not keen on programming to researchers and companies interested in sentiment analysis can utilize the iFeel system as a useful online resource.

As a natural extension of this tool, we would like to add more existing methods for detecting sentiments such as OpinionFinder [18] and also expand the categories of sentiments beyond positive and negative affects, thereby including more sophisticated human moods like guilt, anger, and sarcasm. Futhermore, we would like to offer for future iFeel's users guidelines for better setting weights of methods, based in analysis and comparisons of each one in bigger and different databases.

Acknowledgements

This work was funded by the Brazilian National Institute of Science and Technology for the Web (MCT/CNPq/INCT grant number 573871/2008-6) and grants from CNPq, CAPES and FAPEMIG, as well as the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science, ICT & Future Planning (2011-0012988).

6. REFERENCES

[1] Senticnet 2.0. http://sentic.net/downloads/.

[2] M. M. Bradley and P. J. Lang. Affective norms for English words (ANEW): Stimuli, instruction manual, and affective ratings. Technical report, University of Florida, 1999.

[3] E. Cambria, A. Hussain, C. Havasi, C. Eckl, and J. Munro. Towards crowd validation of the uk national health service. In *ACM WebSci*, 2010.

[4] E. Cambria, A. Livingstone, and A. Hussain. The hourglass of emotions. Lecture Notes in Computer Science. Springer, 2011.

[5] E. Cambria, R. Speer, C. Havasi, and A. Hussain. Senticnet: A publicly available semantic resource for opinion mining. In *AAAI Fall Symposium Series*, 2010.

[6] M. Cha, H. Haddadi, F. Benevenuto, and K. P. Gummadi. Measuring User Influence in Twitter: The Million Follower Fallacy. In *ICWSM*, 2010.

[7] P. S. Dodds and C. M. Danforth. Measuring the happiness of large-scale written expression: songs, blogs, and presidents. *Journal of Happiness Studies*, 11(4):441–456, 2009.

[8] Esuli and Sebastiani. Sentwordnet: A publicly available lexical resource for opinion mining. In *LREC*, 2006.

[9] A. Go, R. Bhayani, and L. Huang. Sentiment140. http://www.sentiment140.com/.

[10] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. Comparing and combining sentiment analysis methods. In *COSN*, 2013.

[11] P. Goncalves, F. Benevenuto, and M. Cha. Panas-t: A pychometric scale for measuring sentiments on twitter. CoRR, 2013.

[12] X. Hu, L. Tang, J. Tang, and H. Liu. Exploiting social relations for sentiment analysis in microblogging. In *WSDM*, 2013.

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

[14] M. Thelwall. Heart and soul: Sentiment strength detection in the social web with sentistrength, 2013.

[15] M. Thelwall, K. Buckley, G. Paltoglou, C. Cai, and A. Kappas. Sentistrength. http://www.sentiment140.com/.

[16] H. Wang, D. Can, A. Kazemzadeh, F. Bar, and S. Narayanan. A system for real-time Twitter sentiment analysis of 2012 U.S. presidential election cycle. In *ACL*, pages 115–120, 2012.

[17] D. Watson and L. Clark. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of Personality and Social Psychology*, 54(1):1063–1070, 1985.

[18] T. Wilson, P. Hoffmann, S. Somasundaran, J. Kessler, J. Wiebe, Y. Choi, C. Cardie, E. Riloff, and S. Patwardhan. Opinionfinder: a system for subjectivity analysis. In *EMNLP*, 2005.