FTC 2016 - Future Technologies Conference 2016
6-7 December 2016 | San Francisco, United States

Fine-Grained Sentiment Analysis of Social Media with Emotion Sensing

Zhaoxia WANG, Chee Seng CHONG, Landy LAN, Yinping YANG, Seng Beng HO and Joo Chuan TONG
Social and Cognitive Computing Department
Institute of High Performance Computing (IHPC)
Agency for Science, Technology and Research (A*STAR)
Singapore
wangz@ihpc.a-star.edu.sg

***Abstract*—Social media is arguably the richest source of human generated text input. Opinions, feedbacks and critiques provided by internet users reflect attitudes and sentiments towards certain topics, products, or services. The sheer volume of such information makes it effectively impossible for any group of persons to read through. Thus, social media sentiment analysis has become an important area of work to make sense of the social media talk. However, most existing sentiment analysis techniques focus only on the aggregate level, classifying sentiments broadly into positive, neutral or negative, and lack the capabilities to perform fine-grained sentiment analysis. This paper describes a social media analytics engine that employs a social adaptive fuzzy similarity-based classification method to automatically classify text messages into sentiment categories (positive, negative, neutral and mixed), with the ability to identify their prevailing emotion categories (e.g., satisfaction, happiness, excitement, anger, sadness, and anxiety). It is also embedded within an end-to-end social media analysis system that has the capabilities to collect, filter, classify, and analyze social media text data and display a descriptive and predictive analytics dashboard for a given concept. The proposed method has been developed and is ready to be licensed to users.**

***Keywords*—sentiment classification; sentiment analysis; opinion mining; social media; social adaptive fuzzy similarity; emotion**

I. Introduction

Social media, such as Twitter, Facebook and Chinese Weibo, is overwhelmingly the go-to platform for internet users to share their comments or experiences towards certain products, services or policies. It is a gold mine for those who appreciate the value of understanding public sentiment.

There are various compelling use cases of social media sentiment analysis: consumers referring to online reviews to help them make better purchase decisions; businesses eager to understand market preferences in order to improve their offerings; politicians aspiring to gauge public response to their policies or speeches. Not surprisingly, one of the hottest areas of research in social analytics is sentiment analysis.

Sentiment analysis aims to understand the sentiment polarity of data [1]. A lot of social media analysis tools are now available to perform such analysis, such as Stanford NLP's natural language processing tool [2], Facebook Insights on Facebook and TweetStats on Twitter. However, these existing analysis technologies focus on finding the aggregate level sentiment such that the sentiment polarity is typically one of two categories (“positive” and “negative”) or three categories (with the addition of “neutral”) [2]. If finer-grained sentiment analysis can be achieved, it will yield more specific and more actionable results with detailed negative emotion subcategories such as anger, sadness, and anxiety or positive emotion subcategories such as happiness and excitement [3].

In this paper, we describe a new method for fine-grained classification of social media sentiment. The actual sentiments as well as detailed emotions were identified in accordance with industry needs. The basis for this method is a series of patents filed in [4] [5] and [6].

The rest of this paper is organized as follows. Section II discusses the existing sensing technologies. Section III presents the proposed methodology of fine-grained sentiment analysis. Section IV examines the performance of the proposed method using real world social media data. Lastly, in Section V, we conclude this study.

II. Existing Technologies

Sentiment analysis methods can be broadly categorized into two types: learning-based and lexical-based [7] [8]. Learning-based method uses known properties derived from labelled training data to make predictions about unlabelled new data. In text data, it derives the relationship between the features of the text segment. Some examples of learning-based methods are the Naïve Bayes (NB) classifier [9] [10], Maximum Entropy (MaxEnt) classifier [11], support vector machine (SVM) [12] [13] and Extreme Learning Machine (ELM) [14] [15].

To be effective, models using such learning-based methods typically require a sufficiently large labelled training dataset [15] [16] to achieve an acceptable classification accuracy [16] [17]. However, in most social media contexts, it is difficult to determine what size of labelled dataset qualifies as being sufficient because the diversity of the social discussion is not known *a priori* [3] [12]. In addition, the labelling task would be costly or even prohibitive [3] [7] [12], not to mention wasteful because the training results could not be readily applied to other datasets.

On the other hand, lexical-based methods typically search a text for sentiment or emotion indicators specified in the existing lexicons used [7] [18] [19] [20]. The effects of the

The work is supported by A*STAR Joint Council Office Development Programme “Social Technologies+ Programme”.

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

indicators are then aggregated in order to derive the dominant polarity of the text. Compared to learning-based methods, lexical-based methods are easier to be applied across different datasets and costly labelling tasks are not required (no training needed).

However, there are shortcomings in the current lexicon-based methods. It is hard to create a unique lexical-basic dictionary for use and test in different applications. Hence, the existing methods use cleaned samples that are manually created. However, such data is different from real-world social media data and only real-world social media data can produce true insight for organizations. The other shortcoming, as mentioned earlier in Section I, is the lack of fine-grained sensing capability [3] that provides detailed emotion identification. These shortcomings of the lexicon based methods are also the limitations of the current learning based methods.

The research on emotion has a long evolutionary history and emotion research activities have increased significantly over the past two decades. One of the earlier efforts in emotion research was the effort of Shaver et al. [21]. Shaver et al. grouped emotions into prototypes on the assumption that different parts of emotion knowledge tend to make up an organized whole [21]. In their experiment, they first selected a group of words and had them rated based on whether the word was an emotion. Using the typical prototyping approach, they managed to develop an abstract-to-concrete emotion hierarchy.

Psychologists Ortony and Turner argued against the view that basic emotions are psychologically primitive [22]. They proposed that all emotions are discrete and independent and are related to each other through a hierarchical structure.

Ekman's emotion model is based on the argument that there are distinctive facial expressions [23]. In this model, the emotions are treated as discrete, measurable, and physiologically distinct. Each of the emotions is a family of related states and this is consistent with Shaver's model [21].

Plutchik enhanced Ekman's biologically driven perspective and developed the "wheel of emotions" [24]. He constructed a wheel-like diagram of emotions to visualize the basic emotions and grouped the primary emotions into a positive vs negative category, e.g., joy versus sadness; anger versus fear; trust versus disgust; and surprise versus anticipation [24] [25].

On the other hand, Alena et al. also took the typical lexicon approach which leveraged and enhanced the above emotion models [26]. They had each emotion word annotated by expert annotators and compiled the words into an emotion dictionary [26].

The above efforts have contributed greatly to emotion research and identification; however, there have been rare research efforts that make use of them to integrate emotion analysis into sentiment analysis and enhance the capability of the sensing technologies.

In this paper, we leverage the above emotion research to develop fine-grained sentiment analysis technologies and implement a fine-grained emotion sensing method to address the limitations of the existing technologies. In addition, we applied the method to a real-world case which provides an answer to the key question of whether public sentiments are useful in support of effective social sensing and policy management. The resulting insights enable decision-makers to formulate strategies and fine-tune the quality of their products, services, and policies.

III. Proposed Technique

Detecting the attitude or emotions of a user with respect to certain topics or certain domains is the aim of sentiment analysis [2] [4] [5] [6]. Various techniques are leveraged for the design of the proposed method [3]. These techniques include: linguistic inquiry and word count (LIWC) method [27], the affective norms for English words (ANEW) approach for assigning normative emotional ratings to text [28], fuzzy logic [29] and emotion theories [21] [22] [23] [24].

The proposed method pays special design attention to the challenges of real-world datasets. It uses an innovative social adaptive fuzzy rule inference technique with linguistics processors designed to minimize semantic ambiguity. This is combined with multi-source lexicon integration and development to derive dominant valence (positive, negative, neutral, mixed) as well as prominent emotions (e.g., anger, sadness, anxiety, satisfaction, happiness, excitement).

A. *Design Features and Components of the Proposed Method*

The backbone of the proposed method is a social adaptive fuzzy inference algorithm that mimics human interpretations of the expression of attitudes and emotions in online social network contexts. There is also a built-in advanced linguistic processing unit that contains the following sub-modules: sentence decomposers, negation handlers, amplifier, diminisher handlers, etc. [4] [5] [6] In addition, the proposed method is empowered by built-in linguistic lexicons from a variety of sources, including a dictionary of emotion words and phrases from Standard English, Internet/social media slang and local languages. It also includes emoticons. With more linguistics-enhanced fuzzy similarity rules to handle sentiment classification and without relying on any training data, it is thus able to achieve the same level of measurement accuracy with less human input than simple lexicon-based and learning-based methods.

The domain knowledge was obtained by using the domain lexicon knowledge extraction algorithm [30] to form domain lexicon dictionaries. In addition, to enhance domain adaptability, an expert user can further configure the domain knowledge through the specification of a seed lexicon. For example, the expert user can add to the lexicon the phrase "salary lower" (in the company review domain), and to remove from the lexicon the word "smart" (as in "smart watch" in the smart phone domain). This can achieve a higher measurement accuracy than simple lexicon-based and learning-based methods.

B. *Social Media Analysis System*

To make the proposed method useful for real-world datasets, we implement it within an end-to-end social media analysis system. The system consists of 6 modules, including social data collectors, noise filters, sentiment & emotion

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

analysis engine module, predictive analyser, results viewer and database. Fig. 1 shows the system's architecture.

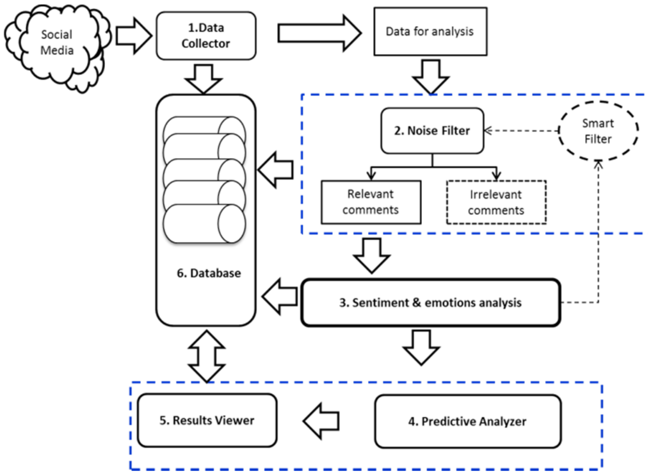

Fig. 1. Social Media Analysis System [4] [5]

The "Data Collector" crawls raw data from various Internet sites, including forums, Twitter, and other blogs. Depending on whether the data sources provide programmatic interface to read data (such as Twitter's REST API based on keywords and Streaming API that reads data constantly), the module is a collection of codes that collects data and passes them to the "Noise Filter" module before processing.

The "Noise Filter/Smart Filter" removes noisy "meaningless data", such as advertisements, useless content which does not include any comment information, and other content-specific noises. Raw data are pre-processed by "Noise Filter" to determine if they are relevant data or irrelevant data. The relevant data are passed to an optional sub-module, "User-defined filter", that allows the user to define rules to further trim out some data. These filtering ensures that data passed to the "sentiment analysis engine" module is relevant to the intended concept for further analysis.

The "Predictive Analyzer" performs the task of predictive analysis of important outcomes such as sales volumes and reputation crisis so that it can be used for important business activities of forecasting, monitoring and action strategizing. It includes two key components, 1) the predictor/feature set and 2) the predictive algorithm pool. The output of sentiment and emotion analysis (i.e., such as positive, negative, neutral and mixed sentiments, and anger, sadness and anxiety emotions) serves as a new predictor/feature on top of existing predictors/features.

Consumer preference analysis, anomaly identification and time-series analysis for sales forecasting will be realized through leveraging the output of the sentiment and emotion analysis engine combined the other results obtained through the predictive algorithm pool.

IV. A Realworld Case Study through the Social Media Analysis System

While understanding the valence of sentiments helps to assess overall public reactions, the understanding of emotion further improves the assessment of the situation, particularly negative emotions requiring attention from decision-makers and crisis managers.

As shown in Fig. 2, the final outcome of any text will be the sentiment categories and fine-grained emotions [2] [23] [24] [25]. Fig. 2 (a) shows the sentiments and Fig. 2 (b) shows the fine-grained emotions the system outputs.

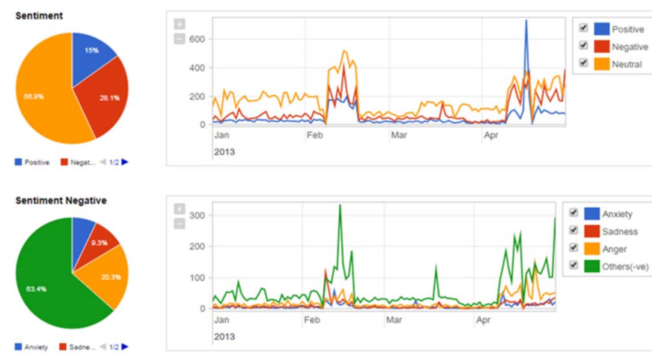

Fig. 2. Positive or negative sentiment can be further broken-down into fine-grained emotions. (a) Sentiments; (b) Break down of negative sentiment into fine-grained emotions

For real-time testing of the proposed method, the interface of real-time data analysis is illustrated in Fig.3. The tweets are used as a test case to illustrate real time data collection, analysis and visualization. The data containing geographic information is displayed in the form of a map.

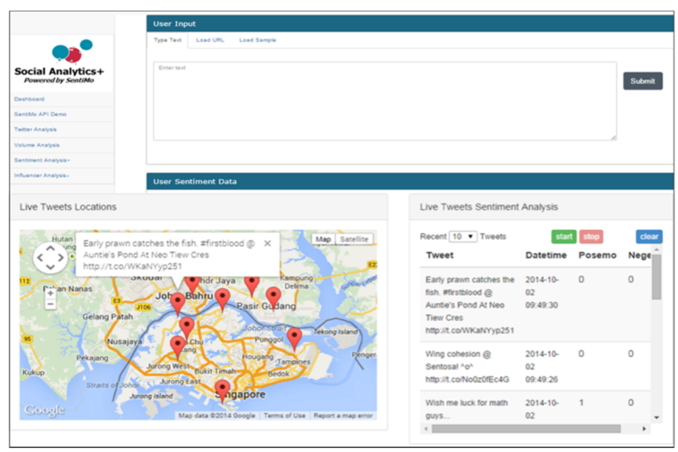

Fig. 3. Part of the interface of the social media analytics system

V. Conclusion

This research describes a social media analytics method that is able to perform fine-grained sentiment and emotion analysis. This research offers new ideas for designing a robust method that leverages adaptive learning capabilities, fuzzy logic, and social science concepts in handling fine-grained sensing classification (sentiments as well as emotions) in textual datasets. There are ample opportunities to apply the proposed method to other sectors such as the healthcare, corporate, leisure, public and private sectors to help them to understand their customers better, identify the relevant risks, and improve their products and services.

978-1-5090-4171-8/16/$31.00 ©2016 IEEE

ACKNOWLEDGMENT

The "SentiMo-Advanced Social Media Analytics" team provided great support for the discussion on issues related to emotion and the algorithm development and implementation. The authors would also like to thank Dr. Kenneth Kwok, Dr Paul Yang and their team for the discussion on issues related to emotion and knowledge building.

The proposed method and system had been developed and the API version is ready to be licensed: https://www.etpl.sg/innovation-offerings/technologies-for-license/tech-offers/2087.

REFERENCES

[1] A. Trilla and F. Alías, "Sentence-based sentiment analysis for expressive text-to-speech," Audio, Speech, Lang. Process. IEEE Trans., vol. 21, no. 2, pp. 223–233, 2013.

[2] R. Socher, A. Perelygin, and J. Wu, "Recursive deep models for semantic compositionality over a sentiment treebank," Proc. Conf. Empir. methods Nat. Lang. Process., vol. 1631, p. 1642, 2013.

[3] Z. Wang, J. C. Tong, and D. Chan, "Issues of social data analytics with a new method for sentiment analysis of social media data," in 2014 IEEE 6th International Conference on Cloud Computing Technology and Science, 2014, pp. 899–904.

[4] Z. Wang, R. S. M. Goh, and Y. Yang, "A method and system for sentiment classification and emotion classification," Patent Cooperation Treaty (PCT) Application, PCT/SG2015/050469, 2014.

[5] Z. Wang, R. S. M. Goh, and Y. Yang, "SentiMo-A Method and system for fine-grained classification of social media sentiment and emotion patterns," Singapore Patent Application10201407766R, 2014.

[6] Z. Wang and J. C. Tong, "ChiEFS-A method and system for Chinese hybrid multilingual emotion fine-grained sensing of text data," Singapore Patent Application No. 10201601413Q, 2015.

[7] P. Gonçalves and M. Araújo, "Comparing and combining sentiment analysis methods," Proc. first ACM Conf. Online Soc. networks. ACM., pp. 27–38, 2013.

[8] B. Yuan, Y. Liu, and H. Li, "Sentiment classification in Chinese microblogs: Lexicon-based and learning-based approaches," Int. Proc. Econ. Dev. Res., vol. 68, pp. 1–6, 2013.

[9] J. Ortigosa-Hernández, J. D. Rodríguez, L. Alzate, M. Lucania, I. Inza, and J. a. Lozano, "Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers," Neurocomputing, vol. 92, pp. 98–115, Sep. 2012.

[10] X. Glorot, A. Bordes, and Y. Bengio, "Domain adaptation for large-scale sentiment classification: A deep learning approach," Proc. 28th Int. Conf. Mach. Learn., pp. 513–520, 2011.

[11] H. Ji, H. Deng, and J. Han, "Uncertainty reduction for knowledge discovery and information extraction on the World Wide Web," Proc. IEEE, vol. 100, no. 9, pp. 2658–2674, Sep. 2012.

[12] B. Gokaraju, S. S. Durbha, R. L. King, S. Member, and N. H. Younan, "A machine learning based spatio-temporal data mining approach for detection of harmful algal blooms in the Gulf of Mexico," IEEE J. Sel. Top. Appl. earth Obs. Remote Sens., vol. 4, no. 3, pp. 710–720, 2011.

[13] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," Assoc. Comput. Linguist., vol. 35, no. 3, 2009.

[14] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," Neurocomputing, vol. 70, no. 1–3, pp. 489–501, Dec. 2006.

[15] Z. Wang and Y. Parth, "Extreme Learning Machine for Multi-class Sentiment Classification of Tweets," Proc. ELM-2015, Springer Int. Publ. 2016, vol. 1, pp. 1–11, 2016.

[16] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," Proc. ACL-02 Conf. Empir. methods Nat. Lang. Process. Assoc. Comput. Linguist., vol. 10, pp. 79–86, 2002.

[17] E. Haddi, X. Liu, and Y. Shi, "The role of text pre-processing in sentiment analysis," Procedia Comput. Sci., vol. 17, pp. 26–32, Jan. 2013.

[18] R. Feldman, "Techniques and applications for sentiment analysis," Commun. ACM, vol. 56, no. 4, p. 82, Apr. 2013.

[19] I. Maks and P. Vossen, "A lexicon model for deep sentiment analysis and opinion mining applications," Decis. Support Syst., vol. 53, no. 4, pp. 680–688, Nov. 2012.

[20] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," World Wide Web, vol. 17, pp. 723–742, Jun. 2014.

[21] P. Shaver, J. Schwartz, D. Kirson, and C. O'Connor, "Emotion knowledge: further exploration of a prototype approach.," J. Pers. Soc. Psychol., vol. 52, no. 6, pp. 1061–1086, 1987.

[22] A. Ortony and T. J. Turner, "What ' s Basic About Basic Emotions ?," Psychol. Rev., vol. 97, no. 3, pp. 315–331, 1990.

[23] P. Ekman, "An argument for basic emotions," Cognition & Emotion, vol. 6. pp. 169–200, 1992.

[24] D. Chafale and A. Pimpalkar, "Review on Developing Corpora for Sentiment Analysis Using Plutchik ' s Wheel of Emotions with Fuzzy Logic," Int. J. Comput. Sci. Eng., vol. 2, no. 10, 2014.

[25] R. Plutchik, "The Nature of Emotions Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice," Am. Sci., vol. 89, no. 4, pp. 344–350, 2001.

[26] A. Neviarouskaya, H. Prendinger, and M. Ishizuka, "Textual affect sensing for sociable and expressive online communication," Affect. Comput. Intell. Interact., pp. 218–229, 2007.

[27] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," J. Lang. Soc. Psychol., vol. 29, no. 1, pp. 24–54, Dec. 2010.

[28] A. P. Soares, M. Comesaña, A. P. Pinheiro, A. Simões, and C. S. Frade, "The adaptation of the Affective Norms for English Words (ANEW) for European Portuguese," Behav. Res. Methods, vol. 44, pp. 256–269, 2012.

[29] J. M. Mendel and D. Wu, "Challenges for perceptual computer applications and how they were overcome," IEEE Comput. Intell. Mag., vol. 7, no. 3, pp. 36 – 47, 2012.

[30] Z. Wang, J. C. Tong, P. Ruan, and F. Li, "Lexicon knowledge extraction with sentiment polarity computation," IEEE Int. Conf. Data Min. Ser. (ICDM), SENTIRE, Accept., 2016.

978-1-5090-4171-8/16/$31.00 ©2016 IEEE