How Do I Look? Publicity Mining From Distributed Keyword Representation of Socially Infused News Articles

Yu-Lun Hsieh

SNHCC, Taiwan International Graduate Program, Institute of Information Science,
Academia Sinica, and Department of Computer Science,
National Chengchi University, Taipei, Taiwan

Yung-Chun Chang, Chun-Han Chu, Wen-Lian Hsu
Institute of Information Science, Academia Sinica, Taipei, Taiwan
Email: {morphe,changyc,johannchu,hsu}@iis.sinica.edu.tw

Abstract

Previous work on opinion mining and sentiment analysis mainly concerns product, movie, or literature reviews; few applied this technique to analyze the publicity of person. We present a novel document modeling method that utilizes embeddings of emotion keywords to perform reader's emotion classification, and calculates a publicity score that serves as a quantifiable measure for the publicity of a person of interest. Experiments are conducted on two Chinese corpora that in total consists of over forty thousand users' emotional response after reading news articles. Results demonstrate that the proposed method can outperform state-of-the-art reader-emotion classification methods, and provide a substantial ground for publicity score estimation for candidates of political elections. We believe it is a promising direction for mining the publicity of a person from online social and news media that can be useful for propaganda and other purposes.

1 Introduction

The Internet has grown into a powerful medium for information dispersion and social interaction, on which one can easily share experiences and emotions instantly. It has become a popular source for sentiment analysis and opinion mining, e.g., movie reviews (Pang et al., 2002; Turney, 2002), product reviews (Hu and Liu, 2004), and other subjects (Turney, 2002; Wilson et al., 2009). Moreover, human feelings can be quickly identified through automatic emotion classification, as these emotions reflect an individual's feelings and experiences toward certain subject matters (Turney, 2002; Wilson et al., 2009). Emotion classification aims to predict the emotion categories (e.g., happy, angry, or worried) to which the given text belongs (Das and Bandyopadhyay, 2009; Quan and Ren, 2009). There are two aspects of emotions regarding a piece of text, namely, the writer's and the reader's emotion. The former consists of the emotions expressed by the author, while the latter refers to the emotions that the readers of the text may possess after reading the text. Recognition of reader-emotion is different from that of writer-emotion and may be even more complicated (Lin et al., 2008; Tang and Chen, 2012). In particular, writers can directly express their emotions through sentiment words; in contrast, reader-emotions possess a more complex nature, as even common words can evoke different types of reader-emotions depending on personal experiences and knowledge of the readers (Lin et al., 2007). For instance, a news article with the title "The price of crude oil will rise 0.5% next week" is just objectively reporting an event without any emotion, but it may invoke emotions like *angry* or *worried* in its readers. In addition, it is possible that more sponsorship opportunities can be obtained from companies or manufacturers if the articles describing a certain product are able to promote greater emotional resonance in the readers. As online commerce becomes more and more prominent nowadays, a growing amount of customers rely on online reviews to determine their purchases. Meanwhile, news organizations observe increasing traffic on their online websites as opposed to paper-based publications. We believe that reader's

74

emotion analysis has a great potential in all domains and applications.

In light of the above rationale, in this work we attempt to capture the perception of readers toward public figures through recognizing reader's emotion from news articles. We propose a distributed emotion keyword vector (DEKV) representation for reader-emotion classification, from which we derive a novel method for publicity mining. It is a practice of monitoring the public opinion toward a certain human subject at a given period of time. Experiments show that DEKV outperforms other text categorization and reader-emotion classification methods; in turn, these results can be used to conduct publicity mining for propaganda and other public relations purposes.

2   Related Work

Articles are one of the most common medium for persons to convey their feelings. Identifying essential factors that affect emotion transition is important for human language understanding. With the rapid growth of computer mediated communication applications, such as social websites and micro-blogs, research on emotion classification has recently been attracting more attention from enterprises (Chen et al., 2010; Purver and Battersby, 2012). In general, a single piece of text may possess two types of emotions: writer-emotion and reader-emotion. The research of writer-emotion investigates the emotion expressed by the writer when writing the text. For example, Pang et al. (2002) pioneered the use of machine learning technique on sentiment classification of movie reviews into positive and negative emotions. Mishne (2005), and Yang and Chen (2006) used emoticons as tags to train SVM (Cortes and Vapnik, 1995) classifiers at the document or sentence level, respectively. In their studies, emoticons are taken as the answer, and textual keywords are considered as features. Wu et al. (2006) propose a sentence level emotion recognition method using dialogs as their corpus, in which "Happy", "Unhappy", or "Neutral" are assigned to each sentence as its emotion category. Yang et al. (2006) adopted Thayer's model (Thayer, 1989) to classify music emotions. Each music segment can be classified into four classes of moods. As for sentiment

analysis, Read (2005) used emoticons in newsgroup articles to extract relevant instances for training polarity classifiers.

On the other hand, the research of reader-emotion concerns the emotions expressed by a reader after reading the text. The writer and readers may view the same text from different perspectives, hence they do not always share the same emotion. Since the recent increase in the popularity of Internet, certain news websites, such as Yahoo! Kimo News, incorporate the Web 2.0 technologies that allow readers to express their emotions toward news articles. Classifying emotions from the readers' point of view is a challenging task, and research on this topic is relatively sparse as compared to those considering the writers' perspective. While writer-emotion classification has been extensively studied, only a few focused on reader-emotion classification. Lin et al. (2007) first described the task of reader-emotion classification on news articles and classified Yahoo! News articles into 8 emotion classes (e.g. happy, angry, or depressing) from the readers' perspectives. They combined unigram, bigram, metadata, and emotion categories to train a classifier for the reader-emotions toward news. Yang et al. (2009) automatically annotated reader-emotions on a writer-emotion corpus with a reader-emotion classifier, and studied the interactions between them. Furthermore, applications of reader-emotion categorization include learning linguistic templates for writing assistance (Chang et al., 2015). One can also collect public opinions toward political issues through emotion classification. Sarmento et al. (2009) used a rule-based method to collect a corpus of online comments for political opinion mining. Fang et al. (2012) extract contents from multiple sources on the same topic and quantify the differences within. An opinion formation framework was developed for content analysis of social media to conduct political opinion forecast (Sobkowicz et al., 2012).

What distinguishes this work from others is that we attempt to test the possibility of inferring publicity, or "likability", of a person by detecting the emotion of the public towards news about that person. Given enough unbiased data, this technique enables for propaganda and maintenance of good public image. Note that we do not aim to predict the probability of a person being elected, as such efforts
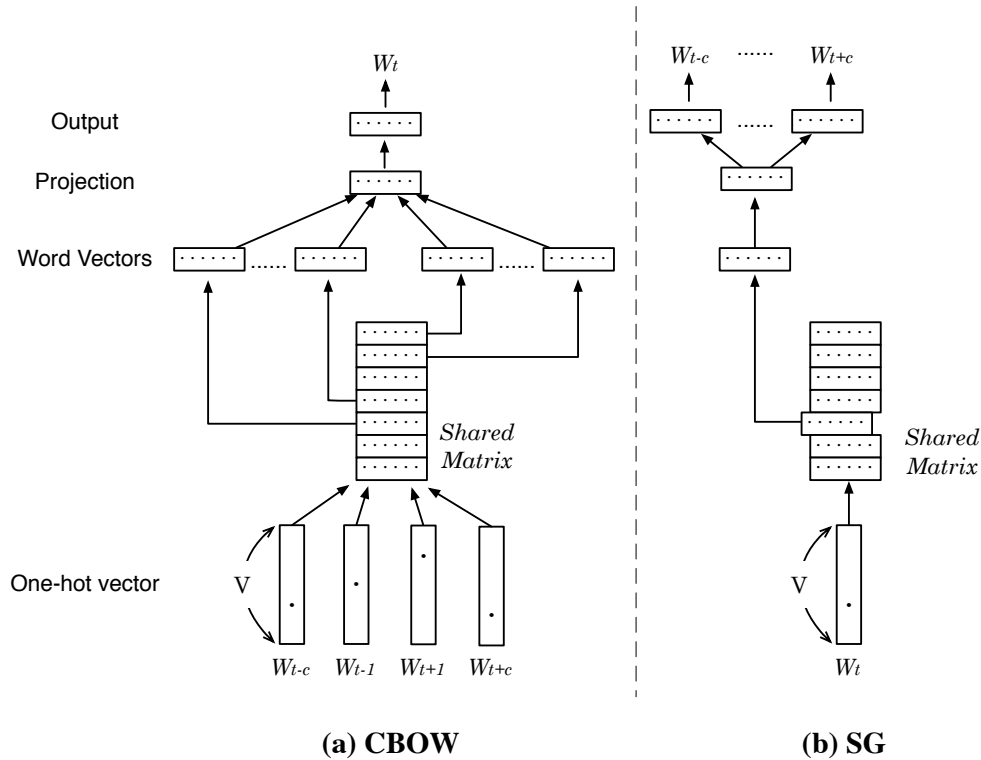
	W_t					W_{t-c} ······ W_{t+c}

Figure 1: (a) The CBOW model uses the context words $W_{t-c}, \cdots, W_{t+c}$ in the window as inputs to predict the current word W_t. (b) The SG model predicts words $W_{t-c}, \cdots, W_{t+c}$ in the context using the current word W_t as the input.

had already been made without showing promising results (Gayo-Avello, 2012).

3 Method

3.1 Distributed Word Representation

Bengio et al. (2003) proposed a neural network-based language model that motivated recent advances in natural language processing (NLP), including two word embedding learning strategies continuous bag-of-word (CBOW) and skip-gram (SG) (Mikolov et al., 2013a). The CBOW method is based on the distributional hypothesis (Miller and Charles, 1991), which states that words occur in similar contexts often possess similar meanings. This method attempts to learn a word representation that can capture the context information for each word. In contrast to traditional bag-of-word models, the CBOW model tries to obtain a dense vector representation (embedding) of each word directly (Mikolov et al., 2013a). The structure of the CBOW model is similar to a feed-forward neural network without non-linear hidden layers, as illustrated in

Fig. 1. It has been proven that this model can learn powerful representation of words and be trained on a large amount of data efficiently (Mikolov et al., 2013a). The SG model, being a simplified feed-forward neural network as well, differs from CBOW in that SG employs an inverse training objective instead for learning word representations (Mikolov et al., 2013a; Mikolov et al., 2013b; Le and Mikolov, 2014). The concept of SG model is illustrated in Fig. 1b. It attempts to predict words in the context by using the current words. In practice, SG tends to be more effective than CBOW when larger datasets are available (Lai et al., 2015).

3.2 Distributed Emotion Keyword Vectors for Reader-Emotion Classification

Building on top of the success of word embeddings, we propose the Distributed Emotion Keyword Vectors (DEKV) to model the reader-emotion of news articles. Chang et al. (2015) demonstrated that keywords are crucial in emotion classification, and motivated us to incorporate the distributed representation approach in the reader-emotion classification

76

$$LLR(w, E) = 2log \left(\frac{p(w|E)^k(1 - p(w|E))^m p(w|\neg E)^l(1 - p(w|\neg E))^n}{p(w)^{k+l}(1 - p(w))^{m+n}} \right) \quad (1)$$

task. To begin, word embeddings are learned from the corpus using the CBOW method. We then find a set of keywords for each emotion category using log likelihood ratio (LLR) (Manning and Schütze, 1999), which is related to the probability of a keyword being specific to this category. LLR value of each word w is calculated as follows. Given a training set with emotion categories, we first define $k = N(w \wedge E)$, $l = N(w \wedge \neg E)$, $m = N(\neg w \wedge E)$, and $n = N(\neg w \wedge \neg E)$, where $N(w \wedge E)$ denotes the number of documents that contain w and belong to emotion E, $N(w \wedge \neg E)$ denotes the number of documents that contain w but does not belong to emotion E, and so on. Then, we employ Eq. (1) to calculate LLR for w in the emotion E.

Finally, a document is represented as illustrated in Fig. 2, in which D_t is a weighted average of keyword vectors, and the weight λ_i for a keyword KW_i is its scaled LLR value. Note that if there is no keyword in a document, we use the average of all word embeddings in this document and compute cosine similarity against all keyword vectors to find the closest ones to represent this document. In this case, the number of keywords that are used to represent this unknown document is the same as that of each category. In essence, each document is projected onto a semantic space constructed by keyword vectors as illustrated in Fig. 3.
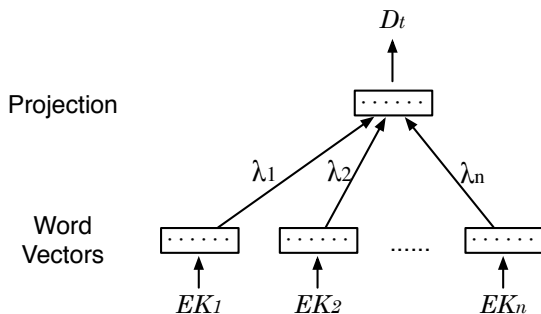

Figure 2: The DEKV model represents each target document D_t as emotion keyword vectors EK_i that are present in this document, weighted by scaled LLR scores λ_i.

4   Mining Publicity from Reader-Emotion

Our approach for mining publicity is by collecting online news articles centered around k specific public figures and determine the reader-emotion towards each of them, with the goal of identifying the public image of these people that can potentially affect how much the general population is willing to support them. We formulate the publicity of a person as a publicity score (PS) with positive or negative notion that can be summarized from identification of reader's emotion of articles. For this purpose, we only consider coarse-grained emotion categories (i.e., *positive* and *negative*). Thus, fine-grained emotion categories like *happy*, *warm*, and *odd* are considered to be "positive", while *angry*, *boring*, *depressing*, and *worried* being "negative". Moreover, PS is not only directly related to the public opinion towards an individual, but also affected by how his or her opponents are viewed. Hence, PS should jointly consider both directions of emotion. We define publicity score PS_i of a person i as:

$$PS_i = (P_i - N_i) + \sum_{j=1, j \neq i}^{k} (\frac{N_j - P_j}{k - 1}), \quad (2)$$

where P_i and N_i denotes the number of documents with positive and negative reader's emotion, respectively. Meanwhile, there are P_j and N_j articles with positive and negative reader-emotion for another person j. We postulate that PS_i also benefits from the negative publicity of other opposing people. However, since the negativity of the person j does not guarantee that the same amount of positivity from the public will automatically divert to a specific person, we divide the negative score of person j by the number of remaining candidates, $k - 1$, before adding that to PS_i. This way, we can quantify the publicity of, e.g. presidential candidates, and examine its relationship with other measurable metrics such as polls.

5   Experiments

We conduct two experiments to test the effectiveness of DEKV. The goal of the first one is detecting the
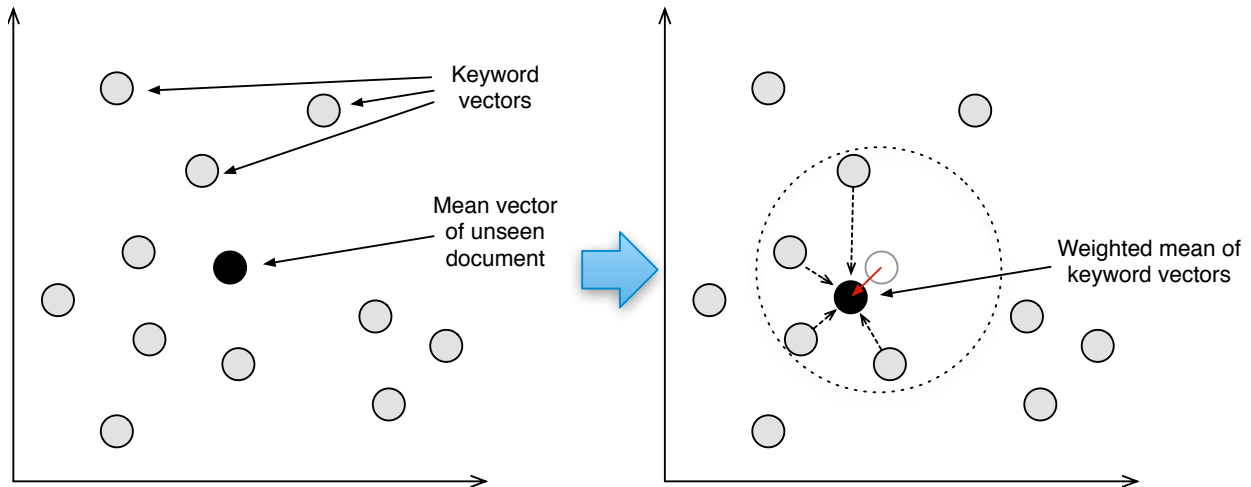
Figure 3: DEKV transforms a document with no keywords as a weighted average of closest keyword vectors.

reader-emotion of a news article, and the second one is inferring the publicity of famous public figures. Details are explained in the following sections.

5.1 Exp. I: Reader-emotion Classification

5.1.1 Dataset

We use a corpus containing 47,285 Chinese news articles[1] for evaluation. It is a very suitable testbed because it contains a socially infused feature of community voting. In particular, a reader of a news article can cast a vote expressing his or her feelings after reading this article with the emotion categories include *angry*, *worried*, *boring*, *happy*, *odd*, *depressing*, *warm*, and *informative*. Furthermore, only those with a clear statistical distinction between the highest vote and others determined by a *t*-test with 95% confidence level are included to ensure the validity of our experiments. The dataset is divided into training and test sets, containing 11,681 and 35,604 articles, respectively. Detail statistics of the corpus is listed in Table 1. Note that the evaluation excludes *informative* for it is not considered as an emotion (Lin et al., 2007; Lin et al., 2008).

5.1.2 Experimental Settings

DEKV is based on embeddings learned from the training set using CBOW with default settings in the toolkit (Řehůřek and Sojka, 2010), and LLR for keywords in each emotion category as weights. Each

Category	#Train	#Test	Total
Angry	2,001	4,326	6,327
Worried	261	261	522
Boring	1,473	1,473	2,946
Happy	2,001	7,344	9,345
Odd	1,526	1,526	3,052
Depressing	1,573	1,573	3,146
Warm	835	835	1,670
Informative	2,001	18,266	20,267
Total	11,681	35,604	47,285

Table 1: Descriptive statistics of the reader-emotion dataset.

article is represented as a weighted average of keywords and classified by linear SVM (Chang and Lin, 2011). Different combinations of the dimension in embeddings and number of keywords are tested, and the best one (500-dimension embeddings with 2,000 keywords/emotion) is compared with other methods described below. First, Naïve Bayes (McCallum et al., 1998) is used as baseline (denoted as NB). Next, we include LDA (Blei et al., 2003) as document representation and an SVM classifier (denoted as LDA). To examine the effect of our keyword extraction approach, an emotion keyword-based model that represents each article as a sparse vector and uses SVM as its classifier, denoted as KW, is also compared. In addition, we implement a method (denoted as CF) in (Lin et al., 2007) that uses extensive features including bigrams, words, metadata, and emotion category words. To inspect the effect of weighting, we also use the average of keyword vectors trained using the same parameters as DEKV, denote as *mean*.

---

[1] Collected from `http://tw.news.yahoo.com`

78

Details of the implementations of these methods are as follows. We employ CKIP (Hsieh et al., 2012) for Chinese word segmentation. The dictionary required by Naïve Bayes and LDA is constructed by removing stop words according to a Chinese stop word list provided by Zou et al. (2006), and retaining tokens that make up 90% of the accumulated frequency. In other words, the dictionary can cover up to 90% of the tokens in the corpus. As for unseen events, we use Laplace smoothing in Naïve Bayes, and an LDA toolkit is used to perform the detection of LDA. Regarding the CF, the words output by the segmentation tool are used. The information related to news reporter, news category, location of the news event, time (hour of publication) and news agency are treated as the metadata features. The extracted emotion keywords are used in place of the emotion category words, since the emotion categories was not released in (Lin et al., 2007).

To evaluate the effectiveness of these systems, we adopt the accuracy measures used by Lin et al. (2007); macro-average (avg_M) and micro-average (avg_μ) are selected to compute the average performance. These measures are defined based on a contingency table of predictions for a target emotion E_k. The accuracy $acc(E_k)$, macro-average avg_M, and micro-average avg_μ are defined as follows:

$$
acc(E_k)
$$
$$
= \frac{TP(E_k) + TN(E_k)}{TP(E_k) + FP(E_k) + TN(E_k) + FN(E_k)}, \quad (3)
$$

$$
avg_M = \frac{1}{m} \sum_{k=1}^{m} acc(E_k), \quad (4)
$$

$$
avg_\mu = \frac{acc(E_k) \times N(E_k)}{\sum_{k=1}^{m} N(E_k)}, \quad (5)
$$

where $TP(E_k)$ is the set of test documents correctly classified to the emotion E_k, $FP(E_k)$ is the set of test documents incorrectly classified to the emotion, $FN(E_k)$ is the set of test documents wrongly rejected, $TN(E_k)$ is the set of test documents correctly rejected, and $N(E_k)$ is the total number of documents in this emotion category.

5.1.3 Results

Table 2 lists performances of all methods. As a baseline, the Naïve Bayes classifier is a keyword statistics-based system which can only accomplish a mediocre performance. Since it only considers surface word weightings, it is difficult to represent inter-word relations. The overall accuracy of the Naïve Bayes classifier is 56.13%, with the emotion "Warm" only achieving 15.09% accuracy. On the contrary, the LDA yields a macro average accuracy of 74.12%, indicating its ability to select important topics for some emotion categories. However, KW is more effective in finding representative keywords using LLR as weights, obtaining 80.79% accuracy overall. Furthermore, it exhibits a more evenly distributed performance among categories than LDA. Next, CF achieves an overall accuracy of 85.69%, which may be attributed to its extensive feature engineering. It also obtains the highest accuracy for the category *boring*. Finally, when comparing *mean* and DEKV, it is clear that using a simple average of embeddings is inferior to weighting by LLR. DEKV obtains the best macro average accuracy of 89.21%, and six out of seven best per-category accuracy. For the purpose of our next task, we combine fine-grained emotions *happy*, *warm*, *odd* into "*positive*", and *angry*, *boring*, *depressing*, *worried* into "*negative*".

Emotion	Accuracy(%)					
	NB	LDA	KW	CF	mean	DEKV
angry	47.00	74.21	79.21	83.71	79.47	**86.31**
worried	69.56	92.83	81.96	87.50	98.33	**98.46**
boring	75.67	76.21	84.34	**87.52**	83.81	85.62
happy	37.90	67.59	80.97	86.27	87.70	**90.86**
odd	73.90	85.40	77.05	84.25	85.41	**86.17**
depressing	73.76	81.43	85.00	87.70	88.28	**91.05**
warm	15.09	87.09	79.59	85.83	92.95	**95.20**
avg_M	56.13	74.12	80.79	85.69	85.58	**89.21**
avg_μ	23.95	80.68	81.16	86.11	87.99	**90.50**

Table 2: Comparison of accuracies from five reader-emotion classification methods. Bold numbers indicate the best performance in each emotion category (row).

To better visualize the effectiveness of our keyword selection method, we present these keywords as a word cloud in Fig. 4. Each keyword is color-coded by its corresponding emotion category, and scaled in size by its LLR score. Through this method, we can easily identify features within each group. For example, as stated in the previous section, we observed that keywords related to "Happy" (in green) are mostly about sports, including terms
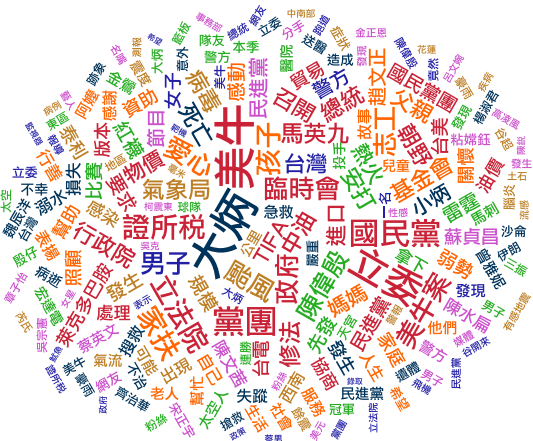
Figure 4: The word cloud generated from reader-emotion keywords. Colors and their corresponding emotion: Red—Angry, Green—Happy, Orange—Warm, Brown—Worried, Dark Blue—Depressed, Magenta—Boring, Light blue—Odd.

such as team names (e.g., "熱火 (Miami Heat)" and "紅襪 (Boston Red Sox)") and player names (e.g., "陳偉殷 (Wei-Yin Chen)", a pitcher for the baseball team Baltimore Orioles). Similar findings had also been revealed previously (Lin et al., 2007). On the contrary, "Angry"-related keywords (in red) consist largely of political parties or issues. For instance, the most noticeable word "美牛 (United States beef)" indicates the controversy of importing beef from the United States to Taiwan, which has been an issue that affects the Taiwan-U.S. relations and causes domestic political unrest. Simultaneously, numerous political terms such as "國民黨 (Kuomintang)", "立法院 (Legislative Yuan)", and "立委 (legislator)" are also keywords that provoke anger. The figure highlights the fact that the extracted emotion keywords are highly correlated with reader-emotions, and including them in the DEKV determine precise reader-emotions. As for the "Depressing" category, keywords are mostly related to social events that involve severe weathers or casualties. The most prevalent word, "大炳 (Da Bing)", refers to a Taiwanese actor who died in 2012, coinciding the time span of our retrieved data. Names of athletes might also show up in this category, owing to the readers' concerns about their performance in major sports events. In addition, the "Warm" category contains words associated with social care, volunteering, and charity.
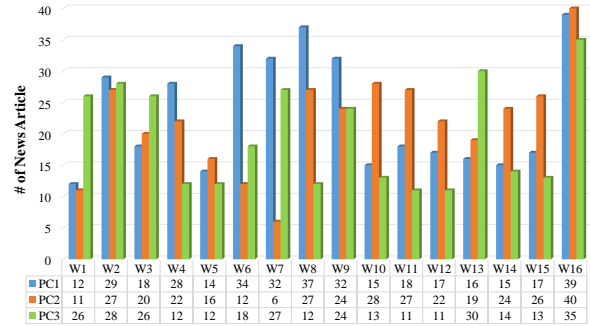

Figure 5: Descriptive statistics of the presidential election dataset. Numbers indicate the amount of news articles about a presidential candidate (PC) per week.

5.2 Exp. II: From Reader-Emotion to Publicity

The purpose of this experiment is to test the effectiveness of publicity score (PS) of a person based on our reader-emotion categorization method to estimate the trend of the poll. We collected 1,036 news articles from October 2015 to January 2016 regarding three presidential candidates (PC) from the same source as the previous experiment. Descriptive statistics about how many articles per PC by week are listed in Fig. 5. Note that they do not overlap with the previous corpus. We used the poll data from the first week as the initial value, and incremented it with PS obtained for each PC every week. These articles are first categorized into "*positive*" and '*negative*' using DEKV, and PS is calculated using (2) defined in Section 4.

	PC 1	PC 2	PC 3
poll/PS	0.20	0.50	0.72
poll/%Positive	-0.44	-0.42	-0.46

Table 3: Comparison of Pearson's r between the poll, publicity score (PS), and the ratio of positive emotion in news articles.

5.2.1 Results

We first examine the Pearson correlation coefficients in Table 3 between the poll and PS as well as the amount of positive emotion in the news articles, defined as the number of positive articles subtracted by that of the negative ones. It shows that the degree of correlation between PS and the poll number is positive and higher than that between a simpler metric, namely, the count of positive and negative articles. As a result, PS can serve as a more suitable

measure of the publicity of a certain subject. Still, we also observe that there is a considerable difference in the coefficients among different candidates. PS for PC 1 appears to be the least correlated, while PC 3 shows a high correlation between PS and poll. Further analysis is required to unveil the reason behind this phenomenon, but we suspect it may be related to the amount of documents for each PC.
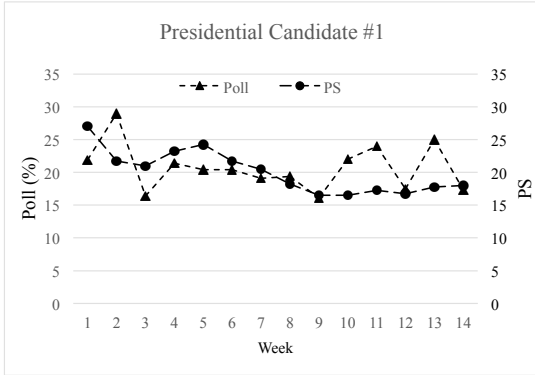

Figure 6: Timeline of the trend in publicity score (PS) and poll for presidential candidate (PC) #1.
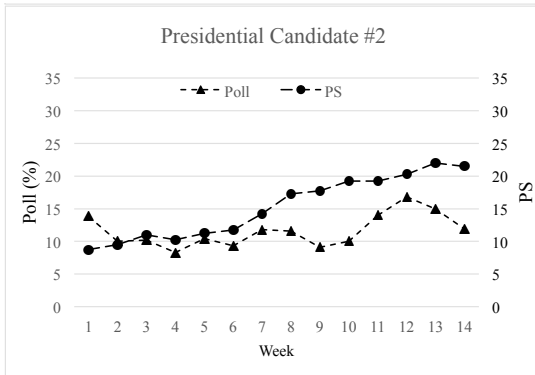

Figure 7: Timeline of the trend in publicity score (PS) and poll for presidential candidate (PC) #2.

Next, we plot PS for each PC in Fig. 6 to 8 for a subjective evaluation. We can see that the direction of increase and decrease (i.e., ups and downs) of the curves roughly align with those of the poll, validating our initial assumption of using the reader's emotion of a news article to quantify the publicity of a person. It also shows that there exists a positive correlation between the poll and PS. In general, PS does not experience sharp turns like the trend we witnessed in the curves of poll, showing that the publicity score is more robust due to its immunity to
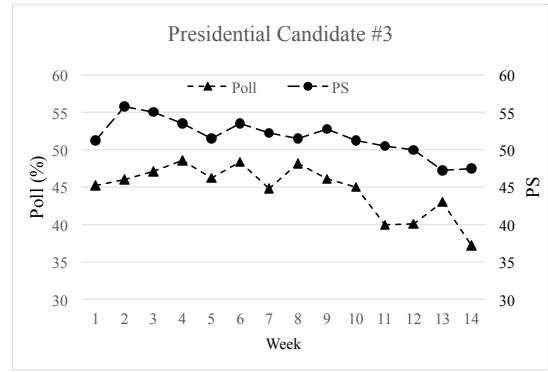

Figure 8: Timeline of the trend in publicity score (PS) and poll for presidential candidate (PC) #3.

the temporary surge in news articles. However, PS is less than optimal for predicting the polls for PC#1, illustrated by the curves in PC#1 being more random than others (e.g., in weeks 2 and 11) and the results in Table 3. Thus, a more sophisticated modeling of the interaction between reader's emotion and a candidate's publicity is worthy of further research.

In sum, our method objectively induce the publicity score through classification of readers' emotion on news events, preserving its accuracy from the fluctuation of sampling bias in non-official polling institutions. Our approach for mining the publicity of public figures through reader's emotion classification provides a promising direction for automated collection of such information online.

6   Conclusion

We propose a novel document representation model, DEKV, for reader-emotion classification, as well as a publicity mining method. Experiments on two Chinese news corpora demonstrate that DEKV outperforms well-known models for reader-emotion detection and can subsequently be related to the publicity of a person. We believe it is an emerging direction for automated collection of social and emotional information online. We also envision its applications on numerous academic as well as business domains. In the future, we will explore different ways to integrate deeper semantics and further investigate the relation between emotion and publicity.

Acknowledgments

References

Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *The Journal of Machine Learning Research*, 3:1137–1155.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3):27:1–27:27.

Yung-Chun Chang, Cen-Chieh Chen, Yu-lun Hsieh, Chien Chin Chen, and Wen-Lian Hsu. 2015. Linguistic template extraction for recognizing reader-emotion and emotional resonance writing assistance. In *Proceedings of the 53rd ACL and the 7th IJCNLP (Volume 2: Short Papers)*, pages 775–780.

Ying Chen, Sophia Yat Mei Lee, Shoushan Li, and Chu-Ren Huang. 2010. Emotion cause detection with linguistic constructions. In *Proceedings of the 23rd International Conference on Computational Linguistics*, pages 179–187.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20(3):273–297.

Dipankar Das and Sivaji Bandyopadhyay. 2009. Word to sentence level emotion tagging for bengali blogs. In *Proceedings of the ACL-IJCNLP 2009 Conference*, pages 149–152.

Yi Fang, Luo Si, Naveen Somasundaram, and Zhengtao Yu. 2012. Mining contrastive opinions on political texts using cross-perspective topic model. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 63–72. ACM.

Daniel Gayo-Avello. 2012. I wanted to predict elections with twitter and all i got was this lousy paper – a balanced survey on election prediction using twitter data. *arXiv preprint arXiv:1204.6441*.

Yu-Ming Hsieh, Ming-Hong Bai, Jason S Chang, and Keh-Jiann Chen. 2012. Improving PCFG chinese parsing with context-dependent probability re-estimation. *CLP 2012*, page 216.

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 168–177. ACM.

Siwei Lai, Kang Liu, Liheng Xu, and Jun Zhao. 2015. How to generate a good word embedding? 07.

Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, pages 1188–1196.

Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. 2007. What emotions do news articles trigger in their readers? In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 733–734.

Kevin Hsin-Yih Lin, Changhua Yang, and Hsin-Hsi Chen. 2008. Emotion classification of online news articles from the reader's perspective. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, volume 1, pages 220–226.

Christopher D Manning and Hinrich Schütze. 1999. *Foundations of statistical natural language processing*. MIT press.

Andrew McCallum, Kamal Nigam, et al. 1998. A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization*, volume 752, pages 41–48.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient estimation of word representations in vector space. In *Proceedings of Workshop at ICLR*.

Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

George A Miller and Walter G Charles. 1991. Contextual correlates of semantic similarity. *Language and cognitive processes*, 6(1):1–28.

Gilad Mishne. 2005. Experiments with mood classification in blog posts. In *Proceedings of the 1st Workshop on Stylistic Analysis Of Text For Information Access (Style 2005)*.

Bo Pang, Lillian Lee, and Shivakumar Vaithyanathan. 2002. Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79–86. Association for Computational Linguistics.

Matthew Purver and Stuart Battersby. 2012. Experimenting with distant supervision for emotion classification. In *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 482–491.

Changqin Quan and Fuji Ren. 2009. Construction of a blog emotion corpus for chinese emotional expression analysis. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, volume 3, pages 1446–1454.

Jonathon Read. 2005. Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL Student Research Workshop*, pages 43–48. Association for Computational Linguistics.

Radim Řehůřek and Petr Sojka. 2010. Software Framework for Topic Modelling with Large Corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, pages 45–50, Valletta, Malta, May. ELRA.

Luís Sarmento, Paula Carvalho, Mário J Silva, and Eugénio De Oliveira. 2009. Automatic creation of a reference corpus for political opinion mining in user-generated content. In *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 29–36. ACM.

Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard. 2012. Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4):470–479.

Yi-jie Tang and Hsin-Hsi Chen. 2012. Mining sentiment words from microblogs for predicting writer-reader emotion transition. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)*, pages 1226–1229.

Robert E Thayer. 1989. *The biopsychology of mood and arousal*. Oxford University Press.

Peter D Turney. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 417–424.

Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2009. Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. *Computational Linguistics*, 35(3):399–433.

Chung-Hsien Wu, Ze-Jing Chuang, and Yu-Chung Lin. 2006. Emotion recognition from text using semantic labels and separable mixture models. 5(2):165–183.

Changhua Yang and Hsin-Hsi Chen. 2006. A study of emotion classification using blog articles. In *Proceedings of Conference on Computational Linguistics and Speech Processing*.

Yi-Hsuan Yang, Chia-Chu Liu, and Homer H. Chen. 2006. Music emotion classification: A fuzzy approach. In *Proceedings of the 14th Annual ACM International Conference on Multimedia*, MULTIMEDIA '06, pages 81–84, New York, NY, USA. ACM.

Changhua Yang, Kevin Hsin-Yih Lin, and Hsin-Hsi Chen. 2009. Writer meets reader: Emotion analysis of social media from both the writer's and reader's perspectives. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology-Volume 01*, pages 287–290. IEEE Computer Society.

Feng Zou, Fu Lee Wang, Xiaotie Deng, Song Han, and Lu Sheng Wang. 2006. Automatic construction of chinese stop word list. In *Proceedings of the 5th WSEAS International Conference on Applied Computer Science*, pages 1010–1015.