Journal of the Korea Convergence Society
Vol. 9. No. 7, pp. 25-31, 2018
ISSN 2233-4890
https://doi.org/10.15207/JKCS.2018.9.7.025

Inference of Korean Public Sentiment from Online News

Andrew Stuart Matteson, Soon-Young Choi, Heui-Seok Lim*
Dept. of Computer Science and Engineering, Korea University

온라인 뉴스에 대한 한국 대중의 감정 예측

Andrew Stuart Matteson, 최순영, 임희석*
고려대학교 컴퓨터학과

Abstract Online news has replaced the traditional newspaper and has brought about a profound transformation in the way we access and share information. News websites have had the ability for users to post comments for quite some time, and some have also begun to crowdsource reactions to news articles. The field of sentiment analysis seeks to computationally model the emotions and reactions experienced when presented with text. In this work, we analyze more than 100,000 news articles over ten categories with five user-generated emotional annotations to determine whether or not these reactions have a mathematical correlation to the news body text and propose a simple sentiment analysis algorithm that requires minimal preprocessing and no machine learning. We show that it is effective even for a morphologically complex language like Korean.

Key Words : Sentiment Analysis, Crowdsourcing, Online News, Emotion Dictionary, Social Emotion Detection, Natural Language Processing

요 약 온라인 뉴스는 기존의 신문을 대체하였고, 우리가 정보에 접근하고 공유하는 방법에 큰 변화를 가져왔다. 뉴스 웹사이트들은 사용자가 댓글을 남길 수 있는 기능을 오랜 시간동안 제공하였고, 그 중 몇몇 뉴스 웹사이트에서는 뉴스 기사들에 대한 사용자의 반응들을 크라우드소싱(crowdsource)하기 시작했다. 감정분석 분야에서는 텍스트에 반영된 감정과 반응들을 컴퓨팅적으로 모델링하기 위한 시도를 하고 있다. 본 연구에서는 뉴스 기사에 대한 반응들이 뉴스 본문과 수학적인 상관관계를 갖는지 밝히기 위해, 사용자로부터 생성된 다섯 가지의 감정 라벨(label)을 사용하여 10가지 카테고리(category)에 해당하는 100,000개 이상의 뉴스 기사들을 분석한다. 본 연구에서는 전처리과정이 최소한으로 필요하고 기계학습이 적용하지 않아도 되는 간단한 감정 분석 알고리즘(algorithm)을 제안한다. 우리는 이 모델이 한국어와 같은 형태론적으로 복잡한 언어에도 효과적이라는 것을 증명한다.

주제어 : 감정분석, 크라우드소싱, 온라인뉴스, 감정사전, 사회적 감정 탐지, 자연어처리

1. Introduction

Most work on sentiment analysis aims to predict the emotions elicited by the author when writing a passage of text[1]. However, predicting public sentiment to news articles is a potentially more difficult task because, except for some opinion pieces, news articles are not always written to evoke particular emotions, but rather simply to convey information. Furthermore, news articles span multiple domains and many sentiment analysis tasks for the Korean language focus

*This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017M3C4A7068189).
*Corresponding Author : Heui-Seok Lim(limhseok@korea.ac.kr)
Received April 25, 2018 Revised May 24, 2018
Accepted July 20, 2018 Published July 28, 2018

on only one domain such as product reviews[2,3], movie reviews[4], or stock market prediction[5]. Further complicating the task is the fact that different readers react differently to the same piece of information. That being said, most readers would consider an act of terrorism to evoke "angry" or "sad" emotions, and an article about philanthropy to evoke "heartwarming" emotions, for example.

In online news, users play the role of both consumer and producer by reading and reacting to articles. Predicting such reader emotions could have deeper implications for marketing where brand image and consumer sentiment is of paramount importance. It could also play a helpful role in the field of public relations for companies writing press releases. However, as it would be intractable for humans to review every possible piece of writing for possible emotional response, it would be helpful to have a method to do so automatically. In this work, we build an emotional dictionary to assist with this task.

Our main contributions are as follows. 1) We propose a totally automatic approach for generating an emotional dictionary which requires no human feature engineering; 2) We analyze the effectiveness of such an algorithm on domain-independent data; 3) Our approach requires no extensive preprocessing; 4) Our algorithm has the ability to predict emotions in text that generally has no explicit expressed emotions.

2. Related Work

Most work on sentiment analysis focuses on classification of short text such as product reviews[6], movie reviews[7], microblogs[8] which are typically a paragraph or shorter. Emotion classification of such data aims to predict the emotion evoked by the writer. Reviews using ordinal rating scores (e.g., 1-5) are typically handled by regression techniques[9,10]. It can also be done by using association rules.[new paper]

News data, on the other hand, is annotated with several subjective emotions (heartwarming, sad, angry, like, desire follow-up article) and presents us with the issue of predicting the reaction experienced by the reader. The length of news articles means that the data is more sparse than short text, and emotions are less explicit and agreeable than with movie or product reviews. Although not all readers will respond the same way to a particular article, several previous research attempts emotion detection with news data. The SWAT system[12,13] uses a supervised approach and develops a word-emotion mapping dictionary to score the word of news headlines. [14] uses a word-emotion dictionary to predict user reactions on a popular Chinese news website.

However, in Korea, there is little prior research on inferring sentiment from large bodies of text such as news. In [5], 1,170 stock news articles are used for constructing an emotion dictionary. The articles were collected 30 per day by hand, limiting the potential size of the corpus. Furthermore, a dictionary of vocabulary related to the domain of stock market news and a reciprocal relation model to handle domain-specific adversative conjunctions are constructed, incurring a high manual labor cost.

In [15], positive, negative, and neutral patterns were extracted based on word frequency to perform opinion mining using Daum movie review site data containing point-based ratings and 100-character review text. However this work also used exhaustive manual labor to construct hand-crafted rules, greatly limiting the scalability of the approach.

Up until now, domestic research has focused on domain-dependent tasks such as classifying product reviews and movie reviews and predicting stock prices. However, when the domain is fixed to a specific field, the emotion dictionary's scope is also limited to that particular domain's lexicon, limiting its potential usefulness in domain-independent tasks.

In this work, we propose a method of constructing a domain-independent dictionary for news data spanning a variety of categories. In addition, since the proposed method does not utilize hand-crafted features, it has the advantage of requiring no extensive human labor beyond initial collection of the training corpus.

3. Model

We assume a corpus of W terms, K emotions, and N training documents for which we would like to find a probability distribution of K emotions. Our objective is to generate a word emotion dictionary. TF-IDF is pre-calculated for each document to account for term frequency and inverse document frequency, similar to [16].

The categorical distribution of E user emotions is denoted by $P(r_i) = (r_{i1}, r_{i2}, \ldots, r_{iE})$ and is normalized such that $|r_i| = 1$. The distribution $P(r_i)$ for every document i is already known.

We define $I_{d_i}(w_j)$ as an indicator function of whether term j occurs in document i, which is a binary value of 0 or 1. This ensures that if term j does not appear in document i, it will not contribute to emotion-term weights for term j.

We define the weight of emotion k for term j as follows:

$$\Theta_{jk} = \sum_{i=1}^{N} \left\{ I_{d_i}(w_j) \alpha_{ij}^{\beta} r_{ik} \right\}$$

where r_{ik} is the normalized probability of emotion k in training document i, which is already specified by our data. α_{ij} represents the pre-calculated TF-IDF weighting constant for term j in document i, and β is a weighting amplification factor.

Using the emotion-term weight Θ_{jk}, we can calculate the probability of r_k conditioned on an unseen document d_m from a held-out testing corpus. TF-IDF is also pre-calculated for every term j on testing document m as γ_{mj}, independently from the training corpus. Absolute IDF values may vary between training and testing sets depending on term distribution. In the case that TF-IDF weighting does not increase performance, weight γ_{mj} can be fixed to a value of 1.

$$r_{mk} = \sum_{j=1}^{W} \left\{ I_{d_m}(w_j) \gamma_{mj}^{\beta} \frac{\Theta_{jk}}{\sum_{z=1}^{E} \Theta_{jz}} \right\}$$

Therefore, the probability mass function $P(r_m)$ for E emotions in unseen document m can be expressed as:

$$P(r_m) = \prod_{k=1}^{E} r_{mk}$$

or, in expanded form:

$$P(r_m) = \prod_{k=1}^{E} \sum_{j=1}^{W} \left\{ I_{d_m}(w_j) \alpha_{mj}^{\beta} \frac{\Theta_{jk}}{\sum_{z=1}^{E} \Theta_{jz}} \right\}$$

Terms appearing in the test document that are missing in the training data are considered OOV (out-of-vocabulary) and do not contribute to the calculation of the emotion probability distribution $P(r_m)$.

To evaluate our system, we label the dominant emotion k for document i as L_i:

$$L_i = \mathrm{argmax_k}(\mathrm{r_{ik}})$$

L_i is then compared to L_m during evaluation.

4. Experiment

4.1 Dataset

132,545 articles were collected from N news website from 2017-1-1 to 2017-11-30 and were utilized to automatically construct a word-level emotion dictionary. The articles span multiple categories, including breaking news, elections, IT/science, entertainment, economy, society, lifestyle/culture, world, politics, and opinion. Depending on training parameters, the number of training documents varied from 14,698 to 92,436. The held-out test set size was fixed at 20,000 documents for consistency and the same test documents were used for all experiments.

4.2 Method

To evaluate the proposed algorithm, we adopt the Acc@k approach as used in [17,18]. Acc@k measures whether or not the predicted emotion appears in the top k emotions of ground truth set E_topK@d. We fix parameter k to 1, also known as the Acc@1 approach which is considered to be the most important metric in the literature [17,18].

Given a document d, ground truth emotion set E_topK@d including the K top-ranked emotions, and the top ranked predicted emotion e_p, Acc_d@K is calculated as:

$$Acc_d@k = \begin{cases} 1 & if \quad e_p \in E_{topk@d} \\ 0 & else. \end{cases}$$

The resulting classification accuracy Acc@k is then calculated by dividing the sum of Acc_d@k values returned from the proposed classifier by the sum of the Acc_d@k values from a classifier returning ground truth. All experiments were performed in Python 3 on Ubuntu Linux 16.04.

4.3 Results & Analysis

For our experiments, we adjust three parameters: Majority Threshold, Minimum Reaction Count, and TF-IDF Weighting Amplification Factor. The model is retrained for each set of attempted parameters. In order to measure the influence of each parameter, the values of other parameters are fixed. Performance is then measured for each set of parameters using the Acc@1 metric.

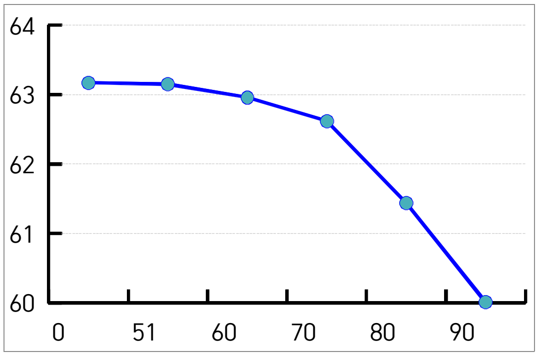


Fig. 1. PerformanceY-axis : %) while adjusting Majority Threshold(X-axis : %) parameter

Majority Threshold specifies the minimum percent of total reactions that the dominant emotion must account for in order to be included in the training set. Fig. 1 shows the impact of the Majority Threshold parameter on performance. We fix Minimum Reaction Count to 10 and TF-IDF Weighting Amplification Factor to 1.0 when adjusting this parameter. As the Majority Threshold is increased, performance appears to degrade. This is because as the Majority Threshold parameter is increased, the number of potential training documents is also reduced as some become unusable. The performance difference resulting from a Majority Threshold of 51% and no Majority Threshold is only 0.02%. We conclude that the number of training documents is the overriding factor in determining performance.

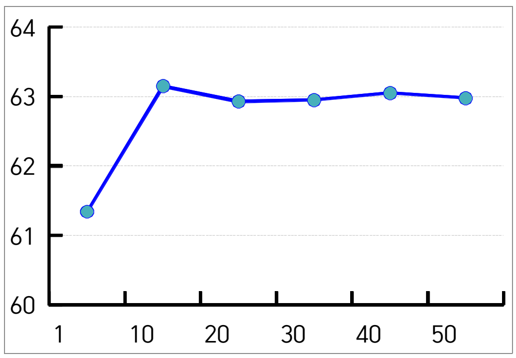


Fig. 2. Performance (Y-axis : %) while adjusting Minimum Reaction Count (X-axis) parameter

The second parameter we adjust is Minimum Reaction Count, which specifies the minimum number of total reactions submitted by users that need to exist in order for the data to be included in the training set. We fix Majority Threshold to 51% and TF-IDF Weighting Amplification Factor to 1.0. As seen in Fig. 2, when Minimum Reaction Count is set to 1 performance is at its worst, suggesting that pruning sparse reaction data can be useful. However, due to the reduction of training set that occurs when pruning some of the more sparse training data, increasing Minimum Reaction Count beyond 10 appears to cause only a degradation in performance.

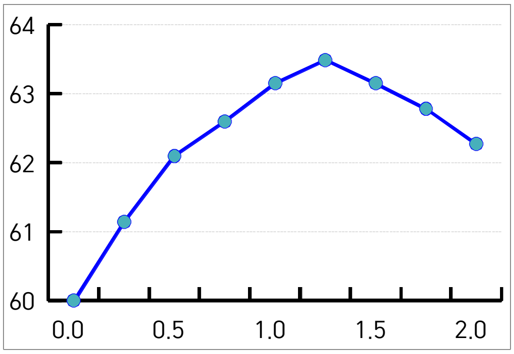


Fig. 3. Performance (Y−axis : %) while adjusting TF−IDF Weighting Amplification Factor (X−axis) parameter

Lastly, we attempt to adjust the TF-IDF Weighting Amplification Factor with Majority Threshold and Minimum Reaction Count fixed to 51% and 10, respectively. This parameter is the factor by which arbitrary word-emotion weights are multiplied by the TF-IDF constants. A value of 0 means that TF-IDF is not considered during emotion dictionary generation. Our experiments shown in Fig. 3 indicate that an amplification factor of 1.25 is ideal, which implies that performing TF-IDF is meaningful for improving performance. However, overamplification of TF-IDF causes a performance decrease in our tests. We suspect that although TF-IDF is one meaningful element, there may also be other latent factors at play.

Controlling for other parameters, our experiments indicate that the best settings should be a Minimum Reaction Count of 10, no Majority Threshold, and a TF-IDF Weighting Amplification Factor of 1.25. The amount of training data appears to be the most crucial factor affecting performance, assuming that data has at least 10 reactions.

5. Conclusion & Future Work

In online news, users not only read but also contribute to news articles by providing their reactions. Predicting such reader emotions could have deep implications for fields such as marketing and public relations. However, doing so is a difficult task because news articles do not always include explicit emotions. Approaches exist for domain-specific fields, but they incur an unreasonable human labor cost and scale poorly, and therefore the training corpus sizes and potential for performance improvement is limited.

The emotional dictionary we built in this work successfully uncovers a correlation between long, domain-independent news article body text and crowdsourced user reaction data. We have proposed a totally automatic approach for producing this emotional dictionary which requires no feature engineering or preprocessing. We have also shown its applicability to predicting user reactions to news articles spanning a wide variety of categories. Performance is shown to be 65% over 5 different reaction types.

Our work focused on creating an emotion dictionary totally automatically at a low cost. While our approach does not utilize machine learning, future work should focus on the potential applications of neural networks or other deep learning architectures to the problem of domain-independent Korean news article sentiment classification, which we believe may be an important factor in increasing classification performance.

REFERENCES

[1] K. H. Lin, C. Yang & H. H. Chen. (2007). What Emotions do News Articles Trigger in their Readers? *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 38(6),* 733-734.

[2] J. S. Song & S. W. Lee (2011). Automatic Construction of Positive/Negative Feature-Predicate Dictionary for Polarity Classification of Product Reviews. *Journal of KIISE: Software and Applications, 38(3),* 157-168.

[3] J. S. Myung, D. J Lee & S. G. Lee (2007). A Korean Product Review Analysis System Using a Semi-Automatically Constructed Semantic Dictionary. *Proceedings of the 19th Annual Conference on Human and Cognitive Language Technology,* 68-75.

[4] C. H. Jeong, J. H. Kim, Y. J. Jeon & H. J. Jeong (2017). Korean Sentiment Dictionary Based on the Reliability of Review Data. *Journal of Korean Institute of Information Scientists and Engineers,* 1965-1967.

[5] J. H. Seo, J. H Cho & J. T. Choi (2015). Design for Opinion Dictionary of Emotion Applying Rules for Antonym of the Korean Grammar. *Journal of Advanced Information Technology and Convergence, 13(2),* 109-117.

[6] X. Fang & J. Zhan. (2015). Sentiment Analysis Using Product Review Data. *Journal of Big Data, 2(1),* 5.

[7] B. Pang, L. Lee & S. Vaithyanahtan (2002). Thumbs Up?: Sentiment Classification using Machine Learning Techniques. *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, volume 10,* 79-86.

[8] A. Pak & P. Paroubek (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *LREc, 10.*

[9] L. Qu, G. Ifrim & G. Weikum (2010). The Bag-of-Opinions Method for Review Rating Prediction from Sparse Text Patterns. *Proceedings of the 23rd International Conference on Computational Linguistics. Association for Computational Linguistics,* 913-921.

[10] S. Baccianella, A. Esuli & F. Sebastiani. (2009). Multi-Facet Rating of Product Reviews. *European Conference on Information Retrieval,* 461-472.

[11] J. S. Kim (2016). Emotion Prediction of Paragraph using Big Data Analysis. *Korea Convergence Society, 14(11),* 267-273.

[12] P. Katz, M. Singleton & R. Wicentowski. (2007). SWAT-MP: The SemEval-2007 Systems for Task 5 and Task 14. *Proceedings of the 4th International Workshop on Semantic Evaluations. Association for Computational Linguistics,* 308-313.

[13] C. Strapparava & R. Mihalcea (2007). SemEval-2007 Task 14: Affective Text. *Proceedings of the 4th International Workshop on Semantic Evaluations. Association for Computational Linguistics,* 70-74.

[14] Y. Rao, J. Lei, L. Wenyin, Q. Li & M. Chen (2014). Building Emotional Dictionary for Sentiment Analysis of Online News. *World Wide Web, 17(4),* 723-742.

[15] J. O. Kim, S. S Lee & H. S. Yong. (2011). Automatic Classification Scheme of Opinions Written in Korean. *Journal of Korean Institute of Information Scientists and Engineers, 38(6),* 423-428.

[16] Y. A. Heo, D. Y. Lee & G. G. Kim (2017). A System for Automatic Classification of Traditional Texts. *Korea Convergence Society, 8(12)*, 39-47.

[17] Bao, S., Xu, S., Zhang, L., Yan, R., Su, Z., Han, D., Yu, Y (2011). Mining Social Emotions from Affective Text. *IEEE Transactions on Knowledge and Data Engineering, 24*, 1658-1670.

[18] S. Bao, S. Xu, L. Zhang, R. Yan, Z. Su, D. Han & Y. Yu, (2009). Joint Emotion-Topic Modeling for Social Affective Text Mining. *Proceedings of the 9th IEEE International Conference on Data Mining*, 699-704.

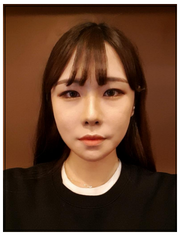
임 희 석(Lim, Heui Seok) [정회원]

▪ 1992년 2월 : 고려대학교 컴퓨터학과 (이학사)
▪ 1994년 2월 : 고려대학교 컴퓨터학과 (이학석사)
▪ 1997년 2월 : 고려대학교 컴퓨터학과 (이학박사)
▪ 2008년 3월 ~ 현재: 고려대학교 컴퓨터 학과 교수
▪ 관심분야 : 자연어 처리(NLP), 뇌신경 언어 정보 처리
▪ E-Mail : limhseok@kroea.ac.kr

앤 드 류(Matteson, Andrew) [정회원]

▪ 2012년 9월 : 미시간 주립 대학교 컴퓨터학과 (이학사)
▪ 2016년 9월 ~ 현재: 고려대학교 컴퓨터학과 소프트웨어 전공 석사과정
▪ 관심분야 : 자연어 처리(NLP), 딥러닝(Deep Learning), 인공지능, 감정분석(Sentiment Analysis)
▪ E-Mail : amatteson@korea.ac.kr

최 순 영(Choi, Soon Young) [정회원]



- 2015년 8월 : 건국대학교 컴퓨터공학과 (이학사)
- 2016년 3월 ~ 현재: 고려대학교 컴퓨터학과 소프트웨어 전공 석사과정
- 관심분야 : 자연어 처리(NLP), 딥러닝(Deep Learning), 인공지능, 감정분석(Sentiment Analysis)
- E-Mail : perfectchl@korea.ac.kr