WordForce: Visualizing Controversial Words in Debates

Wei-Fan Chen   Fang-Yu Lin   Lun-Wei Ku
Institute of Information Science,
Academia Sinica, Taipei, Taiwan.
viericwf@iis.sinica.edu.tw, nis50122806@gmail.com, lwku@iis.sinica.edu.tw

Abstract

This paper presents WordForce, a system powered by the state of the art neural network model to visualize the learned user-dependent word embeddings from each post according to the post content and its engaged users. It generates the scatter plots to show the force of a word, i.e., whether the semantics of word embeddings from posts of different stances are clearly separated from the aspect of this controversial word. In addition, WordForce provides the dispersion and the distance of word embeddings from posts of different stance groups, and proposes the most controversial words accordingly to show clues to what people argue about in a debate.

1 Introduction

Word embeddings have been widely used in deep neural networks and have achieved promising results. Compared to the traditional *n*-gram feature, which represents each document as a high dimensional sparse vector, the word embedding is representing with the low dimensional and dense vector. Hence using embeddings has its merits on decreasing training time and reducing complexity, and many papers have introduced different compositions of word embeddings in their work for comparison (Chen et al., 2015; Lai et al., 2015). However, one drawback of using word embeddings is that human cannot interpret its meaning as when using *n*-gram feature. In previous work, one solution is to visualize the word embeddings by reducing them into two-dimensional vectors on a x-y plot to view the semantic distribution of words. For example, in Iyyer et al.'s work we see people's names would cluster together when they have the same jobs or positions, e.g., presidents of United States, prime ministers or emperors (Iyyer et al., 2014); ScholarOctopus[1] and tsnejs[2] visualize research articles embeddings and word embeddings, respectively; Mikolov also shows the semantics can be calculated using this kind of two dimensional plot (Mikolov and Dean, 2013); the semantic word cloud based on word embedding visualizes the word usage in product reviews (Xu et al., 2016). All these show the distance between word embeddings reveals semantic relations.

In a time that social media becomes part of our life, we attempt to observe the user-dependent word embeddings in a debate to analyze user-dependent semantics. In the past, incorporating meta data to train neural network models for sentiment analysis on product reviews and social media texts has been shown to be effective. For example, our UTCNN integrates users, topics and comments information in Facebook posts (Chen and Ku, 2016); Dong et al. consider topics and add an adaptive layer in their recursive neural network for target-dependent Twitter sentiment (Dong et al., 2014); Tang et al.'s UPNN incorporates users and products (Tang et al., 2015). In this paper, to see how this kind of word embeddings can be further utilized, we consider users who posted or liked the post in the process of training word embeddings in addition to a pure text-based neural network models (Kim, 2014). Such learned word embeddings for the same word would differ among posts when the engaged users are different.

---

[1]http://cs.stanford.edu/people/karpathy/scholaroctopus/
[2]http://cs.stanford.edu/people/karpathy/tsnejs/wordvecs.html

Distribution Statistics	
Intra-group	**Distance**
Supportive	1.26
Neutral	1.47
Unsupportive	1.07
Inter-group	**Distance**
Supportive/Unsupporitve	1.72
Neutral/Supportive	1.70
Neutral/Unsupportive	1.47

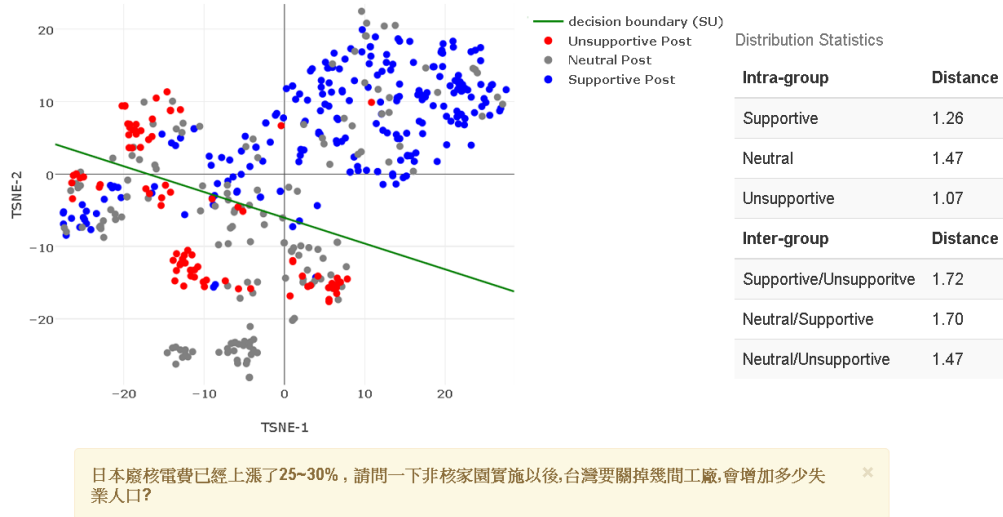日本廢核電費已經上漲了25~30%，請問一下非核家園實施以後,台灣要關掉幾間工廠,會增加多少失業人口?

Figure 1: The WordForce interface showing the results of the controversial word 上漲(rise). The selected post comments that after the implement of non-nuclear policy, the electricity rate in Japan has risen 25-30%; if we follow, how many factories will be closed and how many unemployed people will we have?

Therefore, we may investigate their semantic difference and how they can contribute to the analysis of the stance classification problem in debates.

For this purpose, we present the web-based system, WordForce[3], where users can query an arbitrary corpus word to get its visualization and statistic information. Figure 1 shows the query result of searching the word 上漲(rise) in the nuclear power plant construction debate. The left-hand side shows the two-dimensional visualization including its word embedding in each post and a decision boundary between the supportive and unsupportive stance (if applicable). Supportive/unsupportive posts were those in support of or against anti-reconstruction; neutral posts were those evincing a neutral standpoint on the topic, or were irrelevant. The stance of the post where the word embedding is from is indicated by different dot colors: blue for supportive, gray for neutral, and red for unsupportive. The plot here suggests that the word *rise*, referring to the rise of electric charge in the nuclear power debate, has different semantics between the supportive and unsupportive posts as we expect. The right-hand side shows distribution statistics. Further clicking on any dot will show the original post content below the plot, e.g., after clicking a red dot, the unsupportive post arguing that the abandon of nuclear power will rise the electricity rate shows below.

2   Learning User-Dependent Word Embeddings

To learn the user-dependent word embeddings for stance classification and visualization, we train the 50-dimensional word embeddings via GloVe (Pennington et al., 2014). These embeddings are then transformed via a user-dependent matrix embedding U_k as in equation 1.

$$x'_w = U_k \cdot x_w \tag{1}$$

where x_w and x'_w are the word embeddings of word w trained by GloVe and the transformed word embeddings, respectively. The user-dependent matrix embedding models the user's preference for reading certain semantics where the "user" denotes a pseudo user on behalf of all likers and authors in a given post. Then the transformed word embeddings x'_w are used as the input of a convolutional neural network and fed into a fully connected network to yield the final post stance. The detail descriptions of the proposed neural network model is included in the paper of UTCNN (Chen and Ku, 2016).

We collect data from anti-nuclear-power Chinese Facebook fan groups in one year period of time, including posts and their author and liker IDs. There are a total of 2,496 authors, 505,137 likers and

---

[3]WordForce is available at `http://doraemon.iis.sinica.edu.tw/wordforce`

	Supportive	Neutral	Unsupportive	all
Annotation	7,504	24,816	275	32,595
Stance Classification	.698	.957	.571	.755

Table 1: Annotation results and f-scores of stance classification of Facebook dataset.

32,595 posts. We annotate the stance of all posts as *supportive*, *neutral*, or *unsupportive*. The annotation results are shown in the first row of Table 1. On average, 161.1 users are engaged to one post. The maximum is 23,297 and the minimum is one (the author). Experimental results show that the proposed model achieves good results on the Chinese Facebook fans group material as shown in the second row of Table 1 (Chen and Ku, 2016). For comparison, this model is also tested on the English open benchmark CreateDebate for stance classification and it outperforms the state of the art by achieving the accuracy 0.842 against 0.735 (Sridhar et al., 2015; Chen and Ku, 2016).

3 WordForce

On top of the word embeddings obtained from the state of the art neural network model for stance classification, WordForce visualize these embeddings for debatable issues to provide useful information for research surveys or industrial applications. WordForce can illustrate each corpus word by displaying a two-dimensional word embedding distribution plot as well as the inter- and intra-group distances (dispersion and distance, respectively), where a "group" is a set of word embeddings from posts of the same stance label. Furthermore, with these statistics, WordForce can propose different types of controversial words ,i.e., aspects or events that people of different stance are arguing about.

From Controversial Word Visualization to Suggestion   After training, we gather all the word embeddings from the user-dependent transformation. For each corpus word, we collect their transformed word embeddings x'_w and project them into a two-dimensional space via t-SNE (Maaten and Hinton, 2008). The two dimensions of the t-SNE plot implicitly present latent sentiment or semantic so that similar words would have similar vector representations as in many related work (Iyyer et al., 2014; Melamud et al., 2015).

Now with the positions of embeddings of one word, WordForce can further calculate their intra- and inter-group distance. The intra-group distance (dispersion) of group g is defined as the average Euclidean distance to the group mean shown in equation 2.

$$Dispersion\,(g) = \frac{1}{N_g} \sum_n \|v_{n,g} - \mu_g\| \qquad (2)$$

where N_g is the size (number of dots) of this group, $v_{n,g}$ is the n-th vector, and μ_g is the mean of the group g, respectively. The inter-group distance (distance) is the average link between two groups as in equation 3.

$$Distance\,(g_i, g_j) = \frac{1}{N_{g_i} \cdot N_{g_j}} \sum_{v_n \in g_i, v_m \in g_j} \|v_n - v_m\| \qquad (3)$$

where N_{g_i} and N_{g_j} are the sizes of group i and j, respectively; v_n and v_m are the n-th vector of group i and the m-th vector of group j, respectively. A low dispersion value indicates posts and their engaged users of the same stance group agree in its semantic, while a high distance value indicates posts and their engaged users vary a lot among groups and can be separated. With the dispersion and distance value of each word calculated from its embeddings, WordForce is then able to propose controversial words by ordering their dispersion value ascendingly and the distance value discendingly.

Table 2 shows some words with a high inter-group distance, a low intra-group dispersion or a high TFIDF value, which confirms that WordForce can propose different controversial words in addition to the conventional topical words. WordForce also lists these words for users to see their word embedding distribution plots and statistics.

Controversial Word Type	Example (Translation)
Top high TFIDF	龍門(lonmen), 絕食(hunger strike), 夏天(summer)
Top high distance	日光(solar),廢氣(air pollution), 煙囪(chimney)
Top low dispersion	核融合(nuclear fusion),國庫(exchequer), 偵檢器(radiation-detector)

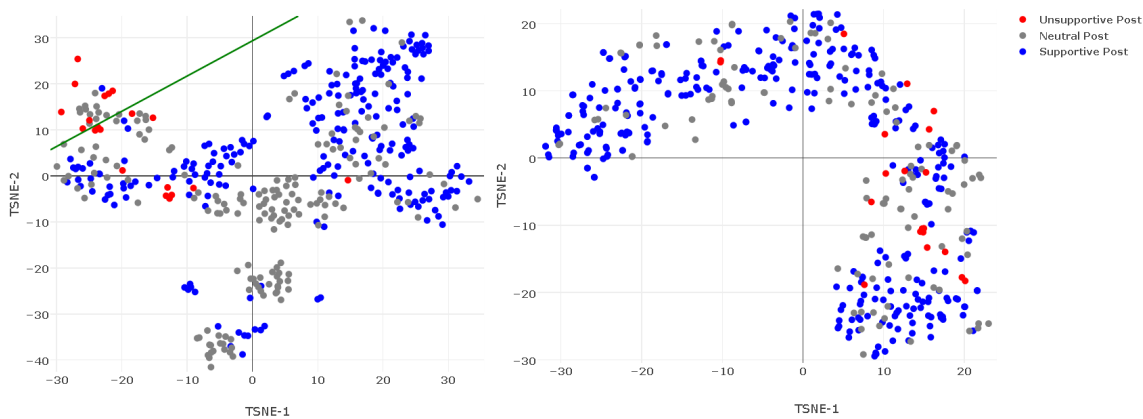Table 2: Example controversial words proposed by WordForce.

Figure 2: Word embedding distribution plots for 絕食(hunger strike) and 廢氣(air pollution).

Discussion We select some cases to illustrate WordForce. Figure 1 shows the plot and statistics of the word 上漲(rise). The dispersion of the neutral group is much larger than that of both the supportive group and the unsupportive groups, and the large inter-group distance tells that supportive to unsupportive posts are more different than neural to supportive or neutral to unsupportive posts. The trend these numbers tell can be easily captured by reading the plot. From the plot we also find that the unsupportive posts are clustered into several sub-groups. These sub-groups represent different related arguments. For example, the sub-group on the far right collects news articles discussing the disadvantages of abandoning nuclear, while the one in the middle includes some personal criticisms. All these observations confirm that WordForce can facilitate deeper analysis.

In Figure 2 we show another two word embedding examples: 絕食(hunger strike) at the right-hand side and 廢氣(air pollution) at the left-hand side. The word 絕食(hunger strike) seems to be unrelated to the nuclear issue but the word embeddings tell differently and are clearly separated: going deeper we find a former politician has organized a hunger strike against the nuclear power. Hence some related posts support the hunger strike to opt for the anti-nuclear, and the others say the hunger strike is a publicity stunt so that to be against the anti-nuclear.

Unlike hunger strike, air pollution is related as the thermal generation supplies most electricity in Taiwan but produces much air pollution. However, the word embeddings from posts of different stances are mixed up. Going deeper we find that both supportive and unsupportive posts express the same opinion towards it: air pollution is a disaster. In supportive posts, users dislike air pollution and suggest to use clean energy such as the solar or hydroelectric power. On the other hand in unsupportive posts, users dislike air pollution either but suggest to use nuclear power as it produce almost no air pollution.

4 Conclusion

We present WordForce, a user-dependent word embedding visualization and analysis system for debate issues, to demonstrate how to analyze transformed word embeddings from the stance aspect. WordForce can provide two-dimensional scatter plots as well as the dispersion and the distance values to demonstrate the word force for debatable topics. In the future, we plan to apply it on the research problems related to more debate issues.

Acknowledgements

References

Wei-Fan Chen and Lun-Wei Ku. 2016. UTCNN: a deep learning model of stance classification on social media text. In *COLING (to appear)*. auspices of the International Committee on Computational Linguistics.

Wei-Fan Chen, Yann-Hui Lee, and Lun-Wei Ku. 2015. Topic-based stance mining for social media texts. In *International Conference on HCI in Business*, pages 22–33. Association for Computing Machinery.

Li Dong, Furu Wei, Ming Zhou, and Ke Xu. 2014. Adaptive multi-compositionality for recursive neural models with applications to sentiment analysis. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence*. AAAI.

Mohit Iyyer, Jordan L Boyd-Graber, Leonardo Max Batista Claudino, Richard Socher, and Hal Daumé III. 2014. A neural network for factoid question answering over paragraphs. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pages 633–644. Association for Computational Linguistics.

Yoon Kim. 2014. Convolutional neural networks for sentence classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1746–1751. Association for Computational Linguistics.

Siwei Lai, Liheng Xu, Kang Liu, and Jun Zhao. 2015. Recurrent convolutional neural networks for text classification. In *Proceedings of the Conference of the Association for the Advancement of Artificial Intelligence*, pages 2267–2273. AAAI.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(Nov):2579–2605.

Oren Melamud, Omer Levy, Ido Dagan, and Israel Ramat-Gan. 2015. A simple word embedding model for lexical substitution. In *Proceedings of the 1st Workshop on Vector Space Modeling for Natural Language Processing*, pages 1–7.

T Mikolov and J Dean. 2013. Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*.

Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1532–1543. Association for Computational Linguistics.

Dhanya Sridhar, James Foulds, Bert Huang, Lise Getoor, and Marilyn Walker. 2015. Joint models of disagreement and stance in online debate. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics*, pages 116–125. Association for Computational Linguistics.

Duyu Tang, Bing Qin, and Ting Liu. 2015. Learning semantic representations of users and products for document level sentiment classification. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, pages 1014–1023. Association for Computational Linguistics.

Jin Xu, Yubo Tao, and Hai Lin. 2016. Semantic word cloud generation based on word embeddings. In *2016 IEEE Pacific Visualization Symposium (PacificVis)*, pages 239–243. IEEE.