Title	
Author(s)	Nguyen, Thanh Huy
Citation	
Issue Date	2019-03
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/15788
Rights	
Description	Supervisor: NGUYEN, Minh Le, ,

Abstract

Sentiment classification on Twitter social networking has been becoming popular in recent years. People express their opinions and feeling about everything on Twitter social networking. These opinions and feeling can be used as useful information for decision making. For example, customers want to know the opinions of other users about a product before making a purchasing decision. Companies want to know the feedback of consumers about a product or the aspects of the product to improve the quality of that product. Therefore, sentiment analysis is playing a big role in the real world and become one of trending research topics in natural language processing. Some previous studies showed the satisfying results of sentiment classification by using traditional machine learning models or lexicon-based approaches. However, these results are on traditional social networks such as forum and review, where texts/ documents are formal, long and easily to interpret. It is still hard to analyze the sentiments of tweets. Tweets are very short and contain many noises (e.g., slang, informal expression, emoticons, mistyping and many words that have no in a dictionary). Traditional methods can not achieve good performance due to the unique characteristics of Twitter social networking. Moreover, most of the traditional methods require laborious feature engineering that is difficult to extract for a specific domain. On the other hand, existing sentiment analysis approaches mainly focus on measuring the sentiment of individual words without considering the semantics of a word and the relationship between words.

In this thesis, we research and develop deep learning methods to classify the sentiment polarities of tweets on Twitter micro-blogging. We not only focus on classifying the sentiment polarity of each tweet by considering textual information but also considering the aspects of each tweet. Three main sentiment analysis tasks are considered (1) Tweet-level sentiment analysis. We introduce a deep learning approach that models the different characteristics (flavor-features) of each word and tries to incorporate them into the deep neural network in order to extract correct sentiment contextual words. Four flavor-features (Word embeddings, Dependency-based word embeddings, Lexicon embeddings, and Character attention embeddings) provide real-valued hints to alleviate the data sparseness and improve sentiment classification performance. Specifically, the data sparseness is reduced by the following two methods. First, we perform data processing and apply semantic rules to deal with noise, negation and specific PoS particles in tweets. Second, we develop the multiple perspectives of each word upon word embeddings for the deep neural network to modeling the structure of tweets. (2) Aspect-level sentiment classification. We propose methods to incorporate aspect information into deep neural networks by using the advantages of multiple attention mechanisms, iterative attention mechanism. In this task, the sentiment lexicon feature is still interpolated into feature vectors and is studied the effect of classifying the sentiment polarities of aspects. (3) Multitask-based aspect-level sentiment classification. We introduce a multi-task learning approach which combines multiple inputs to address the drawbacks of aspect-level data. The multi-task learning called transfer learning allows the model to learn interactive knowledge between many tasks in order to deal with the difficulty in aspect-level data is that existing public data for this task are small which largely limits to the effectiveness of deep learning models. The sentiment lexicon is still considered as a flavor-feature to highlight the importance of aspects and their contexts.

The proposed methods are effective and significantly improve the performance compared to the baselines and the-state-of-the-art models.

Keywords: Tweet-level Sentiment Analysis, Aspect-level Sentiment Analysis, Twitter Social Networking, Deep Learning, Multi-task Learning.