Title	A Joint Model of Term Extraction and Polarity Classification for Aspect-based Sentiment Analysis
Author(s)	Nguyen, Ngoc Gia Hy
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15460
Rights	
Description	Supervisor:白井　清昭，先端科学技術研究科，修士（情報科学）

A Joint Model of Term Extraction and Polarity Classification for Aspect-based Sentiment Analysis

NGUYEN Ngoc Gia Hy (1610434)

School of Information Science, JAIST, hynguyen@jaist.ac.jp

Extended Abstract

With explosion of the Internet in this digital age, more and more people easily access the Internet and share their opinions on social media (e.g., reviews, forum discussion, blogs and social networks). Analyzing and observing these opinions enable individuals and organizations to make their decision easier. However, since there are a lot of websites on the Internet, it is difficult to monitor the huge volume of opinionated text manually. Moreover, it is also known that human analysis may have a bias caused by analyzer's subjectivity toward some services and products. To overcome these limitations, recent researchers have raised a lot of tasks in opinion mining area to build systems that are able to analyze opinionated text automatically.

Aspect-based sentiment analysis (ABSA) is known as a significant task in opinion mining. The task aims to extract explicit aspects of an entity, along with sentiment expressed towards these aspects. To achieve this goal, two subtasks are performed: aspect term extraction (ATE) and aspect polarity classification (APC). ATE is a task to extract aspects of a target (i.e. product or service) from a review text, while APC is a task to classify the polarity of the extracted aspects into the positive, negative or neutral. However, recent work has solved these two subtasks separately or has only focused on either subtask. In addition, a sequential model of two subtasks may cause chain errors from ATE to APC and designing and running two models consume too many resources. In this thesis, we introduce a new problem named aspect term extraction and polarity classification (ATEPC) that is able to deal with two subtasks at the same time. Then, we propose several deep learning models to address this problem.

We first formulate the ATEPC task by considering new output labels. While IOB encoding is used to indicate beginning, inside, and outside of an aspect term in the ATE task and a set of categories {positive, negative, neutral} indicating the polarity of an aspect is used in the APC task, ATEPC uses a new "polarity IOB encoding" to indicate not only aspect term but also its polarity. A set of IOB labels are defined as $\{B_{pos}, B_{neg}, B_{neu}, I_{pos}, I_{neg}, I_{neu}, O\}$. For example, B_{pos} indicates the beginning of an positive aspect term, I_{pos} indicates the inside of an positive aspect term, O indicates the outside of aspect term, etc. By using this IOB encoding, we perform ATEPC task as a sequential labeling problem to extract aspect terms and their polarity.

A new method called BiLSTM+CRF is proposed to solve the ATEPC problem. It is based on a combination of bidirectional long short-term memory networks (BiLSTM) and conditional random fields (CRF). After preprocessing and tokenising a review sentence, each word in the sentence is represented as a continuous vector called word embedding , which are pre-trained from corpora by existing methods. The model then feeds word embedding into bidirectional LSTM in order to learn abstract representation of words in the sentence for not only aspect terms but also their polarity. Finally, the results from BiLSTM are passed as the input of CRF to predict the output sequence for the input sentence.

Next, the attention mechanism is introduced to the model BiLSTM+CRF in order to improve the performance in the ATE and APC tasks. The basic idea of attention mechanism is that the model should focus on the suitable context in the sentence when it decides an output for an target (aspect term). For example, if the model considers the target "voice" in the sentence "The voice of my Moto phone was unclear," the model should pay more attention to suitable context words such as "voice" and "unclear" rather than other unrelated words such as "the", "was." A requirement of the conventional attention mechanism is that it requires a target in order to incorporate new information. The target should be given to guess what words should be paid more attention. It can be applicable for a classification problem or generation problem. However, in a sequential labeling problem like ATEPC, no target is given to a model. Therefore, we propose a contextual attention mechanism that is able to overcome this problem. The basic idea of the contextual attention mechanism is that it treats each word in the sentence as a target and apply the attention mechanism for them. Using this technique, each word in the sentence is represented by not only the information of the word itself but also the information from its words in context. Thus the information of opinion words is directly incorporated into the aspect terms. Moreover, as for out-of-vocabulary (OOV) words, this technique also provides a better representation for them using information from context words instead of using random representation. To sum, the advantage of this attention mechanism is that it not only directly incorporate polarity information of terms in long distance but also synthesize word vectors of OOV words appropriately. These advantages play an important role in boosting the model performance in the APC task.

We carry out experiments to evaluate our proposed methods. Although we propose a joint model of the ATE and APC tasks, the performance of the models are evaluated separately. While the F1-measure is used as an evaluation criteria for ATE, the performance of APC is evaluated with respect to the accuracy. The accuracy of the APC task is defined as a proportion of the aspect terms whose polarity are correctly identified to the total number of the correctly extracted aspect terms. When a set of aspect terms correctly extracted by two methods are different, comparison of the accuracy of these methods is not completely fair. Therefore, for fair comparison of two models, we define micro-accuracy that evaluates the polarity classification of only aspect terms correctly extracted by both models.

Although BiLSTM+CRF is a common model, it performs well when it is applied to the ATEPC task. The experimental results on SemEval datasets show that BiLSTM+CRF outperforms several baseline models in the ATE task and gives a promising result in the APC task by dealing with ATE and APC at the same time. Moreover, the experiments also show that the model using the contextual attention mechanism (CATT+BiLSTM+CRF) have a significant improvement in the APC task compared with BiLSTM+CRF. Furthermore, we visualized the attention parameters trained by the model and confirmed that the contextual attention mechanism was appropriate to incorporate polarity information of words in long distance. In addition, an error analysis is carried out to reveal the major problems of the proposed models.

The contribution of this thesis is summarized as follows. Firstly, we proposed a new framework named ATEPC which could deal with two subtasks ATE and APC in opinion mining simultaneously. Secondly, we proposed the contextual attention mechanism to improve APC performance in ATEPC task. We confirmed that our proposed method outperformed previous work in the ATE task and achieved comparable results in the APC task via several experiments using SemEval datasets.