JID:IJA AID:8247 /FLA [m3G; v1.242; Prn:30/08/2018; 9:59] P.1 (1-10)

International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis ☆

Duc-Hong Pham [a,c], Anh-Cuong Le [b,*]

[a] Faculty of Information Technology, Electric Power University, Ha Noi City, Viet Nam
[b] NLP-KD Lab, Faculty of Information Technology, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[c] Faculty of Information Technology, University of Engineering and Technology, Vietnam National University, Ha Noi City, Viet Nam

ARTICLE INFO

Article history:
Received 28 June 2017
Received in revised form 13 August 2018
Accepted 14 August 2018
Available online xxxx

Keywords:
Aspect category detection
Aspect sentiment classification
Convolutional neural network
One-hot character vector
Word embedding

ABSTRACT

Representing words as real-value vectors and use them as input in deep neural networks is an effective approach in many natural language processing tasks. Currently, some studies use a lower-level representation which is character-based vectors. This paper addresses on how to integrate different representations of input for the problem of aspect-based sentiment analysis. We will propose a joint model of multiple Convolutional Neural Networks (CNNs) in which each individual representation of the input is handled by one CNN. In this work we focus on three kinds of representation including word embeddings from the two methods (Word2Vec and GloVe) and the one-hot character vectors. Our experimental results demonstrate that the proposed model can achieve state-of the-art performance in aspect category detection and aspect sentiment classification tasks.

1. Introduction

Word embeddings represent the words in a vocabulary as real-valued vectors in a multidimensional space. They are trained using a large set of unlabeled data and formulated as real-valued vectors based on the word appearance contexts. Word embeddings can capture syntactic and semantic information without using labeled data, and thus they are usefully applied in many natural language processing (NLP) tasks, such as text classification, information extraction, information retrieval, question answering, sentiment analysis, and machine translation.

There are two well known methods for encoding words as real-value vectors, which are Word2Vec (Mikolov et al. [1]) and GloVe (Pennington et al. [2]). Word2Vec is a "artificial neural network predictive" model with the Skip-gram and CBOW (continuous bag of words) architectures, whereas GloVe is a "count-based" model. The skip-gram model uses the current word to predict context words meanwhile the CBOW model is opposite that the current word is predicted from its context words. GloVe model is performed on aggregated global word–word co-occurrence statistics from a corpus. Therefore these two methods, Word2Vec and GloVe, may encode various aspects of language. Moreover, if using different corpora then they will obviously generate different word vectors where each of them may highly contain complementary information to the others.

☆ This paper is part of the Virtual special issue on Uncertainty Management in Machine Learning Applications, Edited by Van-Nam Huynh.

* Corresponding author.
E-mail addresses: hongpd@epu.edu.vn (D.-H. Pham), leanhcuong@tdtu.edu.vn (A.-C. Le).

https://doi.org/10.1016/j.ijar.2018.08.003
0888-613X/

Please cite this article in press as: D.-H. Pham, A.-C. Le, Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis, Int. J. Approx. Reason. (2018), https://doi.org/10.1016/j.ijar.2018.08.003

Based on a lower level representation some studies (e.g., dos Santos et al. [3], Zhang et al. [4]) have effectively used character level representation instead of word level representation. One-hot character vectors are the widely used character level representation, in which each vector contains only one position with value of 1 corresponding to the character index in the vocabulary, and zeros for all other positions. The dimensionality of the one-hot character vectors equals the character vocabulary size. Actually character level representation can capture arbitrary aspects of word orthography as well as potential sub-words or shape information of words. It means that this representation can represent more information at a more basic level, which is specially useful in cases of sparse data or noisy data.

Sentiment analysis is a type of data mining that aims to determine the polarity of people's opinions about someone or something. The general approach for solving this problem is to use natural language processing (NLP) techniques and computational intelligence models to analyze subjective information from sources such as social networks, Twitter, forums, blogs, etc. For example, Dragoni et al. [5] used linguistics and fuzzy models to formulate the sentiment of documents with considering the domain which a document belongs to; Alsinet et al. [6] used Valued Abstract Argumentation to represent tweet collections as weighted labeled discussion graphs and then develop an automatic labeling system using a Support Vector Machines model. Aspect-based sentiment analysis is a special type of sentiment analysis, its task is to identify the different aspects of entities in textual reviews and to determine the sentiments associated with these aspects. Some previous studies in this area employed traditional approaches based on lexical information (e.g., n-grams as features) and classification machine learning methods (e.g., support vector machines), such as those by Ganu et al. [7], Wagner et al. [8], and Kiritchenko et al. [9]. Recently, many studies have used word embeddings as inputs for neural network models, such as aspect extraction (Poria et al. [10]), aspect-level sentiment classification (Tang et al. [11], Wang et al. [12]), and determining aspect ratings and aspect weights (Pham et al. [13]). However, most of these studies only used one set of the learned word embeddings and they ignored the character-level representation.

In recent years, it has been shown that convolutional neural networks (CNNs) are highly effective and they have achieved state-of-the-art results for some NLP tasks, such as sentence modeling (Kalchbrenner et al. [14]), sentence classification (Lakshmana et al. [15]), and learning semantic representations for web search (Shen et al. [16]). CNNs provide an efficient mechanism for aggregating information at a higher level of abstraction. Some studies have used MCNNs for sentence classification. For example, Kim et al. [17] proposed a multichannel representation architecture based on variable-size filters. However, their multichannel model operates only with a single version of the pre-trained embeddings (i.e., pre-trained Word2Vec embeddings), where one is kept stable and the other is fine tuned by back-propagation. Yin et al. [18] developed this method further by incorporating diverse embedding versions, but their model requires input word embeddings with the same dimensions. Zhang et al. [19] improved this model for sentence classification by treating different word embeddings as distinct groups and applied CNNs independently to each, before concatenating all of the vectors obtained in the classification layer.

In this study we address the problem of how to integrate the different representations (in this work we choose the Word2Vec vectors, the GloVe word embeddings, and the one-hot character vectors) of input to generate a unified representation for the task of aspect-based sentiment analysis. To this end, we will propose a joint model called "Multichannel framework using Convolutional Neural Networks" (MCNN in brief) for multiple information sources, which is inspired from the work in Zhang et al. [19]. This model has the ability to learn the shared representation (i.e. the unified representation) from different input sources (for this work we use different representations as discussed above). Each CNN module (called CNN channel) in this model will handle one individual input representation but all CNN channels are simultaneously trained with the same objective function in only one training phase. It is worth to emphasize that this joint model is different from a trivial combination architecture which just combines the outputs of independent CNN channels.

In the experiments, we used about 52,574 reviews from the domain of restaurant products.[1] We evaluated the effectiveness of our MCNN model by applying it to two aspect-based sentiment analysis tasks, which comprised aspect category detection and aspect sentiment classification. The experimental results showed that our MCNN model performed better than the methods proposed in previous studies such as CNN models (Kim et al. [17]) and the CharSCNN model (dos Santos et al. [3]).

2. Aspect-based sentiment analysis formulation

Aspect-based sentiment analysis involves predicting the aspects of a predefined object (e.g., the *price, food, service, ambience, anecdotes* aspects of the object "restaurant") and the associated sentiments (e.g., *positive, negative*) assigned to each aspect in a certain context (we assume that the context is a sentence). According to [20,21], we formulate the aspect category detection and aspect sentiment classification problems as follows.

Aspect category detection: Given a set of k predefined aspect categories for an entity A, denoted by $\{A_1, A_2, \ldots, A_k\}$, for an input sentence d, we need to predict a binary label vector $a_d \in \mathrm{R}^{k}$. In particular, $a_{di} = 1$ means that sentence d contains the aspect category A_i and $a_{di} = 0$ means that it does not contain the aspect category A_i.

Aspect sentiment classification: Suppose that the aspect task is that defined above and we are given a set of l predefined aspect sentiment labels $O = \{O_1, O_2, \ldots, O_l\}$ (e.g., *positive, negative, neutral*), for the input sentence d, which is determined by

[1] http://spidr-ursa.rutgers.edu/datasets/.

Please cite this article in press as: D.-H. Pham, A.-C. Le, Exploiting multiple word embeddings and one-hot character vectors for aspect-based sentiment analysis, Int. J. Approx. Reason. (2018), https://doi.org/10.1016/j.ijar.2018.08.003