Chapter 8
A Deep Neural Network Model for Cross-Domain Sentiment Analysis

Suman Kumari
Swami Keshvanand Institute of Technology, Management, and Gramothan, Jaipur, India

Basant Agarwal
Indian Institute of Information Technology, Kota, India

Mamta Mittal
https://orcid.org/0000-0003-0490-4413
G.B. Pant Government Engineering College, New Delhi, India

ABSTRACT

Sentiment analysis is used to detect the opinion/sentiment expressed from the unstructured text. Most of the existing state-of-the-art methods are based on supervised learning, and therefore, a labelled dataset is required to build the model, and it is very difficult task to obtain a labelled dataset for every domain. Cross-domain sentiment analysis is to develop a model which is trained on labelled dataset of one domain, and the performance is evaluated on another domain. The performance of such cross-domain sentiment analysis is still very limited due to presence of many domain-related terms, and the sentiment analysis is a domain-dependent problem in which words changes their polarity depending upon the domain. In addition, cross-domain sentiment analysis model suffers with the problem of large number of out-of-the-vocabulary (unseen words) words. In this paper, the authors propose a deep learning-based approach for cross-domain sentiment analysis. Experimental results show that the proposed approach improves the performance on the benchmark dataset.

DOI: 10.4018/978-1-6684-6303-1.ch008

1. INTRODUCTION

Nowadays, online stores are in trend such as such as Amazon (amazon.com), Snapdeal (snapdeal.com). Nowadays numbers of people like to share their views about the product they buy. Whenever any user wants to buy something form online stores the only way to evaluate the product is another user's review about those products. But reading all the review one-by-one and understanding it overall is time consuming. It is attractive if we could have some type of analyser that can analyse reviews automatically. To make this process easy an automatic analyser came into existence which analyse all reviews or opinions and make it easy for the user to make view about the product. This process takes place under the field named as sentiment analysis. Reviewing product is useful for both customer as well as seller; customer can make decision about the product through review whereas seller can understand peoples thinking about the product. If seller sees the negative sentiments it can modify the product and if reviews are positive it can increase the sell accordingly (Jain et al. 2018b).

Sentiment analysis is the field of natural language processing in which computation about opinions, sentiments and emotions expressed in text is analysed (Tai et al. 2015). Sentiment analysis is the way of classification of text where various types of reviews are pre-classified into set of sentiment classes. The document with text can be classified with discrete sentiments value such as rating it on the basis of one to five stars or if the classification is binary, then there are mainly two classes a positive or a negative sentiment. In sentiment analysis, the classifier is trained by using manually labelled user review. For binary classification, the reviews are either divided in positive or negative class or are provided with score of 0 or 1. The whole classification is evaluated on another dataset which is the test dataset. If the classifier is trained using review of one domain and is tested on review of other domain, this is cross-domain sentiment analysis. During the process of training and testing, there are some words which appear in training dataset but do not appear in test dataset. These words which are not in training vocabulary but occur in test dataset are known as unseen words. Cross-Domain sentiment analysis become challenging, because training and testing dataset belongs from different domain, due to which number of unseen words increases and results in decrease of performance of the system. Such as the word '*sharp*', '*durable*' are only related to kitchen domain but will not occur for books in book domain. Dealing with this kind of problem make cross domain sentiment analysis challenging.

For sentiment classification, first, we need to represent words of vocabulary formed during training data in some lower-dimension embedding form. The most popular approach for generation of lower dimension word vector representation is through word embedding. Word embedding is a successful feature representation method in natural language processing (Bengio et al. 2003). The traditional method used for representing word in vector form is one-hot encoding method. Vectors for words were formed in binary format either 1 (if word were found) else 0 (if words were not found). The main limitation of this model is that it does not show the similarity between words. To overcome this limitation word embedding came into popularity. Word embedding forms word representations by considering semantic and syntactic relation between words. Vector representation was formed for only those words which were in vocabulary not for unseen words which are found when evaluation on target dataset is done (Agarwal et al. 2018; Sansanwal et al. 2019).

We proposed a cross-domain sentiment analysis method which has capability of building representation for those unseen words whose synonyms are present in the vocabulary. The proposed method calculates the synonyms of each unseen words and match the calculated synonyms from the words in vocabulary. If a word matches with synonyms the representation of word is provided to unseen word. This method