

Chapter 99

A Literature Review on Cross Domain Sentiment Analysis Using Machine learning

Nancy Kansal

Ajay Kumar Garg Engineering College Ghaziabad, India

Lipika Goel

Ajay Kumar Garg Engineering College Ghaziabad, India

Sonam Gupta

Ajay Kumar Garg Engineering College Ghaziabad, India

ABSTRACT

Sentiment analysis is the field of NLP which analyzes the sentiments of text written by users on online sites in the form of reviews. These reviews may be either in the form of a word, sentence, document, or ratings. These reviews are used as datasets when applied to train a classifier. These datasets are applied in the annotated form with the positive, negative or neutral labels as an input to train the classifier. This trained classifier is used to test other reviews, either in the same or different domains to know like or dislike of the user for the related field. Various researches have been done in single and cross domain sentiment analysis. The new methods proposed are overcoming the previous ones but according to this survey, no methods best suit the proposed work. In this article, the authors review the methods and techniques that are given by various researchers in cross domain sentiment analysis and how those are compared with the pre-existing methods for the related work.

1. INTRODUCTION

Sentiment analysis is also called opinion mining or emotion AI. Sentiment Analysis is used to know about what people think about things they are consuming from watching a movie to purchasing an AC. On the basis of their thinking either positive or negative, producer or holders of service or product get to know whether the given service/product has the future scope or not. For example, there is a new movie released; various social networking sites are the source of thoughts of people who have seen the movie. On the basis of those reviews, publicity of the movie reaches to the makers of the movie and related

DOI: 10.4018/978-1-6684-6303-1.ch099

trend goes on. Whenever there are elections to be held, election outcomes are predicted on the basis of the analysis of sentiments, opinion, and thoughts those are shared by public on various online portals or social networking sites or news. Sentiment Analysis is the study of attitudes of the holder of service towards the consumed service either in the form of love or hate; like or dislike; positive or negative (polarity). This attitude is analyzed from the text that is presented in the form of reviews in word form, sentence form or document form or in the form of ratings given by holder.

Sentiment Analysis can be done using various machine learning algorithms in which a model/classifier is trained using reviews that are annotated with the polarity positive or negative. These annotated reviews can be taken from any domain to train the classifier and the trained classifier is tested for the orientation of text or reviews in same or different domain.

While doing sentiment analysis in single domain, classifier is trained in single domain and is tested for the same domain thus only problem is to annotate the dataset but in cross domain sentiment analysis problem arises that any word/feature having positive meaning in one domain may have negative meaning in other domain; or in one domain, positive sentiments are expressed by some words and in other domain those positive sentiments are expressed using different words (domain specific words in different domains). In the cross domain, the problem of dataset labeling may be a time-consuming and costly process as it is done manually. SentiWordNet can also be a solution as it is an opinion Lexicon derived from WordNet database which is having scores of positive and negative for attributes and hence can tell polarity of the document on the basis of overall polarity of words written in the document, but it is also having a limitation of words. Various methods and techniques have been proposed recently to overcome this problem of labeling of dataset. The aim of the study is to put some relevant studies together in this paper to help the researchers, by comparing the methods proposed in various studies and by also giving a performance comparison of techniques used in studies. This study aims to focus only on the most recent works published during the period 2010 to 2019.

The remainder of this paper is as follows: Section 2 defines some key Terminologies used to understand the study on cross domain sentiment analysis. Section 3 describes some challenges and issues related to sentiment analysis in the cross domain. Section 4 demonstrates the methodology for this study purpose followed by a brief discussion of methods used in researches those the authors have used in this study. Section 5 demonstrates various datasets used by different researches on cross domain sentiment analysis. Section 6 defines some of the baseline methods used to compare performances of their proposed methods by various scholars. Section 7 gives answers to the questions that are aimed at this survey study. It compares the performance of all the baselines proposed by various researches. Section 8 is about some discussion on methods given to the problem of sentiment analysis in the cross domain and the Section 9 concludes the paper.

2. KEY TERMINOLOGY

Here, the authors define some basic terminologies that are used for this review purpose.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

www.igi-global.com/chapter/a-literature-review-on-cross-domain-sentiment-analysis-using-machine-learning/308580

Related Content

Association Rule Mining Based on Hybrid Whale Optimization Algorithm

Zhiwei Ye, Wenhui Cai, Mingwei Wang, Aixin Zhang, Wen Zhou, Na Deng, Zimei Wei and Daxin Zhu (2022). *International Journal of Data Warehousing and Mining* (pp. 1-22).

www.irma-international.org/article/association-rule-mining-based-on-hybrid-whale-optimization-algorithm/308817

Visualizing Multi Dimensional Data

César García-Osorio and Colin Fyfe (2008). *Successes and New Directions in Data Mining* (pp. 236-276).

www.irma-international.org/chapter/visualizing-multi-dimensional-data/29962

Distributed Privacy Preserving Clustering via Homomorphic Secret Sharing and Its Application to (Vertically) Partitioned Spatio-Temporal Data

Can Brochmann Yildizli, Thomas Pedersen, Yucel Saygin, Erkey Savas and Albert Levi (2011). *International Journal of Data Warehousing and Mining* (pp. 46-66).

www.irma-international.org/article/distributed-privacy-preserving-clustering-via/49640

An Efficient Compression Technique for Vertical Mining Methods

Mafruz Ashrafi, David Taniar and Kate Smith (2007). *Research and Trends in Data Mining Technologies and Applications* (pp. 143-173).

www.irma-international.org/chapter/efficient-compression-technique-vertical-mining/28424

Frameworks for Querying Databases Using Natural Language: A Literature Review – NLP-to-DB Querying Frameworks

Syed Ahmad Chan Bukhari, Hafsa Shareef Dar, M. Ikramullah Lali, Fazel Keshtkar, Khalid Mahmood Malik and Seifedine Kadry (2021). *International Journal of Data Warehousing and Mining* (pp. 21-38).

www.irma-international.org/article/frameworks-for-querying-databases-using-natural-language/276763