

Chapter 7

An Improved Cross-Domain Sentiment Analysis Based on a Semi-Supervised Convolutional Neural Network

Lap-Kei Lee

*Hong Kong Metropolitan University,
Hong Kong*

Kwok Tai Chui

*Hong Kong Metropolitan University,
Hong Kong*

Jingjing Wang

Hong Kong Metropolitan University,

Hong Kong

Yin-Chun Fung

*Hong Kong Metropolitan University,
Hong Kong*

Zhanhui Tan

*Hong Kong Metropolitan University,
Hong Kong*

ABSTRACT

The dependence on Internet in our daily life is ever-growing, which provides opportunity to discover valuable and subjective information using advanced techniques such as natural language processing and artificial intelligence. In this chapter, the research focus is a convolutional neural network for three-class (positive, neutral, and negative) cross-domain sentiment analysis. The model is enhanced in two-fold. First, a similarity label method facilitates the management between the source and target domains to generate more labelled data. Second, term frequency-inverse document frequency (TF-IDF) and latent semantic indexing (LSI) are employed to compute the similarity between source and target domains. Performance evaluation is conducted using three datasets, beauty reviews, toys reviews, and phone reviews. The proposed method enhances the accuracy by 4.3-7.6% and reduces the training time by 50%. The limitations of the research work have been discussed, which serve as the rationales of future research directions.

DOI: 10.4018/978-1-7998-8413-2.ch007

1. INTRODUCTION

Internet has become one of the most important tools in our daily life, with the ever-increasing penetration rate. It exceeded 60% this year according to Statista (Statista, 2021). The use of Internet relies more heavily on smartphones compared with computers attributable to the mobility and weight. During pandemic (COVID-19), we have witnessed the escalation of the penetration rate of internet for leisure and online shopping to maintain social-distancing and prevent the outbreak of the pandemic (Chang, & Meyerhoefer, 2021). More and more users begin to use online services, whom tend to read and share review comments on products. Before adding the items into shopping bag, it is often for one to consider the following criteria (i) review comments from other buyers; (ii) reputation of sellers; and (iii) price and quality. Particularly, the review comments contain complex and valuable information which can be effectively analysed via natural language processing (NLP) and artificial intelligence (AI). The research topic is known as sentiment analysis.

The contents can be generally categorized into 3-class, positive, neutral, and negative. In addition, multiple sources, as cross domains could be considered to enhance the analysis. In this chapter, the formulation tackles with 3-class cross-domain sentiment analysis.

1.1 Literature Review

Companies have realized that user-generated contents are useful. However, proper algorithms are needed to be implemented for the analysis. Sentiment analysis has recently received attention which drives various research groups to conduct systematic literature review (Do, Prasad, Maag, & Alsadoon, 2019; Hajiali, 2020; Yadav, & Vishwakarma, 2020). The focus of this chapter is sentiment analysis for product review.

Two network architectures were proposed for the 2-class (positive and negative) sentiment analysis of Amazon reviews (Zhao et al., 2017). The first architecture was constructed by weakly-supervised deep embedding (WDE) and convolutional neural network (CNN) whereas the second merged WDE and long short-term memory (LSTM). Performance evaluation was carried out based on the dataset containing 12k strong labeled and 1.1 million weak labeled reviews. The WDE-LSTM architecture slightly outperforms the WDE-CNN architecture. Both of them increase the accuracy by 3.4-21.7% compared with 11 existing approaches such as CNN-weak, CNN-rand11m, CNN-rand, sentiment-specific word embedding, Naive Bayes enhanced support vector machine, support vector machine, and Lexicon.

More than 14k 2-class reviews for 27 airlines were analyzed using word count analysis, frequency analysis, and topic modeling (Kwon, Ban, Jun, & Kim, 2021).

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/an-improved-cross-domain-sentiment-analysis-based-on-a-semi-supervised-convolutional-neural-network/293154

Related Content

Algebraic and Graphic Languages for OLAP Manipulations

Ravat Franck, Teste Olivier, Tournier Ronan and Zurfluh Gilles (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments* (pp. 60-90).

www.irma-international.org/chapter/algebraic-graphic-languages-olap-manipulations/40398

A Survey of Multidimensional Modeling Methodologies

Oscar Romero and Alberto Abelló (2009). *International Journal of Data Warehousing and Mining* (pp. 1-23).

www.irma-international.org/article/survey-multidimensional-modeling-methodologies/1824

Evolutionary Induction of Mixed Decision Trees

Marek Kretowski and Marek Grzes (2007). *International Journal of Data Warehousing and Mining* (pp. 68-82).

www.irma-international.org/article/evolutionary-induction-mixed-decision-trees/1794

Towards Improving the Lexicon-Based Approach for Arabic Sentiment Analysis

Nawaf A. Abdulla, Nizar A. Ahmed, Mohammed A. Shehab, Mahmoud Al-Ayyoub, Mohammed N. Al-Kabi and Saleh Al-rifai (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1970-1986).

www.irma-international.org/chapter/towards-improving-the-lexicon-based-approach-for-arabic-sentiment-analysis/150252

Anomaly Region Detection Based on DMST

Sulan Zhang and Jiaqiang Wan (2019). *International Journal of Data Warehousing and Mining* (pp. 39-57).

www.irma-international.org/article/anomaly-region-detection-based-on-dmst/223136