Front.Comput.Sci.
DOI

RESEARCH ARTICLE

A Unified Latent Variable Model for Contrastive Opinion Mining

Ebuka IBEKE, Chenghua LIN(✉), Adam WYNER, Mohamad Hardyman BARAWI

Department of Computing Science
University of Aberdeen, UK

Abstract There are large and growing textual corpora in which people express contrastive opinions about the same topic. This has led to an increasing number of studies about contrastive opinion mining. However, there are several notable issues with the existing studies. They mostly focus on mining contrastive opinions from multiple data collections, which need to be separated into their respective collections beforehand. In addition, existing models are opaque in terms of the relationship between topics that are extracted and the sentences in the corpus which express the topics; this opacity does not help us understand the opinions expressed in the corpus. Finally, contrastive opinion is mostly analysed qualitatively rather than quantitatively. This paper addresses these matters and proposes a novel unified latent variable model (contraLDA), which: mines contrastive opinions from both single and multiple data collections, extracts the sentences that project the contrastive opinion, and measures the strength of opinion contrastiveness towards the extracted topics. Experimental results show the effectiveness of our model in mining contrasted opinions, which outperformed our baselines in extracting coherent and informative sentiment-bearing topics. We further show the accuracy of our model in classifying topics and sentiments of textual data, and we compared our results to five strong baselines.

Keywords Contrastive opinion mining, Sentiment analysis, Topic modelling

Received month dd, yyyy; accepted month dd, yyyy

E-mail: *chenghua.lin@abdn.ac.uk* (Corresponding author)

1 Introduction

Recent years have seen a growing interest in text mining applications aimed at uncovering public opinions and social trends. This is partially driven by the fact that the Web now holds a large number of opinionated documents, such as opinion pieces and product reviews, to name a few. An additional driver is that the language one uses to express opinion indicates one's subjective viewpoints; this language can be used to understand and cluster people's opinion based on belief, experience or emotion, rather than facts. Text mining methods are therefore desired for facilitating automatic discovery of subjective viewpoints present in such large amounts of opinionated documents.

We define contrastive opinion mining as the discovery of opinion perspectives held by different individuals or groups, which are related to a given topic but opposite in terms of sentiments. The usefulness of contrastive opinion mining spans across many applications such as discovering the public's stand on major socio-political events [1], observing heated debates over controversial issues where different sides defend their viewpoints with contrasting statements [2], as well as mining issues from product review sites that can serve as an important source of feedback to businesses [3]. For example, there were heated discussions on the web about whether one should install the Mac OS X El Capitan soon after it was released to the public. Table 1 shows some discussions from the Apple Store, where people express highly controversial opinions after upgrading to the system, i.e., some experienced