Identifying the Opinion Orientation of Online Product Reviews at Feature Level: A Pruning Approach

Nilanshi Chauhan, National Institute of Technology, Hamirpur, India
Pardeep Singh, National Institute of Technology, Hamirpur, India

ABSTRACT

This article describes how e-commerce has become so vast that almost every product and service can be purchased online, to be delivered at our doorsteps. This has led to a striking increase in the number of online customers. In an attempt to make the online shopping more appealing and transparent to the online customers, the e-retailers allow their customers to express their opinion about the purchased products and services. Recently, analysis of such online reviews has become an active topic of research. This is because it is of immense concern to various stakeholders vs. online merchants, potential customers and the manufacturers of the particular product or service providers. The present article addresses the problem of summarization of such opinions expressed online and aims to create an organized feature-based summary as a solution. The proposed system depends on the frequency of occurrences of the potential features. A number of pruning methods are applied in order to obtain the final feature set and sentiment analysis has been done for each such feature.

KEYWORDS

Feature Extraction, Feature Pruning, Opinion Words, Sentiment Analysis, Summarization

1. INTRODUCTION

With the increasing popularity of websites such as Amazon, Snapdeal, Flipkart, Jabong, Myntra, and others, anything that one desires to purchase is available online. From electronic gadgets, furniture, clothing to groceries, everything can be bought online. All these items fall under the product domain. Websites such as Trivago, MakeMyTrip, OyoRooms etc. are also quite popular. These serve the service domain. Before purchasing any product or service online, potential customers prefer to find out what the current users of the very same products or services think about. To serve this purpose almost all the online websites provide a platform where customers can post reviews about the purchased items. There could be hundreds of reviews for a particular product. Such an information overload is more likely to frustrate the potential customers rather than helping them. This could have a negative influence on their purchase decisions. Hence there is a dire need to address this problem through efficient organization and summarization of these reviews. Opinion mining at feature or attribute level.
extracts the features of interest (FOI) from the online product reviews and determines the opinion associated with each of them.

All of the users of a product may not have same opinions about different FOIs and therefore the FOIs may vary from user to user. For example, say three customers can buy the same Canon printer but they may have different expectations from it, since a printer comes with a number of different features. If everything that customer_1 cares about is wired printing, then positive reviews may be written about the printing and other features related to it. On the other hand, if customer_2 is more concerned about the air-print and faxing features and is not satisfied with them then negative reviews may be submitted about it. The third customer may be concerned with three features say printing, copying and faxing. The person might write mixed reviews about the product if not satisfied with any one of them. Sometimes an agitated customer may be dissatisfied with the delivery services or has received a defective product or is awaiting the refund which has not been received for over a month. The person might indulge in writing long reviews about the delivery partners, sellers and the refund policy. A potential buyer may get confused and frustrated while going through such reviews. Consequently, an organized feature based summary of online product reviews could be very helpful in providing the customers with, ‘At a glance review’ of the product specific appropriate features. The system proposed in the present research work follows a bag-of-words approach for the task of feature extraction and narrows down the search space by eliminating all the words in the review dataset which are not nouns (NNs) or noun phrases (NNPs). Considering the NNs and the NNPs as the potential FOIs, the present technique subjects them to various filters namely, frequency of occurrence based, distance based and superset/subset based filters, in order to obtain a fine set of non-redundant and important FOIs pertaining to a particular product. The sentiment expressed with respect to each feature in the final feature set is determined using the sentiment classifier package and the opinion lexicon of NLTK.

Rest of the paper is organized as follows. Section 2 presents a brief background of research related to this study. Section 3 outlines the proposed methodology for generating the feature-based opinion summary. Section 4 includes the experimentation and discussion. Finally, conclusions are drawn in section 5. The limitations of the present work are also highlighted along with the scope for future work.

2. RELATED WORK

This study makes use of supervised machine learning approach to achieve the objectives of feature-level opinion mining. Opinion summarization uses machines learning algorithms which can be broadly classified into three categories namely supervised, semi-supervised and unsupervised. Supervised machine learning algorithms require a labeled training data for constructing an inference function. This labeled training dataset is very time consuming to construct or acquire. Semi-supervised learning algorithms use a large unlabeled dataset along with a small labeled training dataset. The unsupervised machine learning algorithms do not require such labeled training datasets. Since the data given to the unsupervised learning algorithms is in unlabeled, it is difficult to evaluate the accuracy of the output produced.

Poria et al. (2016) have suggested a deep learning approach to extract the aspects/features. They have incorporated a seven-layer deep convolutional network to mark each word as either aspect or non-aspect. To identify the sentiment associated with each aspect, the authors have combined the ensemble classifier with a word embedding model. The authors have successfully overcome the limitations of two widely used models in the field of opinion mining namely CRF and linguistic patterns model. The conditional random fields (CRF) is a linear model which needs a large number
Related Content

Service Innovation Analytics: Leveraging Existing Unstructured Data to Assess Service Innovation Capability
[www.igi-global.com/article/service-innovation-analytics/112039?camid=4v1a](www.igi-global.com/article/service-innovation-analytics/112039?camid=4v1a)

CbSSDF: A Two-Layer Conceptual Graph Approach to Web Services Description and Composition – A Scenario Based Solution Analysis and Comparison with OWL-S
[www.igi-global.com/article/cbssdf-two-layer-conceptual-graph/58646?camid=4v1a](www.igi-global.com/article/cbssdf-two-layer-conceptual-graph/58646?camid=4v1a)
Experience with Automatic Product Derivation of Mobile Applications Using Model-Driven Techniques
www.igi-global.com/chapter/experience-automatic-product-derivation-mobile/66463?camid=4v1a

Security Issues of CPS
www.igi-global.com/chapter/security-issues-of-cps/121254?camid=4v1a