

Chapter 17

A Novel Aspect Based Framework for Tourism Sector with Improvised Aspect and Opinion Mining Algorithm

Vishal Bhatnagar

Department of Computer Science and Engineering, Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

Mahima Goyal

Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

Mohammad Anayat Hussain

Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

ABSTRACT

With the growth of e-commerce web sites, the demand of writing reviews on these portals have gained huge popularity. This huge data must be mined to analyze the opinion and for making better decisions in different domains. In this paper, we have proposed an aspect based opinion mining algorithm for the tourism domain. It first determines the aspects, and then extracts the opinion words related to the aspects. The opinion words are provided a score based on the Senti-Wordnet and the final score of each aspect is calculated by the summation of the scores of the opinions. The final score is visualized depicting ranking of scores of different aspects for different hotels.

INTRODUCTION

With the growth of the world wide web in the recent years, the demand of expressing different opinions and experiences on different media has grown exponentially. The different media include reviews, blogs, forum and twitter. This opinionated data, in large volumes, has to be analyzed to make better decisions about different products and services. For instance, customer check reviews of restaurants on web plat-

DOI: 10.4018/978-1-6684-6303-1.ch017

forms before checking to a restaurant. Users also check the review of a product before buying a product and manufacturers need to understand a review to know about the sales performance of a product. The huge information from twitter can also be used to analyze political trends and popularity of a brand. Although, it appears interesting to analyze the available opinionated texts, it is quite challenging to estimate such kind of problems. People write a lot of objective sentences in their review which don't directly affect the opinion of the product. Thus, these sentences must be eliminated while analyzing the reviews.

Most of the authors have focused on classifying the reviews as positive or negative classes (Turney, 2002). For instance, a restaurant review can be classified into positive or negative based on the different data mining techniques. The different machine learning algorithms employed by the authors are Naive Bayes and SVM (Support Vector Machine) (Zhang et al., 2011). However, the aspect or feature based opinion mining (M. Hu and B. Liu, 2004) gives a broader and clearer picture of what the user wants. In this type, different features of the product and services are mined. The opinion is identified for the mined features and results are visualized according to the different features. This type of visualization allows potential customers to look for the features in which they are interested. For example, consider a hotel review in which customers usually comment about the features like food, staff, ambience and the other facilities of the hotel. In this, customer would like to know about the particular features in which one is interested. Thus, the aspect based opinion evaluation gives an insight, depth of these aspects. Moreover, this type of opinion mining has fascinated a lot of authors in the recent times because of its customization according to the aspects.

In this paper, we put forward a novel aspect based opinion mining approach using SentiWordNet by arranging the aspects of different reviews in the tourism domain. A lot of authors have applied opinion mining in product reviews, movie domain and political tweets, but very few have focused on the tourism domain. This domain is not a physical product review, but an intangible service as pointed by Taylor et al. (2014). An algorithm is proposed to extract the explicit aspects of the hotel reviews downloaded from the TripAdvisor site. The opinion is searched for each explicit aspect and those aspects which have an opinion are provided a score based on the Sentiwordnet. The cumulative score for each aspect is calculated and the results are visualized in the form of a graph. For example, consider the sentences- (1) The staff is helpful and friendly. (2) The staff enables you to find your bearings by providing a trishaw facility. In the first sentence, opinion 'helpful' and 'friendly' is extracted for the aspect 'staff'. In the second sentence, no opinion is extracted for the word 'staff'. The different techniques used in opinion mining are sentiment classification, subjectivity analysis, lexicon based, statistical based, dictionary based and semantic based. In this paper, we have used unsupervised dictionary based approach where the dictionary is SentiWordNet to extract explicit

This paper has been classified into different sections. Section 2 elucidates the description of related work. Section 3 describes the proposed framework and its architectural details. Section 4 describes an experimental view in tourism domain. Section 5 illustrates its evaluation and section 6 provides the conclusion with the future scope of the proposed system.

RELATED WORK

Bhatnagar (2010, 2013), Acharjee (2013), and Radhwan (2015) reviewed the different data mining techniques for data analysis. The method of capturing market intelligence has been shown by Li (2013). Ripon et al. (2016) found different automated machine learning techniques and compared them. Opinion

13 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-novel-aspect-based-framework-for-tourism-sector-with-improvised-aspect-and-opinion-mining-algorithm/308494

Related Content

Untangling BioOntologies for Mining Biomedical Information

Catia Pesquita (2009). *Handbook of Research on Text and Web Mining Technologies* (pp. 314-330).

www.irma-international.org/chapter/untangling-bioontologies-mining-biomedical-information/21732

Research on Data Mining and Investment Recommendation of Individual Users Based on Financial Time Series Analysis

Shiya Wang (2020). *International Journal of Data Warehousing and Mining* (pp. 64-80).

www.irma-international.org/article/research-on-data-mining-and-investment-recommendation-of-individual-users-based-on-financial-time-series-analysis/247921

Discovering Hidden Concepts in Predictive Models for Texts' Polarization

Caterina Liberati and Furio Camillo (2015). *International Journal of Data Warehousing and Mining* (pp. 29-48).

www.irma-international.org/article/discovering-hidden-concepts-in-predictive-models-for-texts-polarization/130665

Control of Inductive Bias in Supervised Learning Using Evolutionary Computation: A Wrapper-Based Approach

William H. Hsu (2003). *Data Mining: Opportunities and Challenges* (pp. 27-54).

www.irma-international.org/chapter/control-inductive-bias-supervised-learning/7595

Discovering Hidden Concepts in Predictive Models for Texts' Polarization

Caterina Liberati and Furio Camillo (2015). *International Journal of Data Warehousing and Mining* (pp. 29-48).

www.irma-international.org/article/discovering-hidden-concepts-in-predictive-models-for-texts-polarization/130665