Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 122 (2017) 237–244

Procedia Computer Science

www.elsevier.com/locate/procedia

Information Technology and Quantitative Management (ITQM 2017)

Opinion Ensembling for Improving Economic Growth through Tourism

Charu Puri[a,1], Akhil[a], Gaurav[a], Kush[a], Naveen Kumar[a]

[a]*Dept. of Computer Science, University of Delhi, New Delhi-10065, INDIA*

Abstract

Online reviews had a noticeable impact over the tourism industry, as tourists now make decisions based upon the reviews written by fellow tourists on travel websites. Online reviews directly or indirectly affects the economy of the country through foreign investment or through job opportunities. To process these online reviews as opinions faster, many algorithms were designed which changed the way people viewed online reviews. However, the individual results of these opinion processing algorithms were found biased or varying in some cases, which gave rise to the need of opinion ensemble techniques. We proposed opining ensembling which mine the online reviews extracted from tripadvisor.com. The empirical evaluation of the extracted reviews has resulted in suggestions to improve economic growth through tourism.

Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.

Keywords: Opinion Mining, Tourism, Ensemble, Tripadvisor

1. Introduction

The internet has emerged as a vast retainer of travel data, where users add latest travel reviews every day. TripAdvisor is one of the largest and most visited travel & tourism websites, with the database of more than 60 million members and over 170 million reviews and opinions for hotels, restaurants, attractions and other travel-related businesses [1]. Such online reviews form a sizable part of it and have a variety of information stored in them. Travelers go through these reviews and seek the necessary information they need. Such huge load of data makes it impossible for single person to read it all which gave rise to need for collecting and processing those reviews, and summarizing users relevant information. Several websites are equipped with rating system that grades reviews or other data by stars/numbers or text, while some websites cater both text as well as rating system [2]. Using only numerical rating system does not grant enough data as there are many other review related problems, which makes it hard to assess. Some of them are:

- Lengthy reviews about the hotel makes the reader to ignore the review,
- Difficult to compare hotels with different services offered because of unorganized reviews
- Opinions differ from one user to another
- Overall rating is affected by multiple aspects

1877-0509
Peer-review under responsibility of the scientific committee of the 5th International Conference on Information Technology and Quantitative Management, ITQM 2017.
10.1016/j.procs.2017.11.366

- Some reviews containing response of hotel staff to reviewer's complaints

Opinions have been an important part of our lives and it is a human behavior to analyze these opinions to take decisions. It is the field that deals in determination and classification of opinions or feelings expressed in review [3]. Several algorithms have been designed to process these online reviews as opinions. However, the individual results of these opinion processing algorithms were found biased or varying in some cases, which gave rise to the need of opinion ensemble techniques. Ensemble is an approach that epitomizes the review and excerpt the opinions from the data which gives the main context. Ensemble methods train multiple learners on the provided data set to solve the same problem then combine them to form a single model [3]. Thus, opinions mined through ensemble approach extricate the travelers from decision making process. Henceforth, impacting the economic growth as tourism provides direct and indirect jobs to people.

The main contributions of this paper are: (1) We have extracted the data in form of online reviews from TripAdvisor.com, (2) We proposed an opinion ensembling algorithm, (3) We studied the economic impact on tourism.

The remainder sections of the paper are organized as follows: previous work is expounded in Section 2; tourism and economic growth is discussed in section 3; the opinion ensemble is explained in Section 4; an experimental evaluation of opinion ensemble approaches in the online review is performed in Section 5; and finally, conclusions and future work are discussed in Section 6.

2. Previous Work

In [4] Khan and et al. presented literature survey of opinion mining. They have summarized various machine learning algorithms for sentiment classification from unstructured reviews. They have discussed various applications of opinion mining such as search engines, recommendation systems, email filtering, Web ad filtering, questioning/answering systems. In [5] Cristian Bucurab proposed a system for extracting and summarizing opinions expressed by users on TripAdvisor.com. Sentences are separated into integral units as words using the tokenization process and SentiWordNet evaluates the polarity of the separated words. They have evaluated the platform using text mining domain specific measures such as recall, f-measure, accuracy and precision. In [6] Cambria and et al. have created a compilation of frequently used polarity concepts i.e. common approach with relatively strong positive or negative polarity. They have developed SenticNet, a publicly available semantic resource for opinion mining. It exploits common argumentative techniques, such as blending and spectral activation, together with an emotion categorization model and an ontology for describing human emotions. In [7] Hu and et al. proposed a multi-text summarization technique for identifying the top- k most informative sentences of hotel reviews posted on TripAdvisor.com. To determine the similarity of two sentences the content and sentiment similarities were used. The k-medoids clustering algorithm was used to partition sentences into k groups and identified the top-k sentences. The medoids from these groups were then selected as the final summarization results. In [8] Taylor and et al. have extended Bing Liu's approach to describe customer inclination regarding tourism products by making use of the aspect-based opinion mining approach. They have extracted product reviews from Los Lagos, particularly, hotels and restaurants. They measured the performance of the proposed algorithm and designed and developed an application to extract opinions from these reviews and to generate proposed summarization charts. In [9] Lin and Chao proposed an approach for tourism-related opinion detection and tourist attraction target identification. They have extracted blog articles labeled as in the domestic tourism category in a blogspace. Annotators were used to annotate the opinion polarity and the opinion target for every sentence. They have used machine learning methods to train learners for tourism-related opinion mining. In [10] Basari and et al. have proposed a hybrid method of support vector machine and particle swarm optimization for opinion mining of movie review. A SVM-PSO technique improved the parameters of SVM using PSO. In [11] Li and et al. have developed VisTravel: visualizing tourism network opinion from the user generated content. They have extracted e-tourism User-Generated Content data from Mafengwo, which is one of the Chinese travel social networks. In [12] Akehurst have discussed the importance to systematically identify the type of tourist or traveller who actually writes blogs and what types of trip and stays in destinations are more likely to generate meaningful User Generated Content. In [13] Godnov and Redek have discussed the case study of Croatia for text mining in tourism. Latent