4th Information Systems International Conference 2017, ISICO 2017, 6-8 November 2017, Bali, Indonesia

The Utilization of Filter on Object-based Opinion Mining in Tourism Product Reviews

Aris Tjahyanto*, Bonda Sisephaputra

Department of Information Systems, Institut Teknologi Sepuluh Nopember, Jalan Raya ITS Sukolilo, Surabaya 60111 Indonesia

Abstract

The quality of a tourism products can be assessed using several aspects or objects due to their unique characteristics. The information related to the object can be extracted using object-based opinion mining. Based on the previous research, the implementation of Natural Language Processing (NLP) rules on object-based opinion mining for determining the orientation of the semantic objects showed good result. However, the performance of the objects extraction should be improved. In this study, researchers apply a filter on the objects' extraction process of the hotel and restaurant review data. The utilization of data filter in object-based opinion mining succeeded in obtaining better objects' extraction result due to the utilization of filter that eliminate the unrelated object. The application of filter in the process of objects extraction improve the precision of frequent object approach from 45.7% to 64.49% on of the hotel review and from 44.82% to 64.61% on the restaurant review. For frequent and infrequent approach, the precision was increased from 22.33% to 63.02% on the hotel review and from 21.6% to 65.4% on the restaurant review. For overall extracted object, the usage of filter got better result compared to non-filter classification process. The filtered object approach gave 56.85% accuracy, 60.91% precision, and 79.93% recall on the hotel review, and got 58.85% accuracy, 63.26% precision, and 84.14% recall on the restaurant review.

Keywords: Object-Based Opinion Mining, Opinion Mining, Sentiment Analysis, Tourism, Hotel, Restaurant

* Corresponding author. Tel.: +62-31-5999-944; fax: +62-3-5964-965.
E-mail address: aristj@its.ac.id

1. Introduction

Information about tourism is one of the most accessed information by public. With the availability of this information, public is no longer difficult to find the information about their holiday destination. There are numbers of websites provide information or reviews about the tourism product and also provide feedback feature on it. Then it became a problem when too many feedback occurs on the website, it would be more difficult for the visitor to get the whole picture about the tourism object.

The challenge of information extraction on data review are the number opinions in the examined object and also the features and components inside. On the other hand, there are review that contain no opinion, and also contain irrelevant or unrelated aspects to the components of the object being reviewed [1,2,3,4,5]. For determining the quality of a tourism product, someone can assess it from several aspects. Each tourism product should be evaluated based on its own aspects or objects. Therefore, the suitable method for extracting the information is aspect-based opinion mining. Aspect-based opinion mining analyzes the information in detail, until all aspects can be expressed in one opinion, and a clear picture for a tourism product is obtained. The results of the research indicate that the performance of aspects extraction still can be improved. It was caused by a high percentage of extracted expressions that not fit with the expression of a real aspect domain. As a result, the precision and recall could be decreased when all aspects are considered expressions [6,7,8,9].

The focus of this research is to implement a data filter based on the factors and components related to the domain of hotels and restaurants. Therefore, it is possible to improve the result of aspect extraction by eliminating unrelated or irrelevant objects to the product tourism. In this study the authors will mention the components or attributes of hotel domain and restaurant domain with term "object". The object is extracted from nouns and noun phrases that exist in the review sentence that considered affecting the quality of a tourism product service. The data review was categorized into three classes, that are positive, neutral, and negative. The extracted object will be used to perform classification process using machine learning methods, namely Naïve Bayes, Support-vector Machines (SVM), and Random Forest (RF).

2. Methodology

A general design of object-based opinion mining can be seen on Fig.1. The tourism product review should be prepared and preprocessed for further processing. The next process is finding any objects in the product review and determining whether the review comment is positive or negative. The extracted data will be applied to perform classification using Naïve Bayes, Support-vector Machines (SVM), and Random Forest (RF).
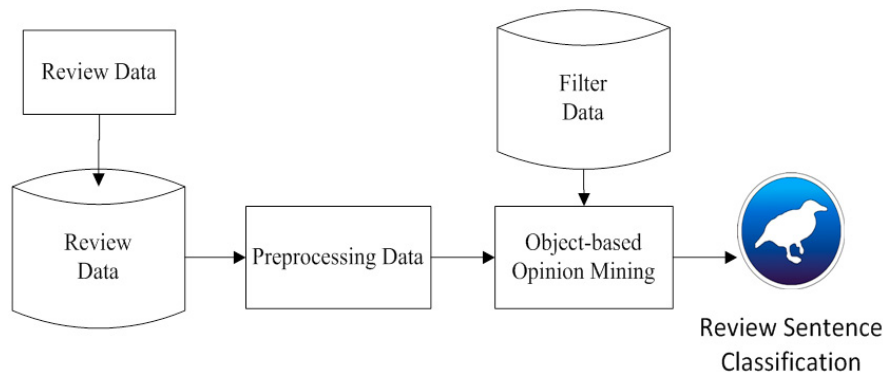

Fig. 1. General design of object-based opinion mining.

In this research, we used the same tourism product review data provided by Marrese-Taylor [1]. We also utilized the English lexicon opinion related with the product review. These data were taken from the website www.cs.uic.edu, that consist of 2.006 positive words and 4.783 negative words. For filtering the review sentence, we determine data