

Accepted Manuscript

Measuring the Quality of Hybrid Opinion Mining Model for E- Commerce Application

G. Vinodhini, R.M. Chandrasekaran

PII: S0263-2241(14)00191-2

DOI: <http://dx.doi.org/10.1016/j.measurement.2014.04.033>

Reference: MEASUR 2840

To appear in: *Measurement*

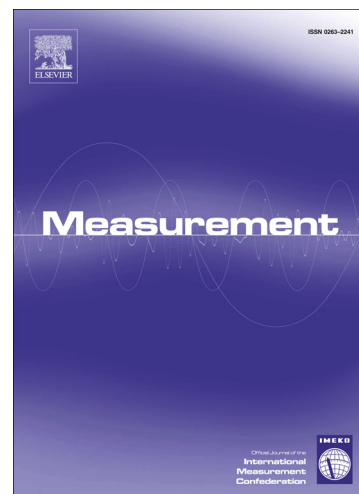
Received Date: 4 September 2013

Revised Date: 5 February 2014

Accepted Date: 23 April 2014

Please cite this article as: G. Vinodhini, R.M. Chandrasekaran, Measuring the Quality of Hybrid Opinion Mining Model for E- Commerce Application, *Measurement* (2014), doi: <http://dx.doi.org/10.1016/j.measurement.2014.04.033>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Measuring the Quality of Hybrid Opinion Mining Model for E- Commerce Application

G. Vinodhini, RM. Chandrasekaran

Department of Computer science and Engineering, Annamalai University, Annamalai Nagar -608002.
g.t.vino@gmail.com

Abstract

With the rapid expansion of e-commerce over the decades, the growth of the user generated content in the form of reviews is enormous on the Web. A need to organize the e-commerce reviews arises to help users and organizations in making an informed decision about the products. Opinion mining systems based on machine learning approaches are used online to categorize the customer opinion into positive or negative reviews. Different from previous approaches that employed single rule based or statistical techniques, we propose a hybrid machine learning approach built under the framework of combination (ensemble) of classifiers with principal component analysis (PCA) as a feature reduction technique. This paper introduces two hybrid models, i.e. PCA with bagging and PCA with bayesian boosting models for feature based opinion classification of product reviews. The results are compared with two individual classifier models based on statistical learning i.e. Logistic regression (LR) and Support vector machine (SVM). We found that hybrid methods do better in terms of four quality measures like misclassification rate, correctness, completeness and effectiveness in classifying the opinion into positive and negative.

Keywords: opinion, classification, unigram, bigram, feature, mining, reviews.

1. Introduction

E-commerce has attracted more and more people to buy and sell products online, customer reviews that describe experiences with product and services are becoming more important in decision making [26, 27]. Potential customers are interested to know the opinions of existing online customers to gather information about the products they plan to purchase, and businesses want to analyze the opinions of the customer of their products to monitor their brand. Customer reviews generally contain the product opinions of customers expressed using features of the product [1,9, 18, 28,40].

Opinion mining is a branch of data mining that analyzes individual subjective opinions such the orientations of the opinions [14, 21, 30, 34 , 37] . Within this broad field, much of the work has been focused on opinion polarity classification, where a text opinion is classified as positive or negative. Specifically, our focus is on feature-based opinion mining, in which the task applies to the sentence level to discover customers opinion about various aspects of a product [44]. Various machine learning classifiers have been used for opinion classification in the literature [24, 29, 33, 43, 48]. Also, many works in machine learning communities have shown that combining individual classifiers is an effective technique for improving classification accuracy[22, 32, 45, 47]. The emerging interest and importance of text sentiment classification in the real world applications, motivates us to perform a comparative study of hybrid methods in opinion classification. This study will greatly benefit application developers as well as researchers in the areas related opinion mining.

In this work, we introduce two hybrid models for opinion classification i.e. PCA with Bagging and PCA with Bayesian boosting models, using product attributes as features for classification model. They are empirically validated using a product review data set containing reviews collected from Amazon reviews. To analyze the relationship clearly two data models are developed. Model I using only unigram product attribute as features for classification . Model II using a combination of unigram, bigram and trigram product attribute as features for classification. The results are compared with two individual statistical model i.e. logistic regression and Support vector machine.

This paper is outlined as follows. Section 2 narrates the background. Section 3 discusses the problem outline used. Dimension reduction technique used is discussed in Section 4. Section 5 presents the various evaluation measures used. The various methods used to model the prediction system are introduced in Section 6. Data source used is reported in Section 7. Section 8 summarizes the results and Section 9 concludes our work.

Download English Version:

<https://daneshyari.com/en/article/7125022>

Download Persian Version:

<https://daneshyari.com/article/7125022>

[Daneshyari.com](https://daneshyari.com)