Accepted Manuscript

Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison

Murtadha Talib AL-Sharuee, Fei Liu, Mahardhika Pratama

PII: S0169-023X(17)30186-6

DOI: 10.1016/j.datak.2018.04.001

Reference: DATAK 1640

To appear in: *Data & Knowledge Engineering*

Received Date: 10 April 2017

Revised Date: 26 March 2018

Accepted Date: 3 April 2018

Please cite this article as: M.T. AL-Sharuee, F. Liu, M. Pratama, Sentiment analysis: An automatic contextual analysis and ensemble clustering approach and comparison, *Data & Knowledge Engineering* (2018), doi: 10.1016/j.datak.2018.04.001.

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.

ACCEPTED MANUSCRIPT

Sentiment Analysis: An Automatic Contextual Analysis and Ensemble Clustering Approach and Comparison

Murtadha Talib AL-Sharuee[a,*], Fei Liu[a], Mahardhika Pratama[b]

[a] *Department of Computer Science and Information Technology, La Trobe University, Bundoora, Victoria 3086, AUSTRALIA*

[b] *School of Computer Science and Engineering, Nanyang Technological University*

Abstract

Product reviews are one of the most important resources to determine public sentiment. The existing literature on review sentiment analysis mostly utilizes supervised models, which usually suffer from domain-dependency and require expensive manual labelling effort to provide training data. This article addresses these issues by describing a completely automatic and unsupervised approach to sentiment analysis. The method consists of two phases, which are contextual analysis and unsupervised ensemble learning. In the implementation of both phases, a sentiment lexicon, SentiWordNet, is deployed. Using effective contextual procedures and modifying the base learning component (the k-means algorithm) results in developing a successful approach to sentiment analysis which can overcome the domain-dependency and the labelling cost problems. The results show that the proposed nonrandom initialization of k-means yields a significant improvement compared to other algorithms. In terms of accuracy and performance, the proposed method is effective compared to supervised and unsupervised approaches. We also introduce new sentiment analysis problems relating to Australian airlines and home builders which could be potential benchmark problems in the sentiment analysis field. Our experiments on datasets from different domains show that contextual analysis and the ensemble phases improve the clustering performance in term of accuracy, stability and generalizability.

Keywords: Text mining, sentiment analysis, unsupervised learning, contextual analysis, ensemble learning, k-means algorithm

1. Introduction

Web development has changed human interaction and communication drastically and has led to an enormous and rapid growth in user-generated data. Thus, a very large number of product reviews is currently available which is rapidly and continuously increasing. Considerable attention has focused on analysing this data in terms of the sentiment it conveys, which has resulted in the emergence of the sentiment analysis (SA) research field. SA involves the computational analysis of user-generated materials, such as reviews, to determine its orientation (positive, negative or neutral). There are two main reasons to automate SA: first, the abundance of online materials is beyond human analysis; and second, public opinion is a significant consideration when governments, institutions, and individuals are making decisions. Many diverse domains and applications can benefit

*Corresponding author.

Email address: al-sharuee.m@students.latrobe.edu.au (Murtadha Talib AL-Sharuee)

Download English Version:

https://daneshyari.com/en/article/6853942

Download Persian Version:

https://daneshyari.com/article/6853942

Daneshyari.com