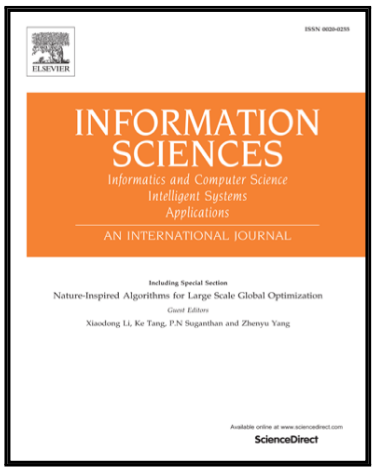
Accepted Manuscript



A Topic-Enhanced Word Embedding for Twitter Sentiment Classification

Yafeng Ren, Ruimin Wang, Donghong Ji

PII: S0020-0255(16)30464-9
DOI: 10.1016/j.ins.2016.06.040
Reference: INS 12314

To appear in: *Information Sciences*

Received date: 4 July 2015
Revised date: 13 June 2016
Accepted date: 23 June 2016

Please cite this article as: Yafeng Ren, Ruimin Wang, Donghong Ji, A Topic-Enhanced Word Embedding for Twitter Sentiment Classification, *Information Sciences* (2016), doi: 10.1016/j.ins.2016.06.040

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

A Topic-Enhanced Word Embedding for Twitter Sentiment Classification

Yafeng Ren, Ruimin Wang, and Donghong Ji

Wuhan University, Wuhan, China

Abstract

Word representation is crucial to lexical features used in Twitter sentiment analysis models. Recent work has demonstrated that dense, low-dimensional and real-valued word embedding gives competitive performance for Twitter sentiment classification. We follow this line of work, and propose a topic-enhanced word embedding for the task, which is generally neglected in previous work. Firstly, we exploit a recursive autoencoder framework to learn topic-enhanced word embedding, where topic information is generated through topic modeling based on an effective implementation of Latent Dirichlet Allocation (LDA). Then we use a uniform framework by adopting Support Vector Machine (SVM) classifier, to compare existing word representation methods with our method. Experimental results on the dataset show that topic-enhanced word embedding is very effective for Twitter sentiment classification.

Keywords: word embedding, recursive autoencoder, latent dirichlet allocation, twitter sentiment classification

1. Introduction

Twitter sentiment classification, which aims to classify the sentiment polarity of a tweet as positive, neutral or negative, has been intensively researched in recent years [16, 17, 18, 24, 29, 43, 49]. Most work follows Pang et al. (2002)' work, building a classifier based on annotated corpus with manually-designed sophisticated features [37].

Download English Version:

https://daneshyari.com/en/article/4944986

Download Persian Version:

https://daneshyari.com/article/4944986

Daneshyari.com