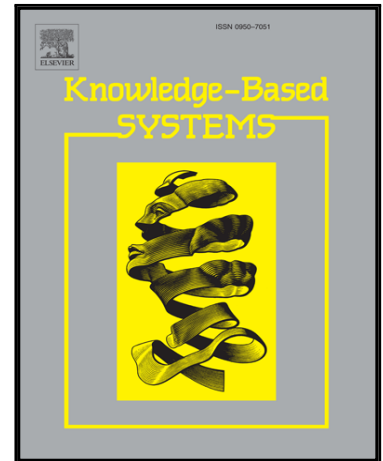


Accepted Manuscript

Weakly Supervised Topic Sentiment Joint Model with Word Embeddings

Xianghua Fu , Xudong Sun , Haiying Wu , Laizhong Cui ,
Joshua Zhexue Huang

PII: S0950-7051(18)30064-9
DOI: [10.1016/j.knosys.2018.02.012](https://doi.org/10.1016/j.knosys.2018.02.012)
Reference: KNOSYS 4220



To appear in: *Knowledge-Based Systems*

Received date: 15 June 2017
Revised date: 29 January 2018
Accepted date: 6 February 2018

Please cite this article as: Xianghua Fu , Xudong Sun , Haiying Wu , Laizhong Cui ,
Joshua Zhexue Huang , Weakly Supervised Topic Sentiment Joint Model with Word Embeddings,
Knowledge-Based Systems (2018), doi: [10.1016/j.knosys.2018.02.012](https://doi.org/10.1016/j.knosys.2018.02.012)

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.

Weakly Supervised Topic Sentiment Joint Model with Word Embeddings

Xianghua Fu, Xudong Sun, Haiying Wu, Laizhong Cui*, Joshua Zhexue Huang

College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, PR. China

fuxh@szu.edu.cn, snowalex@foxmail.com, 2160230425@email.szu.edu.cn, cuiliz@szu.edu.cn, zx.huang@szu.edu.cn

Abstract: Topic sentiment joint model aims to deal with the problem about the mixture of topics and sentiment simultaneously from online reviews. Most of existing topic sentiment modeling algorithms are mainly based on the state-of-art latent Dirichlet allocation (LDA) and probabilistic latent semantic analysis (PLSA), which infer sentiment and topic distributions from the co-occurrence of words. These methods have been proposed and successfully used for topic and sentiment analysis. However, when the training corpus is small or when the documents are short, the textual features become sparse, so that the results of the sentiment and topic distributions might be not very satisfied. In this paper, we propose a novel topic sentiment joint model called weakly supervised topic sentiment joint model with word embeddings (WS-TSWE), which incorporates word embeddings and HowNet lexicon simultaneously to improve the topic identification and sentiment recognition. The main contributions of WS-TSWE include the following two aspects. (1) Existing models generate the words only from the sentiment-topic-to-word Dirichlet multinomial component, but the WS-TSWE model replaces it with a mixture of two components, a Dirichlet multinomial component and a word embeddings component. Since the word embeddings are trained on a very large corpora and can be used to extend the semantic information of the words, they can provide a certain solution for the problem of the textual sparse. (2) Most of previous models incorporate sentiment knowledge in the β priors. And the priors are usually set from a dictionary and completely rely on previous domain knowledge to identify positive and negative words. In contrast, the WS-TSWE model calculates the sentiment orientation of each word with the HowNet lexicon and automatically infers sentiment-based β priors for sentiment analysis and opinion mining. Furthermore, we implement WS-TSWE with Gibbs sampling algorithms. The experimental results on Chinese and English data sets show that WS-TSWE achieved significant performance in the task of detecting sentiment and topics simultaneously.

Download English Version:

<https://daneshyari.com/en/article/6861563>

Download Persian Version:

<https://daneshyari.com/article/6861563>

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