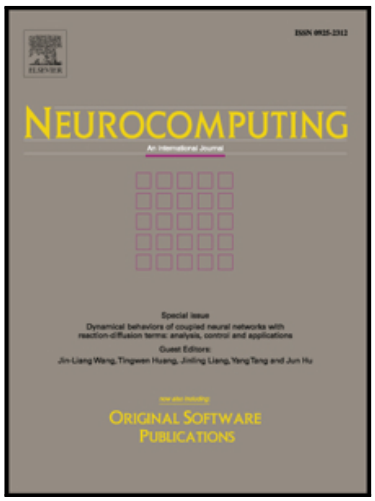

A Multi-Relational Term Scheme for First Story Detection

Yanghui Rao, Qing Li, Qingyuan Wu, Haoran Xie, Fu Lee Wang, Tao Wang

PII: S0925-2312(17)30399-5
DOI: 10.1016/j.neucom.2016.06.089
Reference: NEUCOM 18149

To appear in: *Neurocomputing*

Received date: 31 January 2016
Revised date: 17 June 2016
Accepted date: 28 June 2016

Please cite this article as: Yanghui Rao, Qing Li, Qingyuan Wu, Haoran Xie, Fu Lee Wang, Tao Wang, A Multi-Relational Term Scheme for First Story Detection, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2016.06.089

A Multi-Relational Term Scheme for First Story Detection☆

Yanghui Rao[a], Qing Li[b], Qingyuan Wu[c], Haoran Xie[d,*],
Fu Lee Wang[e], Tao Wang[f]

[a]*School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China*
[b]*Department of Computer Science, & Multimedia software Engineering Research Centre, City University of Hong Kong, Hong Kong*
[c]*School of Management, Beijing Normal University, Zhuhai, China*
[d]*Department of Mathematics and Information Technology, The Hong Kong Institute of Education, Tai Po, New Territories, Hong Kong*
[e]*Caritas Institute of Higher Education, New Territories, Hong Kong*
[f]*Department of Economics, University of Southampton, UK*

Abstract

First Story Detection (FSD) aims to identify the first story for an emerging event previously unreported, which is essential to practical applications in news analysis, intelligence gathering, and national security. Compared to information retrieval, text clustering, text classification, and other subject-based tasks, FSD is event-based and thus faces the challenging issues of multiple events on the same subject and the evolution of events. To tackle these challenges, several schemes for exploiting temporal information, named entity, and topic modeling, have been proposed for FSD. In this paper, we present a new term weighting scheme called LGT, which jointly models the Local element, Global element, and Topical association of each story. An unsupervised algorithm based on LGT is then devised and applied to FSD. We evaluate 4 feature reduction strategies and test our LGT scheme on an online model. Experiments show that our approach yields better results than existing baseline schemes on both retrospective and online FSD.

☆This paper is an extension of our previous work [26]. Compared to our previous work, we have added the following new contents in this paper: (1) the LGT is elaborated and compared theoretically in section 3.2.1; (2) three existing strategies and a newly proposed nonparametric method of feature reduction are included in section 3.2.2; (3) the experimental part is further extended by analyzing the performance with different topic numbers in section 4.4.1; (4) four feature reduction methods are evaluated and compared on two datasets in section 4.4.2; and (5) we have added more discussions on related studies in section 2, and made many improvements on the introduction, method, experimental analysis, conclusion, and future research directions.

*Corresponding author. Tel.: +852 68502340
Email address: hrxie2@gmail.com (Haoran Xie)