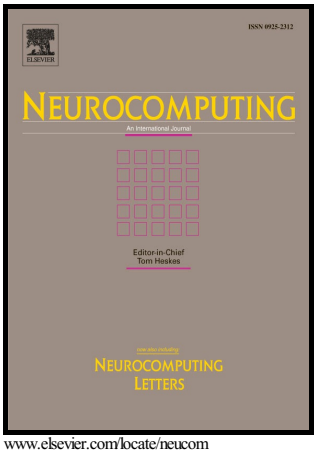

www.elsevier.com/locate/neucom

A Data-Intensive Approach for Discovering User Similarities in Social Behavioral Interactions Based on the Bayesian Network

Kun Yue, Hao Wu, Xiaodong Fu, Juan Xu, Zidu Yin, Weiyi Liu

PII: S0925-2312(16)31070-0
DOI: http://dx.doi.org/10.1016/j.neucom.2016.09.042
Reference: NEUCOM17571

To appear in: *Neurocomputing*

Received date: 19 May 2015
Revised date: 7 September 2016
Accepted date: 17 September 2016

Cite this article as: Kun Yue, Hao Wu, Xiaodong Fu, Juan Xu, Zidu Yin and Weiyi Liu, A Data-Intensive Approach for Discovering User Similarities in Social Behavioral Interactions Based on the Bayesian Network, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.09.042

A Data-Intensive Approach for Discovering User Similarities in Social Behavioral Interactions Based on the Bayesian Network

Kun Yue[1]*, Hao Wu[1], Xiaodong Fu[2], Juan Xu[1], Zidu Yin[1], Weiyi Liu[1]

[1]School of Information Science and Engineering,
Yunnan University, Kunming, China

[2]Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming, China

September 21, 2016

Abstract

Discovering user similarities from social media can establish the basis for user targeting, product recommendation, user relationship evolution and understanding. User similarities not only depend on the topological structure but also the dependence degrees between users. In this paper, we adopt Bayesian network (BN), an important and popular probabilistic graphical model, as the underling framework and propose a data-intensive approach for discovering user similarities. First, upon the massive social behavioral interactions, we give the method for measuring direct similarities between users and the MapReduce-based algorithm for constructing a BN to describe these similarities, called user Bayesian network and abbreviated as UBN. We also give the idea for storing large-scale UBNs in distributed file systems. Then, to measure indirect similarities between users, we give the method for measuring the closeness of user connections in terms of the properties of UBN's graphical structure. Further, we give the MapReduce-based algorithm for measuring the dependence degrees by means of UBN's probabilistic inferences. By combining the above two perspectives of measures, the indirect similarity degree between users can be achieved, while guaranteeing the applicability theoretically. Finally, we give experimental results and show the efficiency and effectiveness of our method.

Keywords: Social behavioral interactions; User Similarity; Bayesian network; MapReduce; Probabilistic inference

1 Introduction

With the rapid development of data acquisition techniques, Web2.0 applications and social networks, more and more user-generated data have been collected. These data include social behavioral interactions (e.g., "follow" or "comment" behaviors w.r.t. user blogs or ratings), and thus reflect user preference, mutual relationship and evolution [4, 12]. Analyzing, understanding and utilizing social media are paid great attention in the paradigms

*Corresponding author. E-mail: kyue@ynu.edu.cn.