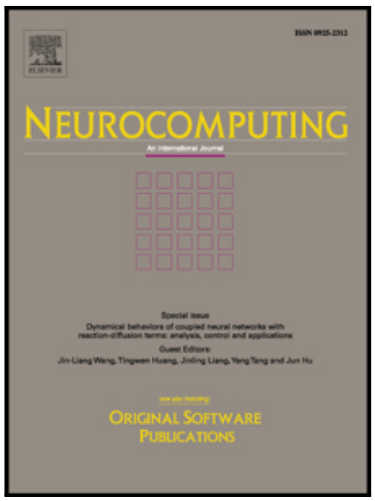

Multi-label Relational Classification via Node and Label Correlation

Zan Zhang , Hao Wang , Lin Liu , Jiuyong Li

PII: S0925-2312(18)30246-7
DOI: 10.1016/j.neucom.2018.02.079
Reference: NEUCOM 19384

To appear in: *Neurocomputing*

Received date: 12 April 2017
Revised date: 1 October 2017
Accepted date: 23 February 2018

Please cite this article as: Zan Zhang , Hao Wang , Lin Liu , Jiuyong Li , Multi-label Relational Classification via Node and Label Correlation, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.02.079

Multi-label Relational Classification via Node and Label Correlation

Zan Zhang[a,b,*], Hao Wang[a], Lin Liu[b], Jiuyong Li[b]

[a]*School of Computer Science and Information Engineering, Hefei University of Technology, Hefei, Anhui, China*
[b]*School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, SA, Australia*

ABSTRACT

Multi-label classification on social network data deals with the problem of labeling nodes in the network (i.e. instances in the data set) with multiple classes. Existing connectivity-based approaches have been used in classification by exploiting the correlations between linked nodes. However, this popular strategy may not always perform well, as it ignores the neighborhood of nodes and the correlations between nodes and class labels. In this paper, we propose a novel multi-label relational classifier which exploits the correlations between nodes and class labels. We first identify similar nodes for each unlabeled node based on local network structure. Then we perform clustering on nodes with known labels. We introduce an aggregated class probability to capture the correlations between nodes and class labels based on the clustering results. Experiments with real-world datasets demonstrate that our proposed method improves classification performance comparing to the existing approaches.

Keywords: Classification with Networked Data, Relational Learning, Clustering Analysis

1. Introduction

With the rapid growth and availability of data in social networks, mining networked data has become an important research task. In this paper, we focus on multi-label classification on networked data, for which instance (represented by a node of the network) may be associated with multiple class labels. Some popular applications of classification on networked data are Wikipedia page categorization based on intra-wiki links, genre identification of movies in a movie-actor network and Amazon book categorization in a co-purchasing network [1,11].

Classification is a well studied topic in machine learning and data mining [24]. However, instances in networked data are interconnected [1], traditional classification approaches, which usually assume that the individuals or instances are independently identically distributed, may not perform well [2].

Relational learning utilizes the correlations between nodes to address the problem of classification on networked data [3,4]. Many studies have shown that relational learning methods achieve better performance than conventional approaches [5,6,7]. However, for a multi-label problem, the dependencies between nodes and their labels are more complex than those in a single-label classification problem, so traditional relational learning methods are facing two challenges in a multi-label environment.

The first challenge is to do with multi-labeled nodes. For instance, in Figure 1, assume that there are two classes, A and B, and the ground truth of the labeling is that

*Corresponding author.
E-mail address: zan.zhang@mail.hfut.edu.cn