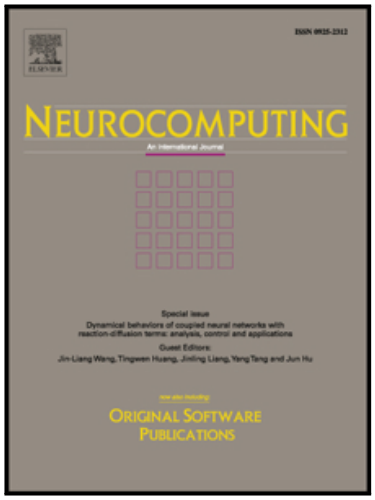
Communicated by Jun Yu

# Accepted Manuscript




Graph Classification Based on Graph Set Reconstruction and Graph Kernel Feature Reduction

Tinghuai Ma, Wenye Shao, Yongsheng Hao, Jie Cao

PII: S0925-2312(18)30321-7
DOI: 10.1016/j.neucom.2018.03.029
Reference: NEUCOM 19431

To appear in: *Neurocomputing*

Received date: 15 October 2017
Revised date: 31 January 2018
Accepted date: 12 March 2018

Please cite this article as: Tinghuai Ma, Wenye Shao, Yongsheng Hao, Jie Cao, Graph Classification Based on Graph Set Reconstruction and Graph Kernel Feature Reduction, *Neurocomputing* (2018), doi: 10.1016/j.neucom.2018.03.029

Available online at www.sciencedirect.com

ScienceDirect

Neurocomputing

Neurocomputing 00 (2018) 1–21

# Graph Classification Based on Graph Set Reconstruction and Graph Kernel Feature Reduction

Tinghuai Ma[a,b,*], Wenye Shao[a], Yongsheng Hao[a], Jie Cao[c]

[a] *School of Computer & Software, Nanjing University of information science & Technology, Jiangsu, Nanjing 210-044, China*
[b] *Jiangsu Engineering Center of Network Monitoring, Nanjing University of information science & Technology, Nanjing 210-044,China*
[c] *School of Economic & Management,Nanjing University of Information Science & Technology, Nanjing 210044, China*

**Abstract**

Graph, a kind of structured data, is widely used to model complex relationships among objects, and has been used in various of scientific and engineering fields, such as bioinformatics, network intrusion detection, social network, etc. Building an automatic and highly accurate classification method for graphs becomes quite necessary for predicting unknown graphs or understanding complex structures among different categories. The kernel method is regarded as a powerful solution to graph classification, which consists of two steps, namely, graph kernel mapping and classification. However, the feature selection process is ignored, and those sub-structures with low discriminative power result in classification accuracy decrease. In order to solve this problem, we propose an efficient graph classification algorithm based on graph set reconstruction and graph kernel feature reduction. First of all, the least discriminative frequent subgraphs and part of the infrequent subgraphs are removed to reconstruct the original graph set. Then we adopt the graph-kernel-based discriminant analysis method to perform feature reduction on the well-reconstructed graph set. At last, the whole framework of the graph classification method is introduced and any commonly used classifiers can be utilized. Extensive experimental results on a series of bioinformatics benchmarks show that our graph classification algorithm demonstrates a significant improvement of prediction comparing with other graph-kernel-based classification approaches.

*Keywords:* Graph classification, Discriminative subgraph, Graph kernel, Kernel discriminant analysis

## 1. Introduction

Graph, a discrete structure, has been widely utilized to represent the complex structural information among objects[1, 2]. The nodes in graphs represent objects while the edges show the connections between pairs of objects[3, 4]. Because of the characteristic of strong flexibility, expressiveness, and with no dimension constraint, graph data mining has been used in extensive applications[5, 6], which range from bioinformatics (e.g., determining whether the DNA protein sequence is mutated)[7, 8] and chemistry (e.g., predicting whether the unknown chemical compounds are toxic)[9], to social networks (e.g., classifying social groups based on their internal structure characteristics)[10]. With the growth in demand for structural data analysis and the popularization of graph databases, there is a great need for building an automated graph classification model, and thus can predict the class of unknown graphs or understand the complex structures between different classes[11]. It is computationally intractable to implement these classification tasks by hand according to the experience of researchers. So more and more attention has been devoted to the development of graph classification algorithms in recent years.

*Corresponding author: thma@nuist.edu.cn