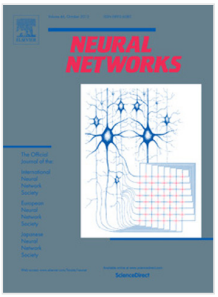
Accepted Manuscript



Attribute-based Decision Graphs: A framework for multiclass data classification

João Roberto Bertini Junior, Maria do Carmo Nicoletti, Liang Zhao

PII: S0893-6080(16)30138-1
DOI: http://dx.doi.org/10.1016/j.neunet.2016.09.008
Reference: NN 3670

To appear in: *Neural Networks*

Received date: 29 June 2016
Revised date: 16 September 2016
Accepted date: 27 September 2016

Please cite this article as: Bertini Junior, J. R., Nicoletti, M. d. C., & Zhao, L. Attribute-based Decision Graphs: A framework for multiclass data classification. *Neural Networks* (2016), http://dx.doi.org/10.1016/j.neunet.2016.09.008

Attribute-based Decision Graphs: A Framework for Multiclass Data Classification

João Roberto Bertini Junior[a,*], Maria do Carmo Nicoletti[a,b], Liang Zhao[c]

[a] *Computer Science Department, Federal University of São Carlos, Rod. Washington Luís Km 235, São Carlos, SP 13565-905, PO Box 676, Brazil*
[b] *FACCAMP, R. Guatemala 167, Campo Limpo Paulista, SP 13231-230, Brazil*
[c] *Department of Computer Science and Mathematics, School of Philosophy, Science and Literature of Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP 14040-901, Brazil*

Abstract

Graph-based algorithms have been successfully applied in machine learning and data mining tasks. A simple but, widely used, approach to build graphs from vector-based data is to consider each data instance as a vertex and connecting pairs of it using a similarity measure. Although this abstraction presents some advantages, such as arbitrary shape representation of the original data, it is still tied to some drawbacks, for example, it is dependent on the choice of a pre-defined distance metric and is biased by the local information among data instances. Aiming at exploring alternative ways to build graphs from data, this paper proposes an algorithm for constructing a new type of graph, called *Attribute-based Decision Graph – AbDG*. Given a vector-based data set, an AbDG is built by partitioning each data attribute range into disjoint intervals and representing each interval as a vertex. The edges are then established between vertices from different attributes according to a pre-defined pattern. Classification is performed through a matching process among the attribute values of the new instance and AbDG. Moreover, AbDG provides an inner mechanism to handle missing attribute values, which contributes for expanding its applicability. Results of classification tasks have shown that AbDG is a competitive approach when compared to well-known multiclass algorithms. The main contribution of the proposed framework is the combination of the advantages of attribute-based and graph-based techniques to perform robust pattern matching data classification, while permitting the analysis the input data considering only a subset of its attributes.

Keywords: Data-Graph construction, Graph-based classification, Multiclass classification, Attribute-based decision graphs, Missing attribute values.

1. Introduction

Graph-based algorithms have been widely studied in several research areas, mainly due to the advantages provided by the ability of topological representation of data. A great deal of tasks related to machine learning (ML) and data mining can be addressed by this kind of algorithms, see (Cook & Holder, 2006), (Bornholdt & Schuster, 2003) for instance. Generally, such a variety of graph-based algorithms can be categorized by the type of the input data they represent. Sometimes, the data are already intrinsically organized as a graph, such as in social networks (Scott, 2000) or protein-protein interaction networks (Vogelstein et al., 2000). However, in many situations, the data set consists of a collection of raw vectors unrelated to each other, such as medical exams of a patient or a client's bank transaction record.

When considering vector-based data, a fundamental step of any graph-based ML technique is converting the original data set into a graph, which is referred to as *data graph*. Then, techniques to process such a graph

*Corresponding author
Email addresses: bertini@dc.ufscar.br (João Roberto Bertini Junior), carmo@dc.ufscar.br (Maria do Carmo Nicoletti), zhao@usp.br (Liang Zhao)