Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 112 (2017) 186–194

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2017, 6-8 September 2017, Marseille, France

A New Feature Selection Method for Nominal Classifier based on Formal Concept Analysis

Marwa Trabelsi[a,*], Nida Meddouri[a], Mondher Maddouri[b]

[a] *Université de Tunis El Manar, Faculté des Sciences mathématiques, physiques et naturelles de Tunis - FST, Laboratoire d'Informatique, Programmation, Algorithmique et Heuristique - LIPAH, 2092, Tunis, Tunisie.*
[b] *College of Buisness, University of Jeddah, Kingdom of Saudi Arabia*

Abstract

The high dimension of data makes difficult to train and test many classification methods. This work aims to present a new filter Feature Selection Method, called H-Ratio, which can identify pertinent features from data. This method improves results of two previous works focusing on nominal classifiers based on Formals Concepts Analysis. The evaluation of H-Ratio shows that this method performs nominal classifiers processing. Our method has an error rate of 5% (∼7% relative improvement over a supervised classification method).

Keywords: Supervised Classification, Formal Concept Analysis, Feature Selection Methods, Nominal Classifiers;

1. Introduction

Classification is a data mining task that aims to get a simple schematic representation of complex data. More precisely, it consists of building a classifier in order to assign a class to each new object. Supervised classification involves a training step to build classifiers [1]. The most known methods of supervised classification such as Decision Trees (DT) and Formal Concept Analysis (FCA), use Feature Selection Methods in training step [2].

Feature Selection Methods have received an increasing attention over the last few years. It has been used for a wide variety of applications. Such methods proceed to eliminate unfavorable features such as noisy, redundant and irrelevant, that can penalize the performance of a classifier [3]. Feature selection contributes thus to reduce the high dimensionality of data and to restrict the input which can contain missing values data into a single or a subset of features [4,5]. In this paper, we propose a new Feature Selection Method to improve the performance of an existing supervised method based on Formal Concept Analysis named Classifier of Nominal Concepts (CNC). This later measures the closure of the most relevant attribute. Thus, the proper selection of relevant attribute leads to an efficient classification and vice versa [3]. The measure proposed here has conducted promising classification results for CNC and DNC (the Dagging of CNC) [6].

* Corresponding author. Tel.: +33-753-711-906.
E-mail address: trabelsimarou@live.com

1877-0509
10.1016/j.procs.2017.08.227

The rest of this paper is organized as follows: we overview existing Features Selection Methods in section 2. Section 3 presents basic notions related to Formal Concept Analysis. We focus mainly on CNC and DNC where we investigate our contribution. In section 4, we describe in detail our new Feature Selection Method. Finally, we show results of experiments in section 5. We compare methods under several conditions.

2. Related Works

As we mentioned in the introduction, Feature Selection Methods rules out irrelevant features and keep only the informative and pertinent ones. We distinguish two main categories of Feature Selection Methods like Wrapper Methods and Filter Methods[4,5]. **Wrapper Methods** maintain a subset of features while **Filter Methods** can select only a single feature which has to be the most relevant in the input set[4,5].

Wrapper Methods, from which we can cite Sequential Forward Selection Methods[7] and Genetic search Methods[8], use a classification method as part of the function evaluating feature subsets. The performance is usually measured in terms of the classification rate obtained on a testing set. In fact, the classifier is used as a black box for assessing feature subsets. Although these techniques may achieve a good generalization, the computational cost of training the classifier for a combinatorial number of times becomes prohibitive for high dimensional data sets. Filter Methods adopt complete independence between the learning machine and the data. They use general characteristics of the data to evaluate attributes and operate independently from any classification algorithm. Filter methods involve generally a non-iterative computation on the data set, which can execute much faster than a classifier training session[4,5].

In order to extract the most relevant feature, we focus in this paper on Filter Methods. Filter Methods estimate the contribution of each feature in classification. To do so, they use different heuristics: Distance Measures, Information Theory Measures or ones that evaluate the dependence between features called Dependence Measures[8].

Table 1 summarizes heuristics and cites for each of them some Feature Selected Methods among the most used.

Table 1. Filter Methods heuristics.

Heuristic	Feature Selected Methods
Information Measures	Information Gain[3], Gain Ratio[3], Mutual Information[9]
Dependence Measures	khi Deux[10], Pearson Correlation[11], Principal Components[12]
Distance Measures	Relief-F (RfF)[3]

2.1. Information Theory Measures

Information Theory Measures present the difference between the prior uncertainty and expected posterior uncertainty using an attribute *att*[8]. The attribute *att* is selected when its informational gain is greater than all attributes.

Among Information Theory Measures, we mention **Information Gain(IG)** which is considered as a popular measure for evaluating attributes. The Feature Selection Method calculates the informational gain of each attribute *att* by contribution to the class Y[3]. This task is based on formula (1):

$$IG(att) = H(Y) - H_{att}(Y) \tag{1}$$

$H(Y)$ calculates the entropy of the class Y using the formula below. In fact, the entropy, which is a mathematical function, corresponds to the information quantity contained or delivered by a source of information[13]. The entropy is used also in other measures of relevance, described below in formula (2):

$$H(Y) = \sum_i -P(v_i)log_2P(v_i) \tag{2}$$