Chapter 14
Feature Reduction with Inconsistency

Yong Liu
Institute of Cyber-Systems and Control of Zhejiang University, China

Yunliang Jiang
Huzhou Teachers College, China

Jianhua Yang
SCI-Tech Academy of Zhejiang University, China

ABSTRACT

Feature selection is a classical problem in machine learning, and how to design a method to select the features that can contain all the internal semantic correlation of the original feature set is a challenge. The authors present a general approach to select features via rough set based reduction, which can keep the selected features with the same semantic correlation as the original feature set. A new concept named inconsistency is proposed, which can be used to calculate the positive region easily and quickly with only linear temporal complexity. Some properties of inconsistency are also given, such as the monotonicity of inconsistency and so forth. The authors also propose three inconsistency based attribute reduction generation algorithms with different search policies. Finally, a "mini-saturation" bias is presented to choose the proper reduction for further predictive designing.

Feature selection is to find the "useful" feature subset from the original features. It is similar to a dimension reduction problem, and normally, after the feature selection, the selected feature set may achieve a superior classifier.

The main solutions of feature selection always try to project the large and high dimensional feature set into a small dimension with a certain constraint, and remove the irrelevant features. From the viewpoint of granular computing (Lin, 1989; Lin, 1997; Lin, 1998a; Lin, 1998b), the high dimensional feature set is a thin granular structure and the feature selection is to find a proper granular structure that can reflect the original feature set.

Many previous feature selection approaches (Bell & Wang, 2000; Koller & Sahami, 1996; Dash & Liu, 2003; Kononenko, 1994; Segen, 1984; Cardie, 1993; Sheinvald, Dom, & Niblack,

DOI: 10.4018/978-1-4666-1743-8.ch014

1990; Blum & Langley, 1997) can be classified into two categories: one is to use a related approximate measure to evaluate the features one by one and adds the features with positive value into the selected feature set, and the other one is to evaluate the subsets of features directly.

There are two disadvantages for the above approaches. The method, which evaluates the feature set one by one, will destroy the internal semantic relation of the original feature set. And the evaluation for the whole subset of features will lead to low efficiency both in temporal complexity and spatial complexity.

To overcome these two problems, many researchers introduce the rough set based reduction into feature selection (Hu, Zhao, Xie, & Yu, 2007; Jelonek, Krawiec, & Slowinski, 1995; Lin & Yin, 2004; Zhong, Dong, & Ohsuga, 2001; Swiniarski & Skowron, 2003). The reduction could preserve the semantic correlation of original features (Jensen & Shen, 2004).

In this paper, we address the two weaknesses in traditional feature selection and introduce a new feature selection approach with rough set based reduction. We propose a new concept named inconsistency which is easy to calculate and can evaluate whether the attribute set is a reduct quickly.

The rest of this paper is organized as follow: Section 2 presents the definitions and concepts related with inconsistency, some properties of inconsistency are also given in this section; Section 3 proposes three inconsistency based reduction algorithms with different search policies; Section 4 presents the "mini-saturation" bias based reduct selection policy to choose the "optimal" one from multiple reducts for further predictive modeling; and finally we conclude this paper in section 5.

1. RELATED DEFFINITIONS AND CONCEPTS

Some related definitions and concepts are presented as follow:

Definition 1 Positive region, *P* and *Q* are two sets in the information system *U(C, D)*, $P, Q \subseteq C \cup D$, then the positive region of *Q* in *P*, denoted as $POS_P(Q)$, can be calculated as:

$$POS_P(Q) = \bigcup_{X \in U/IND(Q)} \underline{P}X$$

Definition 2 Attribute dependency, *P* and *Q* are two sets in the information system *U(C, D)*, $\forall P, Q \subseteq C \cup D$, then the attribute dependency of attribute set *Q* on attribute set *P*, denoted as $\gamma_P(Q)$,can be calculated as:

$$\gamma_P(Q) = \frac{|POS_P(Q)|}{|U|}$$

The attribute dependency can describe which variables are strongly related to which other variables, for example, if $P \subset C$, then $\gamma_P(D)$ can be viewed as the measure between the decision attributes and the condition attributes, which can be implemented in further predictive modeling.

With the definition of attribute dependency, the attribute reduct can be defined as follow:

Definition 3 Attribute reduct, In information system *U(C, D)*, $R \subseteq C$, *R* is the reduct of *C* if and only if

$$POS_R(D) = POS_C(D) \text{ and}$$
$$\forall a \in R,\ POS_{R-\{a\}}(D) \neq POS_R(D)$$

or equivalently $\gamma_R(D) = \gamma_C(D)$ and $\forall a \in R,\ \gamma_{R-\{a\}}(D) \neq \gamma_R(D)$

The essence of attribute reduct is to find a subset *P* from condition set, and the subset *P* can maintain the same discriminability under the instance space. So we can judge whether the set is a reduct by its discriminability under the instance space. So the positive region, which calculates the number of instances that can be discriminable with the attribute set, can be used to find the reduct.

Related Content