

Discovering Semantic Inconsistencies to Improve Action Rules Mining

Zbigniew W. Ras^{1,2} and Angelina A. Tzacheva¹

¹ UNC-Charlotte, Computer Science Dept., Charlotte, NC 28223, USA

² Polish Academy of Sciences, Institute of Computer Science, Ordona 21, 01-237 Warsaw, Poland

Abstract. A new class of rules, called action rules, show what actions should be taken to improve the profitability of customers. Action rules introduced in [3] and investigated further in [7] assume that attributes in a database are divided into two groups: stable and flexible. These reflect the ability of a business user to influence and control their change for a given consumer. In this paper, we introduce a new classification of attributes partitioning them into stable, semi-stable, and flexible. Values of stable attributes can not be changed for a given consumer (for instance *maiden name* is an example of such an attribute). So, stable attributes have only one interpretation. If values of an attribute change in a deterministic way as a function of time (for instance values of the attribute *age*) we call them semi-stable. All remaining attributes are called flexible. Clearly, in the process of action rule extraction, stable attributes are highly undesirable. What about semi-stable attributes? Although, they seem to be quite similar to stable attributes, the difference between them is quite essential. Semi-stable attribute may have many different interpretations but among them only one interpretation is natural and it is called standard. All its other interpretations are called non-standard. In a non-standard interpretation, a semi-stable attribute can be classified as flexible (business user may control its change). In a single database we may easily fail to identify attributes which have non-standard interpretation. Query answering system based on distributed knowledge mining, introduced in [4,5], will be used in this paper as a tool to identify which semi-stable attributes have non-standard interpretation so they can be classified as flexible. This way, by decreasing the number of stable attributes in a database we may discover action rules which would not be discovered otherwise.

1 Introduction

Ras and Wiczorkowska [3] introduced the notion of action rules. Special type of rules can be constructed from classification rules to suggest a way to re-classify customers (objects) to a desired state. In e-commerce applications, this re-classification may mean that a consumer not interested in a certain product, now may buy it, and therefore may fall into a group of more profitable customers.

These groups are described by values of classification attributes in a decision table schema. In [3], all attributes are divided into stable and flexible. In

this paper, a new subclass of attributes called semi-stable attributes is introduced. Semi-stable attributes are typically a function of time, and undergo deterministic changes (for instance the attribute *age*). Different interpretations, called non-standard, of such attributes may exist, and in such cases all these attributes can be treated the same way as flexible attributes. In the algorithm of action rule extraction, presented in [3], attributes which are not flexible are highly undesirable. By identifying which semi-stable attributes have non-standard interpretation, we increase the number of flexible attributes and the same increase the chance to generate more precise action rules.

Assuming that attribute is flexible, we may find a way to change its value for a given object. However, quite often, such a change cannot be done directly to a chosen attribute (for instance to the attribute *profit*). In that situation, definitions of such an attribute in terms of other attributes have to be learned. These new definitions are used to construct action rules showing what changes in values of some attributes, for a given consumer, are needed in order to re-classify this consumer the way business user wants. We may search for definitions of these flexible attributes looking at either local or remote sites for help.

The application of semi-stable attributes to the process of action rules mining involves detection of nonstandard interpretations of semi-stable attributes. At local system level detection is possible, but limited to dependencies existing between local attributes. At distributed information systems level the detection of nonstandard interpretations involves discovering semantic inconsistencies, addressed by Ras and Dardzinska [8].

2 Information Systems and Decision Tables

An information system is used for representing business knowledge. Pawlak [2] gives the following definition:

By an information system we mean a pair $S = (U, A)$, where:

- U is a nonempty, finite set of objects (called customer identifiers),
- A is a nonempty, finite set of attributes i.e. $a : U \longrightarrow V_a$ for $a \in A$, where V_a is called the domain of a .

Information systems can be seen as generalizations of decision tables [2]. Partition of the set of attributes into conditions and decisions is given in any decision table. We assume that the set of conditions is partitioned into stable, semi-stable, and flexible conditions. Attribute $a \in A$ is called stable for the set U if its values assigned to objects from U can not be changed by a business user. An attribute is called semi-stable, if it is a function of time and it is changing in its standard interpretation in a deterministic way. Otherwise, it is called flexible. *Date of birth* is an example of a stable attribute.