



INFORMATION CONTENT OF ASSOCIATION RULES

*Arnost Vesely**

Abstract: Database records can be often interpreted as state descriptions of some world, system or generic object, states of which occur independently and are described by binary properties. If records do not contain missing values, then there exists close relationship between association rules and propositions about state properties. In data mining we usually get a lot of association rules with large confidence and large support. Since their interpretation is often cumbersome, some quantitative measure of their “informativeness” would be very helpful.

The main aim of the paper is to define a measure of the amount of information contained in an association rule. For this purpose we make use of the tight correspondence between association rules and logical implications. At first a quantitative measure of information content of logical formulas is introduced and studied. Information content of an association rule is then defined as information content of the corresponding logical implication in the situation when no knowledge about dependence among properties of world states is at our disposal. The intuitive meaning of the defined measure is that the association rule that allows more appropriate correction of the distribution of world states, acquired under unfair assumption of independence of state properties, contains also larger amount of information. The more appropriate correction here means a correction of the current probability distribution of states that leads to the distribution that is closer to the true distribution in the sense of Kullback-Leibler divergence measure.

Key words: *Association rules, information divergence, information content of logical formulas, information content of association rules*

Received: February 28, 2013

DOI: 10.14311/NNW.2014.24.014

Revised and accepted: May 12, 2014

1. Introduction

In data mining different methods for gaining knowledge are studied and used. The mining of association rules is one of the most successful. The purpose of mining association rules is to discover the associations among data in large databases or data sets, i.e. to find items that imply the presence of other items in the same database records or transactions. Association rules were firstly introduced by Agrawal [1] and then successfully applied by many authors.

*Arnost Vesely, Czech University of Life Sciences, Kamycka 129, Prague, Czech Republic, Tel.: +420 22438 2215, E-mail: vesely@pef.czu.cz