

not shop on the day we collected the data, as well as those that will be performed by customers who have not shopped here yet but one day (in the not too distant future) will. The target population thus is not only much larger than the given data set, but also is typically of infinite cardinality. Examining the whole population is out of the question in many cases of interest, both because of the size of the population, and also because many members of the population, such as those that exist in the future, cannot be available for examination at any cost.

This is not to suggest that the descriptive task is trivial. In general discovering a pattern that characterizes unstructured data is as difficult a task as making predictions about unseen instances. We are merely pointing out that the two problems are distinct, with different issues that need to be addressed. For instance, consider the notion of “interestingness” or “unexpectedness”. A rule that says

being in kindergarten \Rightarrow less than 10 years old

is likely to be highly accurate, but not very surprising. The value of this rule thus would be low in a descriptive context, from a knowledge discovery perspective, while its utility as a predictive rule in an inferential context may very well be much higher.

Making Inferences with Association Rules

The distinction between description and inference is a point worth emphasizing as often the implicit goal of an association rule mining session is inferential rather than descriptive. We look for rules that are expected to hold in a population that typically extends into the inaccessible (for the moment) future, and in any case is far greater than the sample data set we gathered the rules from.

A number of considerations make it unattractive to adopt standard association rules as inference rules mechanically. The former rules are abstracted from a given sample data set, while the latter rules are to be applicable to a larger population.

From Sample to Population

First, we need to take into account variations inherent in the process of sampling. Although the larger the sample size, the higher the proportion of samples that resemble the population, any given sample of any given size is unlikely to have exactly the same characteristics as the population. Giving an exact point value for the rule support and coverage parameters can easily convey an illusion of certainty of the findings.

Note that for the purpose of statistical inference, the sample relevant to a rule $X \Rightarrow Y$ is not the whole given data set Δ , but only that portion of Δ containing X . Thus, even from a single given data set, different rules may require the consideration of different samples (portions of Δ). In addition, the central limit theorem is based on the *absolute number* of instances in a sample, not the proportion it constitutes of the parent population (which is typically infinite).

This cannot be easily modelled in the standard association rule framework. Standard rules of the same coverage are considered to be of equal standing unless we also take

note of their respective sample sizes. The support of a rule $X \Rightarrow Y$ is construed as the *proportion* of XY s in Δ . This proportion is irrelevant here. What we need is the *absolute number* of instances of X in order to establish the degree of certainty, or statistical confidence, concerning the inference from the rule coverage in a sample to the rule coverage in the parent population.

Evaluation

Mining standard association rules is a clearly defined task. The objective there is to generate all rules of the form $X \Rightarrow Y$ which are above some given support and coverage thresholds. The problem of evaluation and validation is thus reduced to one of correctness and efficiency. Correctness in this case is unambiguous. Any algorithm is required to return *the* set of rules meeting the given criteria. Since there is no difference between the set of rules returned by one algorithm and the next, much of the research effort in this area has been understandably focused on efficiency issues, aiming to overcome the challenges imposed by the tremendous size of the data sets involved and the potential number of rules that can be generated. (Mannila, Toivonen, & Verkamo 1994; Savasere, Omiecinski, & Navathe 1995; Agrawal *et al.* 1996; Zaki *et al.* 1997, for example)

In those cases where variations to the standard framework are investigated, the refinements are mostly restricted to imposing additional constraints on top of the support and coverage criteria to pick out the more interesting and relevant rules from the huge pool of acceptable rules. Alternative measures to determine the fitness of a rule include, for instance, correlation, gain, Gini, Laplace, χ^2 , lift, and conviction. These metrics provide grounds to pre- or post-prune the standard association rules in order to arrive at a smaller set of rules. (Silberschatz & Tuzhilin 1996; Brin, Motwani, & Silverstein 1997; Bayardo & Agrawal 1999; Liu, Hsu, & Ma 1999, for example).

The several measures that have been used in the standard association rule literature are not entirely satisfactory as indicators of the quality of an inference rule. *Correctness* is not as well defined in the case of inference. *Efficiency* and the *quantity* of rules are important, but they should be supplementary to a measure of the substance of the rules. *Interestingness*, as we have already noted, is relevant for description but not as much of a concern for inference. In this paper we employ a measure from first principles, namely, comparing rules known to exist (probabilistically) in the parent population to rules obtained from a data set sampled from this population.

Interval Association Rules

The task of deriving predictive rules can be construed as a statistical inference problem. The parent population (all potential transactions) is typically very large if not infinite, and the data set we have at hand (transactions recorded on a given day) constitutes a sample drawn from this population. The problem then can be cast as the problem of projecting the associations found in the sample to justifiably probable associations in the parent population.

