1. INTRODUCTION

Data mining is an important well established research area and Association Rule Mining (ARM) is a very popular topic in the data mining community. The objective of ARM is to identify patterns, expressed as Association Rules (ARs), usually from binary-valued transaction data sets (Fayyad et al., 1996; Bodon, 2003; Coenen et al., 2004a, 2004b; Agrawal et al., 1993). Work has been done on a variety of extensions of the standard (binary-valued) approach to ARM thus allowing for its applicability to quantitative and categorical (non-binary) data (Gyenesi, 2001; Dong & Tjortjis, 2003; Srikant & Agrawal, 1996; Au & Chan, 1999). To deal with quantitative data, values are divided into ranges such that each range represents a binary valued attribute and then labelling the identified range attributes; for example “low”, “medium”, “high”, etc. There are two possible ways for assigning ranges: using crisp bound-
aries or fuzzy boundaries. Fuzzy ARM uses the latter to identify fuzzy ARs. Some earlier works show that more expressive ARs can be obtained using fuzzy ARM than “crisp” methods (Gyenesei, 2001; Kuok et al., 1998; Dubois et al., 2006; Khan et al., 2006). ARM (both fuzzy and standard) algorithms typically use the support-confidence framework to identify “interesting” ARs during the rule generation process. However, this framework has a number of disadvantages, for example, generating a vast AR set many of which are either obvious, subsumed by other rules, or largely redundant. Consequently there are motivations in the data mining community for finding more expressive, succinct or significant and useful ARs. Earlier work (Kuok et al., 1998; Khan et al., 2006) demonstrates this using the certainty measure, which is of note in the context of the work described here.

In this paper we introduce a particular category of a fuzzy ARM application called Composite item Fuzzy ARM (CFARM). CFARM’s objective is to generate fuzzy ARs from “properties” associated with composite attributes (Kim et al., 1989), i.e., attributes or items composed of sets of sub-attributes or sub-items that have a common schema. Image mining is a typical example where different areas of an image has groups of pixels such that each group can be represented by the normalized summation of the RGB values of the pixels in that group. In this case the set of composite attributes \( I \) is the set of groups, and the set of properties \( P \) shared by the groups is equivalent to the RGB summation values (i.e., \( P = \{R,G,B\} \)). We can then express fuzzy sets such as “light”, “medium” and “dark” and find associations between such composite attribute attributes with their properties. Considering the familiar market basket scenario, we can have defined \( I \) as a set of groceries and \( P \) as a set of nutritional properties that these groceries may possess, for example protein, iron, calcium and copper (i.e., \( P = \{Pr, Fe, Ca, Cu\ldots\} \)). Of note is the difference in these two examples. In the shopping basket, \( I \) is constant, i.e., it only represents a categorical list of common properties. In the image mining example, \( I \) is a normalized summation of properties.

Further, a stock control database can have \( I \) as a collection of stock items where \( P \) a collection of stock item properties is common to all items, including for example cost price, sale price, reorder time, etc. Given that we have quantitative attributes that can be partitioned into intervals or ranges, we rename such partitions with linguistic values or in this case, introduce fuzzy sets for these attributes. We are motivated by the fact that the approach described in this paper is a new way of dealing with so-called composite attributes that may potentially have fuzzy features.

The main contributions of the paper are:

1. The concept of CFARM.
2. The potential of ARs from itemset properties.
3. A practical example of the use of CFARM.
4. Employment of certainty factor, a quality measure to produce strong rules.
5. New Fuzzy Apriori-T algorithm for better efficiency.

We also demonstrate that a more succinct set of property ARs (than that generated using a non-fuzzy method) can be produced using the proposed approach.

The paper is organised as follows. In Section 2 we present the background and related work to the proposed composite fuzzy ARM approach described. Section 3 presents a sequence of terms and concepts for the work and Section 4 introduces the CFARM algorithm. The motivation for the work is expanded upon in Section 5 where an example application is described. A complete analysis of the operation of the CFARM algorithm is given in Section 6, and Section 7 concludes the paper with a summary of the contribution of the work and directions for future work.
An Approach to Improve Generation of Association Rules in Order to Be Used in Recommenders

HYBRIDJOIN for Near-Real-Time Data Warehousing
[www.igi-global.com/chapter/hybridjoin-near-real-time-data/70802?camid=4v1a](www.igi-global.com/chapter/hybridjoin-near-real-time-data/70802?camid=4v1a)