

1125-62-3066

**Kelly Bodwin\*** (kbodwin@email.unc.edu), **Kai Zhang** and **Andrew Nobel**. *Coherent Itemset Mining*. Preliminary report.

It is often of interest to find associations between variables based on observed binary data. This problem has been previously studied under the heading of “frequent itemset mining” or “association rule mining”. However, these classical approaches break down in the presence of non-identically distributed samples, which we argue is a common structure in real datasets of interest. In this paper, we propose a new model that accounts for differences in sample behavior while maintaining a common underlying variable dependence structure. We then introduce an algorithm that makes use of this model to identify groups of associated variables, which we refer to as coherent itemsets. The Coherent Itemset Mining (CIM) algorithm relies on an iterative update procedure that adaptively selects variable sets based on statistical testing principles. It is designed to run efficiently for high dimensional data. The CIM algorithm is tested on a variety of simulations as well as real datasets in genetics and text analysis. (Received September 20, 2016)