This thesis deals with a new approach to associative classification, especially fuzzy associative classification. Associative classification integrates Association Rule Mining (ARM), and classification by using ARM algorithms to generate the complete set of classification association rules. These methods mine high quality association rules and build classifiers based on them.

Most real-life datasets have a mix of numerical and binary attributes. And, crisp classifiers handle numerical attributes through sharp partitions, which are used to convert them to binary attributes. But using sharp partitions clustering introduces uncertainty especially at the boundaries of bins or clusters, leading to loss of information.

A better way to solve this problem is to have attribute values represented in the interval [0, 1], instead of just 0 and 1, and to have transactions with a given attribute represented to a certain extent (in the range [0, 1]). In this way crisp binary attributes are replaced by fuzzy ones. Thus, we need to use fuzzy methods, by which quantitative values for numerical attributes are converted to fuzzy values. Doing so ensures that there is no loss of information whatever may be the value of any numerical attribute. Moreover, the inherent uncertainty that is present in numerical data (as far as ARM is concerned) is also appropriately taken care of.

We present a new approach to associative classification (through SEAC and FSEAC algorithms) which relies on the use of the actual association rules for the final classification process. And, the two major hurdles in using these association rules directly, namely their huge volume and redundancy are dealt with appropriately by SEAC and FSEAC. They use a constrained exhaustive approach in order to obtain the best possible results in terms of accuracy, and do not use any greedy approach. They deal with redundant association rules through an effective and simple pruning technique which also helps in cutting down the number of rules which finally form a part of the classifier. The classifiers are built in a two-phased manner so as to achieve maximum accuracy and maximum representation of all possible class labels involved in the domain. They rely on information gain and entropy to build a set of optimal rules. Thus, their accuracies are very good as compared to other popular contemporary state-of-the-art classification algorithms, associative as well as non-associative ones. And, being a fuzzy classifier, FSEAC provides better accuracy as compared to SEAC, as is evident from our experimental studies. This is because of the ability of FSEAC in dealing with numerical attributes in an effective manner through soft partitions.
Using our new approach we show how associative classification can be applied to two real-world problems and domains. I-FAC and another algorithm which is its extension can be used for identifying object classes in images and for finding various visual concepts (like car, face, tree, and beach) in images, by leveraging interest point detectors that can be extracted from images. Another application of our associative classification approach is in Look-alike Modeling for Online User-Targeting of Ads.

To build any fuzzy associative classifier it is imperative to have efficient fuzzy ARM algorithms. There are many efficient crisp ARM algorithms, especially for large datasets which are omnipresent nowadays. But the same is not true for fuzzy ARM algorithms. The fuzzy ARM algorithms in this thesis try to solve this problem so that the actual fuzzy associative classifier building and classification can be done smoothly and quickly. FAR-Miner is meant for mining fuzzy association rules from large traditional transactional datasets (>1M records). We have developed FAR-HD, which is an algorithm for mining Fuzzy Association Rules in very large high-dimensional datasets. It is designed to mine fuzzy association rules from large datasets (>0.5M vectors), with each vector having at least 60 dimensions, or even more.

But, before any fuzzy ARM algorithm can be used, the original dataset (with crisp attributes) needs to be transformed into a form with fuzzy attributes. This pre-processing is done through our algorithm called FPrep, which involves using fuzzy clustering to generate fuzzy partitions. Then, these partitions are used to get a fuzzy version (with fuzzy records) of the original dataset. The fuzzy data is used to mine fuzzy association rules using a suitable algorithm like FAR-Miner.

### 13.1 Future Work

Of late Boosting based Ensemble Classifiers have become very popular and have been shown to be very accurate and efficient. Classifiers like Gradient-Boosted Decision Trees (GBDT) \cite{Fri99} train an ensemble of shallow trees based on the principle of boosting. The same approach can be used in associative classification. Our next endeavour is to build a boosted associative classification (based on SEAC and FSEAC) which trains an ensemble of associative classifiers containing rules with small rule lengths.

Next, with the ever increasing size of real-life datasets, especially those related to the Web, mining association rules quickly becomes a challenge. We plan to extend the FAR-Miner and FAR-HD algorithms such that they can be used in Cloud Computing environments (through Map-reduce tools like Hadoop). The idea is to parallelize the mining process in each partition. Each node on the grid mines rules only from a particular partition. Such parallelized versions of our mining algorithms would help us a lot in hastening the mining process.