CHAPTER - VI

CONCLUSIONS AND LINES FOR FUTURE WORKS

6.1 CONCLUSIONS AND COMMENTS ON RESULTS

In this research we have worked on finding some patterns on temporal datasets. In chapter-I, we have given a brief introduction about the thesis. In chapter-II, we have discussed about the background, preliminaries and other related works. Our works basically starts from chapter-III. In chapter-III and chapter-IV of the thesis, we mainly worked on time-stamp dataset. In chapter-III, we propose two algorithms one for finding locally frequent item sets and other for extracting local association rules from such locally frequent item sets. The first algorithm proposed by us dynamically extracts all the locally frequent item sets along the list of time intervals and each item set is associated with a list of time intervals where it is frequent. The algorithm is basically a modification of well-known A-priori algorithm and it can extract all frequent item sets extracted by other known method plus some extra frequent item sets, which cannot be extracted by others. Experimentally we have shown that our algorithm outperforms the others and the list of time intervals associated with a frequent item set depends upon, the two thresholds minthd1 and minthd2. In extracting local association rules for a locally frequent item set there is a difficulty. For example, if the size of the frequent item
set is \( m \) then it will have \( 2^m - 2 \) numbers of proper subsets and to find association rules for the frequent item set in any of the intervals, we need the supports of subsets in that interval which may not available, rather we would have supports of the subsets in some bigger time interval. So we need a number of passes through the dataset to find the supports which severely increase the I/O operations. So we loosely defined the association rules and proposed an algorithm, which works well in this purpose. The nicety about our algorithm is that it minimizes I/O cost although it compromises in the definition of association rules.

In chapter-IV, we derived some interesting results from the results obtained in the chapter-III viz. periodic patterns, sequential patterns etc. By taking time-stamp as calendar-dates, we can have some interesting patterns namely yearly pattern, monthly pattern, daily pattern etc. In the first part of chapter-IV, we have discussed about the methods of extracting such patterns if they exist. We have also discussed about some variations in periodic patterns viz. partially periodic patterns. In finding the periodicity, if it is found that the patterns are not periodic exactly but holds in enough number of intervals then such patterns are called partially periodic pattern. We propose an algorithm, which uses a set operation called superimposition to extract such partially periodic patterns. In finding the periodicity if the list of time intervals associated with a pattern has large overlapping, the above-mentioned set operation can be used to store the intervals, which turned out to be fuzzy time interval. In the same section, we made our studies in this regard. In certain cases, we may have more than one fuzzy intervals associated with a frequent item set. For example, if a periodic pattern
has two sets of superimposed time intervals, then the set of time intervals will be piled up in two places in the time domain to form two fuzzy time intervals. In some cases it may be more than two.

Up to this we have done intra-item set studies i.e. we tried to find patterns for a frequent item set. However pattern may exist among the frequent item sets. Next we have made inter-item set studies. For example, looking at the nature of the fuzzy time intervals associated with patterns we can do the clustering of corresponding patterns. As the variance of a fuzzy interval is invariant with respect to shifting, then two fuzzy time intervals (belonging to two different places in the same time domain) associated with two different patterns may have similar value of the variances and hence they may belong to the same cluster. In chapter-IV we also report our study made in this regard.

Again based on the ordering of the interval lists associated with item sets, we can extract frequent sequences of item sets. This is one type of sequence mining and it has nothing to do with classical sequence mining problem. In the last part of chapter-IV, we discuss about this. Our proposed algorithm takes as input all the locally frequent item sets and extracts all the frequent sequences of item sets.

In chapter-V, we have discussed about mining maximal frequent fuzzy intervals from temporal interval dataset. A temporal interval dataset is a dataset where each transaction is associated with a time interval of the type \([start\_time, end\_time]\) where \(start\_time\) is the starting time of the transaction and \(end\_time\) is the ending time of the transaction. This portion of our studies is mainly devoted to mining maximal frequent intervals of imprecise boundaries from such dataset.
Our method is a level-wise method. Although it looks like A-priori algorithm, but it has a slight variation from the A-priori algorithm in the sense that the two well-known properties viz. downward and upward closure property of frequent sets does not satisfy.

### 6.2 LINES FOR FUTURE WORKS

Most of the algorithms proposed in this thesis follow level-wise strategy. In future, approaches other than level-wise approaches can be looked for. In chapter-III, we have given the implementation details about algorithm-3.1. In future it can be looked for better implementations of the same. For example, using hash tree instead of trie-data structure, instead of maintaining the intervals as lists, other suitable data structures such as a balanced binary search tree could be used. Also implementations of all the other algorithms discussed in this thesis can be done. In chapter-V, algorithm-5.1 gives maximal frequent fuzzy intervals as output where the fuzzy intervals are specified by user. In future better techniques can be searched which can automatically extract maximal frequent fuzzy intervals. Also clustering of such frequent fuzzy intervals can be done using some statistical measure given in the literature. Modification of all the algorithms can be done so that they can be suitable for mining other types of temporal dataset as well as spatial datasets. Also, modification of all the algorithms can be done so that they can be fit to other frameworks given in statistics rather than traditional support-confidence frameworks. Lastly, different tools viz. Genetic Algorithm, Rough Set Theory, and Neural Network etc. can be used for mining such datasets.

*****