6. Conclusion and future works

6.1. Conclusion

In this paper, we discover association rules along with their multi-dimension patterns in terms of a set of hierarchy trees. We divide database into several meaningful element segmentations, and discover association rules in every meaningful combination of them. We extended the exiting algorithms in [33] to discover such multi-dimension association rules. Our experimental results show that our method is efficient and effective. The advantages of our method are:

1. The rules discovered in our algorithm satisfy a match ratio threshold *m*, thus the rules which only hold in several segmentations will be picked up by our algorithm.

2. We develop a mechanism to produce *multi-dimension patterns* with different granularities related to each dimension automatically. This make our algorithm requires less prior knowledge than the previous methods and hence may discover more unexpected rules.

3. The size of the mining result of an element segmentation R_{Ei} is much smaller than the size of raw data among it. We can store each R_{Ei} for future use. For example, if we want to change the concept hierarchy (such as from Fig. 6.1 to Fig. 6.2.), we need only reproduce each generalized pattern G_i and collect each $R_{Ei}(E_i \subset G_i)$ again, we can obtain the new results very quickly.

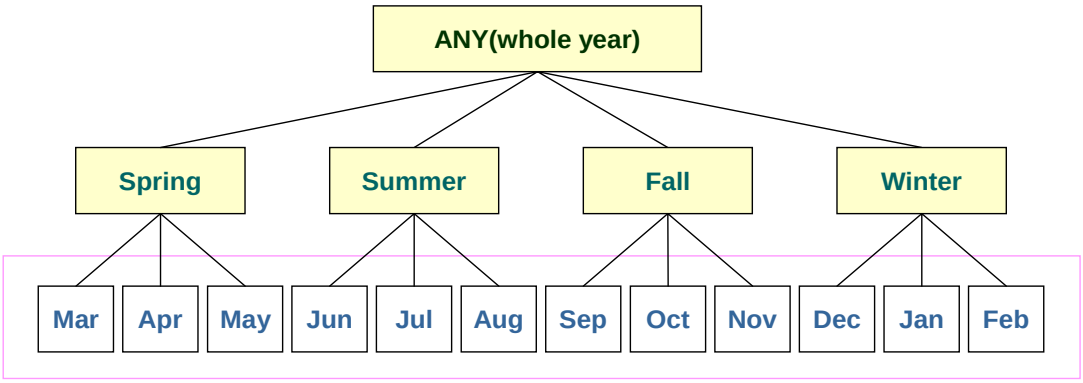

Figure 6.1. Example concept hierarchy I

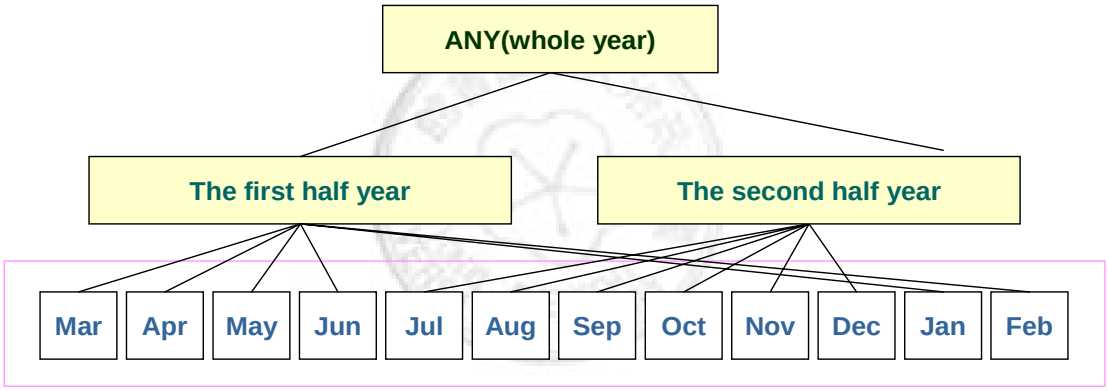

Figure 6.2. Example concept hierarchy II

4. By keeping the rules discovered in each element segmentation, when new data come, instead of taking all the past data into consider, we need only mining rules for the new data. We can produce the new result by combining the rules discovered from the new data with other exiting ones without redundancy scanning the past data. Thus we can implement an incremental mining by our algorithm.

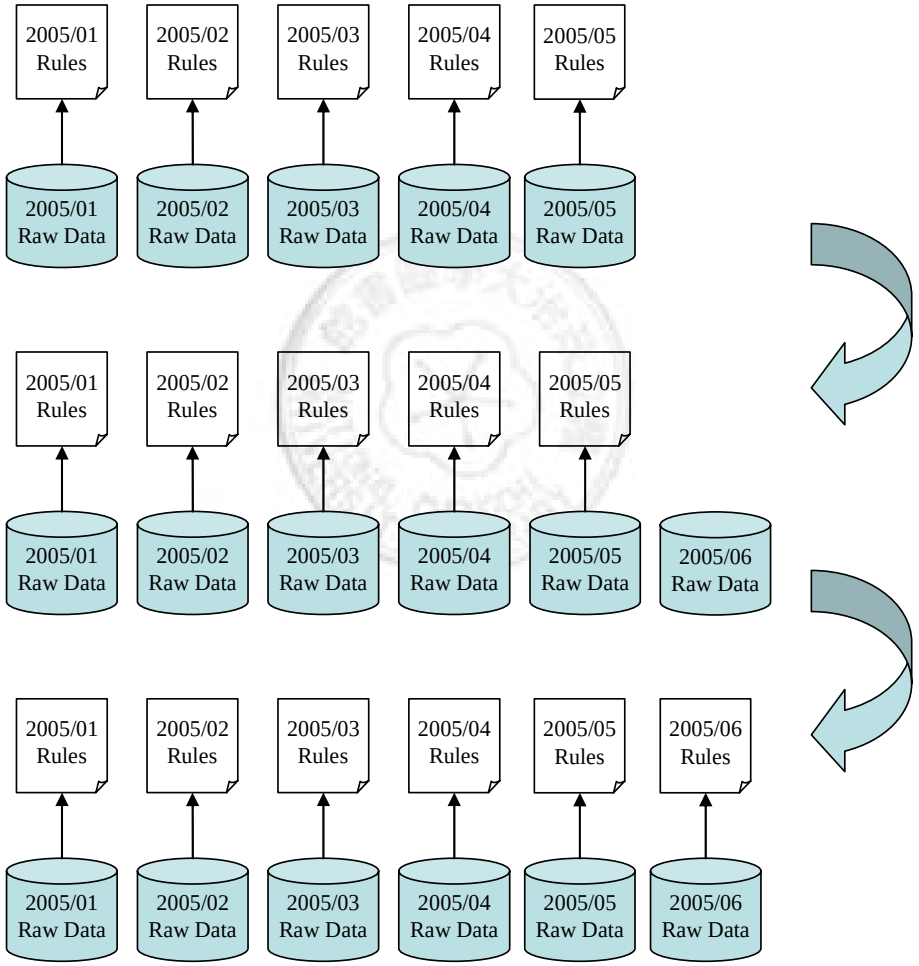

Figure 6.3 incremental mining

6.2. Future works

During the developing of our method, we found some interesting and valuable directions to expand our work:

1. Building the concept hierarchy trees still need background knowledge to determine the granularities of dimension atoms, and control the generalization process. We would like to involve appropriate exiting technologies such as cluster or segmentation tools in CRM to build the concept hierarchy tree in each dimension for different application domain.

2. Discovering large itemsets in each *element segmentation* is independence. We can use different algorithm such as DIC, FP-tree, DHP or other algorithms in different element segmentation, and almost need no synchronize between them. We use the rules in an element segmentation which is much cheaper for network transaction than the raw data to derive the rules in the combination segmentations. The two properties make the potential of our algorithm to fit the distributed even peer-to-peer environment.

3. We would like to design a user interface for representing the rules discovered friendly to help users make use of them.