

Implementation of the Apriori algorithm for association rule mining

¹HarvinderChauhan, ²AnuChauhan
^{1,2}Assistant Professor, P.G.Dept. of Computer Science
 Kamla Nehru College ForWomen, Phagwara (Punjab)

Abstract: With massive amounts of data continuously being collected and stored, many industries are becoming interested in mining association rules from their databases. The discovery of interesting association relationships among huge amounts of business transaction records can help in many business decision making processes. Association rule mining contains some set of algorithms, whenever we mine the rules we have to use the algorithms. Weka, a software tool for data mining tasks contains the famous algorithm known as Apriori algorithm for association rule mining which computes all rules that have a given minimum support and exceed a given confidence. In this paper we are implementing Apriori algorithm using “weather data set” from weka. This paper also gives insights into the association rules mined by this algorithm in the implementation section.

1. Introduction

Association rule mining contains some set of algorithms, whenever we mine the rules we have to use the algorithms. Weka contains the famous algorithm known as Apriori algorithm for association rule mining which searches for interesting relationships among items in a given dataset. This algorithm can compute all rules that have a given minimum support and exceed a given confidence. Confidence and support are two important measures used by this algorithm.

A certainty measure for association rules of the form “A=>B”, where A and B are sets of items is **confidence**. Given a set of task relevant data tuples the confidence of “A=>B” is defined as

$$\text{Confidence}(A \Rightarrow B) = \frac{\text{\#_tuples_containing_both_A_and_B}}{\text{\#_tuples_containing_A}}$$

The support of an association pattern refers to the percentage of task relevant data tuples for which the pattern is true. For the association rules of the form “A=>B” where A and B are sets of items defined as

$$\text{Support}(A \Rightarrow B) = \frac{\text{\#_tuples_containing_both_A_and_B}}{\text{Total_ \#_of_tuples}}$$

2. Implementation of the Apriori algorithm

In order to mine association rules for given data, first we need to convert numeric attributes to nominal because apriori algorithm works with categorical values only. This will be done in wekaexplorer window. Here, we have

loaded weather dataset having 14 instances and 5 attributes. On the basis of information contained in this dataset, weka enable us for searching interesting relationships among items in a given dataset.

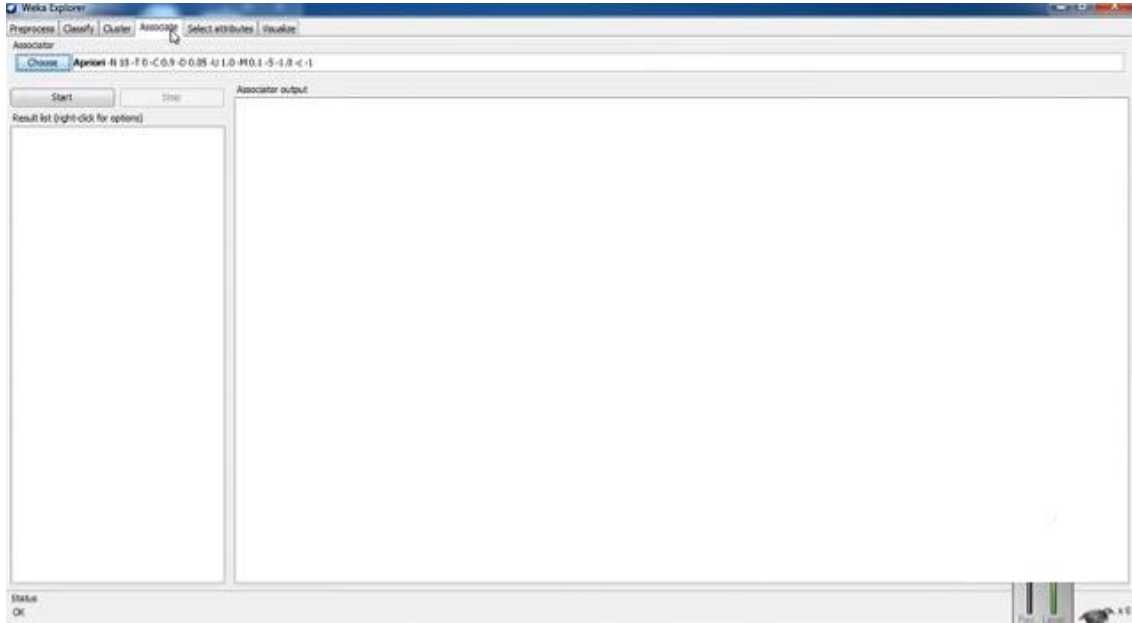


Figure1: Associate panel

In wekaexplorer, click on associate tab. Default algorithm for mining association rules is Apriori algorithm. It has some important values which are shown after double clicking on the parameter list of Apriori as shown in figure1. The window displayed looks like as given in figure2.

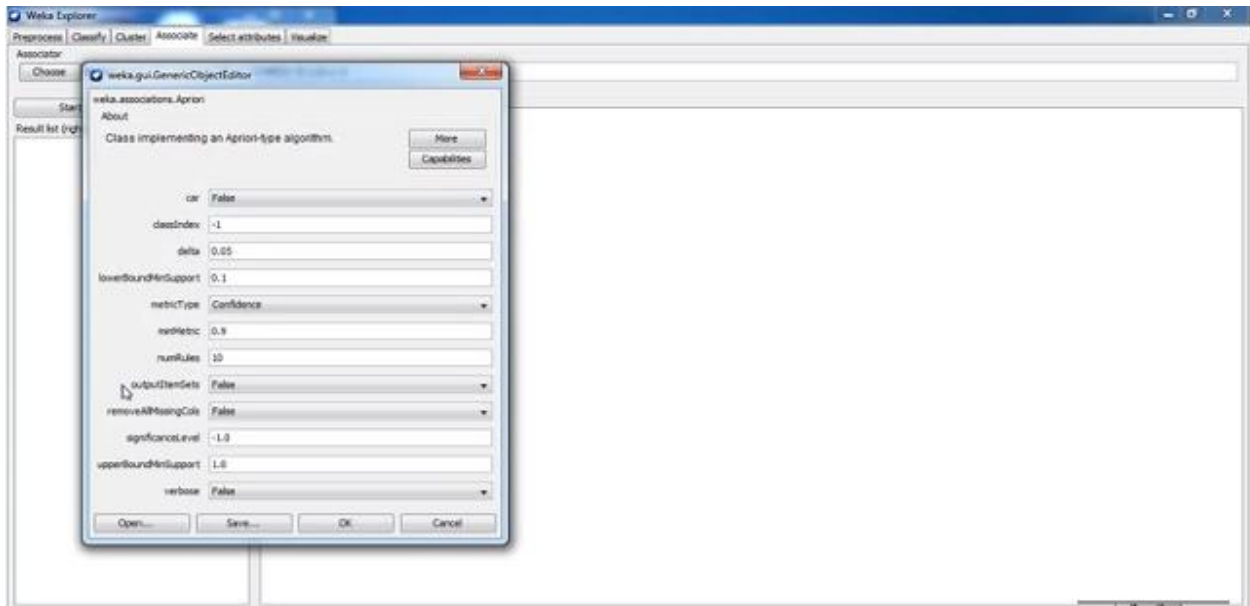


Figure2: Default values of Apriori algorithm

Click on ok button and again click on start button it will produce 10 best rules as given in the following figure



Figure3: Wekarun information for Apriori algorithm

Let us try to understand this output. Here the scheme is Apriorialgorithm. The relation is weather having 14 instances and 5 attributes. The minimum support value is 0.12 and value of confidence is 0.9. From the produced best rules let us try to understand the first rule which states

$$\text{Outlook=overcast } 4 \rightarrow \text{play=yes } 4 \text{ confi}(1)$$

At the left hand outlook is equal to overcast and number of instances are 4 and at right hand side play is equal to yes and number of instances are 4. So number of instances at left side by number of instances at right side is equal to 1 so its confidence is 1. Similarly the remaining rules are produced.

3. Conclusion

In this paper we have shown the step by step implementation of Apriori algorithm for association rule mining. Under the introduction section we have discussed two major measures for association rules which are used by Apriori algorithm. After that, with the help of appropriate snapshots, which we have taken from weka (software tool for data mining tasks) implementation of Apriori algorithm shown using “weather.nominal.arff” data set. Finally, ten association rules which are mined from this algorithm are discussed that helps in many business decision making processes like catalog design, cross marketing and loss leader analysis.

REFERENCES

- [1]. Jiawei Han and Micheline Kamber. Data mining concepts and techniques, second edition, 155-156, 225-230
- [2]. Alex Berson and Stephen J. Smith. Data warehousing, data mining & OLAP
- [3]. Weka tutorials
- [4]. Weka 3.6 data mining software in java