

An Efficient Tree based algorithm for Association Rule Mining

Akash Shrivastava
Computer Science Engineering
Lakshmi Narain College of Technology
Indore, India

Kuntal Barua
Professor, Computer Science Engineering
Lakshmi Narain College of Technology
Indore, India

ABSTRACT

Data mining is a process by which the from raw data information and important patterns are estimated. That involves the intermediate processes to find the data patterns from the input datasets. These processes are pre-processing, algorithm implementation and the testing of developed model. algorithm are used find the data pattern from the data, that may in form of associative, any data structure based or weight based. The proposed work is an effort in order to develop an association rule mining algorithm using the decision tree. Where the decision tree is used first to find the decision rules and according to rules new associate rules are extracted from the data.

General Terms

Apriori Algorithm, Association Rule Mining, Cart Algorithm

1. INTRODUCTION

Machine learning is domain of engineering where algorithms and computer based programs are learnt from data and past experience and provides decisions and analysis based on their experience and learned knowledge. Machine learning introduces stepwise process to learn and utilize the knowledge to solve the real world problems in effective and efficient manner. This process includes data pre-processing, training of algorithm and implementation using any application sometimes the last step is also known as testing of data model.

1.1 Data pre-processing:

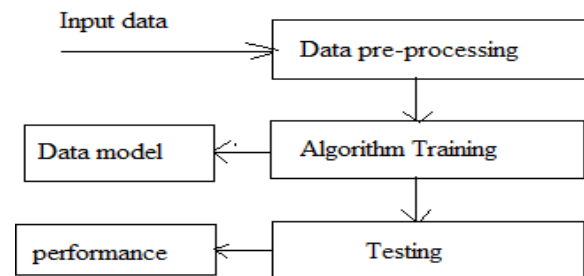
Data in real world is never found in a structured manner, which is hidden between unstructured data or between uneven data sources. in this phase data is separated from the its original format and organized in a desired format to read and utilized for future use.

1.2 Algorithm training:

Using the previous phase data machine prepare a data structure or data model by which decisions are made for providing solutions.

1.3 Testing:

Performance of the system can be measured using this phase of system, here manually or automatically real time or sample data is produces to test the intelligence of trained algorithm.



There are two different kinds of data models are available first transparent and second the opaque data models.

During the observation that is found that the transparent data models are suitable for navigation and data evaluation manually, on the other hand these data model's classification accuracy is poor with respect to the opaque data models. Therefore in this proposed work the accuracy in classification is improved using the opaque data model and transparent data models where for transparent data model decision tree algorithms are used and for the opaque data models Apriori algorithm is hybrid in the proposed work.

The description in the introduction leads to the following key observation:

When we used predictive modeling technique it produces highly accurate models. But we want to use some models simultaneously to increase accuracy as well as comprehensibility

Sometimes accuracy is not the only relevant criterion. Often, there is a need for comprehensible models. When this is the case, the easiest solution is to use a technique directly producing transparent models; most often decision trees. This will, however, normally lead to a loss of accuracy so we required a trade-off between the accuracy vs. comprehensibility. For achieving this trade-off, rule extraction is used, which perform the task of transforming opaque models with high accuracy to transparent models with retaining permissible level of accuracy.

Although several rule extraction techniques exist, the opinion is that no specific technique meets all criteria. Based on the observation we identified problem which is stated below.

Through this we can transform opaque model with accuracy to comprehensible. How to increase or optimized accuracy while using transparent models

How to extract accurate and comprehensible rules from opaque predictive models

How to optimize accuracy using Genetic algorithm.

How to achieve a trade-off between accuracy and comprehensibility.

2. OBJECTIVE

The proposed work includes the following work to accomplish during the proposed research work.

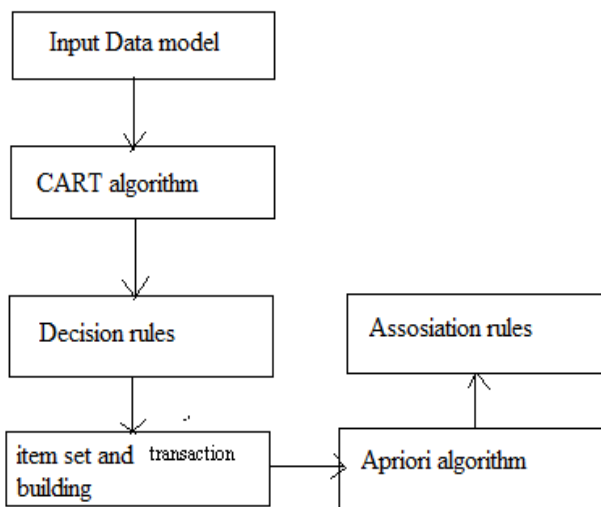
Study of various transparent and opaque models: in this phase various data models are studied in order to find the suitable combination of algorithm by which the proposed work becomes feasible.

Propose and design of the desired data model: in this phase for classification accuracy a new data model is proposed and implemented using the Apriori and CART decision tree.

Performance assessment: in this part the performance of the implemented model is estimated using precision and recall method for justification of the proposed data model

3. SOLUTION DOMAIN

To achieve the desired goal of the proposed data model a hybrid approach is developed for finding the association rules for improving the classification accuracy. The proposed data model is based on the combination of Apriori and CART decision tree. The below given diagram can explain the proposed working model



Using the above given diagram the proposed architecture of the proposed data model can be understand. Input data is first processed using the CART algorithm that provides the decision rules for the dataset. Here the size of data is reduced in terms of instances. In the next the participating attributes are evaluated to design rules, here rules are used as transection for the Apriori algorithm. Using this association rule mining is prepared.

4. EXPECTED OUTCOMES

The proposed work is improvement over the traditional classification and association rule mining algorithm which uses the transparent data model as filter and using the opaque data model the new rules are generated in order to reduce the cost of navigation and improving the classification accuracy

5. REFERENCES

- [1] P. Rajendran, M.Madheswaran,Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm, JOURNAL OF COMPUTING, VOLUME 2, ISSUE 1, JANUARY 2010, ISSN 2151-9617
- [2] A.Martin, M.Manjula and Dr.PrasannaVenkatesan,A Business Intelligence Model to Predict Bankruptcy using Financial Domain Ontology with Association Rule Mining Algorithm,IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 3, No. 2, May 2011 ISSN (Online): 1694-0814
- [3] Murat Kantarcioglu, Robert Nix, and Jaideep Vaidya, An Efficient Approximate for Privacy-Preserving Association Rule Mining,PAKDD 2009, LNAI 5476, pp. 515–524, 2009.c
- [4] MetanatHooshSada,Hamman W. Samuel,Sonal Patel,Osmar R. Zaiane,Fastest Association Rule Mining Algorithm Predictor(FARM-AP),C3S2E-11 2011, May 16-18, Montreal [QC, CANADA]Editors: Abran, Desai, MudurCopyright 2011 ACM 978-1-4503-0626-3/11/05
- [5] Ruihai Dong, Markus Schaal, Michael P. O'Mahony, Kevin McCarthy, and Barry Smyth,THE REVIEWER'S ASSISTANT: RECOMMENDING TOPICS TO WRITERS BYASSOCIATION RULE MINING AND CASE-BASE REASONING, Downloaded 2013-12-12T03:37:02Z