Rules Extraction through Decision Tree Induction Using Information Theory

Myat Mon Khaing
University of Computer Studies, Hinthada
myatmonkhaing01.htd@gmail.com

Thet Thet Aung
University of Computer Studies, Hinthada
thetthetaung86.htd@gmail.com

Hlaing Htake Khaung Tin
University of Computer Studies, Hinthada
hlainghtakekhaungtin@gmail.com

Abstract

Today, data mining is the powerful technology and Decisions, selections in medicine areas and operations are the most essential part of identifying medicaments of a patient. Determination of a delivery type is important because the life of mother and fetus depended to this choice. There are two ways for babies can enter world. Attribute selection measure and decision tree classifiers are famous techniques of classification in knowledge discovery technique. This paper extracts rules on concept learning or training system in dataset. The objective of this paper is to predict categorical class label for the pregnant woman. 20 pregnant women are surveyed as dataset. Depending upon their six attributes, feature selection and the class labeled training tuples are measured using entropy of decision tree induction. The highest information gain among the attributes is the root of tree node, the rules are extracted and finally determine pregnant women are needed Caesarean or not by using ID3 decision tree algorithm.

Keywords*: Information Gain, Attribute selection measure, Decision tree induction, ID3 algorithm, Extraction rule*

1. Introduction

When pregnant women give birth their children with two ways. These two ways are a vaginal birth (normal birth) and surgical delivery by Caesarean section (C-Section). The aim of two birth technique is to safely give birth a well-baby. Giving birth of pregnant woman is considered with their medical conditions (such as mother's blood pressure, heart disease, HIV, other conditions such as sending a very big baby in a mother with a small pelvis, if the baby is not a heads-down position and the baby cannot be afforded to turn normal position before birth), she should be take vaginal birth or C-Section birth. Because some giving birth with vaginal birth have too risky.

Data mining built on classification and forecast methods active of handling large amount of disk-resident data. Mining is discovering for relationships between attributes (features) in class label records, selecting measurable outcomes concerning feature values. Classification used a set of pre-classified data to extract models describing the training data classes. Therefore in this paper work, a data mining system, inducing a decision tree algorithm from training attribute for classification of data, was deployed to mine kinds of attribute values in target dataset and the information gain (entropy) was used to access the expected resources needed to clustering a tuple. Classification design is an essential role of data analysis in data mining field. Evaluation involves splitting and processing of expected information and decision making.

In this paper, research works are described with many sessions. In session 3, the proposed system is described with figure. In session 4, as the data classification, learning step and classification step are briefly described with two figures. In session 4.1, measures attribute selection using ID3 algorithm and pseudo code. In session 4.2, calculates the information gain using information theory. In session 5, real 20 pregnant women records with attributes: Age, Delivery No, Delivery Time, Blood of Pressure, Heart Problem and Caesarean are preprocessed and in caesarean section, four categories 1: Immediate threat to the life of the woman or foetus, 2: Maternal or fetal compromise that is not immediately life threating, 3: No maternal or fetal compromise but early delivery required and 4: Delivery timed to suit woman and staff (elective). In Measuring Attribute Selection Using ID3, the process of retrieving extracting rule information from them involves procedures that analysis the existence of feature interactions and identify relevant features.

This paper works feature with target can be correlated together with other features. The rules are generated from decision tree and ID3 uses information gain (entropy) in this paper.

2. Proposed System

This paper intends to implement the rules extraction from decision tree for case of accouchement. This paper can help determination of delivery type. Before applying the data mining techniques on the dataset, there should be an activity that made this task. The following of Fig.1 depicts the task activity used.

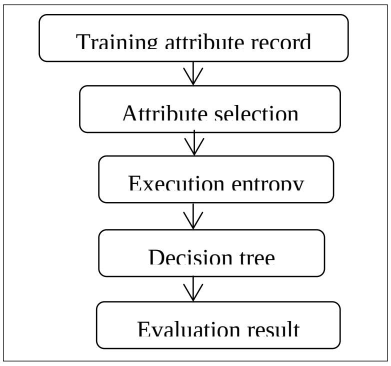

Figure 1. Data Mining task Methodology

3. Methodology

Data classification is the two steps process of finding a model:

Learning step1-Training data are analyzed by classification algorithm. The class studied model (classifier) is built in the form of clustering rules as shown in figure 2.

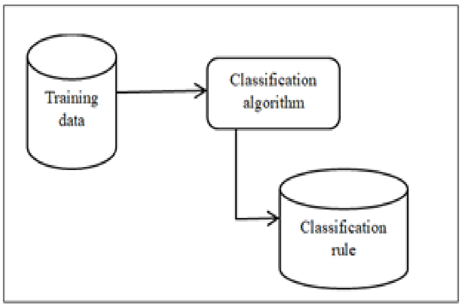

Figure 2. Learning step

Classification step2 -Test data are used to measure the accuracy of classification rules. If the accuracy of the classification is acceptable, the rules can be applied to the classification of new data attributes as shown in figure 3.

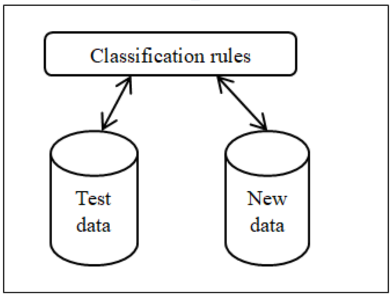

Figure 3. Classification step

3.1. Measuring Attribute Selection Using ID3

This paper is attempts to extract useful rule from different class label of an attribute in partition. Here induction of decision tree is used to compute the highest information gain among attributes values. By the algorithm, it is created for the attribute in training set to extract knowledge that describes the root node of the decision tree. ID3 apply Information gain as its attribute selection measure. In present day's knowledge mining method, a distinguish data classes of finding model in branch outcome of the test is predicted by classifier. The major task of using data mining technique is to discover knowledge from the existing data which is available in the training dataset. The major objective of using data mining technique is to discover knowledge from the existing data class which is available in the training data.

Decision trees are structured in a top-down recursive manner. For learning algorithm are adopted to build model such as ID3. This measure include three strategies: 1) The expected information needed to classify a tuple in partition D , entropy of D 2) The expected information required to classify a attribute from D based on the partitioning by attribute A , 3) The splitting attribute at node N. The best classification is to identify tuples based on data. To calculate the required information for each attribute.

Input:

- D, Data partition,
- attribute-list,
- Attribute selection method,

Output: A decision Tree

Decision Tree Algorithm with Pseudo-code:

```
Begin
Declarations
        string TABLE_HEADING=" Class _labeled  Training data with objects"
        string COLUMN_HEADING=  attribute list
        num classifier value
        num partition
        string QUIT= leafnode
    getReady( )
    while partition < >QUIT
        detailLoop( )
    endwhile
    finishUP( )
End
getReady( )
  output "TABLE HEADING"
  output "COLUMN_HEADING"
  input a list of binary strings
return
detailLoop( )
  create a node N;
  If tuples in D are all of the same class, C, then
    return N as a leaf node labeled with the
    class C;
  if attribute list is empty then
    return N as a leaf node labeled with the
    majority class in D; // majority voting
    apply Attribute_selection_method (D,
    attribute_ list) to find the "best" splitting
    criterion;
    label node N with splitting criterion;
    if splitting attribute is discrete-valued and
      multiway splits allowed then // not
      restricted to binary trees
      attribute list □ splitting
      attribute; // remove splitting attribute
for each outcome j of splitting criterion
          // partition the tuples and grow subtrees   for eachpartition
          let Dj be the set of data tuples in D
          satisfying outcome j; // a partition
          if Dj is empty then
        attach a leaf labeled with the majority
        class in D to node N;
        else attach the node returned by Generate
        decision tree(Dj , attribute list) to node N;
    endfor
    return N;
  finishUp( )
    output N
    output "End of Program"
  return ;
```

In this algorithm, if the attribute values in data partition D is belong to the same class then node N is a terminal and is labeled with that class from first IF statement. Second IF statement is terminating conditions. This implementation system is for classification which includes the training algorithm to construct a model that is subsequently used to identify the feature of an unknown data. This implementation system is for classification which includes the training algorithm to apply a model that is subsequently used to identify the feature of an unknown data Decision tree are supervised classification algorithm.

3.2. Calculating the Information Gain (IG)

Information theory has quantified entropy that key measure of information. Information theory measure information in bits entropy:

$$info\ (D) = - p_1 \log_2(p_1) - p_2 \log_2(p_2) - \ldots\ldots - p_n \log_2(p_n) \quad (1)$$

where P_i is the probability that an arbitrary tuple in D belongs to class C_i and is estimated by $|C_{i,D}| / |D|$. $Info(D)$ is the average amount of information needed to identify the class label of a tuple in D.ID3 uses information entropy. The partition D on some attribute A having v distinct values, $\{a_1, a_2, a_3, \ldots, a_v\}$ from training dataset. Attribute value A can be used to split D into v partitions, $\{D_1, D_2, D_3, \ldots, D_v\}$, where D_j contains outcome a; of A

$$Info_A(D) = \sum_{j=1}^{v} \frac{|Dj|}{|D|} \times Info(D_j) \quad (2)$$

$|D_j/D|$ acts the weight of the j[th] partition. $Info_A(D)$ is the expected information required to classify partitioning by A. The expected information required, the greater the purity of the partition.

$$Gain\ (A) = Info\ (D) - Info_A(D) \quad (3)$$

Information gain is defined as the difference between the original information requirement and the new requirement. *Gain (A)* is the expected reduction in entropy by value of A.

4. Result and Discussion

These datasets are used to compute with feature selection measure on data. In this step, data stored in different attribute value such as Age, Delivery No, Delivery Time, and Blood of Pressure, Heart Problem and Class field Caesarian.

Table1. Attribute Description

No	Age	Delivery No	Delivery Time	BloodofPressure	Heart Problem	Caesarian
1	22	1	Timely	High	apt	No
2	26	2	Timely	Normal	apt	Yes
3	26	2	Premature	Normal	apt	No
4	28	1	Timely	High	apt	No
5	22	2	Timely	Normal	apt	Yes
6	26	1	Premature	Low	apt	No
7	27	2	Timely	Normal	apt	No
8	32	3	Timely	Normal	apt	Yes
9	28	2	Timely	Normal	apt	No
10	27	1	Premature	Normal	apt	Yes
11	36	1	Timely	Normal	apt	No
12	33	1	Premature	Low	apt	Yes
13	23	1	Premature	Normal	apt	No
14	20	1	Timely	Normal	inept	No
15	29	1	Latecomer	Low	inept	Yes
16	25	1	Latecomer	Low	apt	No
17	25	1	Timely	Normal	apt	No
18	20	1	Latecomer	High	apt	Yes
19	37	3	Timely	Normal	inept	Yes
20	24	1	Latecomer	Low	inept	Yes

The feature tests are chosen one at a time in behavior, they are dependent on results of previous tests.

Table2. Categories of emergency for caesarean section

Category	Criterion
1	Immediate threat to the life of the woman or foetus
2	Maternal or fetal compromise that is not immediately life threating
3	No maternal or fetal compromise but early delivery required
4	Delivery timed to suit woman and staff(elective)

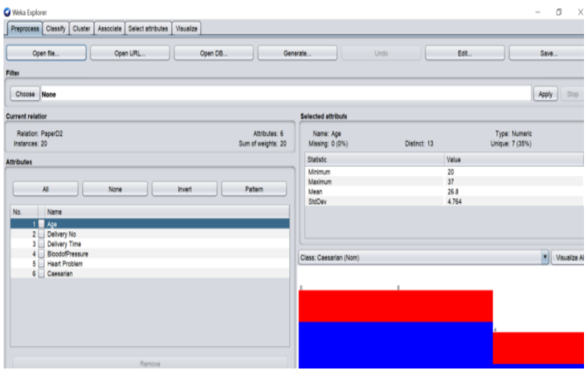

Figure 4. Dataset model using J48 on Weka

Models built using J48.10 for each set of redundant data were obtained with accuracy. The accuracy percentage is rounded to the nearest integer. In Fig.5, accuracy is the percentage of correct scores predicted by the model using a 10-fold cross validation method.

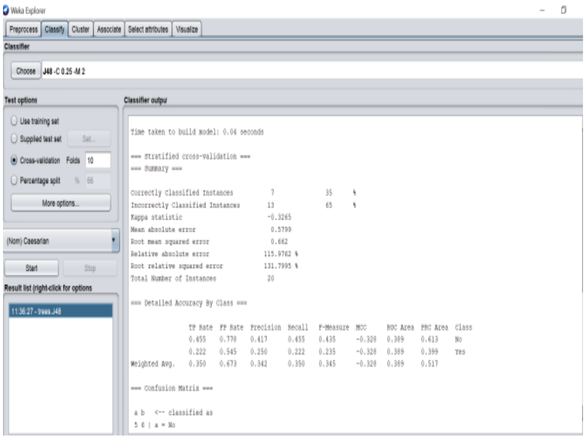

Figure 5. Accuracy obtained with model

IG attribute evaluation is computes the feature value by measuring the IG with respect to the associate class. Evaluation of issues ranking is a classification task. To create a tree, it needs to have a root node first and know that nodes are features. The data analysis is learned and constructed to predict classifier all aspects of given training datasets. In the following tree diagram, tree where each node represents a feature (attribute), each link (branch) represents a decision (rule) and leaf represents an outcome (categorical value). Extracting rules from tree has three main advantages:

1. rule is easier for understand
2. each different branch through the decision tree node set up a distinct rule
3. topmost node is decided not only on entropy value of attribute but also leaf is created with majority class in partition

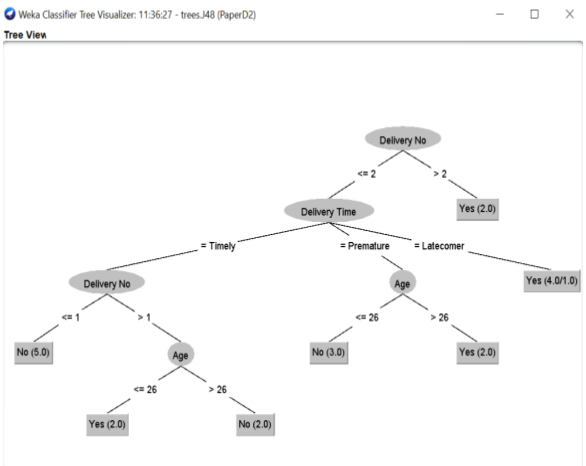

Figure 6. The outcome tree by ID3 in Weka

The resulting classification is expressed as a tree structure. Such a decision tree can help easy to understand and a design is more effective in data representation. Tree root that places high information gain attributes close to the parent root are selected over those above data. Extracting rules defined by the decision; the tree of Fig.6 identifies IF-THEN rules by tracing the branch to a terminal node in the tree.

The rules extracted are as follows:

R1: IF DeliveryNo <=2 AND Delivery Time = Timely AND DeliveryNo <=1
THEN Caesarean= Yes

R2: IF DeliveryNo <=2 AND DeliveryTime = Timely AND DeliveryNo >1
AND Age <= 26 THEN Caesarean= Yes

R3: IF DeliveryNo <=2 AND DeliveryTime = Timely AND DeliveryNo >1
AND Age > 26 THEN Caesarean= No

R4: IF DeliveryNo <=2 AND DeliveryTime = Premature AND Age <= 26
THEN Caesarean= No

R5: IF DeliveryNo <=2 AND DeliveryTime = Premature AND Age >26
THEN Caesarian= Yes

R6: IF DeliveryNo <=2 AND DeliveryTime = Latecomer THEN Caesarean= Yes

R7: IF DeliveryNo >2 THEN Caesarean= Yes

The rules are extracted directly from the tree. Those seven rules provide a ranking for each attribute describing the given training tuple. The attribute values having the score for the depending on the measure score is chosen the best. Thus feature selection provide for selecting the "best" separates a given data partitioned, Splitting rules that determine how the tuples at a given node are to be split.

5. Conclusion

This paper extracts rules from DTs. Decision tree provides certain decision rules. These rules can be specified the best classification from the outcomes of highest information gain. These methods are providing not only for ranking attribute of training tuples but also for splitting attribute score measures in this paper. This paper reflects the tree induction that behavior of feature on the training dataset. Extraction rule has an essential task to apply mining algorithm effectively in real world area. The life of mother and fetus depend on their choice and delivery type. The health conditions are the attributes. If the pregnant women have heart problem, blood pressure and delivery time, this paper determines Caesarean case.

References

[1] J. Han , M .Kamber, J.Pei. *"Data MiningConcept and Techniques*", 3rd ed. , Haryana, 2012.

[2] Gordson S . Linoff , Michael J. A. Berry. "*Data Minig Techniques*", 2008.

[3] Santhosh Kumar, Tanjavur ,Tamil Nadu, Dr. E. Ramaraj, "Modified C4.5 Algorithm With Improved Information Entropy and Gain Ratio ", *International Journal of Engineering Research & Technology,* September -2013.

[4] Farhad Soleimanian Gharehchopogh, Peyman Mohammadi, Parvin Hakimi, "Application of Decision Tree Algorithm for Data Mining in Healthcare Operations: A Case Study", *International Journal of Computer Applications,* August – 2012.

[5] B. Azhagusundari, Antony Selvadoss Thanamani, " Feature Selection based on Information Gain", *IJITEEI* , January -2013.

[6] Bangsheng Sui, " Information Gain Feature Selection Based On Feature Interactions" , Dec – 2013.

[7] Tiruveedula Gopi Krishna, "Evaluation Of Rule Extraction Algorithms", *IJDKP,* Vol.4, No.3, May 2014.

[8] Sabita Barik, Debahuti Mishra, Shruti Mishra, Sandeep Ku. Satapathy, Amiya Ku. Rath and Milu Acharya, " Pattern Discovery using Fuzzy FP - growth Algorithm from Gene Expression Data," *IJACSA*, Vol. 1, No. 5, November 2010.