

Implementing Myanmar Traditional Astrology Using Classification Rules with Decision Tree Algorithm

Thin Thin San
Computer University(Monywa)
thinthinsn52@gmail.com

Abstract

Data Mining is a step in the knowledge discovery process consisting of particular data mining algorithms that, under some acceptable computational efficiency limitations, find patterns or models in data. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. A Tree Classification algorithm is used to compute a decision tree. Decision trees are easy to understand and modify, and the model developed can be expressed as a set of decision rules. This system is intended to implement Myanmar Traditional Astrology by using decision tree algorithm. In this system, training data of astrology are classified by decision trees algorithm and IF-THEN rules. Among Astrologies, five classes of them is represented in this paper. Depending upon the input data, the birthdate, of astrology, the system can classify the type of the astrology that is Aries or Taurus or Gemini and so on.

1. Introduction

Data mining refers to extracting or “mining” knowledge from the large amount of data. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. A classification algorithm takes a dataset as an input and maps each data point in the dataset into predefined classes. A decision tree is usually used to facilitate the mapping. Decision tree Induction is an important technique of classification and have been used for classification in a wide range of application domains. A decision tree is usually used to facilitate the mapping. The main objective of a classification algorithm is to come up with a decision tree to classify the data of the astrology.

2. Classification

Classification is an important technique in data mining. Classification is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown. It can be used for predicting the class label of data objects. Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. Many classification methods have been proposed by researches and are important area of research and of practical by application in a variety of fields, including pattern recognition and artificial intelligence, statistics, vision analysis, medicine and so on.

Classification is the process of finding as a set of models that describe and distinguish data classes or concepts. While classification predicts categorical labels, prediction models continuous-valued functions. Prediction can be viewed as the construction and use of a model to assess the class of an unlabeled sample, or to assess the value or value ranges of an attribute. Data classification is a two-step process. In the first step (model construction), a model is built describing a predetermined of data classes or concepts. The model is constructed by analyzing database tuples described by attributes. Each tuple is assumed to belong to a predefined class, as determined by one of attribute, called the class label attribute. Typically, the model is represented in the form of classification rules, decision trees or mathematical formulae. In the second step (model usage), the model is used for classification[2,28].

3. Decision Tree

Decision tree learning, used in data mining and machine learning, uses a decision trees as a

predictive model which maps observations about an item to conclusion about the item's target value. Decision tree learning, used in data mining and machine learning, uses a decision trees as a predictive model which maps observations about an item to conclusion about the item's target value. More descriptive names for such tree models are classification trees or regression trees. In these tree structures, leaves represent classification and branches represent conjunctions of features that lead to those classifications[9].

In these tree structures, leaves represent classification and branches represent conjunctions of features that lead to those classifications. A decision tree is a flow-chart-like tree structure, where each internal node denotes a test, and leaf node represent classes or class distribution. The top-most node in a tree is the root node. Internal nodes are denoted by rectangle, and leaf nodes are denoted by ovals. Of course, accurate Decision Tree is fundamental to data mining Databases. Decision Tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented decision tree. Decision Tree learning is one of the most widely used and practical methods for inductive inference.

Decision tree generation consists of two phases: tree construction and tree pruning. In order to classify an unknown sample is tested against the decision tree. A path is traced from the root to a leaf node that holds the class prediction for that sample. Decision trees can easily be converted to classification rules. When decision trees are built, many of the branches may reflect noise or outliers in the training data[6].

Decision trees are a special tree structure constructed to help with making decisions. In decision theory and decision analysis, a decision tree is a graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It can be used to create a plan to reach a goal. Another use of trees is as a descriptive means of calculating conditional probabilities. A decision tree is a type of multistage classifier. It is made up of a series of binary decisions that are used to determine the correct category for each class. The decisions can be based on any available characteristic of the dataset. No single decision in the tree performs the complete segmentation of the image into classes. Instead, each decision divides the data into one of two possible classes or groups of classes[10].

4. Classification the Instances of Data with Decision Tree

Instead of representing decisions, leaf nodes represent a particular classification of a data instance, based of the given set of attributes (and their discrete values) that define the instance of data, which is kind of like a relational tuple for illustration purpose. Because Decision Tree can classify data instances into different types, they can be “interpreted” as a “good generalization of unobserved instances” of data, appealing to people because “it makes the classification process self-evident”. They represent knowledge about these data instances.

5. Classification IF-THEN Rules

In rule-based systems, knowledge will be extracted in the form of IF..THEN rules (Ross, 1983; Thompson *et al.*,1986). The efficiency of the expert system lies in generating a minimum number of rules (Thompson *et al.*, 1986). To generate a minimum number of rules, the decision tree algorithm based on the probabilistic concepts of entropy is normally employed (Thompson *et al.*,1986). The decision trees algorithm are the starting points for generating the IF-THEN rules.

The objective of a rule-based system is to develop a classification tree or a decision tree consisting of nodes and branches, and this is used for classifying the observations.

From the decision tree, a set of IF..THEN rules were derived by following each branch from the root to a terminal node. Each rule is a series of conditions consisting of attribute and value pairs, followed by a single conclusion that contains the class and the corresponding class value (Thompson *et al.*, 1986). The intermediate nodes and their branches form the conditions of the rules, the terminal nodes form the rules conclusions. By converting these IF..THEN rules into any high level programming language such as FORTRAN, BASIC, PASCAL, etc., it is possible to develop the so-called rule-based system for carrying out classification analysis. The IF..THEN rules so derived from the algorithm were later translated into QUICK BASIC language statements[11].

6. Constructing Decision Tree

Decision tree induction is a typical inductive approach to learn knowledge on classification. The key requirements to do mining with decision trees are:

1. Attribute-value Description: object or case must be expressible in terms of a fixed collection of properties or attributes. This means that the

continuous attributes are needed to discretize, or this must have been provided in the algorithm.

2. Predefined Classes (target attribute values): The categories to which examples are to be assigned must have been established beforehand (supervised data).
3. Discrete Classes: A case does or does not belong to a particular class, and there must be more cases than classes.
4. Sufficient Data: Usually hundreds or even thousands of training cases.

Most algorithms that have been developed for learning decision trees are variations on a core algorithm that employs a top-down, greedy search through the space of possible decision trees. Decision tree programs construct a decision tree from a set of training cases [1].

7. Astrology

Astrology is a group of systems, traditions, and belief which hold that the relative positions of celestial bodies and related details can provide information about personality, human affairs, and other terrestrial matters. So, Astrology is a complex study. It has ancient origins and is still evolving and developing. It is only one of many tools that can be used to understand life and our existence on Earth.

Astrology is capable of giving much deeper insight but to do this requirement a complete set of birth data – the day, month, year and time of birth together with place of birth on Earth (latitude and longitude)[3]. There are many methods in Astrology to foretell someone's fortune in Myanmar. Among them, the methods of astrology which state the day, month and year is used in this thesis. If the user want to know his/her birthdate being the day, month and year, this system can calculate to get the result by using decision tree algorithm, ID3.

7.1 Types of Astrology

There are many methods in Astrology to foretell someone's fortune in Myanmar. Among them, the methods of astrology which state the day, month and year is used in this paper [7]. There is a number of methods that an astrologer can use to make predictions. The existing methods are Mahabote, Thetyauk, Remainder Prophecy, Horoscope, Zodiac, etc. In Astrology, there are seven methods in Mahabote, eight ways in Thetyauk, twelve means in Remainder Prophecy, four approaches in Horoscope and twelve signs of Zodiac.

8. Extracting Classification Rules From Tree

The knowledge represented in decision trees can be extracted and represented in the form of classification IF-THEN rules. One rule is created for each path from the root to a leaf node. Each attribute-value pair along a given path forms a conjunction in the rule antecedent ("IF" part). The leaf node holds the class prediction, forming the rule consequent ("THEN" part). The IF-THEN rules may be easier for humans to understand, particular if the given tree is very large [2,290].

Rules are easier to understand for human.

In the horoscope (zar tar) of astrology, IF- THEN rules are used as follows:

RULES (1). IF (Remainder 1 == 0) AND (Remainder2 ==0) THEN class = Monk

RULES (2). IF (Remainder 1 == 1) AND (Remainder2 ==6) THEN class = The rich

RULES (3). IF (Remainder 1 == 2) AND (Remainder2 ==5) THEN class = King

RULES (4). IF (Remainder 1 == 3) AND (Remainder2 == 4) THEN class = Court Brahmin.

A rule can be "pruned" by removing any condition in its antecedent that does not improve the estimated accuracy of the rule. For each class, rules within a class may then be ranked according to their estimated accuracy. Since it is possible that a given test sample will not satisfy any rule antecedent, a default rule assigning the majority class is typically added to the resulting rule set.

9. Implementation

This system contains five classes of the astrology that they are Mahabote, Thetyauk, Remainder, Zodiac and Horoscope. Then there are forty three subclasses of the astrology in this system. The training data that are the day, month and year with the prophecy result can classified by using classification rules . There are 49 rules in Mahabote class, 64 rules in Thetyauk, 12 rules in Remainder, 365 rules in Zodiac and 4 rules in Horoscope.

Every body wants to be happy, healthy, lucky and successful in his life. He also wants to know his fortune in the future which will be good or bad. So, he usually goes to an astrologer and asks him about

his future. In this system, the aim is to know easily the user's prophecy of the astrology in a short time.

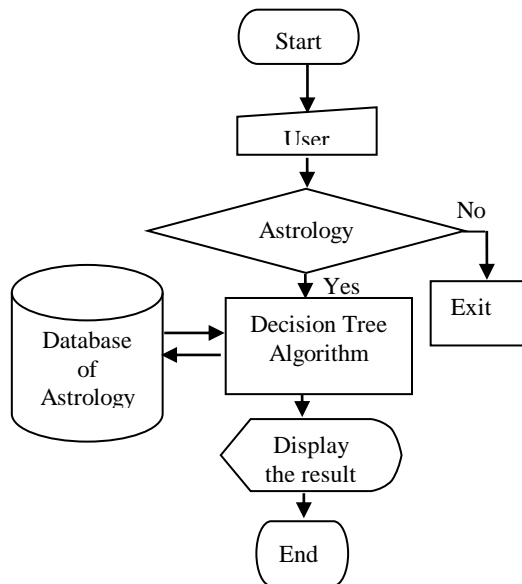


Figure 1. System Flow Diagram

In figure 1. the user chooses his/her birthdate to know his/her result. So this system is calculated by ID3 process that classified the training data and testing data. The Prophecy of astrology is extracted from database of astrology.

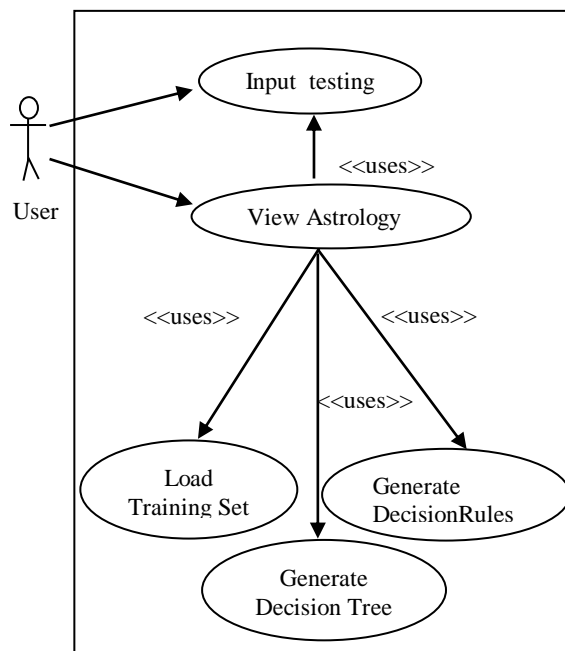


Figure 2. Use Case Diagram of the System

In figure 2, the user requests to view astrology by adding the user's birthdates (testing set). So the load training set, generate decision tree and generate decision tree are used to display the View Astrology.

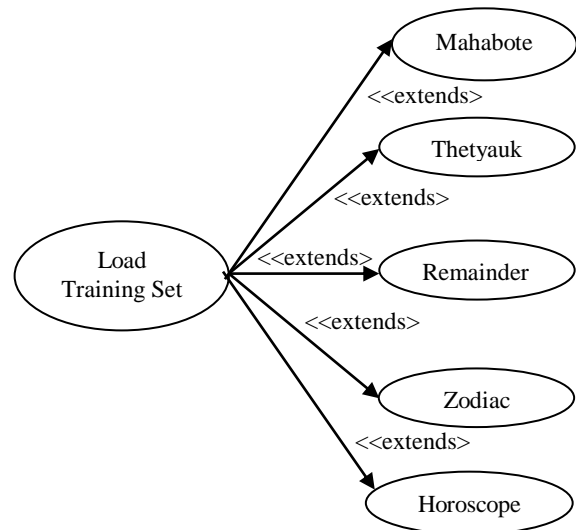


Figure 3. Use Case Diagram of Training Set

In figure 3, the load training set extends the Mahabote System, Thetyauk System, Remainder System, Zodiac System and Horoscope System. They are trained with the data from the astrology database to display the Prophecy of astrology.

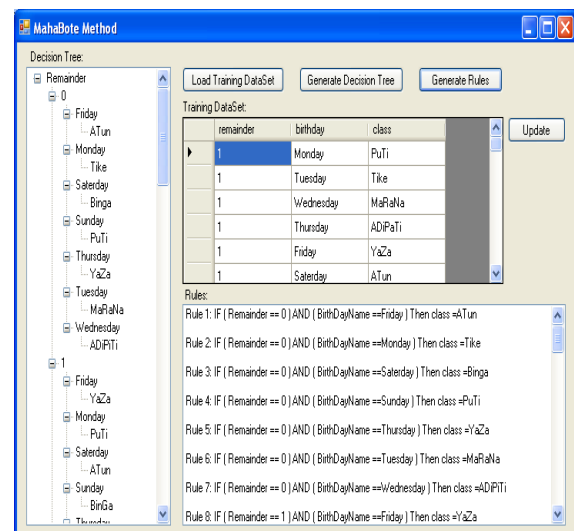


Figure 4. Mahabote Method Form

In figure 4, the user can do the necessary changes to update the training Data by using "Training Dataset Menu Item" for Mahabote System. And then if the user wants to know how to construct the ID3 decision tree by classifying the astrology data, click "Generate Decision Tree" button. Then user can know the decision rules concerning with the attribute data by choosing "Generate Rules" button.

10. Conclusion

Classification is a form of data analysis that can be used to extract models describing important data classes or to predict future data trends. The Decision Tree Classification algorithm provides an easy-to-understand description of the underlying distribution of the data. Classification tree (also known as decision tree) methods are a good choice when the data mining task is classification or prediction of outcomes. Classification tree labels records and assigns them to discrete classes.

This system has presented generating of classification from training datasets of the astrology. This approach demonstrates efficiency and effectiveness in dealing with the astrology data for classification. The aim of this system is intended to present the combination of science (e.g information technology) and traditional astrology and know easily the user's prophecy of the astrology in a short time.

10. References

- [1] A. Bahety, "Extension and Evaluation of ID3 – Decision Tree Algorithm
- [2] J.Han, M. Kamber, "Data Mining: Concepts and Techniques" Morgan Kaufmann, 2001.
- [3] D. L Easks, S. Jarrett, P. Cert (FAA), "An Introduction to Astrology", published by the Victorian Astrologers' Association, 1998
- [4] T. Mitchell, "Machine Learning", ISBN: 978-0-07-042807-2, 1997.
- [5] A. Neymark, "ID3 Algorithm", CS157B – Spring 2007
- [6] P. Ning, T. Michael Steinbach and V. Kumar, "Introduction to Data Mining".
- [7] S. Sein Sann, "Collective Treatise of an Astrologer's Advice", fourth edition, January, 1996
- [8] M. Than, "Myanmar Traditional Collective Treatise of the Humanities and Science", third edition, July 25, 1992
- [9] http://www.cs.uregina.ca/~dbd/cs831/notes/m1/dtrees/4_dtrees.html
- [10] http://en.wikipedia.org/wiki/Decision_tree_learning
- [11] <http://bioinformattics.oxfordjournals.org/cgi/print/12/6/491.pdf>