Agent based Enterological Disease Diagnosis System

Kyaw Khaing Oo
University of Computer Studies, Yangon
kyawkhaingoo05@gmail.com

Abstract

Decision tree learning algorithm is rules for classifying data using attributes and has been successfully used in decision support systems in capturing knowledge. The main task performed in this paper is using inductive methods to the given values of attributes of object to determine appropriate classification according to decision tree rules. Decision tree is mainly used for classification purposes. Decision tree is a classifier in the form of a tree structure. Rules can be easily extracted from the decision tree. This system acquires knowledge from the domain expert who has the special knowledge. System receives symptoms from the users and decides related disease. During these operations, this system uses agents capabilities. Agent can do different services: classification and testing. This system is used the symptoms to classify the disease with the help of ID3 algorithm. It will provide the accurate result to patients by combining the Lab results.

Keywords: Enterological Disease, Expert System, Decision Tree Induction, ID3

1. Introduction

An expert system [1] is software that attempts to provide an answer to a problem, or clarify uncertainties where normally one or more human experts would need to be consulted. Expert systems are the most common in a specific problem domain, and is a traditional application and /or subfield of artificial intelligence. A decision tree (or tree diagram) is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal. Another use of decision trees is as a descriptive means for calculating conditional probabilities. Decision tree[2] are commonly used for gaining information for the purpose of decision-making. The input data of ID3 is known as sets of "training" or "learning" data instances, which will be used by the algorithm to generate the Decision Tree. ID3 algorithm [3] includes Classification Models, also called Decision Trees, from data. Each record has the same structure, consisting of a number of attribute /value pairs. Classification rules represent the classification knowledge as IF-THEN rules and are easier to understand for human users. There are two steps to making productive use of decision trees (1) building a decision tree model, and (2) using the decision tree to draw inferences and make predictions. A number of algorithms for induction decision trees have been proposed over the year ID.3, C4.5, etc.

This paper examines the decision tree learning algorithm ID3.The main task performed in this system is using decision tree induction methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules. Test attributes are selected on the basis of Information gain measure. Such a measure is referred to as an attribute selection measure or a measure of the goodness of split.

2. Related Work

A.Chidanand and W.T.J.Sholom [2] have presented is the code of the Data mining system with from the decision tree and decision rules. G.Minos, H.Dongjoon, R.Rajeev and S.Kyuseok [6] highlighted efficient algorithm for constrating Decision Trees with Constratints.W.Peng, J.Chen implemented of ID3 [4]decision tree learning algorithm .In their approach, decision tree based gaining information for calculating probablites. For example, rules are extracted from the knowledge of experts in the expert system. Khine Moe Nwe (M.C.Sc (Thesis) , 1998, UCSY)[7] have presented the exact diagnosis in sector of fever related

disease in time by developing prototype expert system and applying data driven control strategy for inferencing. Nu Aye Aye Khin (Thesis paper, 2008, UCSY) [8] how rules are focused on the rules-based expert system by applying the combination of mixed chaining approach for inference engine.

3. Expert System

An expert system is a computer program designed to simulate the problem-solving behavior of a human who is an expert in a narrow domain or discipline. An expert system is normally composed of a knowledge base (information, heuristics, etc.), inference engine (analyzes the knowledge base), and the end user interface (accepting inputs, generating outputs). One of the most powerful attributes of expert systems is the ability to explain reasoning. Since the system remembers its logical chain of reasoning, a user may ask for an explanation of a recommendation and the system will display the factors it considered in providing a particular recommendation. This attribute enhances user confidence in the recommendation and acceptance of the expert system [1].

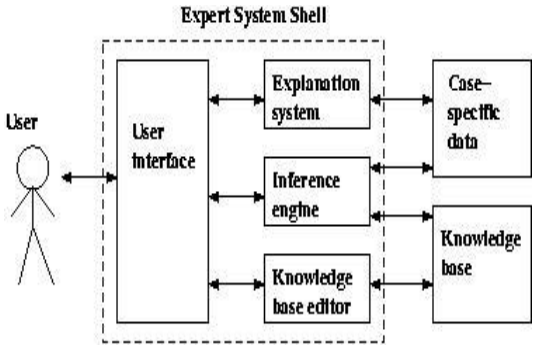

Figure 1. Architecture of Expert System

4. Decision tree Induction

Decision trees[2] are powerful and popular tools for classification and prediction. Decision tree learning algorithm has been successfully used in expert systems in capturing knowledge. The main task performed in this method is using inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules. Decision tree starts with a root node on which it is for users to take actions. Decision tree is a classifier in the form of a tree structure, where each node is either:

A leaf node - indicates the value of the target attribute (class) of examples, or

A decision node - specifies some test to be carried out on a single attribute-value, with one branch and sub-tree for each possible outcome of the test.

From this node, users split each node recursively according to decision tree learning algorithm. Among methods, decision tree learning is attractive for 3 reasons: [5]

1. Decision tree is a good generalization for unobserved instance, only if the instances are described in terms of features that are correlated with the target concept.
2. The methods are efficient in computation that is proportional to the number of observed training instances.
3. The resulting decision tree provides a representation of the concept those appeals to human because it renders the classification process self-evident [2].

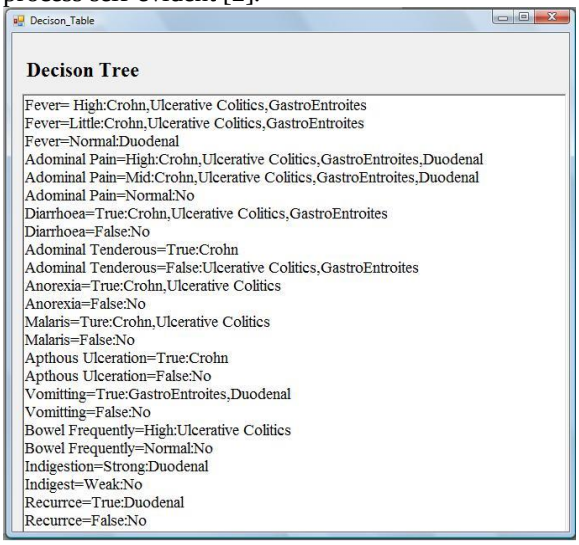

Figure 2. Decision tree for proposed System

5. ID3 Algorithm

The ID3 algorithm[3] can be summarized as follows:

- Take all unused attributes and count their entropy concerning test samples
- Choose attribute for which Information Gain is maximum
- Make node containing that attribute

The sample data used by ID3 has certain requirements, which are:

- Attribute-value description - the same attributes must describe each example

and have a fixed number of values.

- Predefined classes - an example's attributes must already be defined, that is, they are not learned by ID3.
- Sufficient examples - since inductive generalization is used (i.e. not provable) there must be enough test cases to distinguish valid patterns from chance occurrences.

The algorithm is shown below:

Function ID3 (R: a set of non-categorical attributes,C: the categorical attribute, S: a training set) returns a decision tree;

begin

If S is empty, return a single node with value Failure;

If S consists of records all with the same value for the categorical attribute, return a single node with that value;

If R is empty, then return a single node with as value the most frequent of the values of the categorical attribute that are found in records of S;

Let D be the attribute with largest Gain (D, S) among attributes in R;

Let {dj| j=1, 2, .., m} be the values of attribute D;

Let {Sj| j=1,2, .., m} be the subsets of S consisting respectively of records with value dj for attribute D;

Calculate Information Gain from Entropy for attribute D respectively to the trees

ID3(R-{D}, C, S1), ID3(R-{D}, C, S2), .., ID3(R-{D}, C, Sm);

end ID3;

Decision trees are generated from training data in a top-down, general-to-specific direction. The initial state of a decision tree is the root node that is as- signed all the examples from the training set. The process is recursively repeated for each of the new intermediate nodes until a completely discriminating tree is obtained.

5.1. Attribute Selection Measure

Gain measures how well a given attribute separates training examples into targeted classes. The one with the highest information (information being the most useful for classification) is selected. In order to define gain, Entropy concept is very important. A measure used from Information Theory in the ID3 algorithm and many others used in decision tree construction is that of Entropy. Informally, the entropy of a dataset can be considered to be how disordered it is. It has been shown that entropy is related to information, in the sense that the higher the entropy, or uncertainty, of some data, then the more information is required in order to completely describe that data. Entropy measures the amount of information in an attribute.

Given a collection S of c outcomes

$$Entropy(S) = \sum_{i=1}^{C} p_i \log_2 p_i$$

Where p (I) is the proportion of S belonging to class I. S is over c. Log2 is log base 2. Note that S is not an attribute but the entire sample set.

First the entropy of the total dataset is calculated. The dataset is then split on the different attributes. The entropy for each branch is calculated. The resulting entropy is subtracted from the entropy before the split. The result is information Gain. The attribute that yields the largest IG is chosen for the decision node. The ID3 algorithm is run recursively on the non-leaf branches, until all data is classified.

6. Agent Point of View of System

In computer science, a software agent is a piece of software that acts for a user or other program in a relationship of agency. Such "action on behalf of" implies the authority to decide which (and if) action is appropriate. The idea is that agents are not strictly invoked for a task, but activate themselves. The term "agent" describes a software abstraction, an idea, or a concept, similar to OOP terms such as methods, functions, and objects. The concept of an agent provides a convenient and powerful way to describe a complex software entity that is capable of acting with a certain degree of autonomy in order to accomplish tasks on behalf of its user. But unlike objects, which are defined in terms of *methods* and *attributes*, an agent is defined in terms of its behavior [1].

The proposed system was used the following agent. These type of agent are

-User Agent

-Classification Agent

-Test Agent

6.1 User Agent

User agent does the operations on behalf of the system. Patients inquiry disease types by entering symptoms to the user agent. User agent

replies disease type as a system service. To support disease type to user, user agent contact classification agent. To confirm disease type, user agent also gives required information to the Test agent.

6.2 Classification Agent

Classification agent classify disease according to data which is stored in training database. As the feedback of training database, classification agent informed testing ranges to user agent. User agent decides disease type by using Test Agent's confirmation.

6.3 Test Agent

Test agent can test the incoming information. In this system, test agent can get required information from user agent. To test the symptoms, test agent first search related testing in training database. Then, test agent can define disease type.

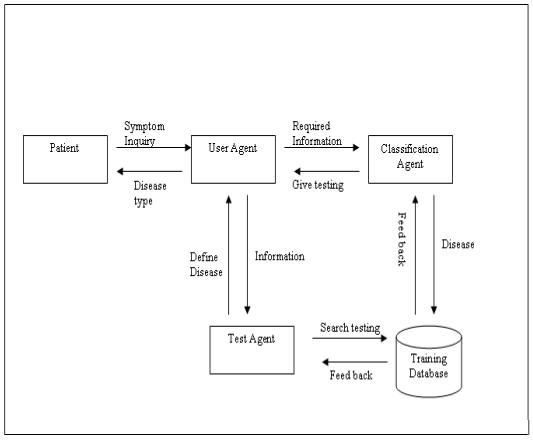

Figure 3. Applied Agents in Proposed System

7. Detailed System Architecture of Proposed System

This system include three component like pre processing which calculate information gain, ID3, testing. At the start of program, the system calculate entropy from the training data to get best attribute. And then, the algorithm check patients symptoms with the ID3 algorithm give requirement testing .When user get test results, the system check with test ranges and define disease .If user's test are not matched with test ranges, the system will give differential diagnosis. The classification process requires the attributes of relevant system. There are eleven attributes and four classes in the system. The attributes are: Diarrhea, Abdominal Pain, Fever, Malarise, Anorexia, Apthous Ulceration, Adominal Tenderous, Vomitting, Bowel Frequently, Indigestion and Recurrence. The classes are Crohn's Disease, Duodenal, Ulcerative Clitics and Gastro Entroites. The set of training data used to produce the learned concept descriptions. The algorithm takes user samples and produces class label as output and confirm the disease according to the testing results.

Structure of proposed system design is described in Figure -1.

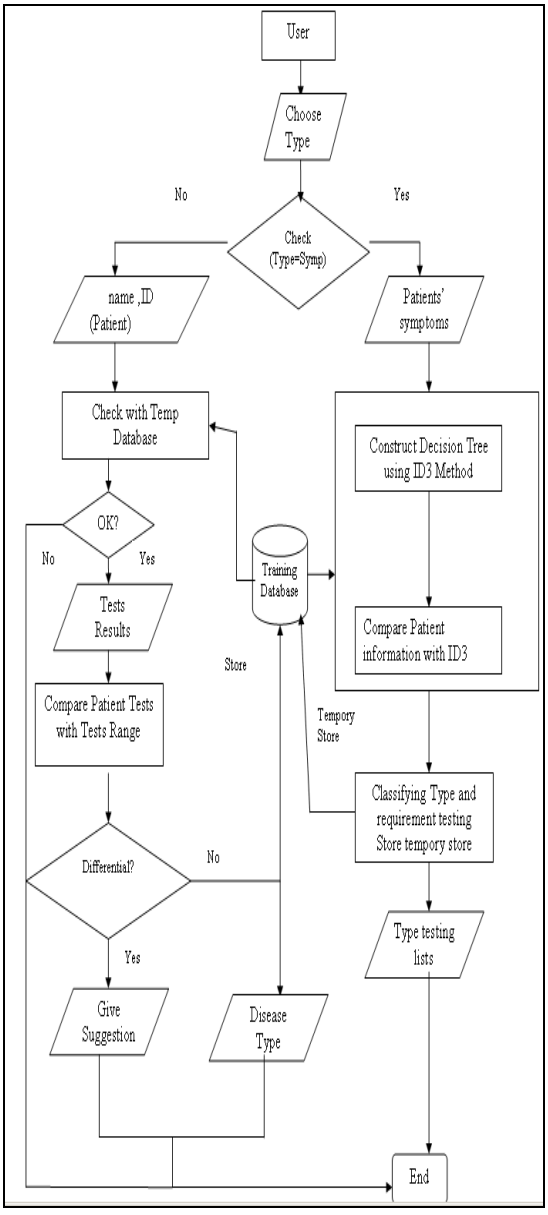

Figure 4. Proposed Detail System Design

8. Clinical features and treatments

Disease has many possible symptoms that are the same as symptoms for other health problems. So, to make a diagnosis, doctor is likely to gather information from multiple sources. Patients probably go through a combination of exams, lab tests, and imaging studies with these goals in mind:

- rule out other health problems that have similar symptoms
- make a clear diagnosis of disease
- determine exactly which part of the digestive tract is affected

The most common symptoms of Crohn's disease are abdominal pain, often in the lower right area, and diarrhea. Rectal bleeding, weight loss, arthritis, skin problems, and fever may also occur. Bleeding may be serious and persistent, leading to anemia. Children with Crohn's disease may suffer delayed development and stunted growth. The range and severity of symptoms varies.

A thorough physical exam and a series of tests may be required to diagnose Crohn's disease. Blood tests may be done to check for anemia, which could indicate bleeding in the intestines.

The doctor may do an upper GI series to look at the small intestine. If these tests show Crohn's disease, more x rays of both the upper and lower digestive tract may be necessary to see how much of the GI tract is affected by the disease. The doctor may also do a visual exam of the colon by performing either a sigmoidoscopy or a colonoscopy. The doctor will be able to see any inflammation or bleeding during either of these exams, although a colonoscopy is usually a better test because the doctor can see the entire large intestine.

Treatment may include drugs, nutrition supplements, surgery, or a combination of these options. The goals of treatment are to control inflammation, correct nutritional deficiencies, and relieve symptoms like abdominal pain, diarrhea, and rectal bleeding.

Ulcerative Colitis is a relapsing and remitting inflammatory disorder of the colonic mucosa. If may affect just the rectum (proctitis) or extend proximally to involve part or all of the colon. The most common symptoms of ulcerative colitis are abdominal pain and bloody diarrhea. About half of the people diagnosed with ulcerative colitis have mild symptoms. Others suffer frequent fevers, bloody diarrhea, nausea, and severe abdominal cramps. Ulcerative colitis may also cause problems such as arthritis, inflammation of the eye, liver disease, and osteoporosis.

Many tests are used to diagnose ulcerative colitis. A physical exam and medical history are usually the first step. Blood tests may be done to check for anemia.A stool sample can also reveal white blood cells, whose presence indicates ulcerative colitis or inflammatory disease. In addition, a stool sample allows the doctor to detect bleeding or infection in the colon or rectum caused by bacteria, a virus, or parasites.

A colonoscopy or sigmoidoscopy are the most accurate methods for making a diagnosis of ulcerative colitis and ruling-out other possible conditions, such as Crohn's disease, diverticular disease, or cancer. The doctor will be able to see any inflammation, bleeding, or ulcers on the colon wall. During the exam, the doctor may do a biopsy, which involves taking a sample of tissue from the lining of the colon to view with a microscope. Sometimes x rays such as a barium enema or CT scans are also used to diagnose ulcerative colitis or its complications.

Gastroenteritis symptoms may include:

- Low grade fever to 100°F (37.7°C)
- Nausea with or without vomiting
- Mild-to-moderate diarrhea:

If the symptoms persist for a prolonged period of time, the physician may want to consider blood and stool tests to determine the cause of the vomiting and diarrhea. As always, taking a thorough history is of great value, as is the physical examination.

A clinical questionnaire containing dyspeptic signs and symptoms was applied. Patients themselves answered the questions, but their answers were confirmed by their parents, whenever necessary.

When the answers were not clear or when the answers given by patients and parents differed, the information was left out of the analysis. The signs and symptoms concerned the ones observed before endoscopic examination.

The differential diagnosis and testing range of Enterological Disease is as shown in figure.

No.	Testing	Range	Class	Differential
1	Hemoglobin	12 to 16 grams	Crohn's Diease	-Drugs Induce
	ESR	1-13mm /hr(for male)		
		1-20mm/hr (for female)		
	White Cell Count	4300 to 10,800 cell per cublc		
	Albumin	3.5 -5 gm /dl		
2	Sigmoidoscopy	Inflammatory bowel disease	Ulcerative Colitics	-Normal -Cancer
3	H.pylori Test in Saliva	H.Pylori Bacteria	Duodental	- Gastritis
	H.Pylori Test in Blood antibody test			
	H.pylori Test in Urea breath Test			
4	Colonscopy	CMV gastro – enteritis or colitis	Gastro-Enteritis	-Normal - Colon cancer Screening -Lactose Intolarence

Figure 5. Differential Diagnosis and Testing Range

The analysis of Enterological Disease is shown in figure.

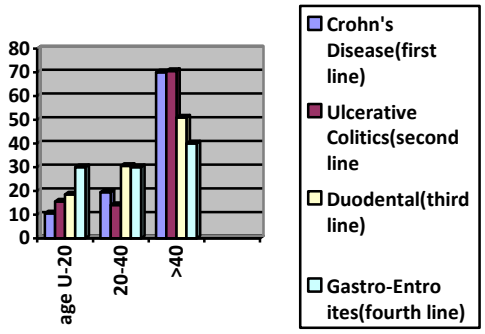

Figure 6. Analysis of Enterological Diseases

9. Conclusion

Decision making process is applied in disease diagnosis. When using decision tree induction, the decision making process itself can be easily validated. Diseases include many symptoms and difficult to decide accurately which disease patient suffer. This system focuses on developing architecture for the diagnosis of diseases by using ID3 method and testing result. It reduces time-consuming, cost, and uses easily without requiring much computer skill.

10. Reference

[1].E Turban,Expert Systems and Applied Artificial Intelligence, California State University.

[2]A.Chidanand and W.T.J.Sholom.:.Watson Research Center Department of Computer Science IBM Research Division "Data Mining with Decision Trees and Decision Rules".

[3]W. Peng, J. Chen and Haiping hou" An Implementation of ID3 -Decision Tree Learning Algorithm".

[4] W. Peng,C Juhua and Z.Haiping: "An Implementation of ID3 Decision Tree Learning Algorithm" Project of Comp 9417: Machine Learning.

[5] G.Giu_rida1, W.Wesley. Chu1 and M. Dominique Hanssens2 "*Mining Classi_cation Rules from Datasets with Large Number of Many-Valued Attributes*".

[6] G.Minos, H.Dongjoon, R.Rajeev and S.Kyuseok. *: "Efficient Algorithms for Constructing Decision Trees with Constraints".*

[7] Khine Moe Nwe ,Computerized Diagnosis System of Fever Related Diseases, M.C.Sc (Thesis), 1998, UCSY.

[8] Nu Aye Aye Khin, Development of Coronary Heart Disease Diagnosis System by Using Mixed Chaining Approach.

[9] Andronache, V., and Scheutz, M. 2004a. Ade - In Proceedings of the Fourth International Symposium "From Agent Theory to Agent Implementation".