Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

Medical Recommendation System using Crawler

Decision Over varied Symptoms and Problems using AI

Harsh Saini
Student
Jaypee Institute of Information Technology
Noida, Uttar Pradesh, India

Saksham Saraswat
Student
Jaypee Institute of Information Technology
Noida, Uttar Pradesh, India

***Abstract*—Internet is virtually the biggest database and informational platform. This electronic market can be used to analyze links that are growing at an exponential rate both as new links and as existing sub links. Due to growing number of variations in the medical symptoms, it is essential to crawl for information to get the necessary results. Prerequisite information about structure of a internet site, is required to obtain a decent assumption about the view of the millions of users/bloggers to enhance result specifications closer to the demand of user, influence decision making strategies and to measure other symptoms, results and techniques. Recommendation System doesn't involve to give definitive results but instead provides the variations to effectively predict medical conditions under the supervision of a Physician.**

Keywords—Decision Making, Implication And Prediction.

I. INTRODUCTION

The Recommendation System involves around the idea of suggesting a type of problem keeping in mind the variations in the symptoms and reducing the simple, usual things as a mandatory structure of the working of the system. As these systems can only suggest, the final decision rest with the user. The automatic system revolves around these important reasons:

-to eradicate usual and simple things which consume a lot of hard work that can be easily done by a machine.

-to point certain anomalies or variations that cannot be pointed out by a human, which can create(in some cases) a difference between life and death.

A medical recommendation system removes the subjective nature of a human involved towards a work that can influence results. The system becomes essential because it is not disturbed by the human specific conditions such as fatigue, stress and anxiety. The systems come with bonuses of Efficiency, Speed and storage of huge amounts of helpful data and to make develop connections and differences between the data to produce effective results. This Paper presents various methods from large domains of Data mining, Artificial Intelligence and Information Retrieval to enhance decision making.

This paper does talk about discussed weak points that are necessary in order to determine the best method for the task. Some of the advantages and disadvantages can be found in the context below which are proved and implemented.

II. DESIGN OF CRAWLER

Artificial Intelligence has mechanisms that if implemented like a human would do something with a perfect, attentive mind, then it is considered as Intelligent[4]. The algorithms of this field can be used to solve problems connected to a large variety of applications including the capacity to make decisions, especially when medical diagnosis or treatment requires intelligent thinking. This ability is a feature of a human being and this mechanism deals with rational approach, not undermining a human's approach as irrational but this mechanism functions to be better than humans. Since, these mechanisms are part of mathematics and use the power of computer science.

Although technology does provide efficiency and speed, features that exceed humans to perform calculations, the technology still is far away human judgment. Hence, Humans are still the best decision makers [5].

Though the artificial intelligence domain provides a large variety of solutions for problems for decision making and reasoning, this paper deals with a single solution:

A. Using the Knowledge Based System

A knowledge based system uses a knowledge base to solve complex problems. The term is broad and is used to refer to many different kinds of system[6]. The one common theme to represent knowledge explicitly via tools such as ontologisms and rules rather than implicitly via a code that a computer does.

The method that we choose to make decisions involve revolve around this decision making, similarly like the very first system based on rules. These expert systems represent facts via rules had

This approach comes to implement the human process of reasoning applied to a representation of knowledge. The important part in this approach is Knowledge base. It consists of a set of tuples of a natural language [4]. The first advantage of knowledge based system: knowledge representation using a language that is very close to the natural one makes them easy to understand and easy to create.

Hence, the mechanism represents the things that an automated decisional system should know in order to have a solution to the problem aroused. The example presented here comes from medical diagnosis [2]. The system requires prior information regarding a solution starting from well-defined states. The initial information constitutes the system's premises and are connected via logical operators, creating the system's rules.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

Suppose: The aim of the system is to diagnose three different types of hepatitis: B, B+D, C. The premises follow the values of lab tests that analyze three markers involved into these diseases: HBsAg, anti-VHD and anti-VHC. These markers can be positive and negative; hence they have a binary representation. Logical first order logic statements are provided to build an effective knowledge base to increase the decision making. Premises are standardized as [4]:

- $\neg$ (negation) $\Rightarrow$ unary operator
- $\wedge$ (conjunction) $\Rightarrow$ binary operator
- $\vee$ (disjunction) $\Rightarrow$ binary operator

The rules have been created for the decisional system according to the standard observed symptom root causes[2] :

$\text{HBsAg} \wedge \neg \text{anti-VHD} \rightarrow \text{Hepatitis B}$

$\text{HBsAg} \wedge \text{anti-VHD} \rightarrow \text{Hepatitis B} \oplus \text{D}$

$\neg \text{HBsAg} \wedge \text{anti-VHC} \rightarrow \text{Hepatitis C}$

The rules can be written in natural language as:

if HBsAg and not anti-VHD then Hepatitis B

if HBsAg and anti-VHD then Hepatitis B $\oplus$ D

if not HBsAg and anti-VHC then Hepatitis C

In fact, the knowledge base has an utility matrix structure [6], which helps the system to be clear and robust.

Hence, this type of decisional system has the main advantage that it is easy to be implemented for a knowledge base with simple rules. The results it offers are very clear; using logic, it decides whether the output is true or false. And it has also a great chance to be accepted by those who are involved in the medical field because its mathematical support is closely related to the natural language, offering a high level of trust to the medical stuff who use it.

A discussion from artificial intelligence domain is incomplete if accuracy of the system is not defined. The simple system, which was presented here, always produces a correct result; obviously, but if the graph becomes too complex, then unexpected results may be encountered.

B. The Probabilistic Vision of Reason

If the information to be calculated from the decisional system is too complex or large to use a knowledge base solution, the probabilistic approach can be a choice. For ex, if a patient is diagnosed with Hepatitis B, the problem arises with the evolution of patient's medical condition. There are three evolutionary types(usual, with relapses and de-compensations) and six severity levels (easy, medium, serious, prolonged, cholestasis, comatose). It might be very useful to have some predictions regarding the evolution of the patient during treatment, but unfortunately the influence of premises(symptoms and laboratory test results) on the patient's evolution is difficult to be represented in clear rules as those required.

The solution can be provided with statistical methods. Thus, in a database will be stored all the standardized premises for as many possible patients. Then, for a new patient, the database is processed by statistical methods to calculate the probability of each evolutionary type and severity level of hepatitis B. Therefore, the solution offered by this method is more accurate. The statistical approach associates a probability to each possible output [6] .

One of the most suitable methods for this is Bayes' theorem:

$$p(D_k \mid S) = \frac{p(S \mid D_k) \cdot p(D_k)}{p(S \mid D_k) \cdot p(D_k) + P(S \mid \neg D_k) \cdot p(\neg D_k)} \quad (1)$$

This theorem describes the connection between two events D_k and S, D_k being a diagnosis(among many possible diseases) from available evidence S, which here is a vector that stores the patient's symptoms and laboratory test results in a standardized form (binary values). Some of the patient's features have a binary form from the beginning: Sex (M/F),living area (town/village), in contact with other infected patients (yes/no), blood transfusions during the last twelve months, any injected treatments during some last months and a lot of symptoms or subjective signs (of anorexia, nausea, queasiness, asthenia, fever, hyper chrome urine, tegument jaundice, arhtralgia, myalgia, skin eruption).

Other features must be processed[2] in order to decide whether they belong to a particular range or not, which may be:

Age, a feature that was divided into four intervals (<20 years old / [20,30) / [30,45) / > 45). Therefore, a patient that is 34 years old will have the following binary values stored in the fields that belong to this feature: 0, 0, 1, 0.

Starting time was split into three ranges ($\leq$ 7 days / [8, 21] / $\geq$ 22 days). If the starting time was 21 days, then the values connected to this feature are: 0, 1, 0.

Jaundice, the color of the skin, was also divided into three intervals ([0, 1] / 2 / [3, 6]) associated to different intensities of color.

Liver is measured in centimeters and has two possibilities ([0, 2) / $\geq$2).

Bilirubin is a laboratory test which was divided into three ranges (<1 / [1, 10) / $\geq$10).

Tymol has two alternatives ([0, 4) / $\geq$4).

Transaminasa glutamicopirúvica

has two intervals ([0, 400) / $\geq$400).

All these ranges are established by a physician, and not by a system's developer, because they have a strong connection to the patients' medical status and to his evolution. If other features must be added to the database, a physician should analyze them in order to decide which will be their standardized form.

As (1) shows, Bayes' theorem expresses a probability based on three other probabilities[1]. This choice could be considered a disadvantage, but it is useful in practice, because the three probabilities can be easily calculated, as (2) to (6) demonstrate. This is because the probability in the casual direction $(p(S \mid D_k))$ is more evident than the probability in the diagnose direction $(p(D_k \mid S))$ [4].

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

The probability $p(D_k)$ is very simple to be calculated (by (2)). It is the frequency of a disease in the statistical population Ω (the number of records r that have diagnosis D_k equal to 1) .

$$p(D_k) = \frac{cardinal\{r \in \Omega \mid D_k(x) = 1\}}{cardinal\Omega} \quad (2)$$

The probability $p(S \mid D_k)$ [3] is calculated assuming the conditional independence of all symptoms for a disease D_k .

This is an important restriction that Bayes' theorem imposes and sometimes it is difficult to fulfill it. In (3), which calculates this probability, n is the number of symptoms stored as binary values in the input vector S and σ_i is a symptom.

$$p(S \mid D_k) = \prod_{i=1}^{n} p(\sigma_i \mid D_k) = \prod_{i=1}^{n} \frac{p(\sigma_i, D_k)}{p(D_k)} \quad (3)$$

The probability $p(\sigma_i, D_k)$ is called Compounded probability [3] and means that both events σ_i and D_k happen at the same time . This probability, calculated by (4), is the frequency of σ_i and D_k in the statistical population Ω .

$$p(\sigma_i, D_k) = \frac{cardinal\{r \in \Omega \mid D_k(r) = 1, \sigma_i(r) = 1\}}{cardinal\Omega} \quad (4)$$

The last probability, $p(S)$, is calculated considering that the diseases are mutually exclusive [1](only one disease is present at a moment) and that it is sure that one disease is present. This restriction is not a problem for the example presented here, because it is obvious that Hepatitis B cannot evolve in different directions for the same patient at the same time, it cannot have two different severity levels, and also it is sure that the disease will have an evolutionary type and a severity level.

Equation (1) is used again in (5) to calculate the probability of not having a particular disease D_k.

$$p(\neg D_k \mid S) = \frac{p(S \mid D_k) \cdot p(\neg D_k)}{p(S)} \quad (5)$$

It can be noticed that both (1) and (5) contain the term $1/\ p(S)$. It is considered a normalization constant [4] and it ensures that the sum of probabilities is 1.

For the system presented, the Hepatitis B has three evolutionary types; the patient will evolve to one of those and nothing else is possible. Therefore, if the term 1/ $p(S)$ is ignored in (1) and (5), the obtained partial probabilities are still in the correct relative proportion. There sum is not 1 anymore, but this case can be solved dividing each one by their sum. Thus, this sum can replace $p(S)$.

$$p(D_k \mid S) = \frac{p(D_k) \cdot \prod_{i=1}^{n} p(\sigma_i \mid D_k)}{p(D_k) \cdot \prod_{i=1}^{n} p(\sigma_i \mid D_k) + p(\neg D_k) \cdot \prod_{i=1}^{n} p(\sigma_i \mid \neg D_k)} \quad (7)$$

The database that was investigated by this system contained 825 patients, 70 of them being reserved for the system's validation. The accuracy of this system is 69.33%. A physician can decide, according to the medical domain where the system can be used. It can be highly effective for the vulnerable places where even simpler results can be evaluated and thereby decisions made.

If, for instance, this type of decisional system is used to predict the evolution of a particular medicine store in a pharmacy, then the accuracy could be enough. But if the system predicts a diagnosis, its accuracy has to be improved.

An important feature of the recommendation system is that it doesn't provide a clear result, but provides with a probability, a recommendation. It quantifies uncertainty, offering a degree of belief. For instance, it is not sure that a patient will develop a serious level of severity regarding hepatitis B, but it is more probably because the system says 0, 0.3077,.7833, 0. Further, a patient gets a review that he might have a 78.33% of chances to have a form of Hepatitis B.

III. CONCLUSIONS

This paper has described important mechanisms of usage of artificial intelligence in expert systems to provide automation, accuracy and efficiency into the domain which can be used to develop recommendation systems for medical variations. The advantages and disadvantages both made it emphasized clearly on the aspects of results for the system.

The benefits provided by these systems are obvious. On one hand they simplify the work, saving time, increasing accuracy and efficiency. On the other hand, an automated system can detect imperceptible things that can hardly be noticed by a physician which is the result of complex computation and reasoning, things that are not evident or that are the effect of too many factors involved.

It is difficult to decide where should be limit between Artificial intelligence and human logic, but automated systems can be of utmost help is a one undeniable truth essential to improved accuracy, efficiency and quality of recommendation system.

ACKNOWLEDGEMENT

We would like to thank our mentor Shariq Murtuza for believing in us and having a hand of support on us. We also extend our special thanks to Dr. Charu Gandhi. We thank our family and friends for their contribution in our lives.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 5 Issue 10, October-2016

REFERENCES

[1] Looks, Moshe and Goertzel, Ben and Pennachin, Cassio, 2005

[2] Yu- Fen Fan, Cheng- Chan Lu , Wen- Chi Chen , Wei- Jen Yao , Hui- Ching Wang , Ting- Tsung Chang , Huan- Yao Lei , Ai- Li Shiau , Ih- Jen Su, 2001

[3] Veerarajan,"Probability, Statistics and Random Processes", ISBN13: 978-0-07-066925-3, Tata McGraw-Hill 2006

[4] McCarthy, J.,"What is Artificial Intelligence",Computer Science Department,Stanford University 2007

[5] Russel & Norvig,"Artifical Intelligence: A Modern Day Approach", Pearson Education, 2003, ISBN:0137903952

[6] Giarratano, Joseph C.; Riley, Gary D.,"Expert Systems : Principles and Programming " 2005