Case Based Reasoning (CBR) for Medical Question Answering System

Setio Basuki[*1], Alfira Rizky[2], Galih Wasis Wicaksono[3]
[1,2,3]Universitas Muhammadiyah Malang/Informatics Engineering
setiobasuki.umm@gmail.com[*1], veraalfirarizky@gmail.com[2], galih.w.w@umm.ac.id[3]

Abstract

In this research, the researchers implement a medical Question Answering System (QAS), a complaint system in the form of sentences or paragraphs of questions about the complaint (illness) suffered by a person. Afterwards, the system will give answer to the questions with answers in the form of diagnosis based on the system knowledge. The system in this study has knowledge of the system obtained based on Case Based Reasoning (CBR) method from the previous cases stored in the database. When there is a new case, the system will perform a matching process using CBR and Sorenson Coefficient calculations to find out which the previous cases have the highest percentage of matches with the new case. Then the selected previous cases will be taken and given to the new case. Testing is processed by using 2 types of testing, expert validation testing with result of 28 data of appropriate test from 30 test data and accuracy testing resulting of 93,33% from the appropriate test data.

Keywords: *Health, Case Based Reasoning (CBR), Sorenson Coefficient*

1. Introduction

Currently, various ways or efforts to improve health are aimed at improving service quality, equity and coverage of health services. The difficulty of reaching the wider range of coverage area to get health care is not an easy task to do. The quality of public health services needs to be continuously improved to enhance the health status of the community due to the realization of healthy conditions for all parties of society [1]. When someone has a question about his health to a doctor, in order to obtain any replies, an individual must spend energy or time to see a doctor and also spend a consultation fee (with the aim of getting answers to all questions above). These issues may be solved to all individuals. For example, in terms of exertion, if someone is in a position which causes him unable to see a doctor, it will be a challenging task to visit a doctor to have consultation. In terms of cost effectiveness, not everyone can afford to pay the service cost of a doctor (health experts). Nowadays, technology becomes inseparable from the life of society, technology being as old as and having important role in human life [2]. Every layer of today's society has started to have a high need even to the point of technological dependence to providing daily assistance. The existing technology uses a lot of expert system, a piece of artificial intelligence accumulating the knowledge and experience processed into a computer to solve problems usually requiring human expertise. Therefore, everyone can solve problems and determine the right solutions to the existing problems [3]. Expert system used in this research is Cased Based Reasoning (CBR), a problem solving technique based on the previous cases. CBR has four stages: retrieval, reuse, revision, and retainment (storing new problems) [4]. The CBR calculation method uses Sorenson Coefficient method. This method calculates the similarity of two binary objects. If a new case is determined to be similar to a previous case then it has the similarity value of 1. However, if it is dissimilar, then the value of its similarity is 0 [5].

Therefore, an automated question and answer system is developed. It takes form of a web-based system that can answer a question from someone who is experiencing health problems and wants to acquire opinions from health experts. This system will provide answers in accordance with the questions given (in accordance with expert system data management capabilities). In case study [6] and [7] applying CBR method and Sorenson Coefficient calculations, the detection was conducted using a checklist of problems based on existing cases, while this study uses free sentences by each user. The existing cases such as cases [6] and [7] have not discussed the problem of complaints in a person's illness to a health expert (doctor).

This system can be accessed anywhere easily (using the internet) and no longer needs to pay for the expert services. Hence, it has potentials to overcome the occurring problems.

2. Research Methods

This study uses a calculation method. This method is Case Based Reasoning (CBR) to perform the main processes in the form of search results from the new data by conducting a search with the highest percentage of processing with the previous cases. Afterwards, pre-processing concept is employed at the stage of the basic Word form for simplifying keyword search in the form of symptoms in the system. The following process is the calculation process of similarity or matching data by using Sorenson's (calculation with binary numbers then figuring out the percentage of the new data matches with the previous data).

The system process flow has several stages. Initially, a new case is given, and then the system will perform data preprocessing on new cases. The results will be compared with the existing data using keywords. Afterwards, the keywords will look for similarities with existing keywords in the base case. The following process utilizes Sorenson's calculation in finding the calculation results with similarity percentage of existing cases. The new cases will be delivered to the existing cases having the highest similarity percentage. Finally, from the selected data, the diagnosis will be taken to be the case with the addition of a new diagnosis of a physician who had already been appointed.

2.1 Case Based Reasoning Method

Case-based reasoning method is employed to facilitate experts in identifying diseases and providing possible solutions by performing problem solving process of adaptation to the new settlement in previous cases [8]. It does not mean replacing the position of the experts. It offers assistance in confirming the sufferers' decision, as it may have a lot of alternatives which must be selected appropriately and accordingly.
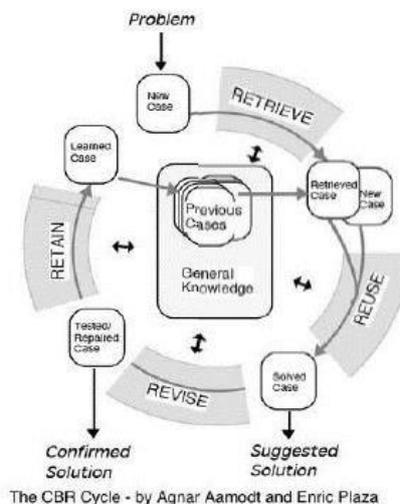

Figure 1. Case Bases Reasoning Process

The method of Case Based Reasoning is a method used to solve problems based on knowledge of previous cases, as presented in Figure 1, that is when there is a new case then it will be utilized to search similar process handling new cases based on the percentage in common with previous cases. Furthermore, this method may provide answers from the previous cases used on new cases. It will also store new cases in the knowledge base, so the system will learn and possess new knowledge. Using this mechanism, the system will grow by inputting the next cases or problem resolutions. This method generally consists of 4 parts or steps as follows:

1. Retrieval.
   Initially data is inputted, and early filtering process will be performed, by which problems is introduced. This process identifies or recognizes new problems or new cases.
2. Reuse.
   In data reuse (system knowledge), the system will perform a database search on previous problems through the identification of new problems. Consequently, it reuses the information

about previously recorded problems having the highest similarity to solve problems or a new case or a new problem.

3. Revision.
   The following stage will evaluate the information with justification return to resolve a problem which occurs on a new case. The system will provide new solutions to new problems or cases.
4. Retainment.
   The system will store problems or new cases into its knowledge base [5].

2.2 Sorenson Coefficient

At the revision stage, Case Based Reasoning (CBR), the process of determining by doing the process of problem solving, will be utilized. In this research, the calculations are performed in the stages of the search results according to percentage similarities between the previous cases and new cases. In order to calculate similarities of the two binary objects, the research employs Sorensen coefficient. The formula used to compute similaritties between two objects x and y is as follows [4].

$$SBC = \frac{2 \times M11}{2 \times (M11 + M01 + M01)} \qquad (1)$$

The Equation 1 searches the percentage of similarity between the previous and the new case to recognize the keywords of each case (between previous cases and new cases), and then the process will be conducted by viewing the keyword matched in each case. There are keywords that exist in the new case. Upon any recognized resemblance, the value of 1 will be given and 0 for any dissimilarity. The comparison of the value emergence of 1 and 0 values between the old cases and new cases will be developed. The following calculation will be conducted with the previous formula percentages [1].

2.3 Data Research

The research data is taken from the website's FAQs for online doctors www.alodokter.com. The case base which is stored in the database is a data comparison with the case of the case to get a diagnosis or solutions in accordance with the similarity between the previous and new cases. There are 100 base cases of the disease by 501 categorizations of diseases consisting of 2 cases (with different types of solution). The data will be stored in the database as data train for new cases in a search for answers in the form of diagnosis. In the base case, there are two main sections namely case data in the form of questions (Table 1) and diagnostic data in the form of answers (Table 2). The following table presents the example of data FAQs in the form of cases and diagnoses.

The test data used is also derived from the same source, the website of www.alodokter.com, but the random nature of the test data is in accordance with the list of diseases that already owned database. The test data and also the case data are taken in the form of a question as in Table 1.

Table 1 Case Data (FAQ)

No	Case (Question)
1	Good afternoon Doc, I'm experiencing something strange like forgetting my childhood memory aged 4 to 11 (like my memory has been erased). It is getting worse like facing difficulties in remembering names/places in the past ¾ month. Moreover, in the last couple of months I have seen bad memories even though they're not clear. Do you think I can be healed?
2	I want to ask, in the last 5 years, I often forget to put things and do planned activities. To make it worse, I believe, I have insomnia and feel difficult to wake up (I feel very exhausted and sleepy and want to continue sleeping) what do you think about my illness?
3	Doc I need to ask you something. In January 2016, my son, aged 8, was diagnosed by his doctor to suffer from bronchitis. He has been medicated, but it happens again. He is coughing and out of breath, but now he was diagnosed to suffer from asthma, cough with phlegm and it hard to remove the sputum. Can you tell me whther his illness is bronchitis or asthma doc? And how to handle this?

Table 2 Data Diagnosis (FAQ)

No	Diagnosis (Answer)
1	Problems such as memory loss or gradual memory loss related to information or experiences are related with insomnia. An individual suffering from insomnia may experience difficulties in learning new information as well as new memories. In general, amnesia is divided by two types: 1. Anretrogade Amnesia Amnesia which occurs when experiencing difficulties in remembering past information, for example memory loss before accident. It is usually difficult to remember short-time memory. 2. Retrogade Amnesia Amnesia which occurs when experiencing difficulties in remembering past information, for example memory loss before accident. This condition basically can be influenced various things which may affect brain performance or function as body defence mechanism to certain events. It may be affected by head injuries (hit, fall down, accident, etc.).
2	Complaints are related to the following causes: • Psychological stress • Alzheimer • Amnesia • Lack of concentration or concentration problems • Nutrition deficiency • Hormonal imbalance • Sleep deficiency • Blood vessel blockage • Bain tumor • Excessive diet • Fatigue • Dehidration
	Bronchitis is a condition in which the main the main channel of the lung or bronchus has an infection which causes inflammation or canal inflammation. Some of the symptoms experienced by people with bronchitis include: • Cough with grayish yellow or green phlegm • Pain in the throat • Out of breath • Nasal congestion or running nose • Pain or discomfort in the chest • Fatigue Bronchitis can be divided into two groups, namely (1) Acute bronchitis, bronchitis which is often…..

3. Results and Discussion

Expert validation testing is the result of the matching system by checking with the doctor's conformity. If it is defined as appropriate, it is considered as accurate in giving an answer in the form of diagnosis of disease against the patient. Therefore, a deeper and specific justification by experts can be conducted. Upon any recognized incompatibility, the system declares as wrong diagnoses and results justified by the expert' judgment as presented in Table 3.
The calculation of the percentage of accuracy:

$$\frac{fx}{fn} \times 100\% = Px \qquad \rightarrow \qquad \frac{28}{30} \times 100\% = 93{,}33\% \qquad (2)$$

Based on the calculation of accuracy with the Equation 2, the result of the calculation of the percentage compliance system is proven that giving answers based on expert validation is as much as 93.33% [9] [10].

Table 3. Experiment Results

Cases	Diseases Prediction (Data Source)	Diseases Prediction (CBR System)	Expert Judgment
Case 1	Amnesia	Amnesia	Match
Case 2	Amnesia	Amnesia	Match
Case 3	Bronchitis	Cold	Not Match
Case 4	measles	measles	Match
Case 5	diarrhea	diarrhea	Match
Case 6	Diabetes	Diabetes	Match
Case 7	Insomnia	Insomnia	Match
Case 8	Vertigo	Vertigo	Match
Case 9	Insomnia	Insomnia	Match
Case 10	Cholesterol	Cholesterol	Match
Case 11	smallpox	smallpox	Match
Case 12	Malaria	Malaria	Match
Case 13	Acne	Acne	Match
Case 14	Acne	Acne	Match
Case 15	Whitish	Whitish	Match
Case 16	Menstruation	Menstruation	Match
Case 17	Menstruation	Menstruation	Match
Case 18	Obesities	Obesities	Match
Case 19	Osteoarthritis	Osteoarthritis	Match
Case 20	Polio	Polio	Match
Case 21	Clod	Esophagitis	Not Match
Case 22	Rheumatism	Rheumatism	Match
Case 23	Rhinitis	Rhinitis	Match
Case 24	Stroke	Stroke	Match
Case 25	Scoliosis	Scoliosis	Match
Case 26	Typhus	Typhus	Match
Case 27	Typhus	Typhus	Match
Case 28	Varices	Varices	Match
Case 29	Vertigo	Vertigo	Match
Case 30	Varices	Varices	Match

4. Conclusion

Based on the research conducted, the conclusions in making System Question Answer of Health Based Case Based Reasoning (CBR) are as follows:

1. Data used is obtained from health website having a question and answer feature with several doctors. The website provides clarity on the basic of disease and the classification of the types of disease in alphabetical order. The explanation of the disease is also completed with reviews about the question and answer to related diseases.
2. The initial data used is data for expert knowledge or as data train. The selected data is disease data having no kind of disease (such as blood cancer, bone cancer, cervical cancer, etc.). Afterwards, the complaints in the form of questions and diagnosis of the doctor in the form of answers are taken and determined according to disease keyword in the form of specific symptoms for each type of disease.
3. For solving the case, new cases are inputted in the form of question then the system will find the symptoms of disease stored and compared with symptoms of disease corresponding to the previous case (data train) and calculated the percentage of compatibility using Sorenson Coefficient calculation.
4. The system is tested to match its result validity and calculates the test data validated or in accordance with the experts. There are as many as 28 correct or appropriate data from 30 test

data. From this test, it can be concluded that this system has a high percentage of validity, that is equal to 93.33%. With this high percentage of validity, this system is considered as feasible to be used by the wider community, especially the patients who require initial actions for the problem or case of disease suffered.

Notations

x : Previous case
y : New case
$M11$: Number of binary attributes, x=1 and y=1
$M10$: Number of binary attributes, x=1 and y=0
$M01$: Number of binary attributes, x=0 and y=1
$M00$: Number of binary attributes, x=0 and y=0
fx : Total valid data
fn : Overall test data
Py : Total not valid data

References

[1]   M. Tiyas and A. Rohmani, *"Relationship between Patient Satisfaction and Patient Interest in Re-utilization of Health Service on Family Doctor Practice,"* Medical Faculty of University of Muhammadiyah Semarang, Vol. 1, No. 1, Pp. 155, 2012.
[2]   E. Durkheim and M. Weber, *"Unit 3,"* Religion, Pp. 23–30.
[3]   A. Fadli, *"Basic Expert System,"* Pp. 1–8, 2010.
[4]   D. Rachmat, *"Implementation of Sorensen Coefficient Method in Determining Dengue Fever Potential Regions (Case Study: Pontianak City)."*
[5]   M. Nugraheni, U. Ahmad, and D. Yogyakarta, *"Design Case-Based Reasoning,"* Study Program of Informatics Engineering, Vol. 6, No. 1, Pp. 612–616, 2012.
[6]   O. Agustina, *"Case Based Reasoning for Preeclampsia Pregnancy Detection Using Sorensen Coefficient Method,"* Malang: University of Muhammadiyah Malang, 2015.
[7]   M. Taufik, *"Case Based Reasoning Using Sorensen Coefficient Method in Determining Tourism Destinations in Lombok, West Nusa Tenggara,"* Malang: University of Muhammadiyah Malang, 2015.
[8]   S. Mulyana, *"Short Review on Case Based Reasoning Development,"* National Seminar of Informatics 2009 UPN "Veteran" Yogyakarta, Vol. 2009, No. semnasIF, Pp. 17–24, 2009.
[9]   A. S. Aribowo and S. Khomsah, *"Case Based Reasoning for Leukemia Early Detection,"* Vol. 2012, No. semnasIF, Pp. 168–174, 2012.
[10] A. P. Putra, A. P. Kurniati, and A. Romadhony, *"Expert System Using Case Based Reasoning, Rule Based Reasoning and Theorema Bayes (Case Study: Common Disease Detection),"* Pp. 2–5.