Journal of
Computer Science & Systems Biology

Salih et al., J Comput Sci Syst Biol 2018, 11:4
DOI: 10.4172/jcsb.1000279

Research Article Open Access

Self-Diagnosis of Diabetes Using CBR Algorithm

Salih NK*[1], Elbashier H*[2], Zang T*[3] Eshtiag A Abd Elrhman[4]

[1]*Electrical and computer engineering department, Engineering College, Karary University, Sudan*
[2]*Computer Science and information technology college, Sudan University of Science and technology, Sudan*
[3]*School of Computer Science and Engineering, Harbin Institute of Technology, China*
[4]*Deanship of Community Service and Continuing Education, Jazan University, Jazan, KSA*

Abstract

The continuously rising cost of medical spending, population size is growing up and increasing their need for healthcare, which requires time saving for laboratory technicians at the examination is a great incentive to create a new approach that helps to provide health care to patients at lower costs with good management. We started to apply the concept of the autonomic system that let the system work without intervention of the user. It has given by implemented and designed Case-Based Reasoning (CBR) algorithm in suitable way to self diagnose diabetes for patients depending on some tests results. The result is implementing new self diagnosis of diabetes system without user intervention which suggests that such a system is valuable both for less experienced clinicians and for experts where the system may function as a second option.

Keywords: Self-diagnosis; Case base reasoning; Autonomic; Diabetes

Introduction

In the medical domain, diagnostic, classification and treatment are the main tasks for a physician. The multi-faced and complex nature of the medical domain such as the psych- physiological domain often requires the development of a system applying several techniques for instance like CBR [1]. In this paper we have proposed method can diagnose new patients according to the similar solution of stored cases. CBR adapts previous solutions for similar problem in solving new problem in hand. For similarity we have used deferent functions that give optimal adaptation. We have summarized our contributions as follows:

- A case based reasoning system is developed to prove that it is possible to diagnose diabetes mellitus using computer system.
- Reduce the time required to come to a decision particularly in an emergency case.
- The system is able to find optimal or near optimal diagnosis result.

We organized this paper by beginning with the case-based reasoning in section II. Depending on CBR to classify the patient status (normal, prediabetic or diabetic) according to specific attributes. In section III we have described the related work for diabetes diagnosis system. We present the conclusion and point of future work in section IV.

Case-based reasoning

Is a psychological theory of human cognition [2,3]. Case-based reasoning (CBR) first formalized in the 1980s following from the work of Schank and others on memory [4], and is based upon the fundamental premise that similar problems are best solved with similar solutions [5,6]. Its idea is to learn from experience. CBR enables utilization of knowledge of previously experienced, concrete problems situations. A CBR system requires a good supply of cases in its case database. The retrieval task starts with a problem description, and ends when a best matching previous case has been found.

A new problem is solved by finding a similar past case, and reusing it in the new problem situation.

Sometimes a modification of the solution is done to adapt the previous solution to the unsolved case. Therefore, there are four key issues in the developing of any CBR system, namely: (A) case representation and identifying key features, (B) indexing and retrieving similar cases from the case memory, (C) measuring case similarity to select the best match, and (D) modifying the existing solution to fit the new problem. It is important to emphasize that CBR also is an approach to incremental and sustained learning; learning is the last step in a CBR cycle [7,8]. CBR has already been applied in a number of different applications in medicine. CBR is appropriate in medicine for some important reasons; cognitive adequateness, explicit experience, duality of objective and subjective knowledge, automatic acquisition of subjective knowledge, and system integration [9].

CBR components

There are several components of CBR:

• Case: the case has two components: the problem description and the solution. So it is defined as an instance of a problem [10].

• Case-base: it contains the experiences and conforms to one of the four sources of knowledge required in a CBR. They are the vocabulary, the case-base, the similarity measure and adaptation containers.

• The first, vocabulary, contains the term which support the others. The case-base comprehends what is in a case and how cases are organized.

• The similarity measure container contains knowledge to determine the similarity between two cases in the retrieval phase.

***Corresponding authors:** Salih KN, Electrical and computer engineering department, Engineering College, Karary University, Sudan, E-mail: nadircom2006@gmail.com

Elbashier H, Computer Science and information technology college, Sudan University of Science and technology, Sudan, E-mail:esho_95@hotmail.com

Zang T, School of Computer Science and Engineering, Harbin Institute of Technology, E-mail:Chinatianyi.zang@hit.edu.cn

Received May 07, 2018; **Accepted** June 11, 2018; **Published** June 20, 2018

Citation: Salih NK, Elbashier H, Elrhman EAA, Zang T (2018) Self-Diagnosis of Diabetes Using CBR Algorithm. J Comput Sci Syst Biol 11: 233-239. doi:10.4172/jcsb.1000279

• The solution adaptation container contains knowledge to adapt past solution to new problem in the reuse stage [10,11].

• Case index: Kolodner identifies indexing with an accessibility problem [12], that is, with the whole set of issues inherent in setting up the case base and its retrieval process so that the right cases are retrieved at the right time.

Case base organization

When there is a new problem, we are going to retrieve all relevant cases from case base, take the appropriate solution of the retrieved case and evaluate this solution, if it is good we will save this problem in the case base, otherwise evaluate it again and do the same stages. Following (Figure 1) shows case base organization.

When there is a new problem, we are going to retrieve all relevant cases from case base, take the appropriate solution of the retrieved case and evaluate this solution, if it is good we will save this problem in the case base, otherwise evaluate it again and do the same stages.

CBR phases

Cycle Aamadot and Plaza [13] identifies four stages of CBR - sometimes referred to as the R4 model - that combine to make a cyclical process:

• Retrieves similar cases to the target problem.

• Reuse past solutions.

• Revise or adapt the suggested solution to better fit the target problem.

• Retain the target and solution in the case base below (Figure 2) shows these phases of CBR.

Similarity functions

Similarity measure is used in problem solving and reasoning to match a previous case (case-base) with the new case to find solution. It select cases that have nearly the same solution that the new case. These similarities can be calculated using these functions describe this function and their equations. Where, d: distance calculated by each function; x: first case; y: second case; n: number of compared cases.

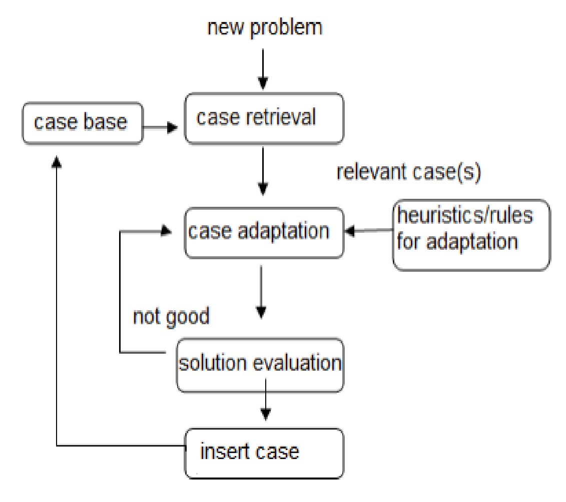

Figure 1: Case base organization.

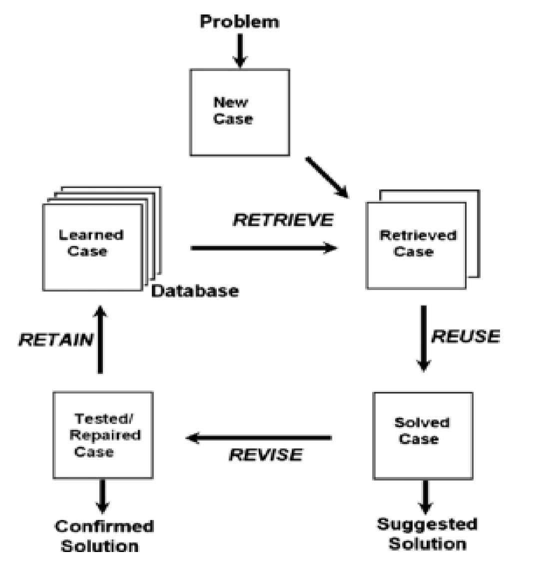

Figure 2: CBR phases.

A distance measure is needed to determine the closeness of Instances. K-nearest neighbor methods are mainly suitable for problems with numerical data (typically real, floating point numbers).

Experimental Results

We have taken a result for every classification (diabetic, prediabetic or normal person). The five attributes used are Age, Gender, Fasting glucose test, two-hour OGTT and HbA1C measurement as the input of the system. Similarity measure is used in problem solving and reasoning to match a previous case of diabetes mellitus with the new problem to find solution. We can depend on accuracy and root Mean squared error (RMSE) to show the different between similarity functions.

Accuracy

Our adaptation algorithm we supposed eight cases for diabetic patients and eight cases for prediabetic patients and eight cases for normal people. The measure accuracy is computed in Equation 1 as follow:

$$\text{Accuracy}=\frac{\text{Correct Diagnosed}}{\text{Total Testing Cases}}\times 100 \qquad (1)$$

For estimating the accuracy rate of the CBR model, dataset is divided into two sets. One of them is training set that is used for model training and another is test set that is used for estimating accuracy of the model. So, 140 of data are allocated to training data and the 50 is allocated to testing data. The output of the system is normal, prediabetic or diabetic. CBR algorithm in this application will bring more than 72% accuracy of diagnose of diabetes mellitus, and its maximum accuracy is 94%. The accuracy for all functions which have been used to calculate similarity distance is explained in Tables 1 and 2.

The following (Figure 3) is a chart that shows the accuracy rate built in x and y axes, x axe determines number of test cases which have been tested using the similarity functions that discussed in previous section of this thesis, and y axe determines the result of testing either correct or not.

Root-mean-square deviation (RMSD) formula: Also called

Function	Formula
Manhattan distance	$d_i=\sum_{i=1}^{n}\lvert x_i - y_i\rvert$
Euclidean distance	$d=\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
Canberra distance	$d_{ij}=\sum_{i=1}^{n}\frac{\lvert x_{ik}-x_{jk}\rvert}{\lvert x_{ik}\rvert+\lvert x_{jk}\rvert}$
Squared Chord distance	$d_i=\sum_{i=1}^{n}(\sqrt{x_i}-\sqrt{y_i})^2$
Squared Chi-Squared distance	$d=\sum_{i=1}^{n}\frac{(x_i-y_i)^2}{(x_i+y_i)}$

Table 1: Similarity functions.

Function	Accuracy
Manhattan	76%
Euclidian	76%
Canberra	94%
Squared Chord	78%
Squared chi-squared	72%

Table 2: Accuracy rate of similarity functions.

root-mean-square error (RMSE) is a frequently used measure of the differences between values (sample and population values) predicted by a model or an estimator and the values actually observed. The RMSD represents the sample standard deviation of the differences between predicted values and observed values.

$$RMSE(X_1,X_2)=\sqrt{\frac{\sum_{i=1}^{n}(X_{1i},X_{2i})^2}{n}} \quad (2)$$

Where, X_1: observed values, X_2: modelled values.

The following Table 3 helps us estimating error rate percentage for all similarity functions which have been applied in this research. C# is the case number from case-base and it value is its similarity value; we took for every status sample of cases.

The following (Table 4) shows the error rate for all similarity functions when diagnosis case is for diabetic patients. The error rate percentage has been calculated using RMSE formula.

The following (Table 5) shows the error rate for all similarity functions when diagnosis case is for pre-diabetic patients. The error rate percentage has been calculated using RMSE formula.

The following (Table 6) shows error rate for similarity functions when diagnosis is for normal patients. The error rate percentage has been calculated using RMSE formula.

Error rate has been calculated using RMSE formula, the results shown in Figure 4, below mentions that the function which gives minimum error is Canberra. The x axe is the number of tested cases and y axe is the error rate percentage for all similarity functions

Function	chosen	C8	C7	C6	C5	C4	C3	C2	C1	
Manhattan	5	3.28	1.46	7.2	0.24	3.2	1.51	1.56	3.66	**Diabetic**
Euclidean	5	2.3	1.06	5.11	0.2	2.29	1.1	1.15	2.6	
Canberra	5	0.31	0.29	0.75	0.13	0.43	0.37	0.38	0.45	
Squared Chord	5	0.189	0.06	0.61	0.007	0.15	0.05	0.05	0.19	
Squared chi-squared	5	0.37	0.12	1.19	0.015	0.31	0.106	0.103	0.38	
Manhattan	3	1.491	1.726	1.361	1.766	0.85	0.261	1.362	0.751	**Prediabetic**
Euclidean	3	1.0039	1.082	0.892	1.118	0.597	0.177	0.932	0.4495	
Canberra	3	0.00479	0.00536	0.003904	0.00531	0.00148	0.000844	0.00306	0.00295	
Squared Chord	3	0.04876	0.08522	0.03853	0.08627	0.02143	0.001613	0.04628	0.02887	
Squared chi-squared	3	0.0965	0.1649	0.07639	0.1678	0.04273	0.00322	0.09202	0.0567	
Manhattan	5	1.23	3.14	2.64	0.495	0.522	2.33	0.541	0.936	**Normal**
Euclidean	4	0.7638	2.15	1.84	0.347	0.33	1.5	0.367	0.5585	
Canberra	2	0.00459	0.01318	0.0117	0.00192	0.00249	0.00657	0.0017	0.00383	
Squared Chord	2	0.05024	1.128	1.106	0.00897	0.01122	0.1328	0.00673	0.0407	
Squared chi-squared	2	0.0983	1.15	1.11	0.01789	0.0223	0.2592	0.01346	0.07967	

Table 3: Cases similarities to calculate error rate.

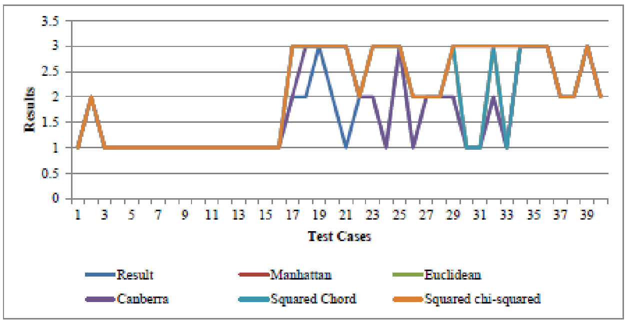

Figure 3: Similarity functions accuracy rate.

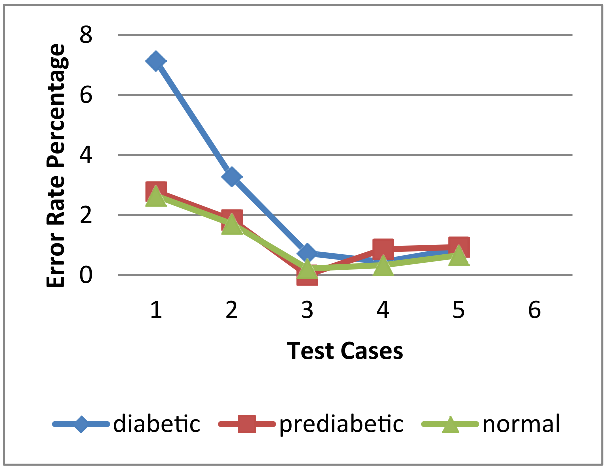

Figure 4: Error rate of similarity functions using RMSE formula.

Function	Error rate
Manhattan	7.138243
Euclidean	3.285647
Canberra	0.731856
Squared Chord	0.441942
Squared chi-squared	0.874691

Table 4: Error rate (diabetic case) calculated using RMSE formula.

Function	Error rate
Manhattan	2.644579362
Euclidean	1.70977712
Canberra	0.219762424
Squared Chord	0.330016281
Squared chi-squared	0.663174237

Table 5: Error rate (Prediabetic case) calculated using RMSE formula.

Function	Error rate
Manhattan	2.783879398
Euclidean	1.844240551
Canberra	0.011447917
Squared Chord	0.859422885
Squared chi-squared	0.934491109

Table 6: Error rate (Normal case) calculated using RMSE formula.

Releted Work

Ambilwade et al. [14] discussed that medical expert systems being used for diabetes diagnosis where the patient's symptoms and other details are inputs and the system diagnose the disease, recommend treatment or drugs which may be prescribed. Artificial neural network used. In this research work, the highest accuracy above 89% is ANN.

Obaloluwa et al. [15] diagnosed diabetes by creating a multilayer feed-forward and trained with back-propagation algorithm which classify patient that are tested positive as binary 1 and patient that are tested negative as binary 0. The use of trained neural network gave recognition rate of 82% on test.

Akteretal [16] have provided a knowledge-based system for diagnosis and management of diabetes mellitus. They believed that preventive care helps in controlling the severity of chronic disease of diabetes. In addition, preventive measures require proper educational awareness and routine health checks. The main purpose of this research was developing a low-cost automated knowledge-based system with easy computer interface. This system performs the diagnostic tasks using rules achieved from medical doctors on the basis of patients' data.

Saeed et al. [17] have provided a rule-based expert system to diagnose all types of diabetes, coded with VP_Expert Shell and tested in Shahid Hasheminezhad Teaching Hospital affiliated to Tehran University of Medical Sciences and final expert system has been presented. The mentioned that some of these patients do not access to the physicians during necessary times. Therefore, such a system can provide necessary information about the indications, diagnosis and primary treatment advices to the diabetics. Since this expert system gathers its knowledge from several medical specialists, the system has a broader scope and can be more helpful to the patients -- in comparison to just one physician.

Conclusion

The success of this work will permit to leverage the development of CBR systems in medicine. It will become possible to develop a web service to federate the CBR process across several domains of medicine. This work will permit patients reuse of CBR systems and develop them. It will also provide the basis for developing a CBR shell for rapid development of CBR systems in medicine.

The above results have provided some indications on the factors affecting the performance of the CBR system, such as the range of values affects the similarity distance between two cases.

Our work will also help new doctor diagnosing this dangerous mellitus, also increase the availability and the number of resources and activities for people with diabetes, their families and other interested parties.

References

1. Begum S, Ahmed MU, Funk P (2009) Case-based systems in health sciences: a case study in the field of stress management. Wseas Transactions on Systems 8: 344-354.
2. Lama P, Zhou X (2013) Autonomic provisioning with self-adaptive neural fuzzy control for percentile-based delay guarantee. Autonomic provisioning with self-adaptive neural fuzzy control for percentile-based delay guarantee. ACM Transactions on Autonomous and Adaptive Systems 8: 121-125.
3. Wang M, Mei W, Jiao J, Jie J, Ma A (2010) Self-Adaptive Mechanism for Software Configuration Based on Case-based Reasoning and Policy. International Conference on Artificial Intelligence and Computational Intelligence.
4. Slade S (1991) Case-based reasoning: A research paradigm. AI Magazine 12: 42-55.

5. Schank RC (1983) Dynamic memory: A theory of reminding and learning in computers and people. Cambridge University Press.
6. Leake DB (1996) CBR in context: The present and future Case-based reasoning experiences lessons & future directions.
7. Aamodt A, Plaza E (1994) Case-based reasoning, Foundational issues methodological variations system approaches. AI Communications 7: 39-59.
8. Kolodner J (1993) Case based reasoning Morgan Kauffman.
9. Geirl L, Rutkowski SS (1994) Integerating consultation and Semi-automatic knowledge acquisition in a proype-based architecture experiences with dysmorphie syndromes. Artif Intell Med 6: 29-49.
10. Lopez B (2013) Case-based reasoning: a concise introduction. Synthesis Lectures on Artificial Intelligence and Machine Learning 7: 100-103.
11. Richter MM (1995) The knowledge contained in similarity measures.
12. Kolodner JL (1996) Making the implicit explicit: Clarifying the principles of case-based reasoning, Case-based Reasoning Experiences lessons and Future Directions.
13. Aamodt A, Plaza E (1994) Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Communications 7: 39-59.
14. Ambilwade RP, Manza RR, Gaikwad BP (2014) Medical expert systems for diabetes diagnosis: a survey. International Journal of Advanced Reserach in Computer Science and Software Engineering 4: 647-652.
15. Olaniyi EO, Adnan K (2014) Onset diabetes diagnosis using artificial neural network. International Journal of Scientific and Engineering Research 5: 754-759.
16. Akter M, Uddin MS, Haque A (2009) Diagnosis and management of diabetes mellitus through a knowledge-based system. In 13th International Conference on Biomedical Engineering.
17. Zeki TS, Malakooti MV, Ataeipoor Y, Tabibi ST (2012) An expert system for diabetes diagnosis. American Academic & Scholarly Research Journal 4: 01-13.