Acta Technica 62 (2017), No. 5B, 803–812

A Diagnosis Method for Elderly Health Assessment Using Case-Based Reasoning

Zhang Lin[2,5], Zhang Jianli[3], Zhang Hao[4]

Abstract. Our country is facing the serious problem of aging population, the research of the elderly health assessment method is very significant for the current situation in China. Therefore, the author will apply the Case -based Reasoning (CBR) method to the elderly health assessment, learn from the professional way to deal with problems, use the case to replace the rules, to avoid the programmer's difficulty of obtaining and expressing the problem of professional knowledge, and use the concept of Inverse Document Frequency (IDF) to design Case organization, Case retrieval and Case retain technology of CBR technology. Design results test shows that the case retrieval accuracy is more than 90%. This means that the design is provides an effective method for the elderly health assessment.

Key words. Case-based reasoning, health assessment, text frequency.

1. Introduction

The sixth national census data in 2010 shows that the elderly people over 60 now in China is about 178 million, 13.26% of the total population, an increase of 2.93 percentage points higher than that in the fifth national census in 2002, which means Chinese aging population problem has become increasingly serious. Therefore, the study of the elderly health assessment problem is becoming urgent and is having practical significance.

[1]Acknowledgment - This research is supported in the collection, analysis and interpretation of data by excellent young talents in colleges and universities support program with No. gxyq2017133, the Colleges and Universities in Anhui Province Natural Science Foundation of Key Projects with No. KJ2016A894, KJ2017A466, the National Natural Science Foundation of China under grant with No. 71331002, 71271072, 71273125 and 71301040.

[2]Workshop 1 - Institute of computer engineering, AnHui SanLian University, HeFei, AnHui,230000, China

[3]Workshop 2 - New Star Institute of Applied Technology NO.451, HeFei, AnHui,230000, China

[4]Workshop 3 - Tongling University TongLing, AnHui,244000, China

[5]Corresponding author: Zhang Lin

Case-based reasoning (CBR) method originated in 1982. Roger Schank's the professor of Yale University said in his book "Dynamic Memory" in 1982: It is a branch of artificial intelligence(AI), which is a kind of AI methods based on empirical knowledge (Case)[1−10]. CBR constructed and formed a rich body of case library, and use the analogy understanding way which is based on case reasoning strategies and mimic human decision-making process in problem solving mechanisms, to effectively solve the problems of unstructured and lack of knowledge.

2. The elderly health assessment method framework based on case-based resoning

The case-based reasoning technology is introduced in the process of the elderly health assessment methods, and the framework based on Case-Based Reasoning is put forward, shown in Figure 1:

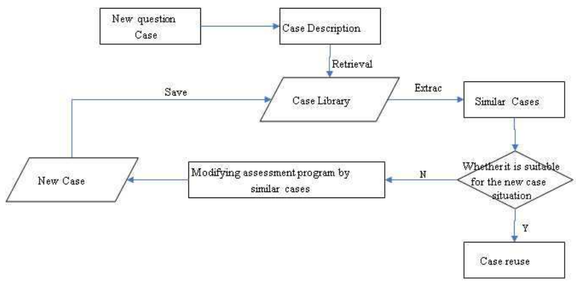

Fig. 1. The elderly health assessment method framework

In the elderly health assessment method based on case-based reasoning, firstly, we should give the case the standard description according to the elderly health assessment index. And then, reflect the health of the elderl in the form of feature vectorsry; finally, search the matched case which is close to the new issues in the case library. If we find the same case or similarity cases within the threshold range, the knowledge of the old case will be directly reused. Otherwise, assessment program will be revised according to the most similar cases to form the new case, which will be saved into the case library.

3. THE KEY TECHNOLOGIES OF ELDERLY HEALTH ASSESSMENT METHOD FRAMEWORK BASED ON CASE-BASED RESONING

Case-based reasoning process includes four key technologies, which are case knowledge indication, case retrieval, case reuse / case modification and case study.

3.1. Case knowledge indication

With the social and economic development and the improvement of physical therapy, medical institutions and local communities at all levels of health services in our country have established a large number of elderly health records. However, because of the local differences, such information isscattered in the time and the space, and it also has the differences of the structure of storage, contents of evaluation features and other aspects. For example, different agencies focus on different health assessment of the elderly. Medical institutions focused on the diagnosis of diseases of the elderly, and community health services have more emphasis on statistic of elder's aspects of mental health and social adaptation. Therefore, it is difficult to compare the different data on the same platform.

Here, we use health indicators statistics table and feature vectors to represent the case of knowledge. Because different agencies have different emphasis, and the data are mostly unstructured. Therefore, we first establish a healthy indicators statistics table, each agency's indicators will be integrated and can be broken down into each evaluated option. For example, Table1 is a community health center's healthy assessment form for the elderly, and Table2 is a hospital's healthy assessment form for the elderly:

Table 1. A community health center's health assessment form for the elderly

1.Marital status	(1)Single (2)Married (3)Remarriage (4)Widowed (5)Other
2.Residence type	(1)Alone (2)Spouses together (3)Children together (4)Spouse and children together (5)Other (specify)
3.Housing type	(1)Building (floor) (2)Lift: Yes No (3)Bungalow (4)Other
……	……
23.Body index	(1)High (2) Higher (3)Normal (4)Lower (5)Low
24.Blood pressure	(1)High (2) Higher (3)Normal (4)Lower (5)Low

Table 2. A hospital's health assessment form for the elderly

1.Body index	(1)High (2) Higher (3)Normal (4)Lower (5)Low
2.Blood pressure	(1)High(2)Higher (3)Normal (4)Lower (5)Low
3.Skin	(1)Moist (2)Dry (3)Rash (4)Nail/Toenail problems (5)Itching (6)Inflammation/Irritation/Ulcers(site) (7)Yellow dye
......	......
9.Respiratory system	(1)Cough (2)Dyspnea (3)Hemoptysis (4)Sputum Chest pain
......	......
25.Circulatory system	(1)Precordial pain (2)Chest tightness, suffocation (3)Arrhythmia (4)Cyanotic (5)Palpitations
26.Digestive system	(1)Loss of appetite(2)Nausea/Vomiting/Vomiting blood (3)Nose/Mouth feeding (4)Abdominal distention (5)Constipation (6)Hematochezia (7)Diarrhea

By integrating and breaking down the two health assessment form's indicators, we can establish health statistics indicators table shown in Table 3:

Table 3. Health statistics indicators table

Index number	Index content
1	Single
2	Married
......	......
6	Live alone
7	Live with spouses together
......	......
117	Skin dry
118	Skin rash
......	......

Assuming that a healthy statistics indicators table has 576 health indicators in all, then each health indicator can be represented by a 576-dimensional feature vector X, X=(x_1, x_2, ..., x_{576}), in the vector X, if the i-th health indicators did not appear then xi=0, otherwise xi=1.

It is easy to find that if the frequency of a health indicators' occurrences in all cases is lower (for example, Hemoptysis, Precordial pain), which means the health indicators for the elderly health assessment is typical. Thus, when making the case retrieval, its weight should be bigger. Conversely, if the frequency of a health indicators' occurrences in all cases is higher(for example, body normal index, skin dry), and these health indicators is difficult to judge the health of the elderly, then its weight to make the case retrieval should be smalerl. So all health indicators which appeared in the cases are set to 1 is unreasonable. This is similar to the information theory Inverse Document Frequency(IDF).

Originally, the similarity between query and web page is calculated by the sum of TF(Term Frequency). Suppose a query contains N keywords, and the TF of them in a specific webpage is TF_1, TF_2,... TF_N, the similarity between this query and this page is:

$$\mathrm{TF_1+TF_2+\ldots\ldots+TF_N} \tag{1}$$

But because the weight of each keyword is different, so when we calculating similarity between query and web page, we will use IDF as weight, so the similarity between query and web page is:

$$\mathrm{TF_1 \bullet IDF_1+TF_2 \bullet IDF_2+\ldots\ldots+TF_N \bullet IDF_N} \tag{2}$$

So-called IDF, in a nutshell, assume that the number of a keyword w's occurrences in a web page is Dw, then the greater Dw is the smaller w's weight is. Vice versa, IDF=log (D/D_w), where D is the number of all pages.

Here, we can learn from the concept of IDF to set the right value of health indicators. Assume that the number of a healthy indicators i's occurrences is D_i, then the weight of Di is log (D/D_i), so D is the total number of the cases in the cases library.

Suppose that there were 100,000 cases in the case library, 50,000 cases among which have skin dry phenomenon, and 800 cases have skin rash phenomenon, namely D=100000, D_{117}=50000, D_{118}=800, then the weight for skin dry index is log (D/D_{117})=log(**??**)=1, and the weight for skin rash index is log(D/D_{118})=log(**??**)=6.96. So, we can get all health indicators' weight, and then we can establish the weighted health statistics indicators table shown in Table4 as follows:

Table 4. Weighted healthy statistics indicators table

Index number	Index content	Weight
1	Single	9.12
2	Married	0.86
......	......	......
6	Live alone	6.34
7	Live with spouses together	1.76
......	......	......
117	Skin dry	1
118	Skin rash	6.96
......	......	......

Then we can obtain the health indicators weight vector Y=(y_1, y_2, ..., y_{576})^T=(9.12, 0.86, ..., 1.72)^T.

Thus, the knowledge of the case can be simply shown by the following 3-tuple: Case = (CA, FV, DT), in which, CA represents the property of classification, including the name, sex, age, the time of establish the case, the geographical area of life and so on, which be used to classify the case (by time, age or geographic classification, etc.); FV represents the healthy indicators' feature vector (the feature vector only have 0 and 1), while carrying case retrieval, we just need to do the matrix multiplication of feature vectors and health indicators' weight vector and constitute a weighted feature vector; DT represents health diagnosis and treatment recommendations, these diagnoses and recommendations can also be used to establish tables through the cases, and each diagnosis and recommendation can be shown by corresponding feature vector.

3.2. Case Retrieval

Case retrieval is the core of CBR, its purpose is to retrieve as few cases as possibel which has reference significance to solve the current problems from a large number of cases, and then as the basis of the problem which is required to solve currently. Common case retrieval strategy are mainly nearest neighbor strategy, induction indexing strategies, knowledge guide strategy and templates retrieval strategy. This paper uses the nearest neighbor strategy, but the similarity is calculated by the law of cosines, not by the Euclidean distance.

During the knowledge representation of the case, we have established the feature vector of health indicators for each case, so we can calculate the angle size between two vectors feature by the law of cosines. Since all weight of health indicators are positive, so the cosine between the two feature vectors is 0-1. If the cosine between the two feature vectors are closer to 1, that namely the angle between two vectors is smaller, thus the two feature vectors indicated by health indicators is closer. Vice versa, the cosine between the two feature vectors are closer to 0, which means the larger the angle between the two feature vectors is, the smaller the velevance of two feature vectors indicated by healthy indicators is.

As we know the cosine of $\angle A$ is:

$$\cos A = \frac{b^2 + c^2 - a^2}{2bc} \tag{3}$$

At this time, if b and c are two vectors starting from A, the above formula can be equivalent to: $\cos A = \frac{<b,c>}{|b|\cdot|c|}$, among which <b,c> represents the inner product of vectors, and |b| and |c| represents the length of vector.

Assuming that the health indicators feature vector of case X is (x_1, x_2, ..., x_{576}), among which xi is 0 or 1, the weighted vector of health indicators is Y = (y_1, y_2, ..., y_{576}) ^{T}, then the weighted feature vector is: (x_1, x_2, ..., x_{576}) $\times$ (y_1, y_2, ..., y_{576}) ^{T} = (x_1y_1, x_2y_2, ..., $x_{576}y_{576}$).

Thus, assuming that the weighted health indicators feature vector of two cases A and B are respectively $(a_1, a_2, \cdots, a_{576})$ and $(b_1, b_2, \cdots, b_{576})$, then the cosine of the angle between A and B is:

$$\cos\theta = \frac{a_1 b_1 + a_2 b_2 + \cdots + a_{576} b_{576}}{\sqrt{a_1^2 + a_2^2 + \cdots + a_{576}^2} \cdot \sqrt{b_1^2 + b_2^2 + \cdots + b_{576}^2}} \tag{4}$$

That means, the smaller the two vectors' value of CosA is, the smaller the similarity degree of the vectors is, on the contrary, the larger the value of CosA is, the closer the two vectors are.

If CosA=1,then,the two vectors completely overlap , that is to say two cases of health indicators are completely the same.

The case retrieval process is shown in Figure 2:

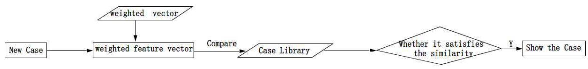

Fig. 2. Case retrieval process

3.3. Case reuse / Case modification

When a new case appeared, we only need to calculate the cosine of new case and each case in the case library, if a , which shows that the current cases and new cases are identical, so the current case can be directly reused. Otherwise, we can descend the order of the case according to the value of $\cos\theta$, and screen the cases which is closer to the new cases, or threshold t, and screened out all cases of $\cos\theta \geq t$ as approximate cases. Because the elderly health assessment gets involve specialized knowledge. Therefore, totally depending on the computer to automatically modify the case is very difficult, which requires manual intervention. However, if DT (health diagnosis and treatment recommendations) in 3-tuple which represents case knowledge is also represented by feature vectors, we can use the DT feature vectors which are screened approximate cases to do Boolean operations "and" first, and to determine the basic treatment recommendations, and then through manual intervention by expert in order to reduce the degree of human intervention.

3.4. Case Study

The results of a new case after a manual intervention was not certainly correct, you need practice to verify, just the new case which is proved to be correct can be added to the case library. When a new case is added to the case library, the weight vector needs to be adjusted, and each weight in weight vector is calculated by log (D/D_i). While re-adjusting the weight vector, due to the addition of a new case, so D'=D+1; if the i-th health indicators appears in the new case, namely $x_i=1$, then D_i'=D_i+1, otherwise D_i'= D_i, thus the new weight of i-th health indicators is log(D'/D_i').

4. Method test

In order to verify the accuracy of the elderly health assessment method based on case-based reasoning put forward in this paper, Visual Studio 2010 is used as the development platform, C# as the development language, and the cases are saved in the SQL Server 2000 to test the accuracy of the method in a simple way. In the process of test, 120 cases are used, including 100 training cases and 20 test cases.

In the test, firstly, according to the training cases and the corresponding evaluation standard, the SQL Server 2000 is used to establish a healthy statistics indicators table which contains 378 attributes (each attribute for a healthy index) and 120 tuples (each tuple is a feature vector X_i) . The type of each healthy indicators in healthy statistics indicators table is Boolean . After establishing the healthy statistics indicators table, the former 100 tuple(training cases) are used as the foundation to calculate the weight of each health indicators in the table, and then a one-dimensional array containing 378 elements is established to hold the weight.

After the preparations all above, we use the later 20 tuple as the test cases, and calculate each vector Angle with the first 100 tuple respectively, and then screen out the cases which meet the threshold.The test shows that the case retrieval accuracy is more than 90%.

5. Conclusion

In this paper, case-based reasoning technology is applied to the elderly healthy assessment to come up with the elderly healthy assessment framework based on case-based reasoning and to provide several key technologies in the reasoning process.

Elderly health assessment method based on case-based reasoning in this paper is easy to understand and is relatively simple to achieve. Through experimental verification, the accuracy is also relatively high. However, the test just verifies the accuracy of the method. As for its effectiveness in practical system applications, it has still been verified, and the key technologies are still have some problems to be solved.

Firstly, each agency's indicators were integrated and then thdy wwere broken down into each evaluated option.This method is simple and is very suitable for most options. However it is not suitable for some indicators with reference data (for example, body temperature, blood pressure, etc.). In these cases, the applicability is not very good. We can not evaluate the indicator through concrete numerical evaluation, but only through a numerical range. For instance, the body temperature can only be evaluated through not fever (36.5-37.2), fever (37.3-38), high fever (>38),but not through specific numerical 37 or 38.2 , so the assessment of the accuracy is not enough.

Furthermore, the proposed method involves the feature vector actually a sparse vector; therefore, in practice, how to make sure the effectiveness and simplify the existing algorithm by sparse matrix algorithms by the same time, improving its efficiency but also one of the direction of the study in the future.

Secondly, when we expressed the knowledge by feature vector, because one index can be decomposed to multiple items, and only on item can be chosen in the same case (for example, Single ??Married??Remarriage and Widowed are all decomposed by marriage status indicators ,only one item can be chosen in a case, the other three cannot be selected at the same time), as a result, the feature vector is actually a sparse vector. In addition,the threshold t which is mentioned in the case reuse technique needs to be set by the professionals , and it will undoubtly to increase the degree of human intervention. Therefore,in the practical application, how to make sure the effectiveness and simplify the existing algorithm at sparse matrix algorithms by the same time, and how to reduce the degree of human intervention, to improve its efficiency are all one of the direction of the study in the future.

References

[1] R. Barletta: *An introduction to case based reasoning.* AI Expert *8* (1991), 43–49.

[2] S. R. Chandler, L. R. Mazlock: *Using existing knowledge sources (cases) to build an expert system.* Expert system *8* (1991), No. 1, 3–10.

[3] J. L. Kolodner: *An introduction to case-based reasoning.* Artificial intelligence review *6* (1992), No. 1, 3–34.

[4] D. Hennessy, D. Hinkle: *Applying case-based reasoning to autoclave loading.* IEEE expert *7* (1992), No. 5, 21–26.

[5] F. Shouri: *An Expert System for predicting gas demand: A case study.* Omega *21* (1993), No. 3, 307–317.

[6] C. Tsatoulis, R. L. Lashyap: *Case-based reasoning and learning in manufacture with the TOL TEC planner.* IEEE Transactions on Systems *23* (1993) 1010–1023.

[7] M. Frank: *Current status of case-retrieval in engineering domains: analysis from the knowledge engineering perspective.* Knowledge-based system *9* (1996), No. 2, 83–91.

[8] S. S. Kyung, H. Ingoo: *Case-based reasoning supported by genetic algorithms for corporate bond rating.* Expert systems with applications *16* (1999), 85–95.

[9] K. Weber, S. Ashley: *Textual case- based reasoning.* The knowledge engineering review *20* (2006), No. 3, 255–260.

[10] P. Enric: *Distributed case-based reasoning.* The Knowledge engineering review *20*, (2006), No. 3, 261–265.

Received November 16, 2017