Collaborative Case-Based Reasoning for Knowledge Discovery of Elders Health Assessment System

Ping Hu¹*, Dong-xiao Gu², and Yu Zhu³

¹Department of Management of Hefei University, HeFei, Anhui 230601, P.R.China
²School of Management at Hefei University of Technology, Hefei, Anhui 230009, P.R.China
³School of Management Engineering at Anhui University of Technology, Wuhu, Anhui 241000, P.R.China

Abstract: The existing Elders Health Assessment (EHA) system based on single-case-library reasoning has low intelligence level, poor coordination, and limited capabilities of assessment decision support. To effectively support knowledge reuse of EHA system, this paper proposes collaborative case reasoning and applies it to the whole knowledge reuse process of EHA system. It proposes a multi-case library reasoning application framework of EHA knowledge reuse system, and studies key techniques such as case representation, case retrieval algorithm, case optimization and correction, and reuse etc.. In the aspect of case representation, XML-based multi-case representation for case organization and storage is applied to facilitate case retrieval and management. In the aspect of retrieval method, Knowledge-Guided Approach with Nearest-Neighbor is proposed. Given the complexity of EHA, Gray Relational Analysis with weighted Euclidean Distance is used to measure the similarity so as to improve case retrieval accuracy.

Keywords: Elders health assessment system, knowledge discovery, knowledge mining, multi-case-based reasoning.

1. INTRODUCTION

Elders Health Assessment (EHA) is the foundation and key of the elders’ health management. Elaborate health assessment not only improves the elders’ health service and care, but also is helpful for the government to optimize the allocation of public health resources and to enhance the social equity [1, 2].

In recent years, many scholars at home and abroad have proposed patterns and methods of Elders Health Assessment. Among them, in the 1980s the Health Management Research Center of University of Michigan began Health Risk Appraisal (HRA) and conducted health assessment [3]. Albaba, Cha et al. [4] studied the elders’ health risk assessment system, which were based on electronic patient records of community elderly apartments. Among the studies of assessment dimensions and programs, American scholar Jr. Ware [5] held that health assessment should consist of at least five dimensions: physical health, mental health, social functioning, role functioning and self-rated health. Rong, Li et al [6] proposed that self-rated health programs should be taken into a comprehensive health assessment system, and China’s fourth National Health Service Survey also contained self-rated health items. Wang Haishun, Li Zhimin, et al [7] built HL7 Protocol-based health assessment system and dynamic generation system on the basis of a disease risk factor model to identify potential disease risk factors with the help of health assessment. Although the above material conducted case studies for their proposed methods, there are still some problems, (1) the current health assessments mostly only consider a certain aspect of elders health, such as certain chronic diseases, mental health of a certain group of people, and there are fewer researches about comprehensive health assessments of sound conditions in physical, psychological, social aspects. And existing assessments are mostly static, and less consider integrity and dynamic nature of elders’ life activity and their health multi-dimension. (2) Some medical and health service agencies have extensive electronic health records or health assessment cases, but current health assessments do not take advantage of these data sufficiently. Meanwhile, the assessment methods are mainly the weighted, fuzzy math, factor analysis, cluster analysis and other conventional statistical methods, and lack of researches and applications of artificial intelligence and data mining.

Patent US 5978785, titled “Object oriented case-based reasoning framework mechanism” [8], Patent US 6701195, titled “Sensor prediction system utilizing case based reasoning”[9] and Patent US 7228329, titled “Methods and systems for providing guidance using case-base reasoning.”[10] invent a new approach called Case-Based Reasoning (CBR) , CBR is an artificial intelligence method based on empirical knowledge (case) to carry on reasoning. It is based on objective cases’ requirements or prompts to retrieve matching cases, and form the solution under its guidance [11]. CBR imitates the analogical understanding in the human decision making process, and is effective to solve problems with unstructured and insufficient knowledge [12].
However, this single-case-library CBR system has low intelligence level, poor coordination, and limited capabilities of assessment decision support, since CBR assessment only considers single-case library reasoning, and is limited within some particular health status (elders’ chronic diseases and psychology). Thus, we propose Collaborative Case-Based Reasoning (CCBR) and apply it to multi-case library resource sharing and knowledge reuse in EHA process. This paper describes cases with XML multi-case representation method, gain knowledge from multi-case library with Knowledge-Guided Approach plus Nearest-Neighbor to provide EHA with auxiliary knowledge support. This article is in the following order: the second part presents a knowledge reuse framework of Elders Health Assessment based on Collaborative Case-Based Reasoning (CCBR-EHA); the third part studies the key techniques of CCBR-EHA, including case representation, retrieval and reuse; the fourth part gives application analysis, and finally makes a conclusion.

2. KNOWLEDGE REUSE FRAMEWORK FOR ELDERS HEALTH ASSESSMENT BASED ON CCBR

Here introduces Collaborative Case-Based Reasoning (CCBR) in Elders Health Assessment (EHA), and proposes knowledge reuse framework of Elders Health Assessment based on Collaborative Case-Based Reasoning (CCBR-EHA).

The basic operating principle of CCBR-EHA is: collaborative cases are used to map the knowledge needs of EHA system, and previous similar assessment programs are optimized and corrected by retrieving, then a new EHA program is drawn, and there is no need to start from the very beginning. The process keeps to the following steps:

Step 1: According to the actual health status of the old person and knowledge needs, define characteristic attributes of a new problem case;

Step 2: Retrieve the case library according to feature attributes, and with CBR and similarity analysis determine the case that is the most similar to the new problem case;

Step 3: Reuse the assessment program that is corresponding to the most similar case as the feasible assessment program;

Step 4: Modify and optimize the assessment program, and examine whether it meets the needs of the new case. If it does not meet the needs, continue to modify and optimize to make it more suitable to the new case;

Step 5: Reproduce the assessment program that meets the requirements and put it into the new case to perfect the new case;

Step 6: Re-examine and evaluate whether the new case is worth saving. If necessary, save it in the case library, and retain the corresponding assessment program;

Step 7: Modify the case indexing and feature weights in the case library.

In the above steps, the new case can be saved in the case library after pre-processing and filtering, and sometimes the previous case and its corresponding assessment program need to be made adaptability revision, add expert knowledge, and organize new cases. It is worth noting that no matter whether case-based reasoning is successful or not, the case indexing scheme and feature weights in the case library will be mended [13].

3. MULTI-CASE KNOWLEDGE REUSE TECHNIQUES IN ELDERS HEALTH ASSESSMENT

3.1. Multi-case Representation

Case representation is the basic work in the knowledge reuse process of EHA, and its representing method is directly related to the efficiency and accuracy of CBR. Since being unstructured has become ubiquitous in case practice and application, in order to represent and store unstructured cases more flexibly, abundantly and completely, this paper uses XML to represent and organize heterogeneous cases in multi-case libraries. Compared with the traditional case representation method, XML-based approach has obvious advantages in describing cases [10], for example, case structure is relatively flexible, case storage is relatively independent, node custom functions in XML documents help to create a new case description and functions, case information is more abundant and complete, heterogeneous case resources can be shared and have cross-platform applications.

In CCBR-EHA, cases are divided into solution case and problem case.

Definition 1 Solution case refers to the case solution that meets the requirements of elders health assessment, and its organization is stored in multiple case bases.

In light of the characteristics of EHA, and according to output-based EHA methods in Bibliography, we represent the solution case of EHA as an eight-feature vector, namely

Case {case-info, elder-info, objective-mix, status-analysis, HA-SWOT, flow-analysis, project-evaluation-analysis, planning-result}

Among them, case-info shows the basic case information, including the case’s number, name, input time and collector, etc.; elder-info is the basic information of the old person’s health status, including the old person’s name, vocation, gender and age, etc.; objective-mix represents the target set, including objectives and development strategies of the health assessment, and the objectives and strategies of the old person’s health assessment, etc.; status-analysis shows the current situation analysis, including the analysis of the old person’s current health status quo, internal and external living environment, access to resources, etc.; HA-SWOT represents SWOT analysis of HA, including the analysis of HA’s strengths, weaknesses, opportunities and threats; flow-analysis represents process analysis and integration, including identifying assessment system function, assessment process analysis and restructuring, identifying key success factors, etc., among which assessment process restructuring includes the process after the restructuring and its mapping table of database files and data structure tables, etc.; project-evaluation-analysis represents program assessment analysis, including HA architecture, the assessment scale, cost analysis, efficiency analysis and
assessment schedule, etc.; at last, planning-result is the old person’s health assessment solution scheme.

XML-based case representation generally uses DTD (Document Type Definition) or XML Scheme to define XML documents’ structure. This paper makes use of DTD to define case structure too. According to the characteristics of XML, DTD-based solution case is expandable, and its structure can make adaptation. For example, the information of a certain tuple can be properly added or deleted according to the case complexity. It is these characteristics of XML knowledge representation that make the case information more abundant and complete, and its structure relatively flexible. In the eight-category information of the solution case, elder-info, objective-mix, status-analysis, HA_SWOT, flow-analysis, project-evaluation-analysis correspond to the case description, each element or sub-element can be selected as characteristic attributes of the case; planning-result corresponds to the case solution.

Definition 2 Problem case is the new case that shows the elders health status quo and knowledge discovery needs, and needs to be solved in CBR-EHA being.

A solution case is already stored in the case library, which has a specific knowledge structure and content description. In order to make the problem case that is formed from elders health status quo and knowledge discovery needs to retrieve and match the similar case (the solution case) in the case library, the problem case and the solution case are required to have the same or similar organizational structure.

3.2. Multi-cases Knowledge Discovery

Case retrieval model and its retrieval strategy have significant impacts on CBR system learning and reasoning performance, whose essence is the similarity matching of the problem case and the solution case [14]. Patent US 7113956, titled “Knowledge management system and method” [15], Patent US 6915297, titled “Automatic knowledge management system” [16] and Patent US 7729924, titled “Virtual knowledge management system” [17] invent a most commonly used method of Nearest-Neighbor Approach, Inductive Algorithm and Knowledge-Guided and so on. According to the development characteristics of health assessment program, this paper uses Knowledge-Guided Approach with Nearest-Neighbor (KGA-NN). However, because case is specific, and difficult to define the key characteristic retrieval attributes, the case retrieval method studied here focuses on retrieval accuracy rather than automatic retrieval efficiency. KGA-NN case retrieval process is shown in Fig. (1). In this method, the retrieval of EHA cases has two stages the preliminary retrieval and optimal case acquisition.

3.2.1. Case Preliminary Retrieval

In this paper, Knowledge-Guided Approach to preliminarily retrieve cases is applied. Knowledge-Guided Approach determines the case retrieval keywords based on the existing knowledge and experience, then retrieves and matches the cases according to these keywords. For example,
health assessment field and health assessment objectives have important influences on EHA program development; thus, the preliminary retrieval can make health assessment field as the first retrieval condition, health assessment objectives are the second retrieval condition. In the preliminary retrieval phase, according to the existing knowledge, experience and assessment needs, select multiple keywords to retrieve the solution cases that are related to the problem cases in the case library until a satisfactory candidate case set is formed, and select the main characteristic attributes to prepare for the next retrieval phase.

3.2.2. Optimal Case Acquisition

The core of the optimal matching retrieval is similarity calculation. Assume Domain \( C^* \) is the set of all solution cases in the case library, \( c_0 \) is inputted problem case to be solved, \( C \) is the set of candidate cases from the preliminary retrieval. Then, \( C \subseteq C^* \)

\[
C = \{case1, case2, \cdots, case_m\} = (c_1, c_2, \cdots, c_m)
\]

Assume \( A \) is the collection of selected characteristic attributes,

\[
A = \{attribute1, attribute2, \cdots, attributen\} = (a_1, a_2, \cdots, a_n)
\]

Firstly, qualify the qualitative attribute value by using elements \((c_j, a_j)\), i.e. the \( j \) attribute value of Case \( c_j (i \leq m) \) is \( v_{ij} \), then

\[
V = (v_{ij})_{m \times n}
\]

In order to eliminate the influence of different dimensional attribute values, \( V \)'s attribute value \( v_j \) is normalized, and a new matrix is obtained

\[
V' = (v'_{ij})_{m \times n}
\]

Considering the complexity of EHA, most knowledge needs only select finite characteristic attributes to analyze, and CCBR system has incomplete information, or “gray” feature [18]. This paper uses Gray Relational Analysis with Weighted Euclidean Distance to calculate the similarity.

In the case set \( C \), the association coefficient between various cases \( c_i (1 \leq i \leq m) \) that have attributes \( a_j (1 \leq j \leq n) \) and the problem case \( c_0 \) is

\[
\xi_{ij} = \frac{\min_{i} \min_{j} |v'_{0j} - v'_{ij}| + \rho \max_{i} \max_{j} |v'_{0j} - v'_{ij}|}{|v'_{0j} - v'_{ij}| + \rho \max_{i} \max_{j} |v'_{0j} - v'_{ij}|}
\]

In this formula, \( v'_{0j} \) is the problem case \( c_0 \)'s normalized attribute value when in the \( j \)th point, \( 1 \leq i \leq m, 1 \leq j \leq n \); \( \rho \) is the resolution ratio, \( 0 \leq \rho \leq 1 \), its value can be set to meet the needs, usually \( \rho \approx 0.5 \). Thus, \( 0 \leq \xi_{ij} \leq 1 \).

Formula (5) determines the gray correlation coefficient matrix, but each characteristic attribute does not have the same importance degree; therefore, characteristic attribute weights must be taken into account. Combined with the specific operating characteristics of the system, Analytic Hierarchy Process is used to determine the normalized characteristic attribute weight vector.

\[
w = (w_1, w_2, \cdots, w_j, \cdots, w_n)^T, \quad \sum_{j=1}^{n} w_j = 1.
\]

The generalized weighted distance between the problem case \( c_0 \) and the solution case \( c_i \)

\[
d_{0i} = \sqrt{r \sum_{j=1}^{n} [w_j (1 - \xi_{ij})]^2}
\]

In the formula, \( r \), the distance parameter, when \( r = 1 \), is the Hamming distance; when \( r = 2 \), is the Euclidean distance, then the similarity between \( c_0 \) and \( c_i \) is defined as the difference between \( 1 \) and Euclidean distance \( (r = 2) \).

\[
sim_{0i} = 1 - d_{0i}
\]

According to the physical meaning of Formula (7), the larger \( \sim_{0i} \) value is, the more similar the corresponding case is with the problem case \( c_0 \). Therefore, \( \sim_{0i} \) is compared \((1 \leq i \leq m)\) to obtain the case as the optimal matching case, which has the largest similarity. In practice, whether the optimal matching case’s assessment program can be the solution to the problem case should be determined according to the set threshold value. Only when the similarity is greater than the predetermined threshold value, can the corresponding case’s assessment program be used as a possible solution program. Another way is to rank similar cases in light of their similarities, then select the optimal matching case from the top-ranking cases according to the elders health needs.

It is necessary to note that to effectively retrieve, there is need to determine the initial weights of case attributes at first. In the retrieval process, if the weights of case characteristic attributes can not be correctly valued, the case retrieval quality will not be guaranteed, and then there is need to adjust and optimize the weights. In EHA retrieval, here classic Delphi method is used to determine the initial weights. Follow system procedures, take anonymous expert comments on EHA, and gain the attributes’ initial weights of EHA cases from repeated questionnaires and common views from people who fill out questionnaires. In the retrieval process, Tabu Retrieval Genetic Algorithm proposed in Bibliography [13] is used to optimize the weights of case characteristic attributes. Self-adaptation and memory function of Tabu Retrieval Genetic Algorithm are used to
3.3. Multi-case Knowledge Reuse

In CCBR-EHA, if the similarity is retrieved as \( \text{sim}_{10} = 1 \), it is basically considered that the solution case’s assessment program meets the knowledge discovery needs of EHA, and can be directly reused. However, in real-world, usually \( 0 \leq \text{sim}_{10} < 1 \), it is necessary to adjust, modify and optimize the solution case’s assessment program gained from the retrieval to fit the knowledge discovery needs of EHA. Because the knowledge involved in the case revision is mostly the tacit expert knowledge, the complete revision automatically made by the system will be very difficult. Therefore, in CCBR-EHA, user intervention is used to make the revision, that is, according to elders’ knowledge discovery needs and the case advice, to modify the entire assessment program. New knowledge can be added in the assessment program, the knowledge that does not meet elders’ knowledge discovery needs can be deleted, biased knowledge can be revised, or a part of knowledge of the assessment program can be replaced with new knowledge, and so on.

The revised assessment program is integrated with the previous description of the problem case, and a new case is reproduced and evaluated whether it is worth saving or not. If there is need to save, the new case is saved as new knowledge in the case library to expand problem solving capacity. In this way, the case study is completed.

4. EXPERIMENT AND APPLICATION ANALYSIS

In order to verify the validity of the proposed method, KGA-NN, JAVA, XML, SQL Server 2000 and DBMS are used as development tools to achieve a Case Base-EHA prototype, which is combined KGA-NN for EHA application. Before Case Base-EHA is practically applied, EHA cases need to be abundantly collected, cases need to be described according to DTD specifications of XML, and stored in multiple heterogeneous case libraries. In Case Base-EHA, to solve the new problem case, at first make use of the defined XML elements and the structure among elements (DTD defined structure among the elements). Then according to elders’ health status quo and assessment needs, make the preliminary retrieval with XML parser to get the candidate case set, and select main characteristic attributes. Then according to main characteristic attributes, calculate the similarity between the candidate case set and the problem case to get the solution case number. From the most similar case number, get the corresponding EHA program. By learning, it is easy to extract related knowledge from the knowledge library or the experts who have the knowledge to form the assessment program, then modify and optimize the EHA solution scheme with the knowledge in the knowledge library. Re-examine and evaluate the optimized case to see whether it is worth saving or not. If it is worthy saving, the optimized case needs to be added into the case library as a new case. In this way, the case study is completed.

The initial threshold value and the initial weights can be obtained with Delphi method by domain experts. To make the expert advice more reasonable and more suitable for the system and the experimental requirements, to get the optimal initial threshold value and the initial weights, a training library is added up to Case Base-EHA system design, in which there are 124 programs to let the experts operate. In Case Base-EHA system, the experts use the data in the training library to have repeated experiments, and then give the threshold value and weights. After several times of aggregation and adjustment, ultimately the initial threshold value is determined \( \lambda_0 = 0.72 \), the initial weight set \( W_0 = (0.03, 0.07, 0.13, 0.14, 0.19, 0.33, 0.03, 0.08) \). Next, 120 EHA programs are used as experiment cases. As the initial case library, 20, 40, 60, 80, 100, 120 cases are respectively put in six sub-systems of Case Base-EHA system to have the retrieval experiment for a hospital EHA program. Each characteristic attribute value is inputted to have the retrieval.

![Fig. (2). Results OF KGA-NN method performance validation.](image)
The retrieved results of six sub-case libraries and the logarithmic graph trend are shown in Fig. (2). The horizontal axis is sub-case library number and case library data volume N (Unit: item), and the vertical axis is the obtained results (item) / maximum similarity value / retrieval time (Unit: second).

As the above figure shows, KGA-NN method can effectively obtain the knowledge of EHA system. With the increase of the data in the case library, the retrieval accuracy is gradually increased (the retrieved maximum similarity value is monotonically increasing), the overall retrieval time is short and remains relatively stable though the data is constantly increasing, which reflects good retrieval efficiency.

The following part is the aided design application of a community EHA program in Case Base-EHA platform. In Case Base-EHA platform, input each problem case’s characteristic attribute values to solve the problem. The obtained results are shown in Table 1.

The community program is integrated with Programs 0012 and 0020, and mildly modified. The community EHA program is obtained in a relatively short time, which greatly reduces the assessment time.

CURRENT & FUTURE DEVELOPMENTS

This paper applies multi-case-based reasoning in EHA and knowledge reuse, puts forward the application framework of EHA knowledge reuse system, investigates case representation, retrieval, optimization and correction, reuse and other key techniques. Because case representation is the basis of CCBR, this paper uses XML-based multi-case semi-structured knowledge representation to describe and organize cases, and proposes specific XML-based case-oriented approach. On this basis, this paper uses Knowledge-Guided Approach with Nearest-Neighbor to retrieve and match cases. In light of the complexity of EHA, Gray Relational Analysis with weighted Euclidean Distance is used to calculate similarity; thus, the quality of multi-case retrieval is guaranteed. Finally the application is introduced. As a preliminary research result, this material has some problems, for example, the prototype system is too simple, the data is small and etc.. Future researches will focus on developing well-functioned CCBR-EHA experimental system and make the retrieval experiment with massive data.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This research is partially supported by the MOE Project of Humanities and Social Science (11YJC630248), Anhui Provincial Soft Science Foundation (12020503037), Anhui Provincial Key Research Institute Project of Humanities and Social Science (SK2013A148), Hefei University of Technology (HFT) Young Faculty Members Innovation Foundation(Feed Fund Program) (2013HGQC0026), as well as HFT Doctoral Science and Research Initiating Fund (2012HGBZ0186). Specially, we are grateful to the senior editor, associate editor, and the anonymous reviewers whose comments have improved this paper considerably.

REFERENCES

[1] A. Y. Lai, H. Ishikawa, T. Kiuchi, N. Mookipil, K. Griva, “Communicative and critical health literacy, and self-management behaviors in end-stage renal disease patients with diabetes on hemodialysis”, Patient Educ. Couns., (In Press), Available online 26 January 2013, pp.189-198.
[2] S. P. Hooker, S. Wilcox, E. L. Burroughs, C. E. Rheuma, W. Courtenay, “The potential influence of masculine identity on health-improving behavior in midlife and older African American men”, J. Men's Health, vol. 9, no. 2, pp.79-88, 2012.
[3] C. Yan, Ai Ding dun, “Development and implication of american health risk evaluation”, J. Chin. Health Manag., vol. 3, no. 4, pp.238-241, 2009.
[4] M. Albaba, S. S. Cha, P. Y. Takahashi, “The elders risk assessment index, an electronic administrative database-derived frailty index, can identify risk of hip fracture in a cohort of community-dwelling adults”, Mayo Clinic Proc., vol. 87, no. 7, pp. 652-658, 2012.
[5] J. E. Jr Ware, “Standards for validating health measures: definition and content”, J. Chronic Dis., vol. 40, no. 6, pp.473-480, 1987.
[6] R. Peng, L. Li Q. He, “Self-rated health status transition and long-term care need of the oldest chinese”, Health Policy, vol. 97, no. 2-3, pp. 259-266, 2010.
[7] X. Xu, G. Gao, “Research and implement of case-based reasoning in a multi-criteria evaluation IDSS”, Comp. Integr. Manuf. Sys., vol. 7, no. 1, pp. 16-28, 2001.
[8] J. V. Mark, “Object oriented case-based reasoning framework mechanism”, U.S. Patent 5978785, December 02, 1999.
[9] B. Michael, H. Lueder, "Sensor prediction system utilizing case based reasoning", U.S. Patent 6701195, March 02, 2004
[10] K. Matthias, “Methods and systems for providing guidance using case-based reasoning”, U.S. Patent 7228329, April 05, 2007.
[11] J. Kolodner, "Case-based Reasoning", Morgan Kaufmann Publishers, pp. 100-112, 1993.
[12] D. X. Gu, C. Y. Liang, "Intelligent technique for knowledge reuse of dental medical records based on case-based reasoning" J. Med. Sys., vol. 34, no. 2, pp. 213-222, 2010.
[13] J. Bajo, J. F. Paz, “Integrating case-based planning and RTW neural networks to construct an intelligent environment for health care”, Expert Syst. with Appl.: An Int. J., vol. 36, no. 3, pp. 5844-5858, 2009.
[14] J. Wang, X. Pan, J. Li, L. Lu, “Knowledge management consultation system based on CBR technologies”, J. Tsinghua Univ. (Sci. and Tech.), vol. 46, no. SL, pp.990-995, 2006.
[15] R. E. Elias, J. D Sipila, A. Young, "Knowledge management system and method,” U.S. Patent 7113956, September 26, 2006.
[16] C. Peilin, "Automatic knowledge management system", U.S. Patent 6915297, July 05, 2005

[17] S. Hemant, "Virtual knowledge management system", U.S. Patent 7729924, June 01, 2010

[18] C-Y. Liang, D-X. Gu, X. Fan, W. Chen, “Research of case retrieval algorithm for decision making with uncertain multiple attributes”. Chin. J. Manag. Sci., vol. 17, no. 1, pp.131-137, 2009.

© Hu et al.; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.