

Case-Based Reasoning for Selecting Study Program in Senior High School

Sri Mulyana, Sri Hartati, Retantyo Wardoyo, Edi Winarko

Department of Computer Sciences and Electronics
Gadjah Mada University
Yogyakarta, Indonesia

Abstract—One of the reasoning methods in expert system is Case-Based Reasoning (CBR). A problem is searching for past cases in the case base with the highest similarity degree. This implies that calculation of similarity degree among the cases is an important aspect in CBR. In this study, an application of computer reasoning system based on CBR is developed for selecting study program in Senior High School (SHS). This application can be used to assist students for selecting study program. The cases used in the study include results of the intelligence test, student's interest, and grades of several subjects. Each case in the case base will be calculated for the similarity degree with new cases entered. Furthermore, it is the cases with highest similarity degree that are recommended as solutions.

Keywords—Case Based Reasoning; Case retrieval; similarity degree; new cases; recommended solutions; selecting study program

I. INTRODUCTION

Case-Based Reasoning (CBR) is a method adopted from knowledge-based system in many domains. CBR uses experiences from previous similar cases to solve new problems. The basic idea of CBR is an assumption that similar problems have the similar solutions.

CBR consists of four basic steps, namely:

- 1) Retrieving the most similar case or cases
- 2) Reusing the information and knowledge in those cases to solve the new problem
- 3) Revising the proposed solution
- 4) Retaining the part of the current experience that is likely to be useful for future problem solving

A new problem is solved by retrieving one or more previously experienced cases, reusing the case in one way or another, revising the solution based on reusing a previous case, and retaining the new experience by incorporating it into the existing knowledge-base. The relationship between these steps is illustrated in Figure-1.

One of the important steps in solving a CBR problem is case retrieval. In case retrieval, similarity degree among the cases becomes the basis for determining a case from the case base. The higher the similarity degree among the cases in the case base with the new cases, the higher the possibility for the solution in the cases to be used as solution for the new cases.

The similarity degree among the cases greatly affects the performance of CBR system, because the solutions in previous

cases can be used as references for solving problem in new cases. Many researchers have used various algorithms in case retrieval. If many cases with different attributes are available in the case base, it will surely cause the retrieval system to face difficulties in finding a relevant case.

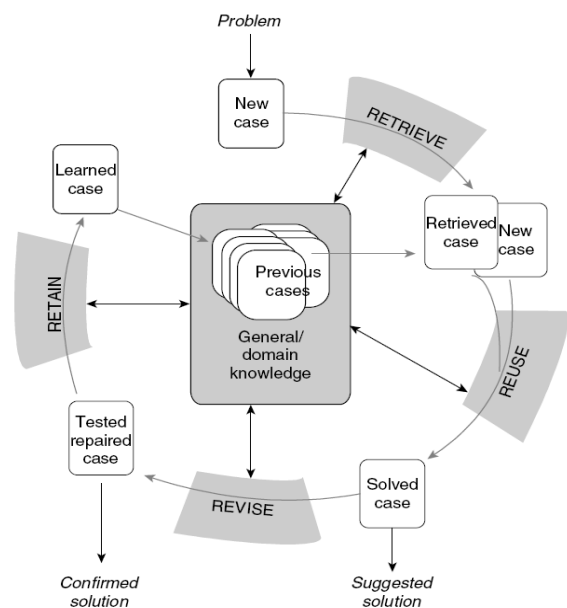


Fig.1. A cycle of Case-based Reasoning[1]

The problem with measuring the case similarity degree is essentially of how to look for similarity between two objects, and it was one of the research focuses in CBR system. The retrieval of similar case with similarity to new cases is an important step of the whole decision making process [2]. In the process, the two objects are measured for similarity using predetermined parameters. However, measuring a similarity degree between the two objects often requires a complex calculation, so the similarity between the two objects can give advantages to the whole problem solving process.

Other problem in CBR is to determine the features of a case to be compared. Determining the features is often very difficult. The explanation of the case is often incomplete such that the resulting compared features are also unable to help in finding the similarity between the two cases compared.

Many approaches can be used in mapping the similarity between two objects. For example, two cases can be

represented as vectors by determining case attribute pairs. The similarity of attribute pairs is usually called as mean weight. The weight allows the attributes to have various levels of importance that can be used in an adaptive learning process.

Retrieval approach is based on surface features and the similarity of each case with the targeted problem is usually represented as real number in the interval (0,1), which is calculated according to the given similarity measure. Usually the case taken is the case most similar to the targeted problem. Such approach is frequently called *K-Nearest Neighbours* (K-NN) approach. CBR system can assure that the retrieval of case most similar to the target problem is to calculate the similarity degree between the targeted problem and each case in the case base [3].

In this study, a system based on CBR for selecting study program in Senior High School (SHS) is developed. This application can be used to assist students for selecting study program when they are starting the second year. The cases used in the study include results of the intelligence test, student's interest, and grades of several subjects. Each case in the case base will be calculated for the similarity degree with new cases entered. Furthermore, it is the cases with highest similarity degree that are recommended as solutions.

II. SOME RESEARCHES ON CBR

The application of CBR method has been developed in various fields, such as CBR for diagnosing infectious diseases [4]. In this study, it is explained that a case has attributes such as temperature, dizziness symptoms including its intensity and area, cough, urination intensity, vomiting and urinating. All these attributes have numeric values from 1 to 4 to show the levels of intensity, i.e. never, low, moderate and high. The study used the Euclidean distance formula to calculate the case similarity degree.

A CBR mechanism for designing special equipment has also been developed using the Euclidean distance formula to calculate the case similarity degree [5]. The formula has also been implemented in other studies such as: CBR extended to the intelligent system configuration [6], a CBR-based software development cost evaluation model [7], quantitative analysis for nonlinear data performance system using CBR [8], and an intelligent decision in GIS- and CBR-based emergence [9]. From the studies, the Euclidean distance method was applied well, because the values of all the attributes were numerical in nature.

In addition to the Euclidean distance formula, there are other methods for determining the case similarity degree, i.e. by hamming distance. This method is also used for the attributes with numerical values. Several researches on the application of CBR using the hamming distance method to calculate the case similarity degree include the a human behavior analysis using CBR [10], recommendation of XML document using CBR [11], the technical application of CBR in a decision support system for using a new energy in rural areas [12].

The method for measuring the similarity degree using the concept of fuzzy set was implemented for selecting optimum

vendors [13]. In this study, the criteria for vendor assessment include: quality, price, service, skill, and reputation. Each of the criteria has sub-criteria with its membership function. An evaluation of the calculation of similarity weight in CBR using a sensitivity analysis was done by Mianghai and Huanmin [14] and implemented in a case of the selection of camera model [15].

III. CALCULATING THE CASE SIMILARITY DEGREE

This research was conducted using a sensitivity analysis on the formulation of method for calculating the case similarity degree. The whole case similarity was calculated by using a synthetic evaluation model, i.e. a mathematical model among values of attributes and their weight coefficients. As an illustration, it is explained as follows:

Let $X(x_1, x_2, x_3, \dots, x_m)$ is input case given by user, $Y(y_1, y_2, y_3, \dots, y_m)$ is a case in the case base, and $W(w_1, w_2, w_3, \dots, w_m)$ is weight coefficient from the attributes of case that satisfies $0 \leq w_i \leq 1$ ($i = 1, 2, 3, \dots, m$) and $\sum_{i=1}^m w_i = 1$.

$Sim(x_i, y_i)$ represents a function of similarity between the i^{th} attribute of Y and the i^{th} attribute of X with real number value in interval (0,1). Assume that each of the attributes is given the weight w_i , the similarity degree between the input case and the targeted case is given by $Sim(X, Y)$, which is formulated by :

$$Sim(X, Y) = \sum_{i=1}^m w_i \times Sim(x_i, y_i) \quad (1)$$

Furthermore, the calculation of $Sim(x_i, y_i)$ was carried out according to the types of case attribute value as follows:

1) *The attribute value in the form of symbols (figures, string):*

$$Sim(a, b) = \begin{cases} 1, & a = b \\ 0, & a \neq b \end{cases} \quad (2)$$

2) *The attribute value in numerical form :*

$$Sim(a, b) = 1 - \frac{|b - a|}{B - A} \quad a, b \in [A, B] \quad (3)$$

where A and B were the lower and upper bound of interval and $A \neq B$.

3) *The attribute value in interval form :*

4) *To calculate the case similarity degree the attributes with interval values are formulated as follows:*

$$Sim([b_1, b_2], a) = \int_{b_1}^{b_2} Sim(a, x) f(x) dx \quad a, b_1, b_2 \in [A, B] \quad (4)$$

where A and B are the lower and upper bound of interval that contain a, b_1 , and b_2 . $Sim(a, b)$ can be calculated based on the equation (3), while $\int f(x)$ shows the probability value

of x at interval (b_1, b_2) that meets $\int_{b_1}^{b_2} f(x) dx = 1$.

Generally, $f(x)$ value is given based on statistical data or expert's authority. If the probability of each attribute value is the same at the interval, then $f(x) = \frac{1}{b_2 - b_1}$.

IV. THE SELECTION OF STUDY PROGRAM IN SHS

The length of education in SHS is for three years. It has three study programs, i.e. Natural Sciences (A₁ and A₂), Social Sciences (A₃) and Language (A₄). Students will select one of the desired study program, but to select it a set of tools is used such as intelligence test results, student interest, and academic competency shown by grades. The process was carried out when students are starting the second year.

The intelligence test was carried out in collaboration with a psychological consulting agency, which include the following aspects:

- a) *General Intelligence (IU), i.e. someone's intelligence level to solve the general problem.*
- b) *Verbal (KV), i.e. someone's competence to speak fluently.*
- c) *Numeric (KB), i.e. some one's competence in numerical and calculating orders.*
- d) *Logic (L), i.e. someone's competence to logically think by using symbols.*
- e) *Mechanic (M), i.e. someone's competence in mechanics.*
- f) *Nonverbal (KNV), i.e. someone's non-lingual thinking competence.*

g) *Interest, i.e. someone's propensity in Natural Sciences, Social Sciences or Language.*

Grades are student's academic achievement values at the semester when the selection of study program was done. For the matching of study program and academic competence, grades for the following subjects were considered:

- a) *Indonesian (bind)*
- b) *English (bing)*
- c) *Mathematics (mat)*
- d) *Physics (fis)*
- e) *Biology (bio)*
- f) *Chemistry (kim)*
- g) *Pancasila Moral Education (pmp)*
- h) *Economics (eko)*
- i) *History (sej)*

Furthermore, a case will be stored in the recorded data structure, which contains the following features:

(iu, kv, kb, l, m, knv, minat, bind, bing, mat, fis, bio, kim, pmp, eko, sej, jur)

V. METHODOLOGY

The procedure applied in this study was to collect data of cases obtained from the records of student's intelligence test results and grades. The data were then stored as a set of cases in the case base. Several case bases were shown in Table-I.

TABLE I. THE CASE BASE FOR SELECTING STUDY PROGRAM INSHS

No	IU	Kv	Kb	L	M	Knv	Minat	Bind	Bing	Mat	Fis	Bio	Kim	Pmp	Eko	Sej	Jur
1	114	65	68	61	68	68	IPA	6	6	6	6	7	6	7	6	7	A3
2	103	76	60	36	79	65	IPA	6	5	6	5	6	6	7	7	7	A3
3	117	76	66	73	75	66	IPA	7	6	7	6	7	7	7	6	6	A2
4	108	54	70	93	73	70	IPA	7	6	6	6	7	7	6	7	6	A2
5	106	54	66	67	64	66	IPA	6	6	5	6	6	7	7	6	6	A3
6	110	54	62	93	58	62	IPA	7	7	7	8	7	7	7	8	7	A3
7	121	76	68	104	87	68	IPA	7	8	7	8	7	7	6	7	6	A1
8	112	65	64	89	70	64	IPA	6	6	6	6	6	7	6	6	7	A2
9	114	60	56	67	54	66	IPA	7	6	6	6	7	6	7	6	6	A3
10	119	92	86	84	64	86	IPA	6	7	7	6	6	7	7	6	6	A1
..
205	100	48	54	67	66	54	IPA	6	7	5	4	6	6	7	6	7	A3

The next step was to select the most appropriate case by applying an algorithm to measure the case similarity degree. From the features of case in the case base, it can be stated that only features of interest had symbol (string) values, while others had numerical values. Each case in the case base will be calculated for the similarity degree with new cases entered. Furthermore, it is the cases with highest similarity degree that are recommended as solutions.

VI. RESULTS AND DISCUSSION

In this study, the case base containing previous cases stored as described above was used. The following were the procedures applied and the test results:

The first step was to provide a new case input in the form as shown in Figure- 2.

Fig.2. Case Input Form

For example, a new case was given, i.e. a student with the following grades:

- Iu = 107
- Kv = 67
- Kb = 89
- L = 78
- M = 65
- Knv = 48
- Indonesian = 7
- English = 6
- Mathematics = 7
- Physics = 6
- Biology = 7
- Chemistry = 7
- PMP = 6
- Economics = 7

The results of the new case input with the above data were shown in Figure- 3.

After completing the data of student interest, it will give the results as shown in Figure- 4.

Figure- 4 shows that students with intelligence test results, interest and grades as previously entered were recommended to select Study program A1.

Fig.3. Example of the cases entered

Fig.4. Results of study programselection

The result was obtained by calculating the similarity degree between case input and all the cases in the case base. The results from a calculation of the case similarity degree were presented in table-II asfollow:

TABLE II. CASE SIMILARITY VALUES

No	IU	Kv	Kb	L	M	Knv	Minat	Bind	Bing	Mat	Fis	Bio	Kim	Pmp	Eko	Sej	Jur	Sim
1	114	65	68	61	68	68	IPA	6	6	6	6	7	6	7	6	7	A3	0,823
2	103	76	60	36	79	65	IPA	6	5	6	5	6	6	7	7	7	A3	0,753
3	117	76	66	73	75	66	IPA	7	6	7	6	7	7	7	6	6	A2	0,861
4	108	54	70	93	73	70	IPA	7	6	6	6	7	7	6	7	6	A2	0,900
5	106	54	66	67	64	66	IPA	6	6	5	6	6	7	7	6	6	A3	0,789
6	110	54	62	93	58	62	IPA	7	7	7	8	7	7	7	8	7	A3	0,826
7	121	76	68	104	87	68	IPA	7	8	7	8	7	7	6	7	6	A1	0,809
8	112	65	64	89	70	64	IPA	6	6	6	6	6	7	6	6	7	A2	0,854
9	114	60	56	67	54	66	IPA	7	6	6	6	7	6	7	6	6	A3	0,822
10	119	92	86	84	64	86	IPA	6	7	7	6	6	7	7	6	6	A1	0,773
..
67	117	60	84	64	75	48	IPA	7	6	7	6	7	6	7	7	7	A1	0,909
..
205	100	48	54	67	66	54	IPA	6	7	5	4	6	6	7	6	7	A3	0,724

VII. CONCLUSION

In this study, CBR method was implemented for selecting study program in SHS. By providing the new case in the form of intelligence test results, student interest, and grades, the

system could provide the recommended choice of study program. These results were based on the calculation of the similarity degree between input case and the cases from the case base with highest similarity degree.

REFERENCES

- [1] A. Aamold dan E. Plaza, "Case-based Reasoning: foundation issues, methodological variation and System approach", *AI Communication* 7(1), pp. 39-59, 1994
- [2] D. Leake, "Case-Based Reasoning : Experiences, Lessons, and Future Directions", Menlo Park : AAAI Press/MIT Press, 1996
- [3] R. L. Mantaras, D. Mcsherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Falting, M. L. Maher, M. T. Cox, K. Forbus, M. Keane, A. Aamodt, I. Watson, "Retrieval, reuse, revision and retention in case-based reasoning", *The Knowledge Engineering Review*, Vol. 20:3, pp. 215-240, Cambridge University Press, United Kingdom, 2006
- [4] M. Denis, and A. Jasmin, "Applying Case-based reasoning for mobile support in diagnosing infective diseases", *International Conference on Signal Processing Systems*, IEEE Computer Society. Pp 779-783, 2009
- [5] W. Chen, P. Lou, Z. Shen, "Case-based reasoning and Intelligent Variation Approach in Fixture Design", *Second International Symposium on Intelligent Information Technology Application*. IEEE Computer Society, pp 833-837, 2008
- [6] H. F. Jun, C. S. Ying, X. Zhen, S. Ping, "Extended Case-based reasoning for Intelligent System Configuration", *Second International Conference on Computer Modeling and Simulation*. IEEE Computer Society, pp 307-309, 2010
- [7] G. Jiang, Y. Wang, H. Liu, "Research on Software Cost Evaluation Model based on Case-based Reasoning", *Second WRI World Congress on Software Engineering*. IEEE Computer Society, pp 338-341, 2010
- [8] J. W. Keung, and T. Nguyen, "Quantitative Analysis for Non-linear System Performance Data using Case-based Reasoning", *Asia Pacific Software Engineering Conference*. IEEE Computer Society, pp 346-355, 2010
- [9] L. Z. Wei and L. Li, "Research on Emergency Intelligent Decision Based on GIS and Case-based Reasoning", *International Conference on System Science, Engineering Design and Manufacturing Information*. IEEE Computer Society, pp 231-234, 2010
- [10] W. Boehmer, "Analyzing Human Behavior using Case-based Reasoning with help of Forensic Question", *24th International Conference on Advanced Information Networking and Application*. IEEE Computer Society, pp 1189-1194, 2010
- [11] S. Vacharaskunee and S. Intakosum, "XML Document Recommendation by Using Case-Based Reasoning", *11th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing*. IEEE Computer Society, pp 121-126, 2011
- [12] Q. F. Sun, L. Tan, L. Wang , C. C. Yu, "Application of Case-Based Reasoning Technology in Decision-Making System of Rural New Energy Utilization", *3rd International Conference on Information Management, Innovation Management and Industrial Engineering*. IEEE Computer Society, pp 235-239, 2010
- [13] F. Na, W. Jinmin, L. Ping, M. Suchang, "Optimal Vendor Selection Using Fuzzy Case Based Reasoning", *Fourth International Conference on Intelligent Computation Technology and Automation*. IEEE Computer Society, pp 1082-1084, 2011
- [14] Y. Mianghai and X. Huanmin, "The Weight Calculation in the Case-Based Reasoning Based on Sensitivity Analysis", *International Conference on Electrical and Control Engineering*, IEEE Computer Society, pp 3119-3121, 2010
- [15] H. Shi, M. Xin, W. Dong, "A Kind of Case Similarity Evaluation Model Based on Case-based Reasoning", *IEEE International Conference on Internet of Things, and Cyber, Physical and Social Computing*. IEEE Computer Society, pp 453-457, 2011