Design and Realization of Case-indexing Model Based on AHP

Fugui Chen, Baojian Zhang, Linfeng Bai
School of Information Engineering, Henan Institute of Science and Technology
Xinxiang, China
E-mail: Cfg7710@hist.edu.cn

***Abstract*—The importance of case-indexing method selection in the CBR system is analyzed, and the defaults of traditional case indexing methods are pointed out. A model of case-retrieving based on AHP is presented and the basic principle and process of CBR and AHP are introduced. With the development environment of CGI, the prototype system is developed, and the core source code and running interface are shown to confirm the effectivity and feasibility of this model.**

***Index Terms*—case-based reasoning, analytic hierarchy process, cases indexing**

I. Introduction

CBR, Case-based Reasoning is the process of solving new problems based on the solutions of similar past problems. It is a popular method in artificial intelligence because it is very simple and reasonable. Especially, in dealing with complex issues of multi-attribute decision-making, CBR is often the preferred method. CBR is not only a powerful method for computer reasoning, but also a pervasive behavior in everyday human problem solving; all reasoning is based on past cases personally experienced. In fact, CBR has many advantages, the most important one is that it can simulate the human thinking to solve problem and make decision. The principle figure of CBR is shown as Figure.1:

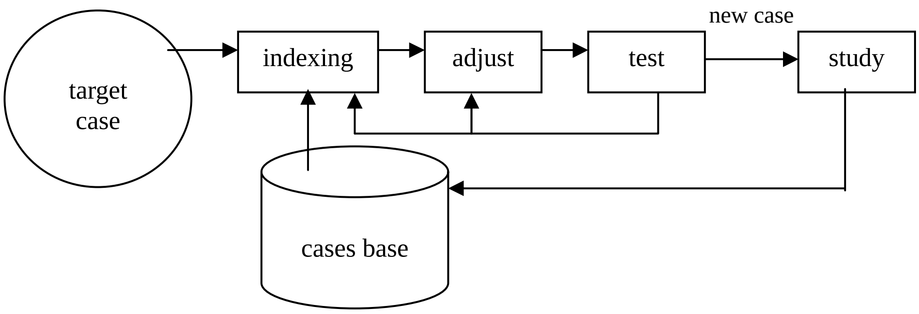


Figure1: The principle figure of CBR

The process of CBR includes four steps, and they are retrieval, reusing, revising and retain. To develop CBR system, the four steps are the core work. Moreover, retrieval is the most important process. Given a target problem, retrieve cases from memory those are relevant to solving it. A case consists of a problem, its solution, and, typically, annotations about how the solution was derived. So, the most important thing is that there are sufficient cases in case base. Of course, how to retrieve the cases from the case base is the key issue for CBR system. If there is no effective method for case retrieval, CBR system will become failure.

When indexing the case base, how to decide the similarity between the cases is very important. So, CBR system always is called similarity searching system. There are three typical CBR searching strategies, and they are nearest adjacent indexing method, inductive indexing, and knowledge guide method. Nevertheless, these methods are only suitable for the cases with qualitative attributes; they aren't competent for the cases with quantitative attributes, especially for the cases with fuzzy quantitative attributes.

However, in practical applications of CBR system, there are large numbers of cases with quantitative attributes and qualitative attributes. Therefore, in this paper, we adopt semantic distance to score the similarity degree between the attributes of the two cases, and adopt AHP algorithm to score the similarity degree between the problem case and the cases in the case base.

doi:10.4304/jsw.5.8.851-857

II. THE DISTANCE BETWEEN THE ATTRIBUTES OF CASES

In order to describe the similarity between the cases, all of their attributes must be taken into account synthetically, whereas, which kind of method is adopted to describe the similarity becomes most important. In this paper, we don't score the similarity directly, but adopt the distance to score the similarity between the cases.

There are several methods to describe the distance, and they are as follows:

1) absolute distance: $d_{ij}=\sum_{k=1}^{N}|v_{ik}-v_{jk}|$ (1)

v_{ik} and v_{jk} represent the distance between the *k*th attribute of the *j*th and *i*th case respectively.

2) euclidean distance：$d_{ij}=\sqrt{\sum_{k=1}^{N}(v_{ik}-v_{jk})^2}$ (2)

In fact, it is the inter-space geometry distance.

3) Michael Karpinski distance: $d_{ij}=[\sum_{k=1}^{N}|v_{ik}-v_{jk}|^q]^{1/q}$ (3)

4) For qualitative attributes, it is difficult to describe by a concrete number, because they are fuzzy. If adopt the semantic distance to score the distance between the attributes of different cases. In semantics, a fuzzy inter-zone number ［a, b］/ CF means that the possibility of the fuzzy number being in [a, b] is CF. Of course, if the fuzzy inter-zone is the maximum, then, the CF is sure to be 1. In semantics, a fuzzy centric number (c, r)/CF means that the possibility of the fuzzy number being In the zone which c is center and r is the radius is CF. Of course, if the r is the maximum, CF is 1.

The semantic distance between the two fuzzy number A, B: ［a1, b1］ / CF1 and［a2, b2］ / CF2 is defined as:

$$SD(A, B)=(wa\,\sigma\,|a1-a2|^u+wb\,\sigma\,|b1-b2|^u+wCF\,\sigma\,|CF1-CF2|^u)^{1/u} \quad (4)$$

In the formula (4), wa ⩾ 0, wb ⩾ 0, wCF ⩾ 0, and wa+wb+wCF=1;u ⩾ 1, and they are all integer. When CF1=CF2=1, if u=1, then:

$$SD(A, B)=wa\,\sigma\,|a1-a2|+wb\,\sigma\,|b1-b2| \quad (5)$$

In fact, the value of u is decided according to the actual condition. Of course, the fuzzy operator "σ" can be multiplication.

III CASES DESCRIPTION MODEL

A. Description of similarity between attributes of cases

The attributes of cases are composed by quantitative and qualitative attributes. If the cases are composed by n attributes. In order to describe the case easilly, we assume the m attributes are quantitative, and the n-m attributes are qualitative.

Now, suppose C0 is problem case, and it is composed by n attributes, and it is described as fellows:

C0=(C01, C02, ..., ［d0k , b0k］, ..., ［d0m, b0m］) (6)

Suppose that Cp and Cq are the cases in the cases base, of course, Cp and Cq are two different cases. and they are described as follows respectively:

Cp=(Cp1, Cp2, ...,［dpk , bpk］, ...,［dpm, bpn］) (7)
Cq=(C01, Cq2, ..., ［dqk , bqk］, ..., ［dqn, bqn］)

The semantic distance of j1 th attribute of Cp and C0 is:

SD(apj1, a0j1)=wa*｜dpj1-d0j1｜+wb*｜b pj1-b0j1｜ (8)

In the same way, the semantic distance of j1 th attribute of Cq and C0 is:

SD(aqj1, a0j1)=wa*｜dqj1-d0j1｜+wb*｜b qj1-b0j1 (9)

B. construction of the model

AHP(The Analytic Hierarchy Process)was presented by American operational research expert T.L.Satty in 1977. AHP is a decision-making method combining with quantitative and qualitative analysis. The AHP provides a comprehensive and rational framework for structuring a decision problem, for representing and quantifying its elements, for relating those elements to overall goals, and for evaluating alternative solutions. It is used around the world in a wide variety of decision situations, in fields such as government, business, industry, healthcare, and education. In AHP, the complex problem is divided into several factors, and the factors are separated into hierarchical and ordinal structure by control relationship. Users of the AHP first decompose their decision problem into a hierarchy of more easily comprehended sub-problems, each of which can be analyzed independently. The elements of the hierarchy can relate to any aspect of the decision problem—tangible or intangible, carefully measured or roughly estimated, well- or poorly-understood—anything at all that applies to the decision at hand.

In the topside, it is the target layer, the middle layer is rule layer, and the bottom layer is approach layer. The importance degree of every factor is determined by comparing each other. The detailed process is as follows:

For the target layer, it will compare the importance of every criterion in the rule layer according to the given target, and then it will get the judgment matrix, marking the judgment matrix as RB. The comparison formula of comparative importance of all criteria is shown as Table.1:

Table.1 The comparison formula of comparative importance of all criteria

The Fundamental Scale for Pairwise Comparisons		
Intensity of Importance	Definition	Explanation
1	Equal importance	Two elements contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favor one element over another
5	Strong importance	Experience and judgment strongly favor one element over another
7	Very strong importance	One element is favored very strongly over another; its dominance is demonstrated in practice
9	Extreme importance	The evidence favoring one element over another is of the highest possible order of affirmation
Intensities of 2, 4, 6, and 8 can be used to express intermediate values. Intensities 1.1, 1.2, 1.3, etc. can be used for elements that are very close in importance.		

All of the approaches will compare each other for the goal of every criterion in the rule layer, and then will get a series of matrix:

$$R_C^k = \begin{bmatrix} r_{11}^k r_{12}^k r_{13}^k \cdots\cdots r_{1M}^k \\ r_{21}^k r_{22}^k r_{23}^k \cdots\cdots r_{2M}^k \\ \cdots\cdots\cdot \\ r_{N1}^k r_{N2}^k r_{N3}^k \cdots\cdots r_{NM}^k \end{bmatrix} \quad k=1, 2, 3, \ldots\ldots, N$$

(10)

N is the number of the criterion in rule layer, and M is the number of the approach.

For the N matrixes, normalize every column in the matrix. Namely, let every element divide the summary of all element in its column.

$$r_{ij}' = r_{ij} \Big/ \sum_{i=1}^{n} r_{ij} \qquad (11)$$

The values of r_{ij}' are values of elements in the judgment matrix after the normalization processing.

Then, for R_C^k, sum every row, the value is the relative weight of every approach. Of course, we will normalize the result.

$$w_i = \sum_{j=1}^{n} r_{ij}' \Big/ \sum_{i=1}^{n}\sum_{j=1}^{n} r_{ij}' \qquad (12)$$

According to the result obtained above, we can obtain the weight vector of the rule layer relative to the target layer. It is denoted as w_B. By the same way, we can get a series of weight vectors of all the approaches relative to every criterion, and they are denoted as w_C^i. Then, let w_B multiply the matrix composing of the w_C^i, and the result is denoted as R_{WC}. From the R_{WC}, we can get the last rank for all of the approaches.

In this paper, we built a CBR indexing model based on AHP. In the model, the target is finding the case most similar with the problem case from the case base. All of the case attributes are treated as criterion.

IV. INSTANCE

For example, when the buyer makes a purchase decision, he always takes the former decision as the reference. If we compare the process with the CBR model, they are very similar. So, to the purchase decision, we regard it as a problem case, and regard the former decisions as the cases in the case base. The decision result is the most similar case indexed from the case base, and the buyer can make the last decision with reference of the most similar case. Of course, the buyer can adapt the decision result, then, store the decision as a new case in the case base.

Due to the limited space for this paper, so, for simplicity, the problem case is denoted as C0, and there are 4 cases in the case base. We assume that the case include 5 attributes, and they are price of goods(A1), the distance with the company (A2), grade of goods(A3), reputation of the vendor(A4) and quality of goods(A5).At the same time, we assume the A1, A2 and A3 are quantitative attribute; A4, A5 are qualitative attributes.

After a series of data conversion, the detailed data are shown as the Tab.2:

Table.2. The detailed information of every case

	B1	B2	B3	B4	B5
C0	150	880	7000	(3, 5)	(7, 8)
C1	120	200	5000	(8, 9)	(4, 6)
C2	220	1100	1500	(5, 7)	(3, 6)
C3	115	400	6000	(4, 6)	(5, 7)
C4	300	450	10000	(7, 8)	(8, 9)

The semantic distance of C0 with C1, C2, C3 and C4 are denoted as SD (0, 1), SD(0, 2), SD(0, 3)and SD(0, 4)respectively.

According to the formula (5), after the calculating, the detailed data of the semantic distance C0 with C1, C2, C3 and C4 are shown as Tab.2. To the formula (1), taking the value of w*a* and w*b* as 0.4 and 0.6 respectively. The detailed information of semantic distance between C0 and other cases is shown as Tab.3.

Table.3 The detailed information of semantic distance between C0 and other cases

	B1	B2	B3	B4	B5
SD(0, 1)	30	680	2000	4.4	2.4
SD(0, 2)	70	220	5500	2.0	2.8
SD(0, 3)	35	480	1000	1.0	1.4
SD(0, 4)	150	430	3000	3.2	1.0

As for the attributes, the judgment matrix of comparative importance is getting by AHP, and the judgment matrix is shown as follows:

$$R_B = \begin{bmatrix} 1 & 2 & 3 & 1/5 & 1/7 \\ 1/2 & 1 & 2 & 1/6 & 1/8 \\ 1/3 & 1/2 & 1 & 1/7 & 1/9 \\ 5 & 6 & 7 & 1 & 1/2 \\ 7 & 8 & 9 & 2 & 1 \end{bmatrix}$$

According to the formula (11), after the normalization processing, we can get the following matrix:

$$R_B^{'} = \begin{bmatrix} 0.072 & 0.114 & 0.130 & 0.057 & 0.076 \\ 0.036 & 0.057 & 0.087 & 0.047 & 0.061 \\ 0.024 & 0.290 & 0.043 & 0.041 & 0.059 \\ 0.361 & 0.343 & 0.304 & 0.285 & 0.266 \\ 0.506 & 0.457 & 0.391 & 0.570 & 0.532 \end{bmatrix}$$

According to the formula (12), after the calculation, we get the result of weight value: W_B =(0.090, 0.059, 0.091, 0.304, 0.456)

According to the Tab.3, as to criterion B1, after the comparison each other, we get the judgment matrix as follow:

$$R_C^1 = \begin{bmatrix} 1 & 3/7 & 6/7 & 1/5 \\ 7/3 & 1 & 2 & 7/15 \\ 7/6 & 1/2 & 1 & 7/30 \\ 5 & 15/7 & 30/7 & 1 \end{bmatrix}$$

According to the formula (11) and formula(12), after calculation, we can get the weight value result of the 4 cases to the criterion B1.

W_C^1 =(0.105, 0.246, 0.123, 0.526)

Likewise, we can get the weight value result of the 4 cases to the criterion B2, B3, B4 and B5, and they are shown as follow:

W_C^2 =(0.375, 0.122, 0.265, 0.238)

W_C^3 =(0.174, 0.478, 0.087, 0.261)

W_C^4 =(0.415, 0.189, 0.094, 0.302)

W_C^5 =(0.315, 0.368, 0.184, 0.132)

Finally, let the matrix W_B multiply the matrix constituted by W_C^i. After the calculation, we can get the last result, and they are:

(0.317, 0.332, 0.146, 0.243).

Now, we can conclude that the most similar case with C0 is C3. So, the buyer can make decision according to C3. Of course, the buyer can adjust the C3 according to the actual condition.

V. REALIZATION

The prototype is developed in the development of environment of CGI, and the database is Sybase; the development language is C language. The system class diagram is shown in Figure 2.

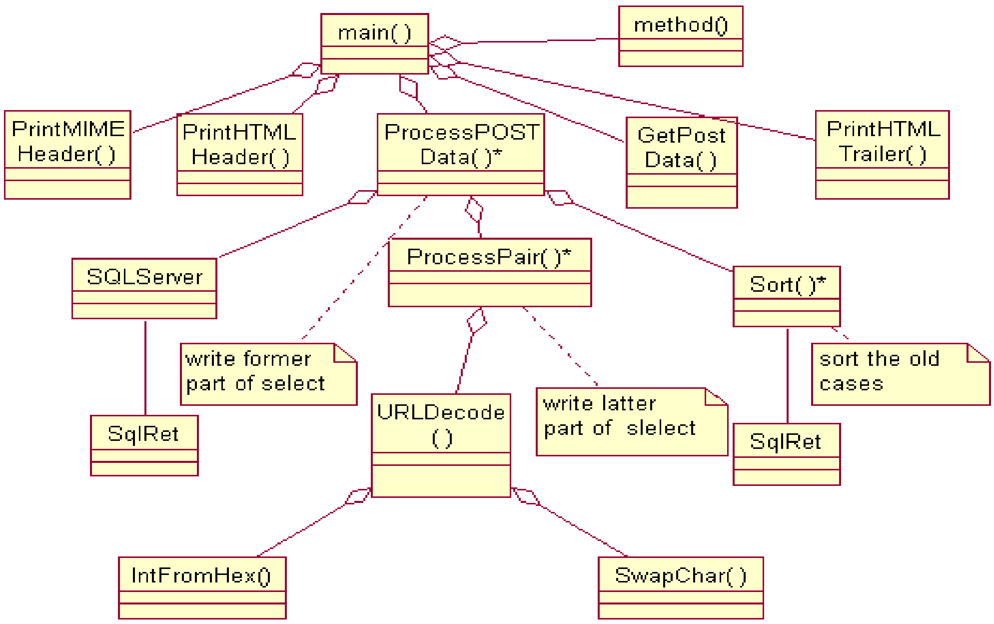


Figure.2. The system class diagram for CBR system

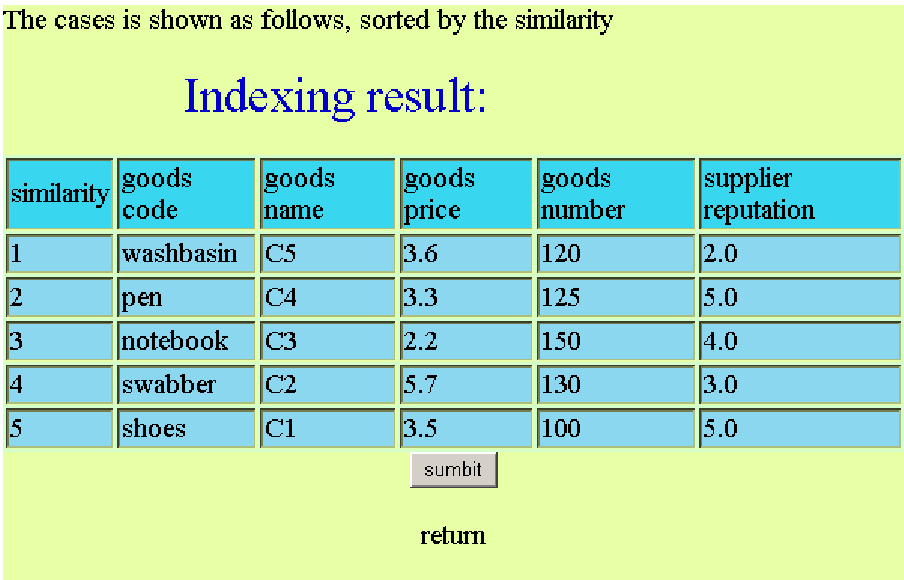
The core source code is shown as follows:

```
void ProcessPOSTData()
{        ……
/* Begin - Creat a SQL statement without WHERE clauses or SQLserver command. You may write SQL statement here */
/*Examples follow as: */
/*SELECT * FROM */
/*Rest parts of SQL statement will be written in */
/* ProcessPair() function. */
/* strcpy(sqltemp, "select * from ");*/
/* End-------------------------------------- */
……
SQLserver s(Sqlserver, a->Username, a->PassWord, Database);
{……
sort(*sqlret, jiage, shuliang, shengyu, zhiliang1, zhiliang2);
}
void ProcessPair(char * VarVal, char sqltemp[])
{
……
/* Begin - Assemble rest parts of SQL statement into the SQL statement. */
if (strcmp(VarVal , "tablename") == 0)
{strcat(sqltemp, pEquals);
strcat(sqltemp, ";");}
/* these Let the SQL statement complete, and tablename is the table which needs select */
if (strcmp(VarVal , "jiage") == 0)
jiage=atof(pEquals);
if (strcmp(VarVal , "shuliang") == 0)
shuliang=atof(pEquals);
if (strcmp(VarVal , "changshangshengyu") == 0)
shengyu=atof(pEquals);
if (strcmp(VarVal , "zhiliang1") == 0)
zhiliang1=atof(pEquals);
if (strcmp(VarVal , "zhiliang2") == 0)
zhiliang2=atof(pEquals);
/* These let all of the value transmited from webpage kept in the corresponding variables */
/* End ---------------------------------------------------*/
    ……
}
void sort(SqlRet sqlret, float price, float number, float repute, float quality1, float quality2)
{
……
    for ( i=0;i<sqlret.GetRow();i++)
    {
      records++;
      priceones[i]= sqlret.GetItem(i, 2);
      numberones[i]=sqlret.GetItem(i, 3);
      reputeones[i]=sqlret.GetItem(i, 4);
      quality1ones[i]=sqlret.GetItem(i, 5);
      quality2ones[i]=sqlret.GetItem(i, 6);

    }
    for (i=0;i<records;i++)
    for (j=0;j<records;j++)
    {
```

```
 pricesquare[i][j]=fabs(priceones[j]-price)/(fabs(priceo
 nes[j]-price)
+fabs(priceones[i]-price));
if (pricesquare[i][j]>=criticalvalue)
      pricesquare[i][j]=1;
else
      pricesquare[i][j]=0;
                  ……
    }
          ……
for (i=0;i<records;i++)
         for (j=0;j<records;j++)
         {
               if (pricesquare[i][j]==1)
               priceones[i]++;
          }
         for (i=0;i<records;i++)
         {
               sortall[i]=
               priceones[i]*powerall[0]+numberones[i]*powera
               ll[1]+reputeones[i]*powerall[2]
               +(quality1ones[i]*powerquality[0]+quality2ones
               [i]*powerquality[1])
               *powerall[3];
          }
     for (i=0;i<records;i++)
         {
          max=0;
          for (j=0;j<records;j++)
          if (!checked[j]&&sortall[j]>max)
                 max=sortall[j];tag=j;
          sortlast[p]=tag;p++;checked[tag]=true;
                           }
         ……
 }
```

The running interface is shown as follows:



similarity	goods code	goods name	goods price	goods number	supplier reputation
1	washbasin	C5	3.6	120	2.0
2	pen	C4	3.3	125	5.0
3	notebook	C3	2.2	150	4.0
4	swabber	C2	5.7	130	3.0
5	shoes	C1	3.5	100	5.0

Figure.3. The running interface of CBR system

VI. Conclusion

AHP is a very mature and stable method. Moreover, CBR is a scientific decision method. According to the instance presented above, the case indexing model for CBR system based on AHP is an effective and feasible method. Of course, AHP is improved after the research of experts and scholars. For example, the fuzzy AHP is a more scientific method. In our subsequent work, we will adopt fuzzy AHP in CBR system.

Acknowledgment

The authors gratefully acknowledge the Projects Supported by Scientific Research Fund of Hunan Provincial Education Department(08A009 and 09C271) for supporting this research.

This research is supported by the construct program of the key discipline in Hunan province; project supported by Provincial Natural Science Foundation of Hunan(01JJY2157)support the research; research fund of Hunan Institute of Engineering(0850) supporting the research.

REFERENCE

[1] Juarez, Jos e M.; Guil, Francisco; Palma, Jose.Temporal similarity by measuring possibilistic uncertainty in CBR. Fuzzy Sets and Systems, v 160, n 2, p 214-230, January 16, 2009

[2] Zhang Bensheng, Yu Yongli. Hybrid Similarity Measure for Retrieval in Case-based Reasoning System. System Engineering Theory and Practice.2002, 3:131-136

[3] Shi Zhongzhi. Knowledge finding. Beijing: Tsinghua University Press, 2002:85-102

[4] Zhong Sisheng, Wang Zhixing. A Model for Mixed Attributions Cases Indexing. Chinese Journal of Software.1999, 10:521-526.

[5] LI Pei-gen, ZHANG Jie.Reconfiguration and Control of agile intelligent manufacturing system. Mechanical Industry Press, Beijing:, 2003.

[6] Satty, T.L. Application of Analytic Hierarchy Process in Resource-allocating, Management and Confliction-analysing.Translated by XU Bolin, Coal Industry Press, Beijing, pp.4-15, 1988

[7] Chang Da-yong, Zhang Li-li. Math Model for Economy management. Beingjing:Institute of Beijing Economy Management Press, pp.146-154, 1995

[8] Saaty, Thomas L. (2008-06). "Relative Measurement and its Generalization in Decision Making: Why Pairwise Comparisons are Central in Mathematics for the Measurement of Intangible Factors - The Analytic Hierarchy/Network Process". RACSAM (Review of the Royal Spanish Academy of Sciences, Series A, Mathematics) 102 (2): 251–318.

[8] Bhushan, Navneet; Kanwal Rai. Strategic Decision Making: Applying the Analytic Hierarchy Process. London: Springer-Verlag. ISBN 1-8523375-6-7, January 2004.

[8] Forman, Ernest H.; Saul I. Gass.The analytical hierarchy process—an exposition. Operations Research. Jul, 2001, 49 (4): 469–487.

[9] Steiguer, J.E.; Jennifer Duberstein, Vicente Lopes. The Analytic Hierarchy Process as a Means for Integrated Watershed Management, in Renard, Kenneth G., First Interagency Conference on Research on the Watersheds, Benson, Arizona: U.S. Department of Agriculture, Agricultural Research Service, October 2003, pp. 736–740

[10] Berrittella, M.; A. Certa, M. Enea, P. Zito. An Analytic Hierarchy Process for the Evaluation of Transport Policies to Reduce Climate Change Impacts, Fondazione Eni Enrico Mattei (Milano), January, 2007.

[11] Grandzol, John R. Improving the Faculty Selection Process in Higher Education: A Case for the Analytic Hierarchy Process, August 2005, IR Applications 6.

[12] Atthirawong, Walailak; Bart McCarthy. An Application of the Analytical Hierarchy Process to International Location Decision-Making, in Gregory, Mike, Proceedings of The 7th Annual Cambridge International Manufacturing Symposium: Restructuring Global Manufacturing, Cambridge, England: University of Cambridge, September, 2002, pp. 1–18.

[13] Dey, Prasanta Kumar. Analytic Hierarchy Process Analyzes Risk of Operating Cross-Country Petroleum Pipelines in India. Natural Hazards Review 4 (4): 213–221, November 2003.

[14] Lippert, Barbara C.; Stephen F. Weber. HIST 1.0; Decision Support Software for Rating Buildings by Historic Significance. National Institute of Standards and Technology, NISTIR 5683, October, 1995.

[15] Larson, Charles D.; Ernest H. Forman.Application of the Analytic Hierarchy Process to Select Project Scope for Videologging and Pavement Condition Data Collection, 86th Annual Meeting Compendium of Papers CD-ROM, Transportation Research Board of the National Academies.January, 2007.

[16] Agnar Aamodt and Enric Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches, " Artificial Intelligence Communications 7 (1994): 1, 39-52.

[17] Rule-induction algorithms are procedures for learning rules for a given concept by generalizing from examples of that concept. For example, a rule-induction algorithm might learn rules for forming the plural of English nouns from examples such as dog/dogs, fly/flies, and ray/rays.

[18] Roger Schank, Dynamic Memory: A Theory of Learning in Computers and People (New York: Cambridge University Press, 1982).

[19] Janet Kolodner, "Reconstructive Memory: A Computer Model, " Cognitive Science 7 (1983): 4.

[20] Michael Lebowitz, "Memory-Based Parsing, " Artificial Intelligence 21 (1983), 363-404.

[21] Bill Mark, "Case-Based Reasoning for Autoclave Management, " Proceedings of the Case-Based Reasoning Workshop (1989).

[22] Trung Nguyen, Mary Czerwinski, and Dan Lee, "COMPAQ QuickSource: Providing the Consumer with the Power of Artificial Intelligence, " in Proceedings of the Fifth Annual Conference on Innovative Applications of Artificial Intelligence (Washington, DC: AAAI Press, 1993), 142-151.

[23] Acorn, T., and Walden, S., SMART: Support management automated reasoning technology for Compaq customer service. In Proceedings of the Tenth National Conference on Artificial Intelligence. MIT Press. (1992).

[24] Cheetham, W., Goebel, K., Appliance Call Center: A Successful Mixed-Initiative Case Study, Artificial Intelligence Magazine, Volume 28, No. 2, (2007). pp 89 – 100.

[25] Hinkle, D., and Toomey, C. N., CLAVIER: Applying case-based reasoning on to composite part fabrication. Proceeding of the Sixth Innovative Application of AI Conference, Seattle, WA, AAAI Press, (1994). pp. 55-62.

[26] Cheetham, W., Tenth Anniversary of Plastics Color Matching, Artificial Intelligence Magazine, Volume 26, No. 3, (2005). pp 51 – 61.

[27] Watson, I. Gardingen, D. (1999). A Case-Based Reasoning System for HVAC Sales Support on the Web. In, the Knowledge Based Systems Journal, Vol. 12. no. 5-6, pp.207-214