퍼지 및 지능시스템학회 논문지 2003, Vol. 13, No. 3, pp. 366-370

Hybrid Intelligent Web Recommendation Systems Based on Web Data Mining and Case-Based Reasoning

Jin Sung Kim

School of Business Administration, Jeonju University
Hyoja-Dong 3-1200, Wansan-Ku, Jeonju, Jeonbuk, 560-759, Korea
E-mail : kimjs@jeonju.ac.kr

Abstract

In this research, we suggest a hybrid intelligent Web recommendation systems based on Web data mining and case-based reasoning (CBR). One of the important research topics in the field of Internet business is blending artificial intelligence (AI) techniques with knowledge discovering in database (KDD) or data mining (DM). Data mining is used as an efficient mechanism in reasoning for association knowledge between goods and customers' preference. In the field of data mining, the features, called attributes, are often selected primary for mining the association knowledge between related products. Therefore, most of researches, in the arena of Web data mining, used association rules extraction mechanism. However, association rules extraction mechanism has a potential limitation in flexibility of reasoning. If there are some goods, which were not retrieved by association rules-based reasoning, we can't present more information to customer. To overcome this limitation case, we combined CBR with Web data mining. CBR is one of the AI techniques and used in problems for which it is difficult to solve with logical (association) rules.

A Web-log data gathered in real-world Web shopping mall was given to illustrate the quality of the proposed hybrid recommendation mechanism. This Web shopping mall deals with remote-controlled plastic models such as remote-controlled car, yacht, airplane, and helicopter. The experimental results showed that our hybrid recommendation mechanism could reflect both association knowledge and implicit human knowledge extracted from cases in Web databases.

Key Words : Internet Business, Recommendation, KDD, Data Mining, CBR, Association knowledge.

1. 서 론

In the field of Internet Business, recommendation systems can serve as intermediaries between the buyers and the sellers, creating a 'cyber marketplace.' In addition, recommendation facility is becoming an integral part of Internet Business. Therefore, recommendation systems should have more intelligent facilities can lowers the buyer"s cost and reduce the time to acquire information (Changchien & Lu, 2001; Cho et al., 2002; Hui & Jha, 2000). In this purpose, many of companies tried to reduce the inefficiencies caused by information search and retrieval costs. Recently, most of companies have a customer service department called as CRM (Customer Relationship Management) center that provides direct one-to-one marketing, advertisement, promotion, and other relationship management services (Cho et al., 2002; Choy et al., 2002; Hui & Jha, 2000; Kannan & Rao, 2001; Kim et al., 2002; Lee et al., 2002; Song et al., 2001). Especially, in these companies, marketing managers should know and predict the customer's intentions for purchase and future behaviors. If there is insufficient understanding of a customer"s behavior, it can lead to problems such as low profit. In this case, many researchers used Web mining as a new technology, which can reduce the risk of prediction. Web data mining has the same objective with data mining in that both attempt to search for valuable and meaningful knowledge from databases. However, Web data mining has the differences from data mining in that the former is more unstructured task than the latter. Source of such difference is based on the characteristics of Web documents or Web log files which represent unstructured relationships with little machine-readable semantics while data mining is aimed at dealing with more structural database. With the advent of CRM issues in Internet Business, most of the modern companies operating Web sites for several purposes are now adopting Web mining as a strategic way of capturing knowledge about potential needs of target customers (Cho et al., 2002; Hui & Jha, 2000; Lee et al., 2002).

In previous literatures, many of researchers used machine learning technologies, decision trees, and other

접수일자 : 2003년 2월 6일
완료일자 : 2003년 6월 13일
"이 논문은 2002년도 한국학술진흥재단의 지원에 의하여 연구되었음" (KRF-2002-003-B00099)

statistical methodologies to suggest the effective solutions for data mining, CRM (Cho et al., 2002; Choy et al., 2002; Hui & Jha, 2000; Kannan & Rao, 2001; Kim et al., 2002; Lee et al., 2002). Basic (Web) data mining, however, has potential limitations such as poor interpretability and stiffness in reasoning (Lee et al., 2002). If the recommendation system has no association (logical) rule, it may provide no additional information to customer (Aha, 1991; Chiu, 2002; Choy et al., 2002; Finnie & Sun, 2002; Fyfe & Corchado 2001; Hui & Jha, 2000; Lee et al., 2002; Schirmer, 2000). Therefore, we may say that the older data mining techniques has the limitation in the quality of reasoning and environmental adaptability. To overcome these limitations, we suggest hybrid Web recommendation mechanism based on Web data mining and CBR. Web data mining can extract the explicit knowledge from databases such as IF-THEN logical rule. Then, CBR is an intelligent knowledge extraction mechanism that shows the implicit knowledge. Therefore, our proposed mechanism can improve the reasoning ability and environmental adaptability of Web-based recommendation systems.

2. Research methodology

Among the data mining techniques, association rules mining algorithm has been popular in marketing fields (Lee et al., 2002). Therefore, we applied the association rule mining technique to Web mining task. From a data pre-processing viewpoint, the Web log data poses the following challenges, 1) large errors, 2) unequal sampling, 3) missing values. To remove these noises included in data set, we applied pre-processing techniques to Web log databases. Association rules are similar to IF-THEN logical rules in which a condition clause (IF) triggers a conclusion clause (THEN). In addition, association rules include the support and confidence (Agrawal et al., 1993). First of all, association rule mining is derived from data mining approach, which is known as efficiently discovering logical rules from large databases. The association rule mining focused on finding association rules like "$X \Rightarrow Y$ with support A% and confidence B%, where X and Y are sets of items",

Table 1. Pseudo code of association rules mining algorithm

```
C_k : Candidate transaction set of size k
L_k : Frequency transaction set of size k
L_j = {frequent items};
For (k=1; L_k != φ ; k++) Do Begin
C_{k+1}=Candidates generated from L_k ;
For Each transaction t in database Do
  Increment the count of all candidates in C_{k+1}
  that are contained in tL_{k+1}=candidates in C_{k+1} with
min_support
End Return L_k ;
```

from the transactions in database. The meaning of this association rule is that transactions in the database, which contain the items in X tend to also contain the items in Y. The support A% of the rule $X \Rightarrow Y$ is the probability 'Pr(X∩Y)'. Then, the confidence B% means the probability 'Pr(X∩Y)/ Pr(X)'. In other words, the support represents the usefulness of the discovered association rules, while the confidence indicates certainty of the detected association rules. The pseudo code of association rules mining algorithm was shown in Table 1.

Our hybrid recommendation mechanism was composed of four phases as shown in Figure 1.

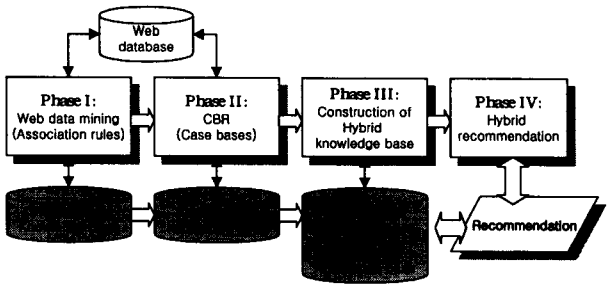

Figure 1. Research methodology

The first phase of hybrid recommendation mechanism is Web data mining to extract association rules from the Web log databases. In the second phase, after the extraction of the association rules, we used CBR to extend the quality of reasoning and recover the limitation of rule-based reasoning. Cases extracted from the customer database may imply customer's cognition for products and predict the customer's future behavior. Through this phase, CBR shows significant promise for improving the effectiveness of complex and unstructured decision-making. The third phase of our hybrid recommendation mechanism is construction of hybrid knowledge base. In this phase, two different knowledge bases were combined into one hybrid knowledge base. The key features to combine these two different knowledge bases are customer's profile and product information. Final fourth stage of our proposed recommendation mechanism is to apply cooperative inference procedures to the hybrid knowledge base and suggest the recommendation information.

3. Data & implementation

In the implementation process, we developed the prototype system using the VBA language and ASP in a Web environment. We called this prototype system as HYRES (HYbrid REcommendation Systems). HYRES was composed of five components as shown in Figure 2. Five components of HYRES are 1) Rule & case generator, 2) Hybrid knowledge base, 3) Cooperative inference engine, 4) Justifier, and 5) User interface.

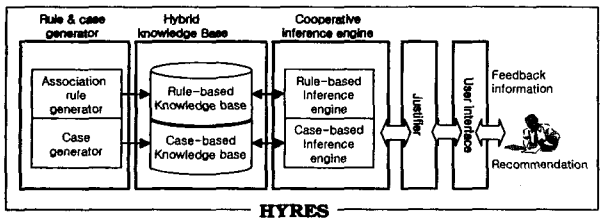

Figure 2. The HYRES system structure

3.1. Phase I: Web data mining (association rule mining)

Web log database was composed of production information (car, tank, helicopter, glider, yacht and ship) and customers' profile. Web log data was collected from Web-based RC (remote-controlled plastic model) shopping mall. The original Web log data was contaminated by several type of irrelevant and redundant information such as slashes (/, \), file name suffixes (htm, html, gif, jpg, jsp, etc.), and other information for query communications (&, =, ≪=, ?, etc.). To mine a meaningful set of association rules from the Web database, the first step to be done is to cleanse the original Web log data so that the preprocessed Web log data may become more traceable (Lee et al., 2002). Figure 3 shows preprocessed Web database.

	A	B	C	D	E	F	G	H	I
1	Name	Sec-No.	Address	Position	Registration	Sex	Phone	E-Mail	Products
8	M. Kwon	618-1405	n APT 108-406, Mannyun-Dong, Da	Independent enterprise	11 26 2001 3:47PM	Male	kmk@itech.cc		RC Car GuideBook
9	S. Kwon	805-1802	vn 1303, Hyunggok-Dong 170-1, Kur	Independent enterprise	10 26 2001 9:42PM	Male	4-456-5	mi66@dau	MASTER 30%(1G)(Airplane, He
10	Jung H. Ke	405-1379	f101, Nonhyun-Dong 190-16, Kangn	Undergradute student	Aug 27 2001 2:19AM	Male	-9616-7	mh@korea	MSX3
11	S. Kwon	805-1802	vn 1303, Hyunggok-Dong 170-1, Kur	Independent enterprise	10 26 2001 9:42PM	Male	4-456-5	mi66@dau	12V Battery(7A)
12	K. Kim	412-1623	Changwon Kyungnam	Businessman	02 20 2002 5:15PM	Male	5-278-4	isi@empal	Booster Charger
13	ae H. Kwo	702-1775	1, 272-14, Suhan-Myun, Boeun-gur	Undergradute student	26-Jun-01	Male	3) 544-2	1c@hanm	RC Car GuideBook

Figure 3. Preprocessed Web database

APRIORI algorithm (Agrawal et al., 1993a, 1993b, 1997) was known to yield a set of association rules. Based on the preprocessed Web database in Figure 3, the corresponding association rules were extracted with a threshold of 20% confidence. Table 2 shows an excerpt of the derived association rules.

Table 2. Example of association rules from Web log database

Pocket Booster (Checker)	<= RC car guidebook (4.67%, 20%)
ACE 2000	<= 7.2V low speed charger(4.67%, 40%)
7.2V low speed charger	<= ACE 2000 (3.74%, 50%)
15% SM15 (1G)	<= 7.2V low speed charger (4.67%, 20%)

3.2. Phase II, III: CBR & Construction of hybrid knowledge base

Aamodt & Plaza (1994) proposed four-phase CBR mechanism, which could assist the decision makers in detecting similar cases (Figure 4). The first step of CBR is retrieval of former cases from case base. After the retrieval, extracted former cases were matched with new case. If there is a need for revision of former case to resolve the new problem, extracted former cases were revised through the revision process. The final step of CBR is retention of tested and repaired case.

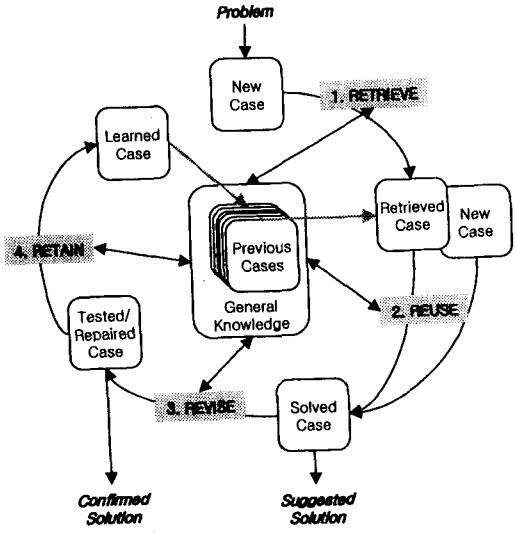

Figure 4. Four-phased CBR (Admodt & Plaza, 1994)

The key concept of CBR is the similarity measure to detect similar cases from case base. The concept of similarity used in CBR play a fundamental role in many fields of pure and applied science. The simplest CBR algorithm is CBL1(Case Based Learning 1). Its pre-process linearly normalizes all numeric feature values (Aha, 1991). CBL1 defines the similarity of cases C_1 and C_2 as

$$Similarity(C_1, C_2, P) = \frac{1}{1+\sqrt{\sum_{i \in P} Feature\_dissimilarity(C_1, C_2)}}$$

(Note: C_1 = new case, C_2 = compared case extracted from case base)

Where P is the set of predictor features and

$$Feature\_dissimilarity(C_1, C_2)$$

$$\begin{cases} (C_{1i} - C_{2i})^2 & \text{if } \textit{feature } i\textit{'s values are numeric} \\ 0 & \text{if } C_{1i} = C_{2i} \\ 1 & \textit{otherwise} \end{cases}$$

C_1 and C_2 means the cases extracted from case base, which has key features (attributes) such as sex (male/female), age, a matter of concern and interest, career, job, and etc. When you adopt the CBR mechanism in case extraction, you can use these features

Table 3. Equations for CBR

Similarity = Customer's characteristics + probability of success

$$\text{Probability of success} = \frac{\text{former frequency of the purchase}}{\text{possible maximum frequency of the purchase}} \times \text{constant}$$

$$\text{Customer's characteristics} = \sum_{i=1}^{n} CS_i$$

$$CS_i = weight \times \left(1 - \frac{|\ \text{coustome's } characteristic - \text{selected customer's } characteristic\ |}{\text{maximum degree of } characteristic}\right)$$

instead of C_1 or C_2 to compute the similarity. Equations for CBR were summarized in Table 3.

The HYRES supports a CBR mechanism shown in Table 3, and transforms the case extraction results into case-based knowledge base. Figure 5 summarizes five cases extracted from case base.

	A	B	C	D	E	F	G
1	Age	Age class	Sex	Position code	Experience	Interest	Products
12	18	1	1	1	18	1	ST-36
13	23	2	1	5	14	1	Crytal case
14	20	2	1	2	14	1	RC car guidebook
15	35	3	1	5	13	1	GLINT 25L
16	20	2	1	3	13	1	Nitro quake

Figure 5. Example of case
(*Experience: month experienced, Interest: 1=car/tank, 2=yacht/ship, 3=airplane/helicopter)

After the extraction of association rules and related cases, rule-based knowledge base and case-based knowledge base were combined using the customer's key features and other profile.

3.3. Phase IV: Hybrid recommendation

After the construction of hybrid knowledge base, HYRES ready to execute cooperative inference. In this sense, the HYRES has cooperative inference engine, which uses rule-based inference and CBR concurrently. In this phase, HYRES may suggest the hybrid recommendation results to customers. Table 4 shows the Web customer's brief profile and preference to validate the performance of HYRES.

In the first step, customer will try to search and select the 'Guidebook for remote-controlled car.' Then, he might get the additional information suggested by association rule miner. If customers want to get a more information, nonetheless, how can you react this request?

If your Web site has a data-mining driven inference engine, you couldn't give additional information. In addition, that customer couldn't develop his ability to control the remote-controlled car. Therefore, the Web site may lose his potential loyal customer and opportunity. In this case, HYRES can react to customers' requests intelligently. Figure 6 shows the hybrid recommendation systems' dialog box and results of recommendation. As shown in figure 6, customer could find more additional information describing other products suggested by HYRES. Finally, recommended products are 'Pocket Booster (interconnection checker for battery and plug)', 'Tool Box (container including starter, fuel tank, controller, etc.)', 'FORD FOCUS WRC 4WD Set (full set of remote-controlled car suitable for beginners)', and 'STARTER BOX (box including starter of on/off-road car)'. This information is very important

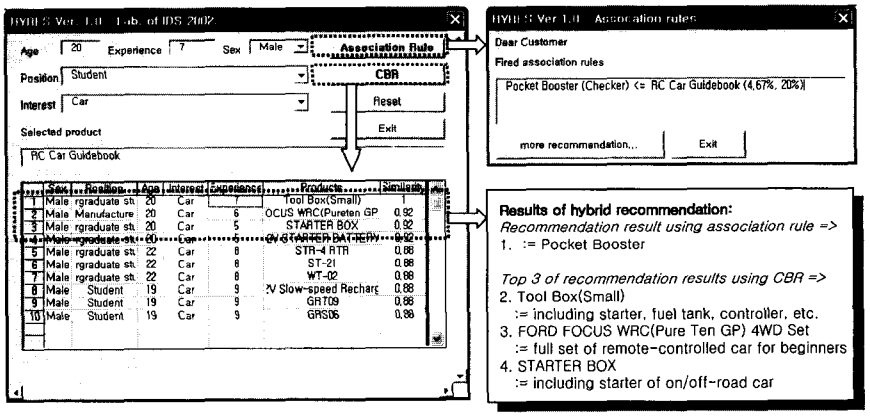

Figure 6. Hybrid recommendation results of HYRES

Table 4. Customer"s profile

Customers profile
Birth: February 1983 / Sex: Male / Position: Student / Experience (career): 7 months / Interest (a matter of concern): Remote-controlled Car

Customers preference:
Buying the Guidebook for remote-controlled Car

to beginners. Therefore, customer can take more products that he didn't expected and suitable for beginners.

4. Conclusions

In this study, we suggested a hybrid recommendation

mechanism. The proposed mechanism consists of the four phases-Web data mining, CBR, Construction of hybrid knowledge base, and Hybrid recommendation. In addition, we developed a prototype hybrid recommendation system HYRES and proved the adaptability of that systems. These hybrid recommendation mechanism and systems and were based on Web data mining and CBR, which was aimed at expand the recommendation ability.

This study has shown how the implicit knowledge (association rule mining) and tacit knowledge (CBR) within Web database can be brought together to create intelligent inference (recommendation) engine used in Web recommendation. Therefore, it is expected that our proposed hybrid recommendation mechanism will have a significant impact on the research domain related to CRM. Nonetheless, further research topics still remains. First, the basic technology of Web data mining should be improved with other artificial intelligence technique. Second, there are need to transform the customer"s deep knowledge into hybrid knowledge base, so that more complicated decision problems could be analyzed efficiently.

References

[1] Aamodt, A. and Plaza, E., "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *Artificial Intelligence Communications*, IOS Press, 7(11), pp. 39-59, 1994.

[2] Agrawal, R., Imielinski, T., and Swami, A.,, "Database Mining: A Performance Perspective," *IEEE Transactions on Knowledge and Data Engineering*, 5(6), pp. 914-925, 1993.

[3] Aha, D.W., "Case-Based Learning Algorithm," *Proceedings of the Case-Based Reasoning Workshop*, pp. 147-158, May 1991.

[4] Bonchi, F., Giannotti, F., Gozzi, C., Manco, G., Nanni, M., Pedreschi, D., Renso, C., and Ruggieri, S., "Web Log Data Warehousing and Mining for Intelligent Web Caching," *Data & Knowledge Engineering*, 39, pp. 165-189, 2001.

[5] Changchien, S.W., and Lu, T.C., "Mining Association Rule Procedure to Support On-Line Recommendation by Customers and Products Fragmentation," *Expert Systems with Applications*, 20, pp. 325-335, 2001.

[6] Chiu, C., "A Case-Based Customer Classification Approach for Direct Marketing," *Expert Systems with Applications*, 22(2), pp. 163-168, 2002.

[7] Cho, Y.H., Kim, J.K., and Kim, S.H., "A Personalized Recommender System based on Web Usage Mining and Decision Tree Induction," *Expert Systems with Applications*, 23(3), pp. 329-342, 2002.

[8] Choy, K.L., Lee, W.B., and Lo, V., "Development of a Case Based Intelligent Customer-Supplier Relationship Management System," *Expert Systems with Applications*, 23(3), pp. 281-297, 2002.

[9] Finnie, G., Sun, Z., "Similarity and Metrics in Case-Based Reasoning," *International Journal of Intelligent Systems*, 17, pp. 273-287, 2002.

[10] Fyfe, C., and Corchado, J.M., "Automating the Construction of CBR Systems using Kernel Methods," *International Journal of Intelligent Systems*, 16, pp. 571-586, 2001.

[11] Hui, S.C. and Jha, G.., "Data Mining for Customer Service Support," *Information & Management*, 38, pp. 1-13, 2000.

[12] Kannan, P.K. and Rao, H.R., "Introduction to the Special Issue: Decision Support Issues in Customer Relationship Management and Interactive Marketing for e-Commerce," *Decision Support Systems*, 32(2), pp. 83-84, 2001.

[13] Kim, E., Kim, W., and Lee, Y., "Combination of Multiple Classifiers for the Customer's Purchase Behavior Prediction," *Decision Support Systems*, 34(2), pp. 167-175, 2002.

[14] Lee, K.C., Kim, J.S., Chung, N.H., and Kwon, S.J., "Fuzzy Cognitive Map Approach to Web-mining Inference Amplification," *Expert Systems with Applications*, 22, pp. 197-211, 2002.

[15] Schirmer, A., "Case-Based Reasoning and Improved Adaptive Search for Project Scheduling," *Naval Research Logistics*, 47, pp. 201-222, 2000.

[16] Song, H.S., Kim, J.K., and Kim, S.H., Mining the Change of Customer Behavior in an Internet Shopping Mall, *Expert Systems with Applications*, 21, pp. 157-168, 2001.

김진성(Kim, Jin Sung)

김진성 (金珍成)은 현재, 전주대학교 경영학부 전임강사로 재직중이다. 주요 관심분야는 퍼지와 인공지능 기법을 이용한 의사결정지원시스템, 인터넷 비즈니스, 웹 기반 의사결정지원시스템 등이다.

Phone : 063-220-2932
Fax : 063-220-2052
E-mail : kimjs@jeonju.ac.kr