Designing a Course Recommendation System on Web based on the Students' Course Selection Records

Ko-Kang Chu, Maiga Chang[1] and Yen-Teh Hsia[2]
kirk@ms2.hinet.net, maiga@ms2.hinet.net, hsia@ice.cycu.edu.tw
Dept. of Electronic Engineering, Chung-Yuan Christian Univ.
[1]Dept. of Special Education, Chung-Yuan Christian Univ.
[2]Dept. of Information and Computer Engineering, Chung-Yuan Christian Univ.
#22, Pu-Jen, Pu-Chung Li, Chung-Li, 320, Taiwan

Abstract: There is a web-based course recommendation system constructed by this paper. A course recommendation system is used to provide students some suggestions when they have trouble in choosing courses. This paper designs the recommendation system based the Prediction Methodology proposed by us previously. The prediction methodology integrates both Data Mining techniques and Graph Theory, which will be still revealed in this paper for readers to get the idea about how can we make appropriate course suggestions for students. Results in this paper are generated by applying both our prediction mechanism and the questionnaire to two classes (class 2001 & 2002) in the past two school years The questionnaire is used for verifying our predicting methodology either useful or not.

1. Introduction & Research Backgrounds

Students take a course rather than other courses always involves many issues, including teachers (who teaches), time (when), easy or difficult (easy to pass?), *etc*. This paper focuses on the relation between course categories and students' preferences, since most of students dislike waking up in the early morning and there is also few popular teachers in school, not to mention students have not enough resource to know a course he/she didn't take yet. It is worth to note that because of the prerequisite courses are always have to choose either way, those mandatory courses should no be taken into consideration when analyzing students' preference.

However, what we have is only the course selection records of students, how can we know which course categories that a student like or dislike? So, how about class all courses into different categories previously. Unfortunately, as we know that each course covers more than one category such as the Fuzzy course in Computer Science. The Fuzzy course should at least belong to three categories, including Mathematics, Artificial Intelligence (AI for short) and Research. Hence, if a student chooses Fuzzy, it is hard to tell what kinds of course categories that he/she really likes.

Once again, we can assign some sorts of weights to course for indicating the relatedness between course and categories. Example 1 below shows this possibility.

Example 1. Given 6 courses and its related categories

Fuzzy: AI (95%), Research (85%), Mathematics (70%)
Neural Networks: AI (90%), Research (85%), Mathematics (70%)
Computer Mathematics: Research (90%), Mathematics (90%)
Graph Theory: Research (100%), Mathematics (80%)
Statistics: Mathematics (100%)

Data Mining: AI (95%), Research (90%)

There are several students take different courses above:

Ken: Fuzzy and Neural Networks

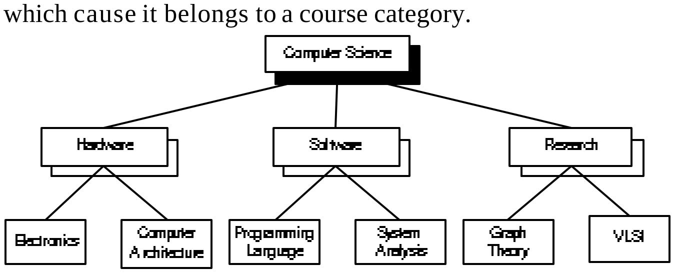
Jane: Fuzzy and Statistics

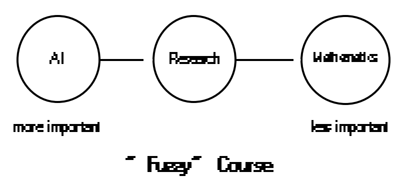
John: Data Mining and Neural Networks. □

What's happen? It will be not easy to know which category that Ken really likes and perhaps the category weights for each course are total different from Ken's view. Therefore, the objective of this paper is to construct a web-based course recommendation system that only depends on the courses chosen by students.

In this paper a prediction methodology is revealed. The prediction methodology integrates both Data Mining techniques and Graph Theory. Data Mining is a method for extracting rules (knowledge) from data and currently used in many different fields include Artificial Intelligence, Business and Education *etc.*.(Han et al., 2001) In AI field, data mining can deduce rules from facts, for example "If Bird can fly Then Sparrow can fly." (Agrawal et al., 1993) Data mining techniques are also important and useful to Business, since by analyzing the sales transactions in marketplace the relations between products and customers could be built.(Ozden et al., 1998)(Miller et al., 1997)(Srikant et al., 1996) After the relations are found out, it is possible to provide appropriate products to appropriate customers. A small number of researches focused on course selection made by students, analyzing the correlation between courses.(Mannila et al., 1994)(Klemettinen et al. 1994)

The most of simple and famous data mining technique is Apriori algorithm, proposed by Agrawal and Srikant in 1994.(Agrawal et al., 1994) Based on Apriori algorithm, the association rules, "A $\rightarrow$ B" for example, can be found from large databases. The rule "A $\rightarrow$ B" means if purchase item A will result in a high probability of item B being purchased. But as (Fig. 1) shown below, each course has several characteristics which cause it belongs to a course category.

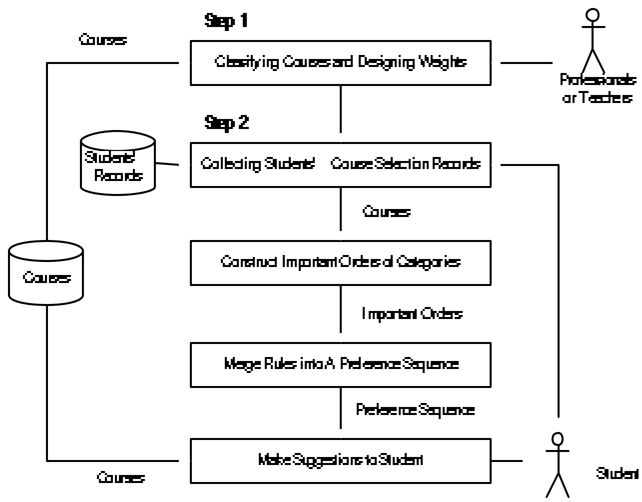




analyzed and its related algorithm is also designed.(Chu et al., 2001)(Chu et al., 2002) In order to let our course recommendation system cleared with readers, part of the prediction methodology is still expressed in Section 2. Besides, the recommendation process for designing the experiment system in this paper will be also analyzed by Section 2. Section 3 describes the web-based course selection system and evaluates the results collected from both system and questionnaire. A simple conclusion and possible future works are done in Section 4.

2. Prediction Methodology & Recommendation Process

Although this paper focuses on the web-based course recommendation system, the fundamental prediction theory is still needed to describe at first. The whole recommendation can be thought as (Fig. 3).



$$CM = \begin{bmatrix} f_{AI}(Fuzzy) & f_{Research}(Fuzzy) & f_{Math}(Fuzzy) \\ f_{AI}(NN) & f_{Research}(NN) & f_{Math}(NN) \\ f_{AI}(CMath) & f_{Research}(CMath) & f_{Math}(CMath) \\ f_{AI}(GTheory) & f_{Research}(GTheory) & f_{Math}(GTheory) \\ f_{AI}(Statistics) & f_{Research}(Statistics) & f_{Math}(Statistics) \\ f_{AI}(DM) & f_{Research}(DM) & f_{Math}(DM) \end{bmatrix}$$

$$CM = \begin{bmatrix} 0.95 & 0.85 & 0.70 \\ 0.90 & 0.85 & 0.70 \\ 0 & 0.90 & 0.90 \\ 0 & 1 & 0.80 \\ 0 & 0 & 1 \\ 0.95 & 0.90 & 0 \end{bmatrix}. \square$$

After the relative relatedness function and the category matrix have been introduced, the important order of categories, l_{course}, can be constructed. Here, we use two symbols to indicate different orders: “>“ means more important and “~” means undecided (still can not decide at present time).

Example 3. (Follows Example 2) The important order of categories

for course Neural Networks (NN for short):

$$f_{AI}(NN) = 0.9 > f_{Research}(NN) = 0.85 > f_{Math}(NN) = 0.70$$
$$\Rightarrow l_{NN} = AI > Research > Math$$

for course Computer Mathematics (CMath for short):

$$f_{Research}(CMath) = 0.90 \sim f_{Math}(CMath) = 0.90$$
$$\Rightarrow l_{CMath} = Research \sim Math . \square$$

Step 4 is the most important and difficult part in the recommendation process. In this step, we need to merge all important orders from step 3 into a preference sequence for the student. There are four merge rules proposed by us:

Rule 1. Transitivity: $l_a > l_b , l_b > l_g \rightarrow l_a > l_b > l_g$

Rule 2. Inconsistency: $l_a > l_b , l_b > l_a \rightarrow l_a \sim l_b$

Rule 3. Undecided after merging: $l_a , l_b , l_a \cap l_b = \varnothing \rightarrow l_a \sim l_b$

Rule 4. Union: $\begin{matrix} t_m \\ t_m > l_a \end{matrix} \Rightarrow t_m > l_a$; $\begin{matrix} t_m \\ l_a > t_m \end{matrix} \Rightarrow l_a > t_m$

Rule 5. Implicitly more important: $t_m , t_m \sim l_a \rightarrow t_m \bar{>} l_b$

Example 4.

$t_1 > t_2 , t_1 > t_3$ are merged into $t_1 > t_2 \sim t_3$;

$t_1 > t_3 , t_2 > t_3$ are merged into $t_1 \sim t_2 > t_3$. $\square$

If we think that a perfect preference sequence can be generated depends only on those merge rules above, then some different sequences derived from same sources will surprise us.

Example 5. Given $l_a = t_1 , l_b = t_1 > t_3 , l_g = t_2 > t_3$

1st sequence: merge l_a and l_b first

$$l_a = t_1 , l_b = t_1 > t_3 \rightarrow t_1 > t_3$$

$$t_1 > t_3,\ l_g = t_2 > t_3 \to t_1 \sim t_2 > t_3$$

2nd sequence: merge l_b and l_g first

$$l_b = t_1 > t_3,\ l_g = t_2 > t_3 \to t_1 \sim t_2 > t_3$$

$$t_1 \sim t_2 > t_3,\ l_a = t_1 \to t_1 > t_2 > t_3 . \square$$

According to Example 5's demonstration, Graph Theory is taken into consideration. A directed graph (DG) is an important order, and two graphs can be merged simply without side-effect suchlike Example 5 shown. (Fig. 4) shows the graph-form expression of the important order. Although there is also an algorithm is designed for merging two directed graphs, we ignore in this paper please see (Chu et al., 2002) instead. So, once the preference sequence of student has been generated, appropriate courses are able to provide for students.

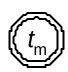
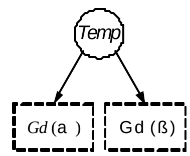
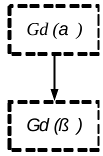
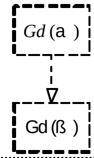
?=	*Dd* (?)	Description	?=	*Dd* (?)	Description
t_m		***SINGLE***	*a ~ b*		Undecided (an extra Temp node to link a and ß; if ?>*a*, and ? > ß, then the Temp node can be replaced by a ? node)
a > b		more important	*a* $\gtrsim$ *b*		implicitly more important

Figure 4: Graph-form expression of the important order

3. Results and Evaluation

In order to accomplish the course recommendation system on Web, the prediction methodology mentioned in previous Section should be integrated into the process of course selection as (Fig. 5) shown below. Based on the (Fig. 5), the architecture with snapshots of the course recommendation system is designed in (Fig. 6). (Fig. 7) and (Fig. 8) represent the step 4 and step 6 in (Fig. 5).

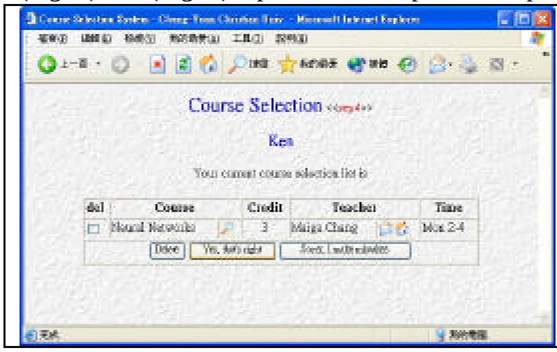


Figure 7: Confirmation step (step 4)

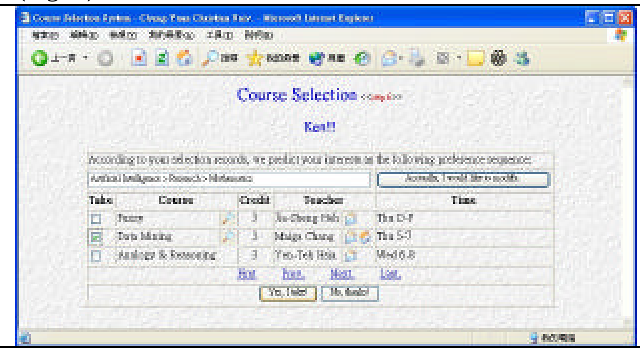
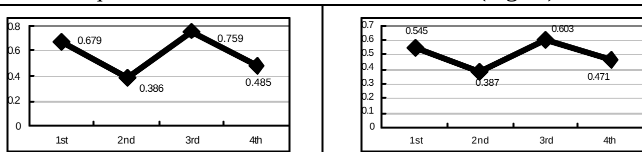


Figure 8: Course Recommendation step (step 6)

Via the course selection system we collected course selection records of 4 consecutive school terms from senior college students in the Department of Information and Computer Engineering, and also conducted a questionnaire for the same two classes of students (the class of 2001 and 2002). About the class of 2001 who have graduated already, although the system has analyzed the course selection records of 127 students, out of 83 questionnaires we sent out, only 34 have responded. About the class of 2002, the system has analyzed the course selection records of 102 students, and all 102 of them have responded to our questionnaire.

With the help of professors in the dept. of ICE, all of the information science-related courses (excluding the prerequisite courses) are divided into six categories, including research, theory, mathematics, hardware, software and network. The accuracy rate of the preference sequence between our deduction outcomes and students' questionnaire returns can be seen in (Fig. 9).



(a) (b)

Figure 9: Accuracy rate for preference sequence of (a) class 2001; (b) class 2002

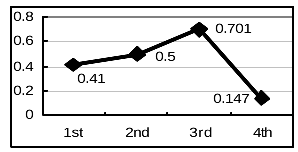


Figure 10:
Accuracy rate of the graduate school candidates among class 2002

The trends of accuracy rate on these two figures are similar. Both figures indicate the 3rd term has the highest accuracy rate. It contradicts the general presumption of the closer a term is to graduation, the more likely a course is selected based on a student's personal preference; in other words, the 4th term should have the highest accuracy rate. Instead, we found the 4th term in both (Fig. 9)(a) and (Fig. 9)(b) declined sharply in accuracy rate. The drop, according to a subsequent analysis, can be explained as most students attending the last term in school have fewer prerequisite courses and more elective courses to choose from, therefore they are more inclined to deviate from their previous pattern and try something different. In addition, because some of them are applying for or already being accepted by graduate schools, they tend to follow graduate school instructor's guidance, which sometimes deviate considerably from their prior perception of the order of relative importance, to take courses more related to their future studies.

In order to verify whether those who plan to go to graduate school straight after college are the main cause for the drop of accuracy rate at the 4th term, we targeted specifically such students (only 13 people) among the class 2002 (102 people) and did an analysis. The result is shown on (Fig. 10). Obviously, for those who intend to go to graduate school, the accuracy rates for the first three terms increase steadily and peak at the 3rd term.

4. Conclusion

A web-based course recommendation system is constructed in this paper according to our previous research results. Actual course selection records of two classes (the class of 2001 and 2002) during two academic years are collected through the course selection system. By using our recommendation process, the preference sequence of students can be generated. With the sequence, the most appropriate courses then are able to find out for making suggestions to students. However, since we don't know how close the preference sequence between we generated and students' thoughts, a questionnaire is also designed for evaluating purpose. Finally, after compared the deduction outcome with the survey result, our system and methodology achieved a reasonable level of accuracy rate. It is worth to note that our system is not only can used for university or college, but also available for any level of education.

From now on, there are still several works we can do:

1. How to find course categories classified by students?
2. Time series analysis should be made, which means can we predict the students' change of interest from year to year.
3. Course development: can we develop or plan a series of courses depends on the student's major interests? (sort of professional education.)

References

[1] Agrawal, R., & Imielinski T., & Sqami A. (1993). Mining Association Rules between Sets of Items in Large Databases. *Int'l Conf. Management of Data (SIGMOD'93)*, ACM, Washington, DC. 207-216.

[2] Agrawal, R., & Srikant, R. (1994). Fast Algorithms for Mining Association Rules. *Int'l Conf. Very Large Databases (VLDB'94)*, Santiago, Chile. 487-499.

[3] Chu, Ko-Kang, & Hsia, Yen-Teh. (2001). Formula of Relative Importance for Modeling User Behavior. *Proceedings of the Sixth Conference on Artificial Intelligence and Applications (TAAI 2001)*, Nov. 9, 2001, Kaohsiung, Taiwan. 444-449.

[4] Chu, Ko-Kang, & Hsia, Yen-Teh. (2002). Deducing the Importance of Course Categories from Course Selection Records. *Chung-Yuan Journal*, 30 (3), 401-410.

[5] Han, J., & Fu, Y. (1995). Discovery of Multiple-Level Association Rules from Large Data Bases, *Int'l Conf. Very Large Databases (VLDB'95)*, Zurich, Switzerland. 420-431.

[6] Han, J., & Kamber, M. (2001). *Data Mining Concepts and Techniques*. Morgan Kaufmann.

[7] Klemettinen, M., & Mannila, H., & Ronkainen, P., & Toivonen, H., & Verkamo, A. I. (1994). Finding Interesting Rules from Large Sets of Discovered Association Rules. *Int'l Conf. Information and Knowledge Management*, 1994, Gaithersburg, MD. 401-408.

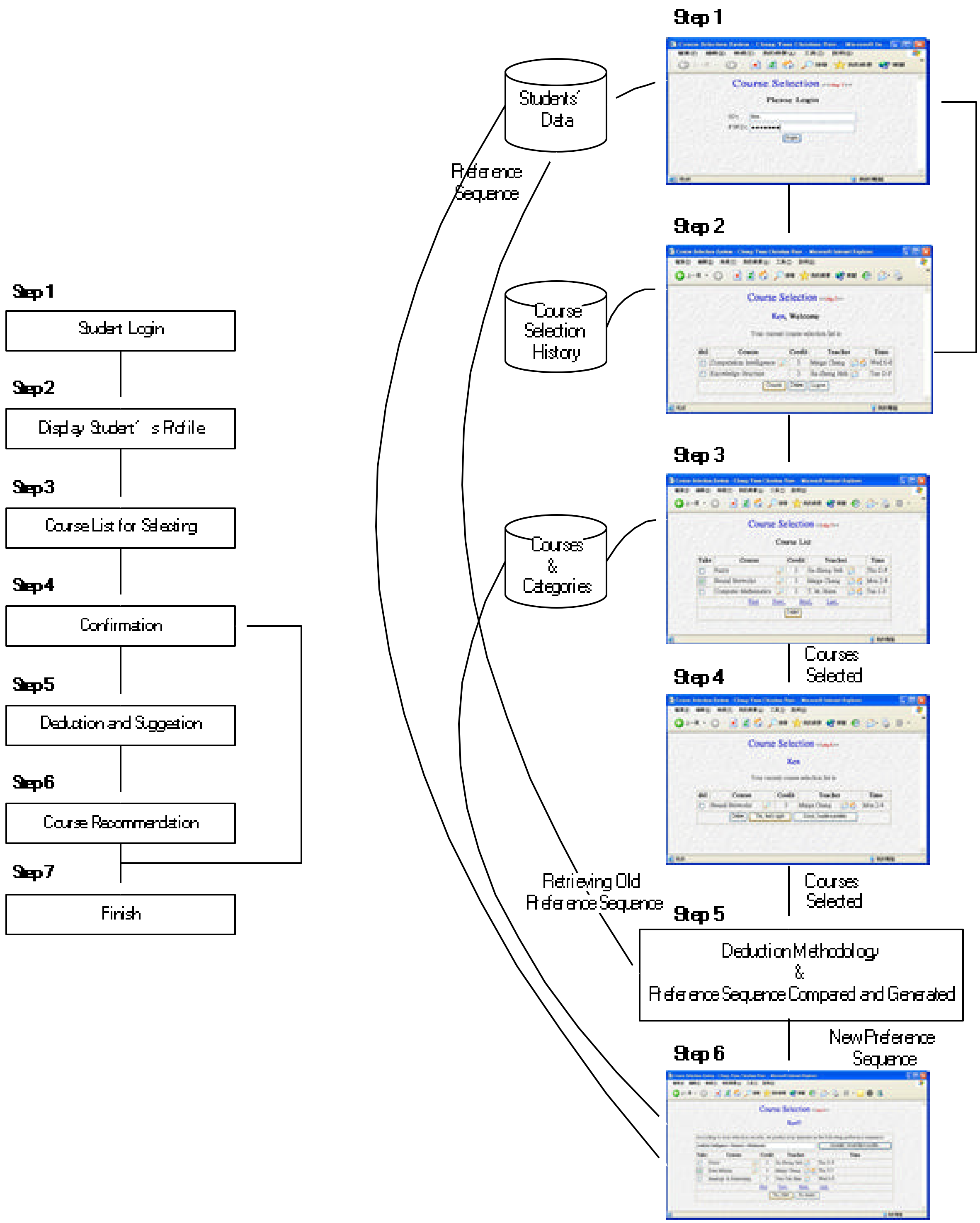
[8] Miller, R. J., & Yang, Y. (1997). Association Rules over Interval Data. *Int'l Conf. Management of Data (SIGMOD'97)*, ACM, Tucson, AZ. 452-461.

[9] Mannila, H., & Toivonen, H., & Verkamo, A. I. (1994). Efficient Algorithms for Discovering Association Rules. *AAAI'94 Workshop Knowledge Discovery in Database (KDD'94)*, *AAAI*, Seattle, Washington. 181-192.

[10] Ozden, B., & Ramaswamy, S., & Silberschatz, A. (1998). Cyclic Association Rules. *Int'l Conf. Data Engineering (ICDE'98)*, Orlando, FL. 412-421.

[11] Srikant, R., & Agrawal, R. (1995). Mining Generalized Association Rules. *Int'l Conf. Very Large Data Bases (VLDB'95)*, Zurich, Switzerland. 407-419.

[12] Srikant R., & Agrawal R. (1996). Mining Quantitative Association Rules in Large Relational. *Int'l Conf. Management of Data (SIGMOD'96)*, Montreal, Canada. 1-12.


Step 1
Student Login
Step 2
Display Student's Profile
Step 3
Course List for Selecting
Step 4
Confirmation
Step 5
Deduction and Suggestion
Step 6
Course Recommendation
Step 7
Finish
Students' Data
Preference Sequence
Course Selection History
Courses & Categories
Courses Selected
Retrieving Old Preference Sequence
Deduction Methodology & Preference Sequence Compared and Generated
New Preference Sequence
Course Selection