The 2008 International Conference in Management Sciences and Decision Making
Taipei, R.O.C., June 28, 2008

Faculty Development by Student Questionnaire Analysis: A Class Partition Problem

Shigeichi Hirasawa†, Takashi Ishida††, and Masayuki Gotoh‡

† School of Science and Engineering, Waseda University,
3-4-1, Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan
email: hira@waseda.jp

†† Media Network Center, Waseda University,
3-4-1, Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan
email: t.ishida@aoni.waseda.jp

‡ Faculty of Environmental and Information Studies, Musashi Institute of Technology
3-3-1, Ushikubo-nishi, Tsuzuki-ku, Yokohama-shi, Kanagawa 224-0015, Japan
email:goto@yc.musashi-tech.ac.jp

Abstract: Recently, universities are faced to introduce useful and effective programs for class management and to improve the quality of education. To perform these activities, student questionnaires are often used. By establishing a class model, we have evaluated student characteristics, the degree of satisfaction and final scores, and their relationships for a set of students or subsets of them. To improve the degree of satisfaction of the students and to increase the effectiveness of education, we are going to try to partition the students of the class into a few subclasses depending on their interested areas, levels, or motivation before beginning the class.

Based on the probabilistic latent semantic indexing (PLSI) model, we developed new document classification and clustering methods. These methods show good performances for a small set of documents. Applying these methods to the students for the class: "Computer Engineering" at the second year of our department, we discuss on partition by the contents of topics into Class G (generalist): wide and shallow technical topics, and Class S (specialist): deep technical and professional topics, by only using the initial questionnaire. Although we know that most of all students do not decide their future jobs yet at their second academic year, we evaluate the difference of partition of the classes between results given by an automatic partition method, those by student's own choice and those by estimated their jobs which they got after graduation from the university. Although the results obtained by the automatic partition method can not explain enough their jobs after graduation, it would be still useful to assist or guide the students. We show the effectiveness of the questionnaire by deriving the student implicit characteristics and we can obtain useful information which leads to faculty developments.

Keywords: student questionnaire, classification algorithm, faculty development, PLSI model

1 Introduction

In recent two decades, declining birth rate brought competitive intensity among Japanese universities not only in research activities but in educational activities. For many universities in Japan, the evaluation of the quality assurance of education programs by Japan Accreditation Board for Engineering Education (JABEE) has become important.

To improve the class management, our university introduced the student questionnaire system using Web-site. It is, however, not enough for improving in detail with taking into account the special situation of the class. Especially, there have been existing difficult problems for class management in the class: "Computer Engineering" at the 2nd academic year, Department of Industrial and Management Systems Engineering, Waseda University, since the students in this class have different qualities of a prior knowledge, interested areas, motivation, experiences, and levels in computer skills [4]. Moreover, their jobs after graduation have many kinds of fields in business. Therefore, we have designed the student questionnaire only applicable to this class [4],[9],[10]. The student questionnaire has two types of replies: item type and text type [5]. The former is fixed format such as answered by selected numbers, symbols, and yes or no. The latter, free format answered by text.

In this paper, we focus upon the class partition problem. It arises to this class, because we intend to perform effective education by supplying different contents of topics to different sorts of students. We discuss on partition into two classes which have two contents of topics and are represented as Class G (generalist): wide and shallow technical topics, and Class S (specialist): deep technical and professional topics by only using the initial questionnaire.

Based on the probabilistic latent semantic indexing (PLSI) model [8], we have proposed the classification[1] and clustering algorithms [5],[13]. These methods exhibit good performances for a small set of documents [6]. Besides the proposed algorithm, we have also developed the extraction algorithm of important sentences, feature words, and feature sentences used for the text part [12],[14],[15]. The traditional statistical techniques can also be applicable to the item part.

We have applied the student questionnaire for these six years (2002-2007), where the contents of the student questionnaire are the same except for the first year (2002). The students at the 2nd year in 2003 graduated in March 2006 as Bachelors, and those in part, in March 2008 as Masters. Then we know the type of business after their graduation. Estimating the kind of their jobs from the type of business, we would try to guess their true choice from Class G or Class S.

By analyzing the student questionnaire, coincident rates between results by the automatic partition, those by student's own choice, and those by student's estimated true choice are derived. Although the results obtained by the automatic partition method can not explain enough their future jobs, it would be still useful to assist or guide the students. As a result, we can obtain useful information which leads to faculty developments.

In section 2, we represent the target class and its partition problem. Methods for analyzing the student questionnaire are briefly described in section 3. Section 4 discusses results of analysis. Conclusions are given in section 5.

2 A Class and its Partition Problem

2.1 Target class

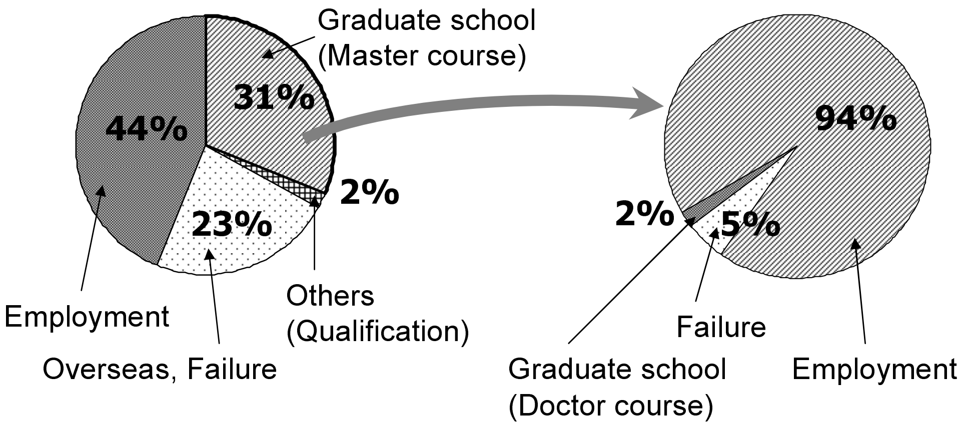
The target class of this paper is summarized in Table 2.1. Since the students who study at this class are those of the department of industrial and management system engineering, their future jobs are existing in many type of business, such as trading, banking, finances, consultants, and manufacturing which includes electric or electronic companies, automobile manufacturing companies, software development companies and so on.

In this paper, we will analyze the student questionnaire in 2003. Since the object students were at the 2nd academic year in 2003, they graduated in March 2006 from undergraduate school receiving their Bachelor Degree in Engineering, and a part of them also graduated in March 2008 from graduate school receiving their Master Degree in Engineering. After graduation, they had their jobs. The paths of object students in March 2006 and in March 2008 are shown in Fig. 2.1. The types of business of the object students are shown in Fig. 2.2. Table 2.2 indicates the name of major companies in which the object students have joined.

[1] We use the term in this paper "partition" rather than "classification", since the students must choose a subset of the class, and cannot choose two or more subsets of the class. Strictly speaking, partition implies mathematically as follows: Let a set A has its subsets A_i, where $A = \bigcup_i A_i$. Then A_i is a partition of A iff $A_i \cap A_j = \phi$ or $A_i = A_j$ for $i \neq j$.

Table 2.1: Target Class

Class name	Computer Engineering
Credit	2 units (90 min lecture/week, at Spring Semester)
Subject	Obligatory at the 2nd academic year
Students	Department of Industrial and Management Systems Engineering
Topics (at present)	1. Fundamental concept of computer (Neumann architecture, etc.) 2. Computer architecture (stack machine, instruction set, binary system, processor architecture, etc.) 3. Hardware (Boolean algebra, logic design, combinatorial circuit, etc.) 4. Software (operating system, Kernel, Unix, etc.)



(a) The paths of object undergraduate students in March 2006.

(b) The paths of object graduate students in March 2008.

Figure 2.1: The paths of object students

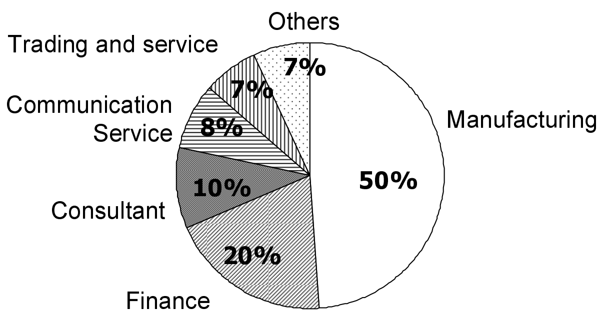


Figure 2.2: Business areas of object students

Table 2.2: List of major companies

[Manufacturing]
- Canon Inc.
- Nihon Unisys, Ltd.
- Suntory Limited
- Sharp Inc.
- Sony Corp.
- Toshiba Corp.
- TORAY Ltd.
- IBM Japan Ltd.
- NEC
- Nissan Motor Co., Ltd.
- Fujitsu Ltd.
- Honda Motor Co., Ltd.
- Matsushita Electric Industrial Co., Ltd.
- Mitsubishi Electric Corp.
- Astellas Pharma Inc.

[Consultants]
- Accenture
- CSK Systems Corp.
- Deloitte Touche Tohmatsu. Japan Inc.
- The Japan Research Institute, Ltd.
- Nomura Research Institute, Ltd.
- Pricewaterhouse Coopers, International Ltd.
- Mitsubishi Research Institute, Inc.

[Finance]
- The Goldman Sachs Group, Inc.
- The Bank of Tokyo-Mitsubishi UFJ Ltd.
- Sumitomo Mitsui Banking Corp.
- Mizuho Bank, Inc.
- Nomura Secureties Co., Ltd.

[Communication Services]
- NTT Data Corp.
- Nippon Telphone and Telegraph East Corp.

[Trading and Services]
- East Japan Railway Company
- Hakuhodo Inc.
- Mitsui and Co. Ltd.

[Others]
- Kashima Corp.
- Nikkei Corp.
- The Mainichi Newspapers

2.2 Class model

To design the student questionnaire, we have proposed a class model for the target class as shown in Fig. 2.3 [4],[10].

The implicit characteristics of each student are essentially measured by questionnaire, while explicit characteristics are objectively given by numerical data of the class. These characteristics generate explanatory variables. Then each student yields his or her final score and the degree of satisfaction as the result of the class. The degree of satisfaction is also measured by questionnaire. The final score and the degree of satisfaction play as criterion variables in this model. Besides these variables, we can expect to get some information regarding to the class management such as partition of the class.

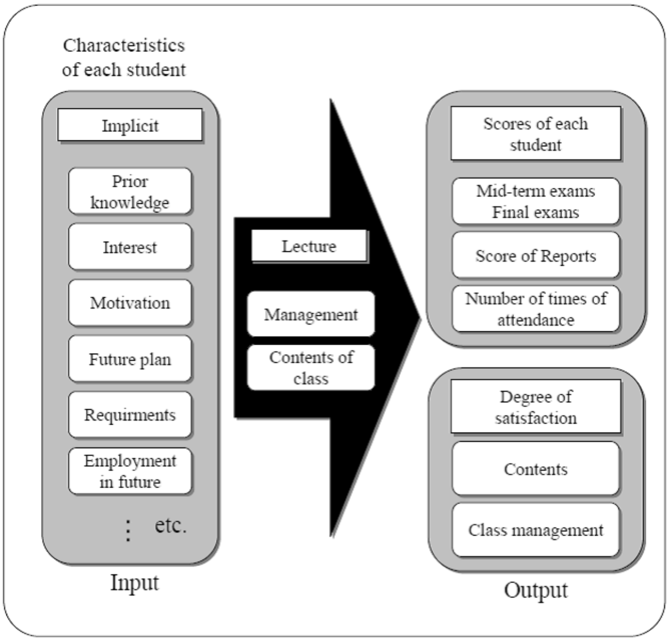


Figure 2.3: Class model

Table 2.3: Contents of topics

Class	Contents
Class G	- History of computers, fundamental concepts in computer - Basics of architecture - Basics of hardware - Basics of software - Applications of information technology (information transmission systems, computer networks and internet, information secrity and PKI, data base, information retrieval system, AI) etc.
Class S	- Architecture (binary system, stack machine, processor architecture, memory architecture) - Hardware (logic design, logical circuit, automaton) - Software (operating system, UNIX, language processor) etc.

2.3 Class partition

The facts shown in Figs. 2.1 and 2.2 lead to investigate the possibility of class partition to increase the degree of satisfaction.

Usually there exist many differences between each student such as in a prior knowledge, interested areas, motivation, experiences and levels before beginning the class. Hence a proper partition of the class depending on features such as the future plan of each student is desirable. The partition shown in Table 2.3 can be considered depending on the contents of topics of the lecture, where G stands for a generalist, S, for a specialist by predicting his or her future job.

2.4 Design of questionnaire

A questionnaire is applied to the target class. It consists of the initial questionnaire (IQ) and the final questionnaire (FQ). Scores of technical report (TR) submitted every week, and those of the midterm exam (ME) and final exam (FE) are explicit characteristics of each student. The data of the class, and the contents of a questionnaire and their examples are shown in Table 2.4 and in Table 2.5, respectively. The time schedule for the class is depicted in Fig. 2.4.

We analyze them by using statistics, data mining, and information retrieval techniques which include partition of a set of documents.

Table 2.4: Data of class

Exercise	Contents
Initial Questionnaire (IQ)	
Item type	7 questions (4-20 sub-questions each)
Text type	5 questions (250-300 characters in Japanese and 100 in Chinese each)
Midterm Exam (ME)	5 subjects
Technical Reports (TR)	11 times (each 1-2 subjects)
Final Exam (FE)	5 questions
Final Questionnaire (FQ)	
Item type	6 questions (6-21 sub-questions each)
Text type	5 questions (250-300 characters in Japanese and 100 in Chinese each)

Table 2.5: Contents of questionnaire

Exercise		Examples (sub questions)
IQ	Item-type	✓ For how many years have you used computers? ✓ Do you have a plan to study abroad? ✓ Can you assemble a PC? ✓ Do you have a qualification related to information technology? ✓ Write 10 technical terms in information technology which you know.
	Text-type	✓ Write about your knowledge and experience on computer. ✓ What kind of work will you have after graduation? ✓ What do you imagine from the name of this class subject name?

Exercise		Examples (sub questions)
FQ	Item-type	✓ Could you understand the contents of this lecture? ✓ Was the midterm test difficult? ✓ Was it easy to read the handwritings on the white-board? ✓ Do you think the contents of this lecture to be useful to yourself? ✓ Do you want to finish this course even if it is optional? ✓ Which are you interested in applied technology or the fundamentals of computers? ✓ Which do you choose class (S) or class (G)?
	Text-type	✓ Do you want to be a member of laboratories related to the information technology? ✓ In the future, will you get a job in industries related to the information technology? ✓ Did your image on computers change after taking this lecture?

This questionnaire is made in WEB form, and it is on the following Web Site.
http : //hirasa.mgmt.waseda.ac.jp/users/comp-eng/

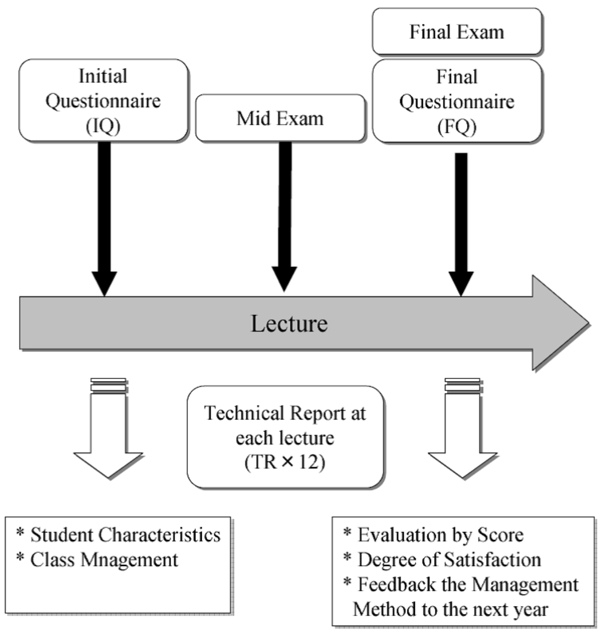


Figure 2.4: Time schedule for class

3 Methods for Analysis

3.1 Document set

Documents called in this paper imply the replies of student questionnaire.

The documents with fixed formats are represented by an item-document matrix $G = [g_{mj}]$, where $g_{mj} \in \{0, 1\}$ is the selected reply of the item m (i_m) in the document j (d_j). The documents with free formats are also represented by a term-document matrix $H = [h_{ij}]$, where $h_{ij} \in \{0, 1, 2, \cdots\}$ is the frequency of the term i (t_i) in the document j (d_j). The dimensions of matrices G and H are $I \times D$, and $T \times D$, respectively, where the number of the total documents is D, that of the total items, I, and that of the total terms, T. Both matrices are compressed into those with smaller dimensions by the probabilistic decomposition in PLSI model [2],[8] similar to the single valued decomposition (SVD) in LSI (latent semantic indexing) model [1]. The (latent) states are denoted by $z_k \in \mathcal{Z}$ ($k = 1, 2, \cdots, K$). Introducing a weight λ ($0 \le \lambda \le 1$), the log-likelihood function corresponding to the resultant matrix A:

$$A = \begin{bmatrix} \lambda G \\ (1-\lambda)H \end{bmatrix} = [a_{ij}]$$
$$(i = 1, 2, \cdots, I+T, \quad j = 1, 2, \cdots, D) \tag{3.1}$$

is maximized by the EM algorithm [8]. Then we obtain the probabilities $\Pr(z_k)(k = 1, 2, \cdots, K)$, and the conditional probabilities $\Pr(t_i|z_k, i_m)$, and $\Pr(d_j|z_k)$. Using these probabilities, $\Pr(i_m, d_j)$ and $\Pr(t_i, d_j)$ are derived, and we decide the state for d_j depending on $\Pr(z_k|d_j)$.

The similarity function between z_k and $z_{k'}$, $s(z_k, z_{k'})$ is defined by [11]:

$$\begin{aligned} s(z_k, z_{k'}) = \sum_i \Big\{ & h\big[\alpha \Pr(t_i|z_k) + (1-\alpha)\Pr(t_i|z_{k'})\big] \\ & -\alpha h\big[\Pr(t_i|z_k)\big] - (1-\alpha) h\big[\Pr(t_i|z_{k'})\big]\Big\} \end{aligned} \tag{3.2}$$

where $0 \le \alpha \le 1$ and $h[x] = -x \log x$ ($0 \le x \le 1$).

Assume that pairs (i_m, d_j) and (t_i, d_j) are generated independently, and also assume that i_m and t_i are generated independently of d_j conditioned on z_k. We construct the matrix A so that the above assumptions hold. Based on the good performance for a relatively small document set [2] discussed in the previous paper [5],[6],[13] and the further improvement of it [11], we have used the following algorithm.

3.2 Partition algorithm [5]

The algorithm is constructed strongly dependent on the fact and property that the EM algorithm usually converges to the local optimum solution from starting with an initial value. Hence we use a representative (pseudo) document as the initial value for the EM algorithm.

Suppose a set of documents $\mathcal{D}$ for which the number of classes is K, where the K classes are denoted by $C_1, C_2, \cdots, C_K$.

(1) Choose a subset of documents $\mathcal{D}^*$ $(\subset \mathcal{D})$ which are already categorized. Compute representative document vectors $\boldsymbol{d_1^*}, \boldsymbol{d_2^*}, \cdots, \boldsymbol{d_K^*}$:

$$\begin{aligned} \boldsymbol{d_k^*} &= \frac{1}{n_k} \sum_{\boldsymbol{d_j} \in C_k} \boldsymbol{d_j} \\ &= (a_{1k}^*, a_{2k}^*, \cdots, a_{(I+T)k}^*)^{\mathrm{T}} \end{aligned} \tag{3.3}$$

where n_k is the number of selected documents to compute the representative document vector from C_k and $\boldsymbol{d_j} = (a_{1j}, a_{2j}, \cdots, a_{(I+T)j})^{\mathrm{T}}$, where T denotes the transpose of a vector. Set the initial values as:

$$\Pr(z_k) = \frac{1}{K}, \tag{3.4}$$

$$\Pr(d_j|z_k) = \frac{1}{D}, \tag{3.5}$$

$$\Pr(t_i|z_k) = \frac{a_{ik}^* + \alpha}{\sum_{i'}(a_{i'k}^* + \alpha)}, \quad (\alpha > 0). \tag{3.6}$$

(2) Compute the probabilities $\Pr(z_k)$, $\Pr(d_j|z_k)$ and $\Pr(t_i|z_k)$ which maximizes the log-likelihood function corresponding to the matrix A by the Tempered EM (TEM) algorithm, where $|\mathcal{Z}| = K$.

(3) Decide the class $C_{\hat{k}}$ for d_j as

$$\max_k \Pr(z_k|d_j) = \Pr(z_{\hat{k}}|d_j) \Rightarrow d_j \in C_{\hat{k}}. \tag{3.7}$$

□

[2]Note that algorithms used in this paper are required to exhibit good performance to a set of a small number of documents, since the number of the students in a class will be usually at most 200.

3.3 Extraction algorithm of important sentences [14]

A document is composed of a set of sentences. Measure the similarities between a sentence and the other sentences, and compute the score of the sentence by the sum of the similarities. Then choose a sentence which has the largest score as the important sentence in the document.

3.4 Extraction algorithm of feature sentences and feature words [12]

Let $\Pr(t_i|z_k) - \Pr(t_i)$ be the score of t_i, and let the sum of the scores of t_i's which appear in a sentence be the score of the sentence. Then choose the words which have the larger scores as the feature words. Similarly, choose a sentence which has the larger scores as the feature sentence in the class.

4 Questionnaire Analysis

It is easy to decide by students themselves whether they choose Class G or Class S. The reason why we apply student questionnaire (by initial questionnaire (IQ)) is to extract the student characteristics which are not awaked by them and to try to adequately partition the classes depending their latent properties at the beginning of the class.

We use the term "job" as the kind of occupation such as:

(S) circuit design, mechanical design, electric design, production management, quality control, software development, system engineering, R&D, and so on, and

(G) sales, accounting, personal management, services, and so on.

The former (S) is a type of engineering or technology, while the latter (G) is not the type of them. Hence (S) would require professional skills in computer, and (G), does not so much.

On the other hand, we use the "business" as the kind of company such as

(a) trading, finance, banking, service, securities market, consultation, general construction, and so on, and

(b) electric manufacturing, automobile manufacturing, precision instrument manufacturing, system integration, software development, and so on.

4.1 Estimation of the job

After graduation, the students join companies. Although we know only the name of companies in which they joined, the job of each student is estimated by the authors according to their experience. As the result, there is a difficulty to estimate the job from the name of company, which means estimation of (G) or (S) from (a) or (b), where (a) and (b) are classified by the name of companies such as Canon Inc., IBM Japan Ltd., NEC, Toyota Motor Corp., Accenture, Nomura Research Institute Ltd., East Japan Railway Co., Kashima Corp., and so on. This difficulty arises an ambiguity of decision in the label (G) and (S). In other words, the name of company does not directly fix on the job. For example, we know one of the graduated students joins Sony Corp., but we don't know his job. His job may be production management, or may become sales. As another example, she joins Tokyo Mitsubishi UFJ Bank, but her job may be a teller, or may be a staff at information service division. The authors' experiences avoid an error in estimation by their knowledge on the jobs of graduated students as possible as they can.

4.2 Results of partition

By using IQ, partition algorithm is performed, where the representative documents for Class G and Class S are generated by those of graduated students in March 2007. The results obtained by it are indicated as "Automatic Partition (AP)". By using FQ, each student replies his own choice of Class G or Class S. The results obtained by it are indicated as "Student's Own Choice (SOC)". By the estimation of the job of the graduated each student, the results obtained by it are indicated as "Student's Estimated Choice (SEC)".

Table 4.1 shows the list of the number of partitioned students between AP and SEC. Table 4.2 also shows that between SOC and SEC.

Table 4.1: Numbers of partitioned students between AP and SEC

		SEC		Total
		G	S	
AP	G	20	19	39
	S	17	30	47
	Total	37	49	86

AP: Automatic Partition
SEC: Students Estimated Choice

Table 4.2: Numbers of partitioned students between SOC and SEC

		SEC		Total
		G	S	
SOC	G	30	24	54
	S	7	28	35
	Total	37	52	89

SOC: Sutudent's Own Choice

4.3 Results of extracted important sentences

The extraction of the important sentences of the documents is examined for the students appeared in the elements except for diagonal elements in Table 4.1. The results are shown in Table 4.3.

Table 4.3: Extracted important sentences

(a) AP vs. SEC

(AP, SEC)=(Class G, Class S)

[IQ]	- I think that what is necessary is just to be able to master a computer. - What I am reminded of from the term "computer" is a personal computer. - I would like to be able to master a computer.
[FQ]	- It was meaningful that the knowledge of the computer was able to be acquired. - In the future, I think that I will associate with a computer for a long time. - I thought that it was not so difficult to understand the structure of a computer.

(AP, SEC)=(Class S, Class G)

[IQ]	- I would like to decompose by myself or to set up a personal computer. - I am very interested in the content of the class.
[FQ]	- I did not think that this class was not much important for myself. - I was not able to acquire the impression that this field was interesting. - Although it is not interested in a computer, I think that knowledge is required.

(b) SOC vs. SEC

(SOC, SEC)=(Class G, Class S)

[IQ]	- I would like to be able to master a computer. - Since I was imagining that I used a personal computer in this lesson, it differed from prior imagination.
[FQ]	- My view about a computer changed by having studied the principle of the computer. - From now on, I will associate with a computer for a long time. - The content of the class was difficult. - It was serious to have understood the content of the class. - I am interested in how to use a computer.

(SOC, SEC)=(Class S, Class G)

[IQ]	- I would like to understand the principle of a computer. - It is required to understand a principle, in order to master a computer.
[FQ]	- I would like to study a computer more and to obtain a deeper understanding. - In order to master a computer, it is helpful to know the structure.

4.4 Discussions

(1) It is shown that the coincident rate between AP and SEC is approximately 58.1% by IQ only (Table 4.1), and that between SOC and SEC, 65.1% by FQ (Table 4.2). The method for partitioning the class is probably not accurate enough, although the rate of the latter is slightly improved.

(2) It can be explain that the above improvement is brought by learning the subjects, since FQ is performed at the end of the class.

(3) Table 4.2 suggests us that the student at the 2nd academic year do not decide their future jobs. Hence they do not awake whether professional skill is required or not in future.

(4) From the view-point of the hypothesis testing, under the hypothesis H_0 : Two variables are independent, H_0 for Table 4.1 cannot be rejected, while H_0 for Table 4.2 can be rejected (See Appendix A).

(5) Although the coincident rates are not large, partition is still useful to guide the students by the suggestions: There are cases such as

  (i) Even though the student becomes a generalist, he who interested in computers, would chose Class S (Table 4.3 (a)).

  (ii) There are many cases such that if the student wanted to learn only the method for using computers, he who graduated as a Master, will join an industry as a specialist (Table 4.3 (a)).

  (iii) If the student who wanted to be a specialist, could not be interested in computers, he will becomes a generalist (Table 4.3 (a)).

  (iv) In contrast to (iii), there is a case such that the student who was interested in such as the structure of computers, will go to professional in engineering (Table 4.3 (a)).

  (v) If the student who chose Class G, changed his idea by learning the principle of computers, he becomes a specialist (Table 4.3 (b)).

  (vi) Even if the student felt that the lecture was difficult, he will become a specialist (Table 4.3 (b)).

  (vii) Since recent students usually chose easy way, there is a case that he who want to become a specialist, joins the Class G.

(6) Most of all students state that they will satisfy fruitful and interested contents of the lecture, and their choice of the Class S or Class G depends on the topics. Therefore, the contents of topics are very important.

As an additional experiments, if we use FQ, we can partition the students into Class G and Class S with high coincident rate by weighting the following items.

1. [IQ] Prior knowledge (technical term).
2. [FQ] The range of the theme is suitable?
3. [FQ] I would like to study about a logic circuit.
4. [FQ] I would like to study about cache memory.

5 Concluding Remarks and Future Works

Collecting documents obtained by student questionnaire for these six years, we analyze the graduated student questionnaire by trace back to their 2nd academic year. It is necessary to collect data at least four years for taking account the estimated their jobs. The results obtained in Section 4 are not accurate enough to use automatic partition of the class, but it is still useful to assist and to consult the students. We know that almost all students do not decide their future jobs yet in their 2nd academic year. It proves, however, that students are sound and have some robustness in their future plan, in a sense that they are going to learn not only their future job but their unsophisticated thirst for knowledge.

As a trial in the latter part of the semester in 2007, we have partitioned the class into two classes whose contents of topics are shown in Table 5.1, where A stands for application, and B, basic. The results obtained by the student questionnaire are summarized in Fig. 5.1. From this figure, it is worth to note that:

(1) The reason for the choice of the course is strongly dependent on their contents of interested topics. This corresponds to the previous result, i.e., the degree of satisfaction depends on the contents of the lecture [7].

(2) The degree of satisfaction for 90% of the students is in high (including in very high). This suggests us that we have to update the topics so that we let the students be always interested in.

(3) The 2/3 students support the introduction of the course system. This leads us to introduce the class partition into Class G and Class S.

Table 5.1: Contents of topics for Course S and Course G

Course	Contents	
Course G	1. Fundamental concept of computer (Neumann architecture, etc.)	5. Applications of information technology (information transmission systems, computer networks, Internet, information security and PKI, etc.)
Course S	2. Computer architecture (stack machine, instruction set, binary system, processor architecture, etc.)	3. Hardware (Boolean algebra, logic design, combinatorial circuit, etc.) 4. Software (operating system, Kernel, Unix, etc.)

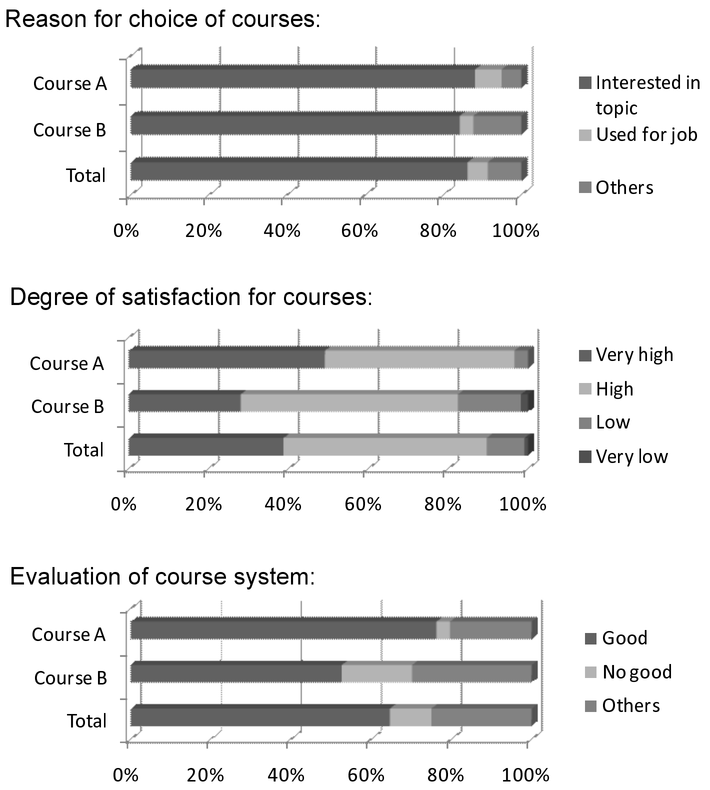


Figure 5.1: Evaluation of course system from FQ

Based on the results of the above trial, we will introduce the class partition in next year, although careful and detailed investigations are required as further works.

Interpreting the questionnaire in Japanese into in Chinese [7], we can apply it to students in R.O.C. This is also remained as a future study.

Acknowledgement

The authors would like to thank Professor Fu-Yih Shih at Leader University, R.O.C. and Professor Wei-Tzen Yang at Tamkang University, R.O.C. for their helpful activities for appling the student questionnaire in R.O.C.

They also thank Mr. M. Kumoi at Waseda University for his valuable work to this study and Dr. H. Yagi at University of Electro-Communications for his helpful joint investigation to this study.

A part of this research was supported by the Grant of the Telecommunication Advancement Foundation (TAF).

References

[1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, New York, 1999.

[2] D. Cohn, and T. Hofmann, "The missing link - A probabilistic model of document content and hypertext connectivity," *Advances in Neural Information Processing Systems* (NIPS ×13), MIT Press 2001.

[3] M. Goto, T. Ishida, and S. Hirasawa, "Representation method for a set of documents from the viewpoint of Bayesian statistics," *Proc. IEEE 2003 Int. Conf. on SMC*, pp.4637-4642, Washington DC, Oct. 2003.

[4] M. Gotoh, T. Sakai, J. Itoh, T. Ishida, and S. Hirasawa, "Knowledge discovery from questionnaires with selecting and descripting answers," (in Japanese) *Proc. of PC Conference*, pp.43-46, Kagoshima, Aug. 2003.

[5] S. Hirasawa, and W. W. Chu, "Knowledge acquisition from documents with both fixed and free formats," *Proc. IEEE 2003 Int. Conf. on SMC*, pp.4694-4699, Washington DC, Oct. 2003.

[6] S. Hirasawa, and W. W. Chu, "Classification methods for documents with both free and fixed formats," *Proc. 2004 Int. Conf. Management Sciences and Dexision Making*, pp.427-444, Taipei, R.O.C., May 2004.

[7] S. Hirasawa, F-Y. Shih, and W-T. Yang, "Student questionnaire analyses for class management by text mining both in Japanese and in Chinese," *Proc. 2007 IEEE International Conference on System, Man and Cybernetics*, pp. 398-403, Montreal, Canada, Oct. 2007.

[8] T. Hofmann, "Probabilistic latent semantic indexing," *Proc. of SIGIR'99*, ACM Press, pp.50-57, 1999.

[9] T. Ishida, J. Itoh, M. Gotoh, T. Sakai, and S. Hirasawa, "A model of class and its verification," (in Japanese) *Proc. of 2003 Fall Conference on Information Management, JASMIN*, pp.226-229, Hakodate, Nov. 2003.

[10] T. Ishida, M. Gotoh, and S. Hirasawa, "Analysys of student questionnaire in the lecture of computer science," (in Japanese) *Computer Education, CIEC*, vol.18, pp.152-159, July 2005.

[11] M. Nagao, H. Yagi, and S. Hirasawa, "Document clustering methods based on probabilistic latent semantic indexing with feature of words," (in Jappanese) *The 29th Symposium on Information Theory and its Applications (SITA2006)*, Hakodate, Hokkaido, Japan, Nov. 28 - Dec. 1, 2006.

[12] J. Itoh, T. Ishida, M. Gotoh, and S. Hirasawa, "A method for extracting important sentences using co-occurrence similarities between words ," (in Japanese) *Forum on Information Technology 2002*, pp.83-84, Tokyo, Sept. 2002.

[13] J. Itoh, T. Ishida, M. Gotoh, T. Sakai, and S. Hirasawa, "Knowledge discovery in documents based on PLSI," (in Japanese) *Forum on Information Technology 2003*, pp.83-84, Ebetsu, Sept. 2003.

[14] J. Itoh, T. Sakai, and S. Hirasawa, "A method for extracting parts of important sentences from Japanese documents using dependency trees," (in Japanese) *IPSJ, Tech. Rep. Natural language processing*, 158-4, pp.19-24, Nov. 2003.

[15] T. Sakai, J. Itoh, M. Gotoh, T. Ishida, and S. Hirasawa, "Efficient analysis of student questionnaires using information retrieval techniques," (in Japanese) *Proc. of 2003 Spring Conference on Information Management, JASMIN*, pp.182-185, Tokyo, June 2003.

[16] T. Sakai, T. Ishida, M. Gotoh, and S. Hirasawa, "A student questionnaires analysis system based on natural language expressions," (in Japanese) *Forum on Information Technology 2004*, N-021, pp.325-328, 2004.

Appendix A: Hypothesis Test for Tables 4.1 and 4.2

Let the null hypothesis H_0 be:

$$H_0 : \text{Two variables are independent},$$

then the probability of the occurrence such as the results shown in tables is p, from χ^2 examination under H_0, where p is give by 0.159 in Table 4.1, and by 0.000886 in Table 4.2. Hence H_0 for Table 4.1 cannot be rejected with smaller probability than 0.159. On the other hand, that for Table 4.2 can be rejected with smaller probability 0.01 by χ^2 hypothesis test.