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

Faculty Development by Student Questionnaire Analysis: A Class Partition Problem

Shigeichi Hirasawa, Waseda University, Japan
Takashi Ishida, Waseda University, Japan
and
Masayuki Gotoh, Musashi Institute of Technology, Japan

1. Introduction

(1) Questionnaire Analysis System
[7][9][10][15][16] …Fig. 1.a

--Extraction algorithm for important sentences [14]…3.3

--Extraction algorithm for feature words, and feature sentences [12]…3.4

--Document classification and clustering algorithm using PLSI [5][6][11][13]…3.2

(2) Faculty Development by Student Questionnaire [9][10]
… Fig. 1.b

--Class model: Computer engineering [4]…Fig. 2.3

--Questionnaire design [4][6]

--Applying the student questionnaire for these six years, and also in Taiwan [8]…Fig. 1.c

Questionnaire Analysis Model

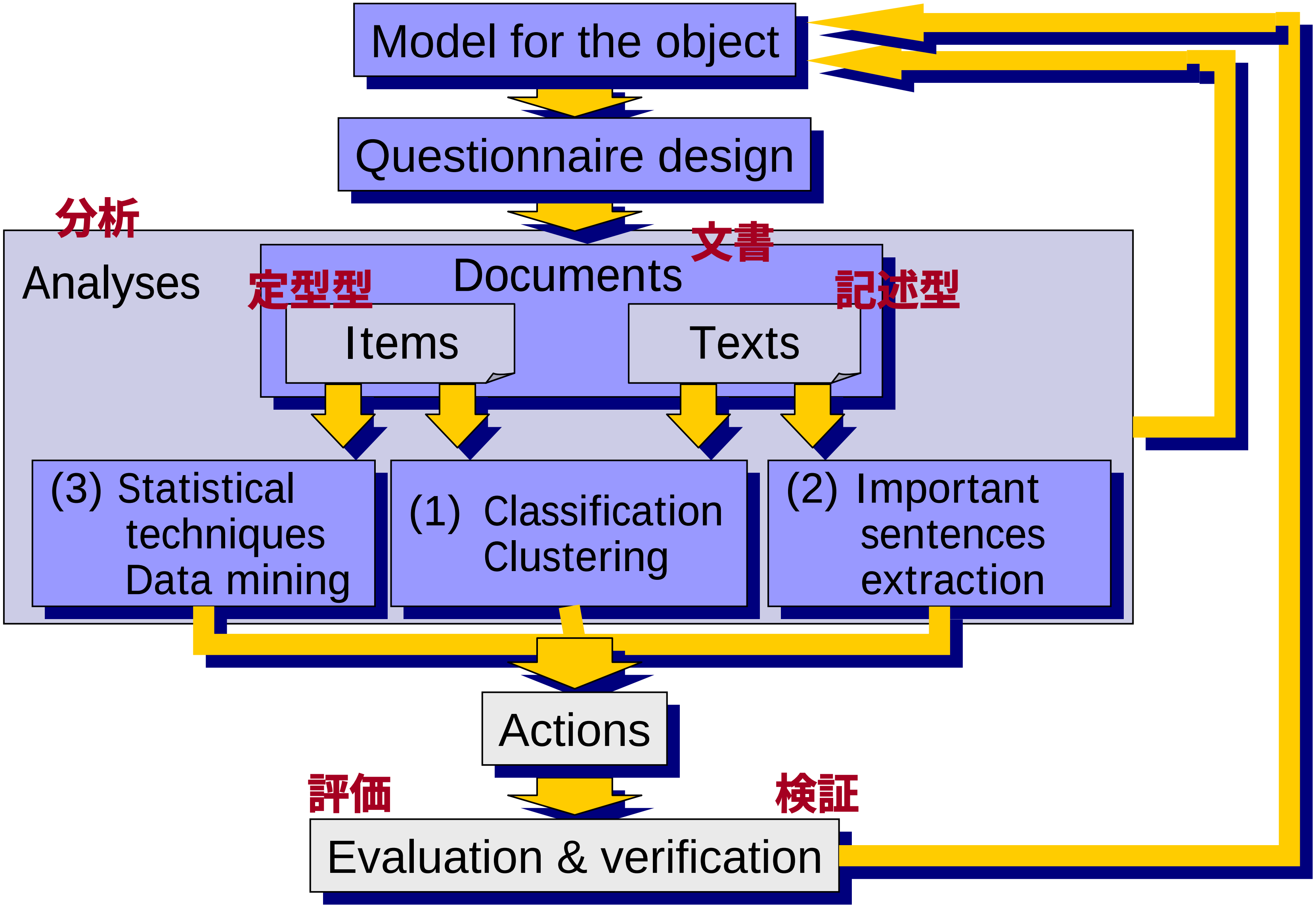


Fig. 1.a: Questionnaire analysis system [7]

Student Questionnaire

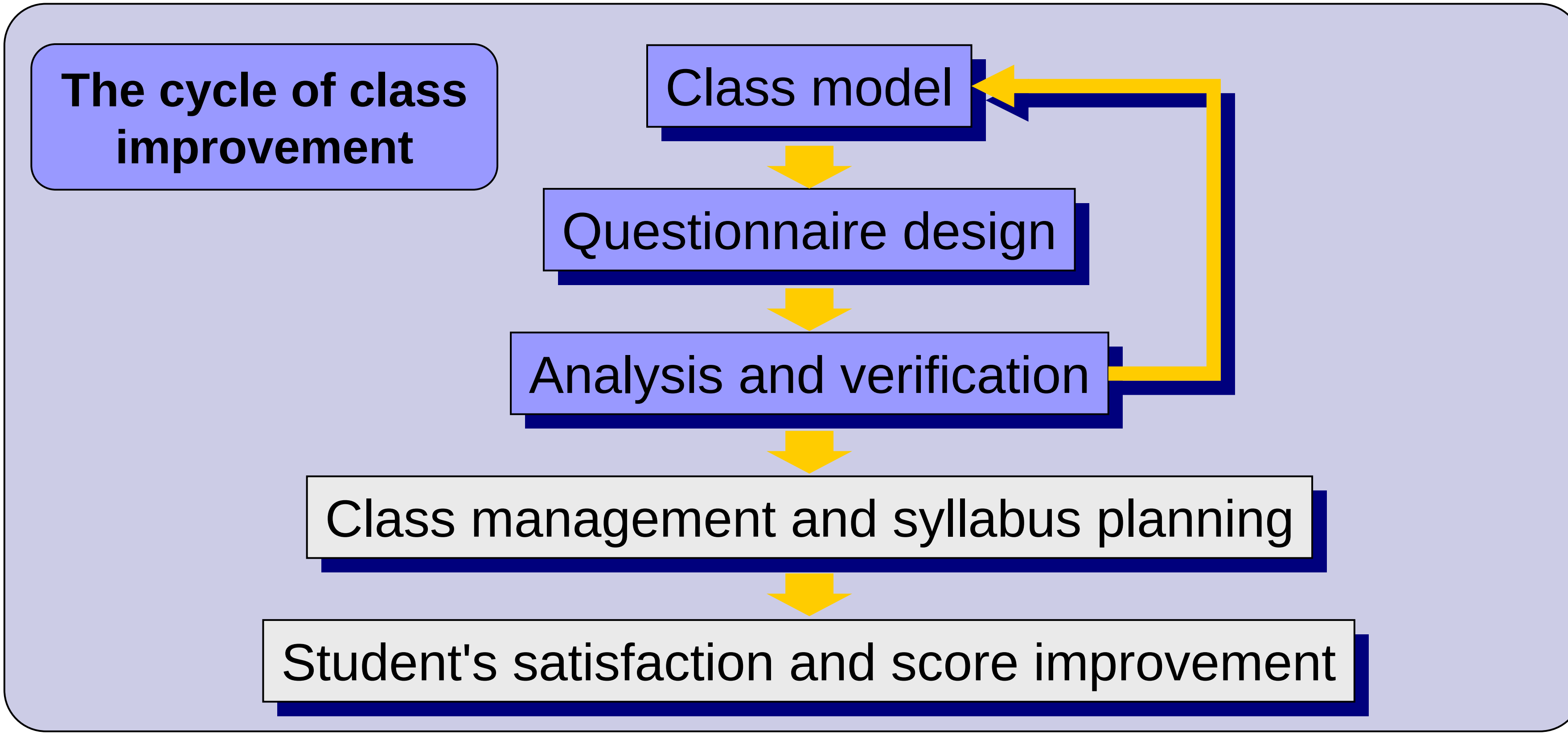


Fig. 1.b: Faculty Development by Student Questionnaire [10]

Questionnaire

- Fixed format (multiple choice questions: Items)
- Free format (Texts)

We have derived useful information for class management:

- The method for letting students be ***interested in*** computer engineering
- The method for letting students be ***satisfied in*** studying computers.

etc.

Another important purpose of the questionnaire is to **partition** a set of students by extracting their hidden consciousness **before beginning the class**.

Hypothesis:

A student who will become a **specialist** (**generalist**) related to computer engineering as his/her future job should choose **Class S** (**Class G**).

- Two groups will be required for different knowledge in computers.
- In this paper, the conclusion of this problem will be:

  It is difficult to partition the students into a generalist group and a specialist group depending on their future job at beginning of the class.

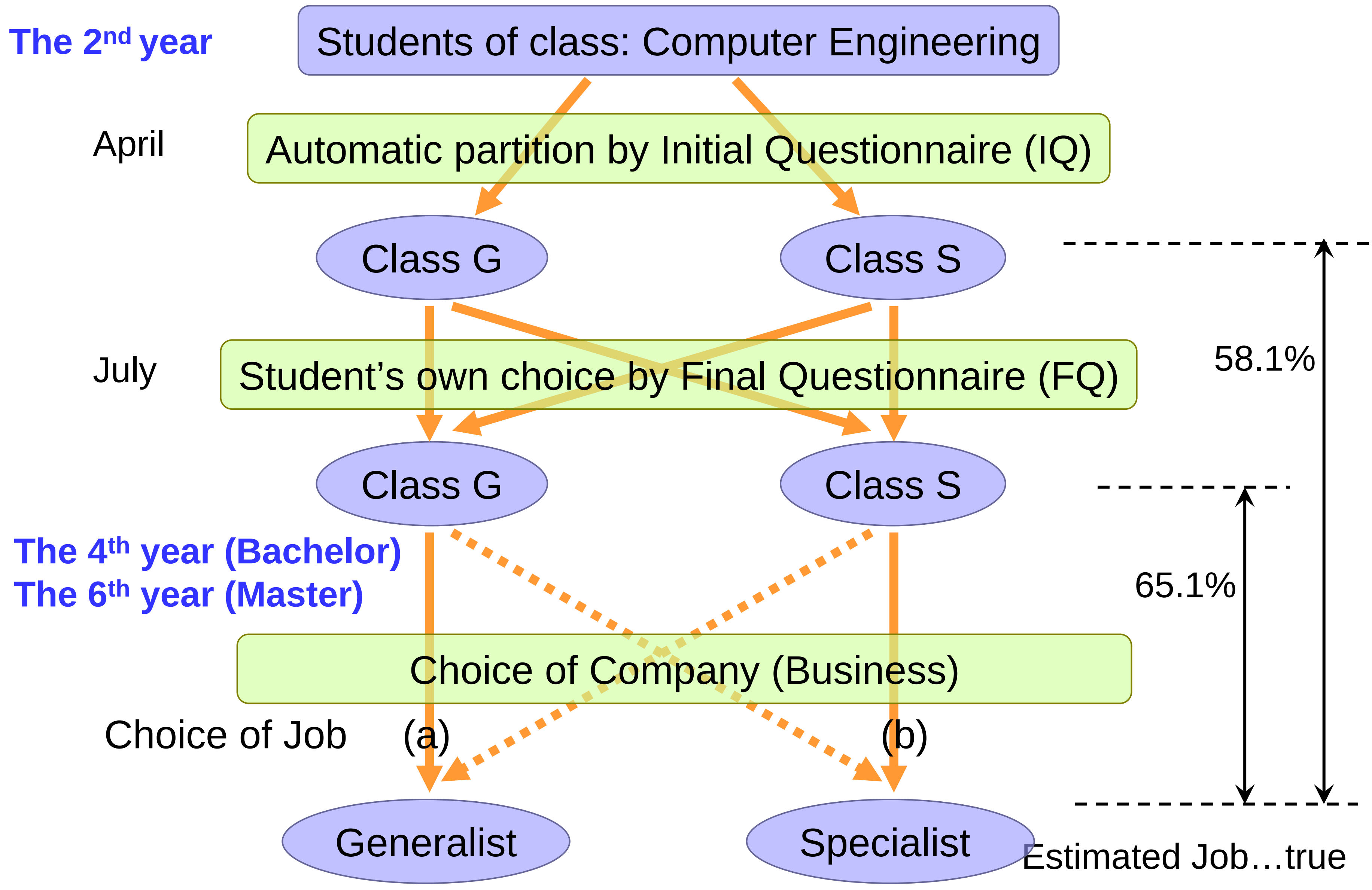


Fig. 1.c: Transition of students

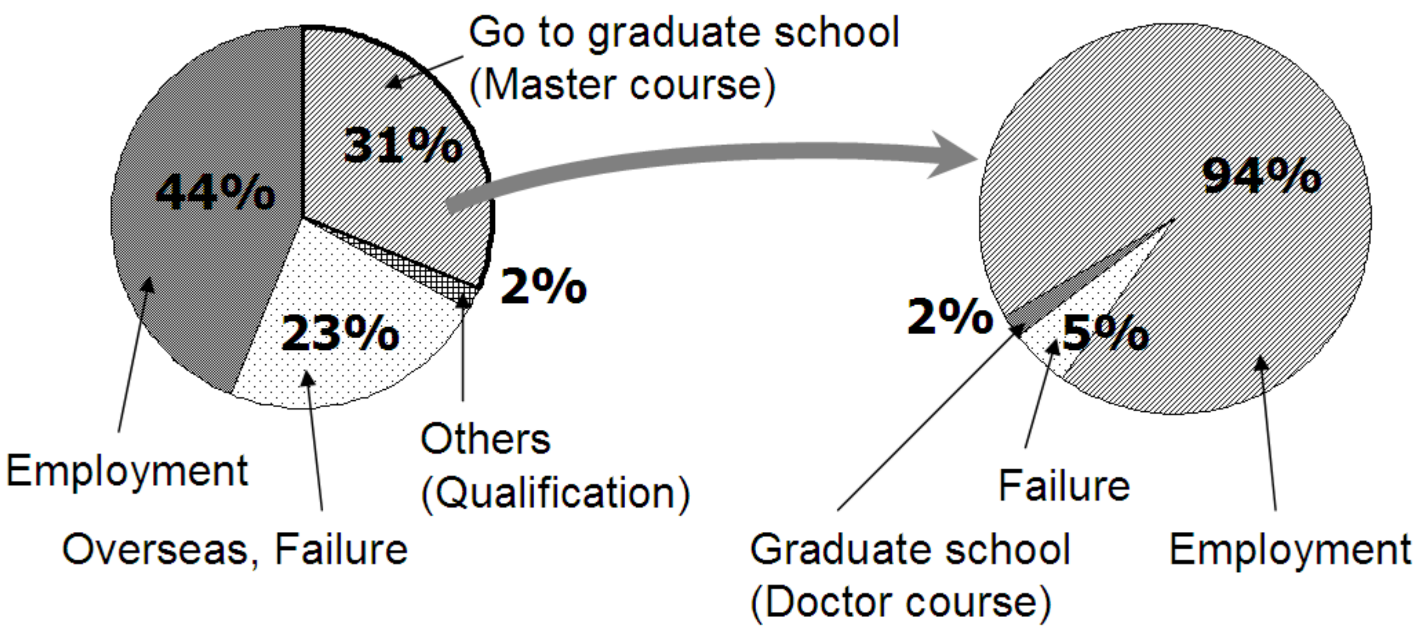
2. A Class and its Partition Problem

2.1 Target class (present)

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.)

2. A Class and its Partition Problem



(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

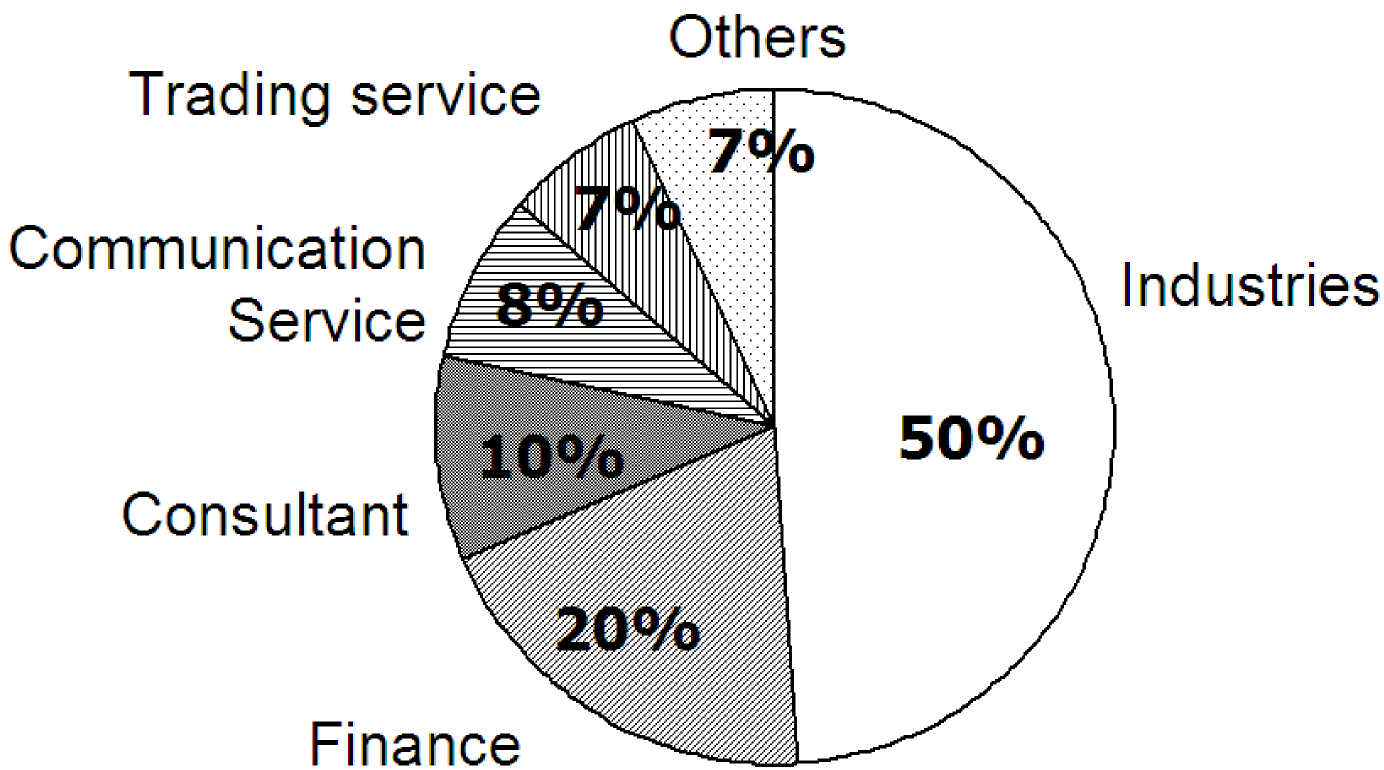


Figure 2.2: Business areas of object students

2. A Class and its Partition Problem

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. A Class and its Partition Problem

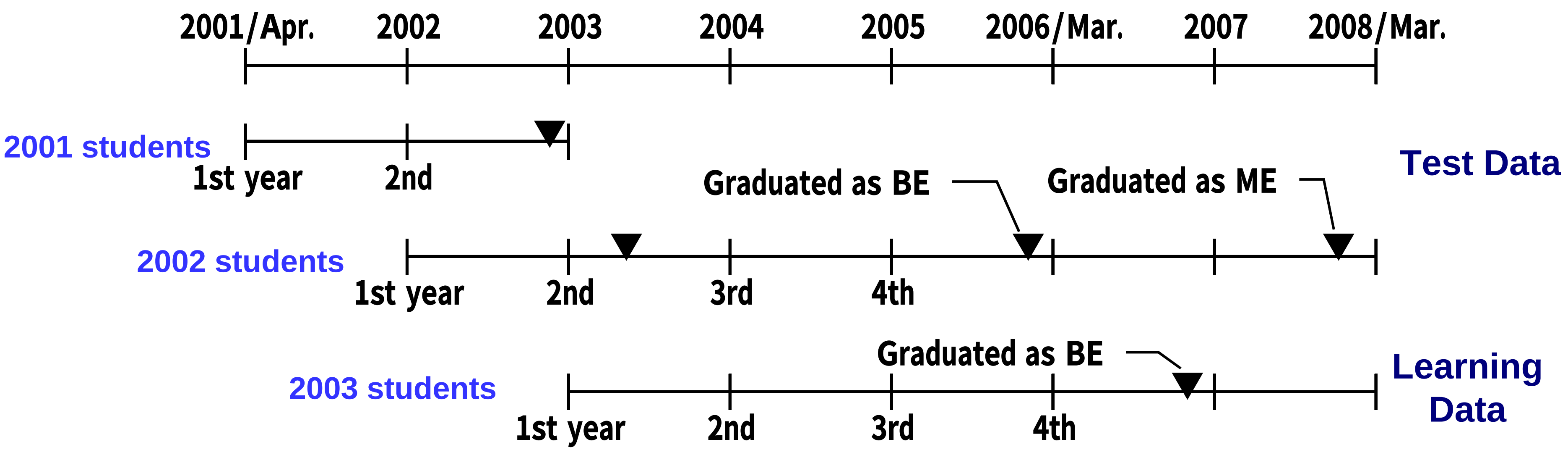


Fig. 2.a: Collected data

2.2 Class model

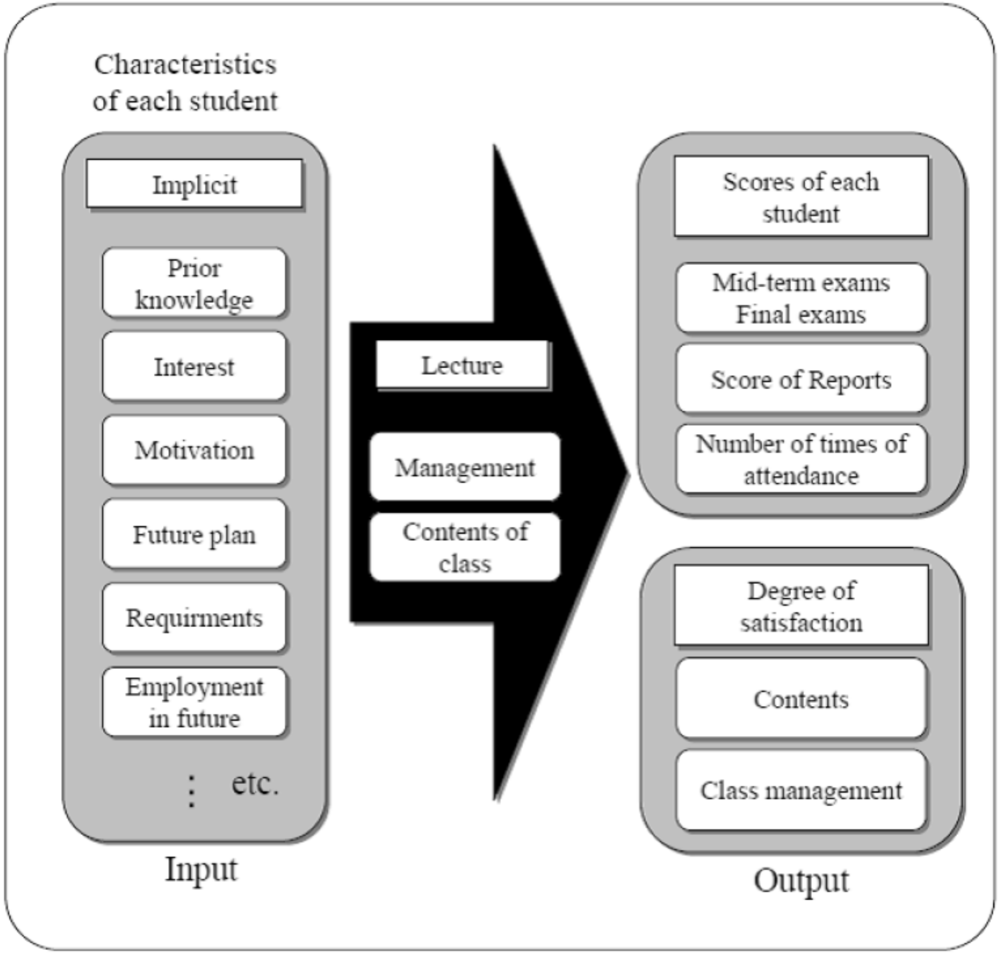


Figure 2.3: Class model

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.

2.4 Design of questionnaire

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)

2. A Class and its Partition Problem

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/

2. A Class and its Partition Problem

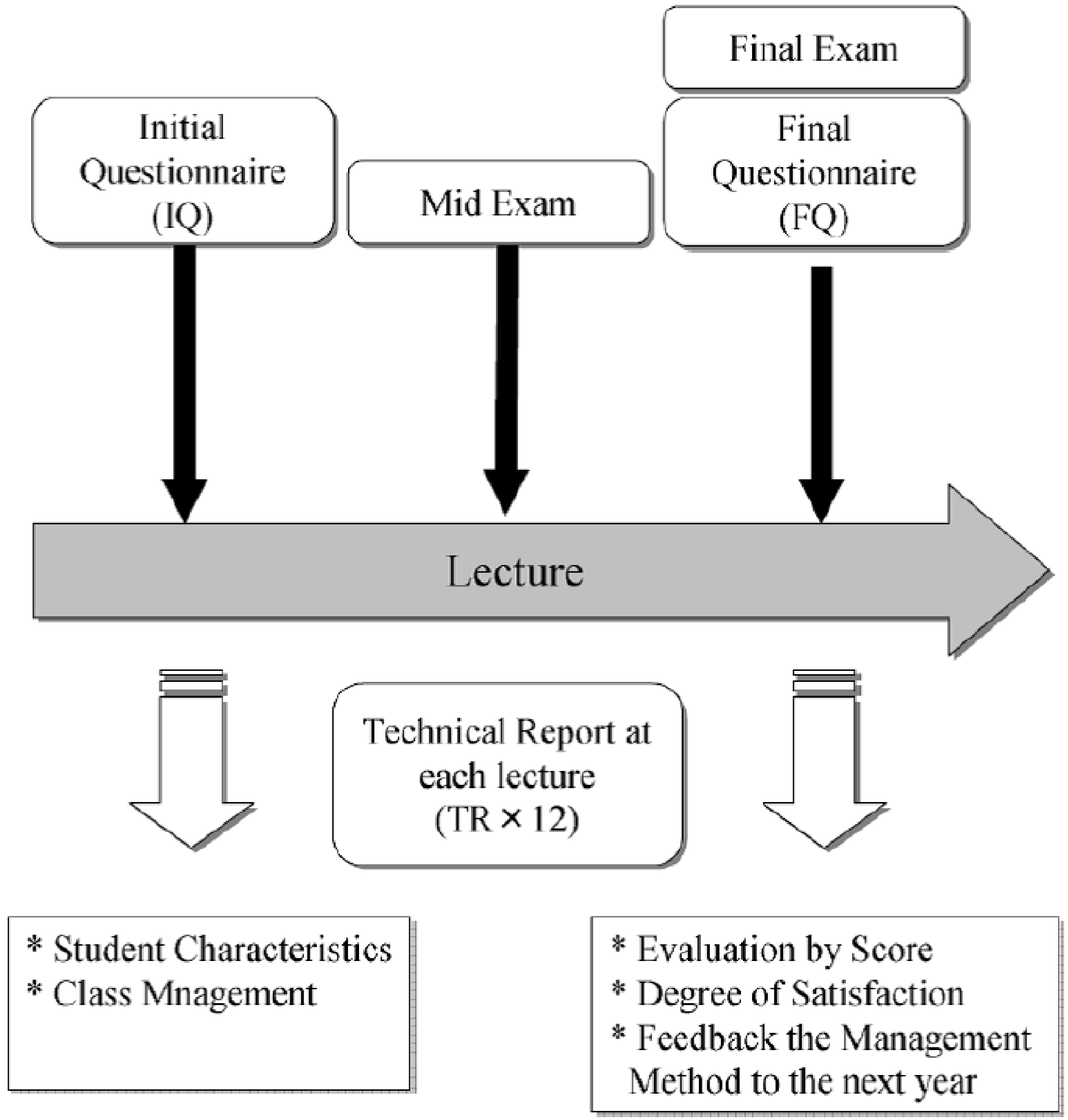


Figure 2.4: Time schedule for class

3. Methods for Analysis

Document Set

Format		Example in paper archives		matrix
Fixed format	Items	- The name of authors - The name of journals - The year of publication - The name of publishers	- The name of countries - The year of publication - The citation link	$G \in \{0,1\}^{I \times D}$
Free format	Texts	The text of a paper - Introduction - Preliminaries .... - Conclusion		$H \in \{0,1,2,\cdots\}^{T \times D}$

$G = [\, g_{mj} \,]$: An item-document matrix

$H = [\, h_{ij} \,]$: A term-document matrix

d_j : The j-th document

t_i : The i-th term

i_m : The m-th item

g_{mj} : The selected result of the m-th item (i_m) in the j-th document (d_j)

h_{ij} : The frequency of the i-th term (t_i) in the j-th document (d_j)

The Probabilistic LSI (PLSI) Model

A) $$A = [a_{ij}] = \begin{bmatrix} \lambda G \\ (1-\lambda)H \end{bmatrix}, \quad a_{ij} = tf(i,j)$$

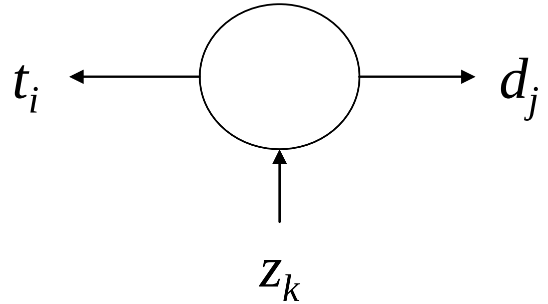
the number of term t_i in document d_j

B) Reduction of dimension by latent class (similar to SVD)

C) Latent class (state model based on factor analysis)

(i) an independence between pairs (t_i, d_j)

(ii) a conditional independence between t_i and d_j



D) Similarity function:

$$s(z_k, z_{k'}) = \sum_i \Big\{ h\big[\alpha \Pr(t_i|z_k) + (1-\alpha)\Pr(t_i|z_{k'})\big] - \alpha h\big[\Pr(t_i|z_k)\big] - (1-\alpha) h\big[\Pr(t_i|z_{k'})\big] \Big\} \quad (2)$$

where $0 \leq \alpha \leq 1$ and $h[x] = -x \log x$.

PLSI Model

[PLSI Model]

Let a term-document matrix $A = [a_{ij}]$ be given by only $tf(i,j)$ of eq.(1). Then the probabilities $\Pr(d_j)$, $\Pr(t_i|z_k)$, and $\Pr(z_k|d_j)$ are determined by the likelihood principle, i.e., by maximization of the following log-likelihood function:

$$L = \sum_{i,j} a_{ij} \log \Pr(t_i, d_j) \qquad (4.1)$$

EM Algorithm

[EM algorithm]

According to eq.(1), the maximum value of eq.(4.1) is computed by alternating E-step and M-step until it converges.

E-step:

$$\Pr(z_k|t_i,d_j) = \frac{\Pr(z_k)\Pr(t_i|z_k)\Pr(d_j|z_k)}{\sum_{k'}\Pr(z_{k'})\Pr(t_i|z_{k'})\Pr(d_j|z_{k'})} \quad (4.2)$$

M-step:

$$\Pr(t_i|z_k) = \frac{\sum_j a_{ij}\Pr(z_k|t_i,d_j)}{\sum_{i',j} a_{i'j}\Pr(z_k|t_{i'},d_j)} \quad (4.3)$$

$$\Pr(d_j|z_k) = \frac{\sum_i a_{ij}\Pr(z_k|t_i,d_j)}{\sum_{i,j'} a_{ij'}\Pr(z_k|t_i,d_{j'})} \quad (4.4)$$

$$\Pr(z_k) = \frac{\sum_{i,j} a_{ij}\Pr(z_k|t_i,d_j)}{\sum_{i,j} a_{ij}} \quad (4.5)$$

Then we have the probabilities $\Pr(d_j), \Pr(t_i|z_k)$, and $\Pr(z_k|d_j)$. □

Partition Algorithm [5]

The EM algorithm usually converges to the local optimum solution from starting with an initial value.

K: The number of categories (C_1, C_2, ... , C_K)

(1) Choose a subset of documents $\mathscr{D}^* \ (\subset \mathscr{D})$ which are already categorized and compute representative document vectors $\vec{d}_1^*, \vec{d}_2^*, \cdots, \vec{d}_K^*$:

$$\vec{d}_k^* = \frac{1}{n_k} \sum_{\vec{d}_j \in C_k} \vec{d}_j \tag{3}$$

where n_k is the number of selected documents to compute the representative document vector from C_k and $\vec{d}_j = (a_{1j}, a_{2j}, \cdots, a_{Dj})^T$, where T denotes the transpose of a vector.

(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 TEM algorithm, where $|\mathscr{Z}| = K$

(3) Decide the state $z_{\hat{k}} (= C_{\hat{k}})$ for $\vec{d}_j$ as

$$\max_k \Pr(z_k|\vec{d}_j) = \Pr(z_{\hat{k}}|\vec{d}_j) \Rightarrow d_j \in z_{\hat{k}} \tag{4}$$

If we can obtain the K representative documents prior to classification, they can be used for $\vec{d}_k^*$ in eq. (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.

Extraction algorithm of feature sentences and feature words

Let $\Pr(t_i|z_k)-\Pr(t_i)$ be the score of t_i, and 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 category or the cluster.

4. Questionnaire Analysis

"Job" : 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,

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.

"Business" : as the kind of ***company*** such as:

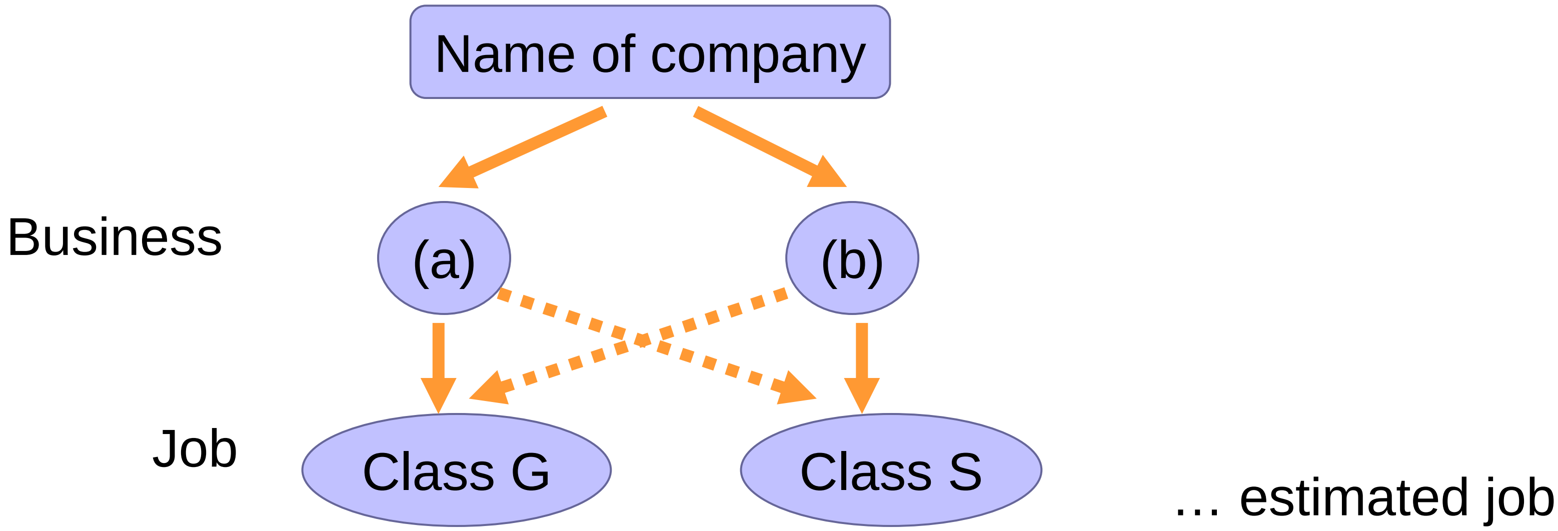
(a): trading, finance, banking, service, securities market, consultation, general construction, and so on,

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

4.1 Estimation of the job

We know only the name of companies in which they joined, such as:

Canon Inc., IBM Japan Ltd., NEC, Toyota Motor Corp., Accenture, Nomura Research Institute Ltd., East Japan Railway Co., Kashima Corp., Sony Corp., Tokyo Mitsubishi UFJ Bank, and so on.



4.2 Results of partition

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

AP: Automatic Partition
SEC: Students Estimated Choice

58.1%

Table 4.1: Numbers of partitioned students between AP and SEC

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

SOC: Sutudent's Own Choice

65.1%

Table 4.2: Numbers of partitioned students between SOC and SEC

4.3 Results of extracted important sentences

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.

Table 4.3: Extracted important sentences

(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 Discussion

(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 students at the 2nd academic year do not decide their future jobs. Hence they do not awake whether professional skill is required or not in their 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).

4.4 Discussion

(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 become 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)).

4.4 Discussion

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

(ii) 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.

5. Concluding Remarks and Future Works

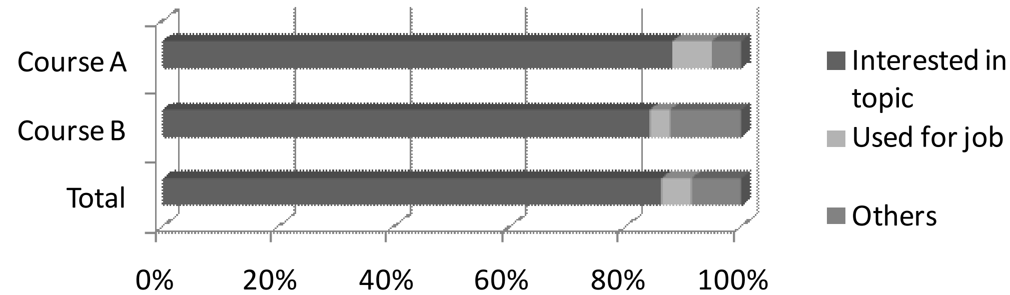
- Collecting documents obtained by student questionnaire for these six years, we analyze the graduated student questionnaire by trace back to their 2nd academic year. It is necessary to collect data at least four years for taking account the estimated their jobs.
- The results obtained in Section 4 are 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.

Trial case in 2007: Course system

A: Application
B: Basic

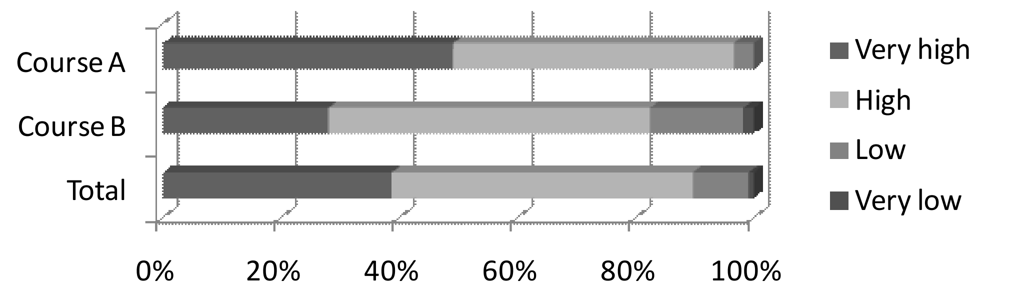
Reason for choice of courses:

	Course A	Course B	Total
Interested in topic	52	48	100
Used for job	4	2	6
Others	3	7	10
Total	59	57	116



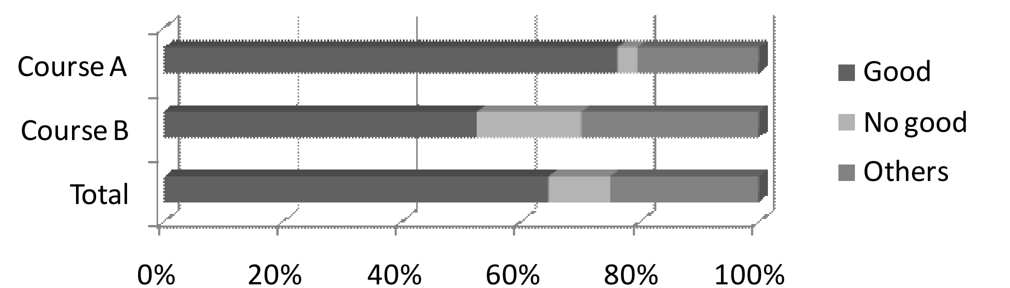
Degree of satisfaction for courses:

	Course A	Course B	Total
Very high	29	16	45
High	28	31	59
Low	2	9	11
Very low		1	1
Total	59	57	116



Evaluation of course system:

	Course A	Course B	Total
Good	45	30	75
No good	2	10	12
Others	12	17	29
Total	59	57	116



(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.

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.