Study on Student Classification Using Logistic Learning Curve and GNP-based Class Association Rule Mining

Rong ZHANG

February 2014

Waseda University Doctoral Dissertation

Study on Student Classification Using Logistic Learning Curve and GNP-based Class Association Rule Mining

Rong ZHANG

Graduate School of Information, Production and System
Waseda University

February 2014

Chapter 1 Introduction

1.1 Background and Objective of this Research

Remedial education is a term derived from developmental education. It originally refers to special preparatory courses offered by American community colleges to some incoming students who lack specific skills and abilities in one or more academic areas [1][2]. Many American universities and colleges using remedial education as a device to assure the stability of higher education system [3][4][5][6].

In Japan, the population of the younger generation is experiencing rapid decrease and more and more universities/colleges, especially privately-owned universities/colleges have to make attempts to ensure their annual quotas. Remedial education is an inevitable result of the “open access” policy regarding student recruitment [7][8][9][10] and the relaxing education policy of MEXT (Ministry of Education Culture, Sports, Science and Technology) carried out about 10 years ago [11]. As a result, “college-level courses are beyond the skill level of many entering freshmen” and “remedial coursework is required to bring the students’ skill level up to the point where they can succeed academically at college-level work” [12].

Programs for remedial education have been established to strengthen students’ basic scholastic skills in some common subjects with the support of MEXT [13][14][15][16][17]. Many private universities and colleges are actively getting involved in remedial education with the hope that it can help to compensate for the deficiency in students’ academic learning. Web-based learning is widely used for remedial education and the efficiency of remedial education is largely promoted through corresponding courses [18][19][20][21][22]. But, due to the large discrepancy in students’ scholastic achievement between the top and bottom students [23][24][25][26], how to select qualified candidate students for an established remedial course has become a very challenging task for teachers. Allocating students to the appropriate courses which suit their individual levels leads to more productive educational outcome [26][27]. Without doubt, how to classify students in conformity with the content levels lays the foundation for the success of any applicable remedial education course [24][25][28].

1.2 Previous Study on Student Classification

Student classification is viewed as a dispensable part of the issue connecting with the improvement of educational quality [29]. Danilowicz and Kukla [30] emphasized the significance of capturing the features of students in different categories through the construction of web pages so that an effective didactic process might be advanced. Abedi [31] studied the issue of student classification within a social context and reported that less than 10% of the variance of the classification for English language learners could be explained by students' efforts and other factors such as ethnicity, socio-economic status and teacher and parent opinion tend to be more responsible for the outcome. Wood [32] criticized the incongruence caused by inconsistence in the identification concepts and procedures for student classification. He insisted that the test scores should be taken into account to reach the consonance in various criteria for differentiating learning-disabled students. Baker, Corbett and Wagner [33] pointed out the time-consuming weakness of human observations and operation in student classification and made suggestions about a "low-fidelity" test-based method to detect a gold standard of behavior patterns.

Although many researchers have devoted to the solution of the problem with various perspectives, previous studies show many disadvantages:

(1) Most of the papers don't focus on the topic of student classification. They discuss student classification within the frameworks of other topics and student classification is mentioned as an issue closely related to the educational effectiveness. In fact, researchers have realized that student classification is the very important action that teachers should take if they aim at more desirable educational outcome. But, they can't make any constructive suggestions because there is no quantitative analysis and no details about student classification are available.

(2) Most of the papers focus on one or few factors affecting the result of student classification. Student classification should be conducted based on their original scholastic abilities and learning progress. But in the previous studies, student classification tends to be related to one or a very limited number of independent factors, which disregard the prerequisite that the combination of factors actually exert greater influences on the output of any educational activity. As far as the author is concerned, there is no study which has given a general description about the features of students with different learning progresses by using a profound database. As Berger (1997) indicated, student classification must be made according to multiple measures which may not correlate highly with one another and thus reliable result may be achieved [34].

(3) Very few studies have reported works on student classification using engineering devices and making proposals from quantitative perspectives. Student classification has been carried out to serve different purposes based on qualitative analysis. Seldom are there any engineering skills and

techniques applied to the solution of this problem because it is considered as an issue more related to education and sociology than engineering. In this study, the author makes a preliminary attempt to adopt engineering methods to solve the practical problem of student classification.

Without any doubt, student classification within the framework of English remedial education has not been sufficiently investigated [35][36][37]. How much students benefit from a remedial course should have the maximum priority when the issue of student classification is discussed. More quantitative analysis on the relationship between the students' learning progress and student classification will throw lights on the instructional methodologies and pedagogic validities in English education [38][39][40][41][42][43][44][45].

1.3 Commonly-accepted Method for Student Classification

Student classification has not been taken seriously in the case of English remedial education, though the great discrepancy in students' scholastic competence has been noted as the major factor preventing teachers giving effective teachers. Many factors have been used as criteria for student classification. Sometimes, teachers divide students based on their registered grades. Sometimes all the students are just divided in average to keep the balance in class size. Sometimes, the students are even told to make self-assessment and choose to go to any class they want to. Other factors, such as age, gender, nationality, one's native language, have been used as criteria for student classification and not much evidence and sound reasoning have ever been given on how effective student classification could be using these methods.

Among all the methods, the most widely-accepted method is to use students' pre-course test scores (*PRTS*) obtained through the placement test. Pre-course testing is usually carried out beforehand for student classification, but rather than relying on accurate calculation, teachers tend to determine a score criterion by intuition and experience. Pre-course testing score, which indicates students' scholastic achievement before the course study, is commonly used as the major standard for classification rather than their progress during the course study. Students are divided into different groups for one reason or another without a definite standard. Sometimes, the mean score is used as the standard when dividing the students into two classes [46]; in other situations, the percentage of correct answers might be applied [47]. This traditional method based on human operation and observations has been criticized by Baker, Corbett and Wagner [33] as time-consuming weakness and low in fidelity.

The traditional practice for student classification lacks in evidence showing the interactions among students, the classification criterion, the relationship between student classification and the education results. Its disadvantages can be summarized as the following:

(1) From the perspective of students' progress in learning:

The traditional methods are based on the pre-course test score, which indicates the background knowledge students have already mastered before the course study. It ignores the significant point in student classification that students' progress during the knowledge learning process at the remedial course should be taken into account.

(2) From the perspective of capturing the features of students in different classes

The traditional methods give descriptions of the whole situation by showing the mean score, but tend to ignore the features of students in the respective classes. That is, the teacher makes judgment based on the situation from a whole without paying adequate attention to the need of students in different groups. As mentioned above, students with a great variety of learning needs are attending the remedial courses. Special attention should be given to the change in the scholastic competence of students in different classes in order to achieve more effective educational outcome.

(3) From the perspective of the educational outcome evaluation:

The traditional methods lack in proof showing the role of student classification in their learning progress. Teachers use a completely different set of questions for the post-course testing, which is supposed to reflect the contents students need to master during the course study. Because different sets of questions are used for testing, no meaningful comparison can be made to investigate the changes in the amount of knowledge students have. That is, there is no meaningful data for the evaluation of both this classification method and the educational outcome.

Any effective student classification method should "accurately predict learner performance, not just evaluate prior learning" in English study [48]. Any method, which fails to pay enough attention to the benefits of students and tends to ignore the very basic issue of educational results, should not be considered proper for student classification. Attempts have been made to use inquiries to help with the student classification process because they better capture the features of a certain group of students [49]. There is not a big collection of papers on the topic of student classification and most researches are just fragmentary. As far as the author is acknowledged, there is still no research showing convincing results on student classification from a quantitative perspective by using sound engineering tools.

The criterion for evaluating a method for student classification should focus on how much students are benefited from learning the course. Students' progress in academic learning is the important index implying the nature of this issue. Although students' post-course scholastic performances are influenced by many factors, such as course objectives, level of learning tasks, instruction methods and assessment [15][28][32], two reciprocal aspects are considered primarily significant for desirable outputs from remedial education [50]: (1) Selection of proper candidate students for a prepared remedial course; (2) Selection of appropriate contents with the most suitable level for a group of students. That is to say, students should be divided into proper categories so that they might be exposed to the teaching materials and pedagogical methods which best suit their features and scholastic levels [51].

1.4 Contents of this Dissertation

This dissertation focuses on the following topics about student classification:

(1) Student classification should be conducted based on students' learning progress in addition to their original scholastic competence.

(2) Students' learning progress can be predicted by differentiating the logistic learning curve based on their pre-course test scores, i.e., students' original scholastic competence;

(3) GNP-based class association rule mining is proposed for student classification based on students' learning progress, which suggests that student classification can be conducted through class association rule mining using the discovered class association rules for students in the respective classes, i.e., the qualified class and the unqualified class.

(4) Advanced GNP-based class association rule mining is proposed to improve the classification accuracy using the attribute selection process through Genetic Algorithm (GA) with consideration of the negative attributes.

The essence of this dissertation is as follows:

Chapter 2 reports the process for data collection and conducts a basic statistical analysis for two important indexes: Pre-course Test Score (*PRTS*) and Post-course Score Change (*PSC*) using the observed data. According to the previous studies, Japanese students' learning performances are affected by three aspects ---- individual commitment to the academic work (including motivation, strategies and efforts), intercultural communication awareness and competence, skills in computer operation and acceptability of electronic utilities [38][39][40][41][42]. They are quantifiable through inquiries in this study. The optimized prediction model through multiple regression analysis shows that students' pre-course test score (*PRTS*) is the only decisive factor for Post-course score changes (*PSC*). But due to the unsatisfactory prediction result, the relationship between pre-course test scores (*PRTS*) and post-course score changes (*PSC*) needs further investigation in detail using the differential curves of the various learning curve models in Chapter 3.

Chapter 3 discusses the possibility of fitting the training data to the various forms of learning curve models. Differential curves of exponential curve, modified exponential curve, Gompertz curve, logistic curve and modified logistic curve are employed to demonstrate the relationship between students' pre-course test scores (*PRTS*) and post-course score changes (*PSC*). Parameters of the differential curves are estimated by the minimum root mean square method. Consequently, it has been found that the differential curves of the learning curves with S-shape show relatively small minimum mean square error for the prediction on students' post-course score changes (*PSC*). These s-shaped learning curves, with their differential curves as a bell-shape, are eligible for explaining the relationship between pre-course test scores (*PRTS*) and post-course score changes (*PSC*). Although the differential curve of the modified logistic learning curve shows a little higher prediction accuracy, logistic learning curve is adopted as an applicable model for student classification in this study because it is simple and the most widely accepted.

Chapter 4 shows the attempts to conduct student classification though data mining using advanced GNP-based class association rule mining through attribute selection using Genetic Algorithm (GA). Genetic Network Programming (GNP) is effective in dealing with dense databases for classification, because it reuses the fixed number of nodes constituting a directed graph and saves both space and time for association rule mining. Sufficient important association rules are extracted through GNP and a high accuracy is reached for classification. It is different from other black-box structure techniques, such as neural network and SVM (Support Vector Machine). The extracted class association rules offer information on the frequent attributes and attribute combinations which have significant influence on students' progress in English remedial learning.

The unique point of the method used in this chapter is the combination of attribute selection through GA and the class association rule extraction through Genetic Network Programming (GNP). Optimal Attribute Set through Genetic Algorithm (OASGA) is created to carry out attribute selection before the execution of GNP. Attribute selection through Genetic Algorithm (GA) refers to the preprocessing technique of data before data mining. It prunes less relevant information and discovers high-quality knowledge. Thus, attribute selection increases the possibility that predictor attributes given to the mining algorithm become relevant to the class attribute. Because OASGA conducts class association rule mining with small data subsets rather than the original big database, simple but important class association rules are obtained for classification. Genetic Network Programming (GNP) for class association rule mining undertakes generation at the lower level of the whole system while the attribute selection through GA generates at the higher level. Simulation result using the testing data shows that the classification accuracy is largely improved when classification is made using the proposed method compared with the conventional GNP-based class association rule mining.

Chapter 5 makes another attempt to improve the classification accuracy of GNP-based class association rule mining for student classification by taking into account the negative attributes as an addition. In order to discover negative class association rules, the negative aspects of all predictor attributes (65 attributes) are added to the original database and thus a database with 130 prediction attributes is newly created after the amplification. Comparison is made for the prediction accuracy between the original database with 65 prediction attributes and the amplified database with 130 prediction attributes. The simulation result shows that the latter achieves a higher prediction accuracy for student classification because the negative aspects of the prediction attributes give a more accurate and fruitful description about the students with desirable learning progress and thus lead to the extraction of more important class association rules. Detailed exploration of the important attributes and their combination are also made through the observation of the extracted class association rules. Important rules hold the maximum strength and reliability for the explanation of students' features in the respective classes. The frequently appearing attributes and attribute combinations in the important rules have special implications for the improvement of pedagogical instructions in English remedial programs.

Chapter 6 is the conclusion of this study. Consequently, the following two methods are proposed as solutions to the problem of proper student classification: (1) When students' pre-course test scores (*PRTS*) are basically used, the differential curve of the logistic learning curve can function as a convenient tool for student classification, because it reveals the interaction between the pre-course test scores (*PRTS*) and the post-course score changes (*PSC*); (2) Student classification can also be conducted through data mining by using inquiries because GNP-based class association rule mining can find the features of students in the respective classes, i.e., with desirable and undesirable learning progresses. Students' class labels can be predicted by using these discovered features. Important attributes and their combinations obtained from the extracted class association rules provide teachers with hints on the improvement of teaching effectiveness.

This thesis highlights the importance of student classification in web-based English remedial education, discusses the features of students with admirable academic achievement in the remedial course and proposes a more accurate criterion for student classification through the practice of scientific streamlining and structural optimization rather than human observation. Colleges and universities have a responsibility to promote the remedial education program and make it a success. Therefore, teachers should allocate students to classes whose contents best suit their respective levels [52]. Students should go to the classes where they can get the most out of the teaching and learning activities, which is the most important criterion for successful matching of prepared classes and students.

Chapter 2 Fundamental Statistical Analysis

The prerequisite for effective student classification lies in the point that all students should be benefited by the remedial course and their scholastic competences could be improved to a large extent [53]. That is, students' benefits in learning should always have the maximum priority when discussing the issue of student classification. Therefore, students' post-course score changes are used as a criterion for student classification in addition to their original scholastic abilities. Students' learning benefits refer to their learning progress, because it is an important index for the measurement of the effectiveness of any educational activity. In this chapter, students' learning progresses in English remedial course study are investigated in detail through fundamental statistical methods. Multiple regression analysis is conducted to find out important predictor attributes which exert great influence on post-course score changes *(PSC)* based on the optimized prediction model. As a result, students' pre-course test score (*PRTS*) proves to be the only decisive predictor attribute responsible for the prediction of their post-course score changes (*PSC*). Because prediction result using the testing data fails to show a satisfactory accuracy, more detailed examination is conducted in Chapter 3 to discover the interdependence between *PRTS* and *PSC* because *PRTS* is the most closely related to students' learning progress in the remedial course study.

2.1 Data Collection

All of the data used for training and testing in this study were collected by the author. 240 sample data were obtained at a web-based English remedial education course adopted by Nishinippon Institute of Technology, Japan. 240 sample data were collected, with 65 predictor attributes included in each record. The collected 240 sample data are divided into two sections: training data and testing data. 120 records are selected randomly from the 240 sample data and form the training data for this study. The left 120 data are used as testing data. In order to discover useful models for student classification, quantitative analysis is carried out for the training data and the validity of the proposed models is examined using the testing data.

Pre-course testing and post-course testing were held before and after the course study, respectively, thus, students' English abilities before the course study and their changes after the course are indicated through two fundamental indexes, i.e., students' Pre-course Test Scores *(PRTS)* and their Post-course Test Scores *(PTS)*. The former stands for the prior knowledge students have mastered before receiving remedial education and the latter signifies the total amounts of knowledge or scholastic attainments they achieve after the course study. The changes in students' score points indicate the changes in the amount of knowledge they have learned. The details about the system used for the English remedial education in this study are indicated as in Appendix I.

For the convenience of analysis, the gap between a student's pre-course test score and his/her post-course test score is defined as Post-course Score Change *(PSC)*, which can be calculated by subtracting pre-course test scores from post-course test scores as Eq.2.1.

$$PSC = PTS - PRTS, \qquad (2\text{-}1)$$

where,

PSC is the post-course score change.

PTS is the post-course test score.

PRTS is the pre-course test score.

The same questions were used for both the pre-course testing and the post-course testing, so that the changes in students' English proficiencies are precisely observed. The pre-course test score (*PRTS*) suggests students' scholastic competence, i.e., the amount of knowledge they have mastered in English before the course starts. In fact, pre-course test score (*PRTS*) is the factor that teachers tend to pay most attention to and is often treated as a criterion for student classification. As mentioned in the previous section, pre-course test score (*PRTS*) is often used as the tool for student classification, but the commonly-accepted practice is to divide all the students into two groups according to their mean or average score. Pre-course test score (*PRTS*) indicates students' scholastic competences before the course starts rather than the changes brought about by educational effectiveness. Therefore, student classification using their pre-course test scores (*PRTS*) is based on the amount of knowledge a student has obtained already before the course study. Teachers are not paying enough attention to students' benefits in learning when using pre-course test score (*PRTS*) as the only criterion for student classification. As far as this topic is concerned, there is still no quantitative evidence showing that student classification using pre-course test score (*PRTS*) as the only criterion is beneficial to students or the teaching effectiveness has been improved.

The post-course score change *(PSC)* refers to the newly-acquired knowledge a sample student obtained through the web-based learning. Therefore, it is most suggestive for the effectiveness of this web-based remedial education program. The post-course score change *(PSC)* is expected to be above zero, which means students' total post-course scholastic achievements increase after the study comparing with the starting point of the course. When a student's post-course score change *(PSC)* is below zero, it indicates that his/her total scholastic achievement decreased after the remedial course study and the student was not benefited from the web-based learning. The change in the total amount of knowledge a student has is the factor showing the efficiency of a remedial educational program. Therefore, more attention should be paid to this factor when discussing the issue of student classification. In this study, students' pre-course test score (*PRTS*) is used as a basis for selecting qualified students for the established remedial course. That is, students with $20 \leqq PRTS \leqq 100$ are considered as candidate students for the remedial course based on the observed data. In addition to *PRTS*, students' post-course score changes (*PSC*) are taken into account as another classification point which emphasizes the important role of students' learning progress in student classification.

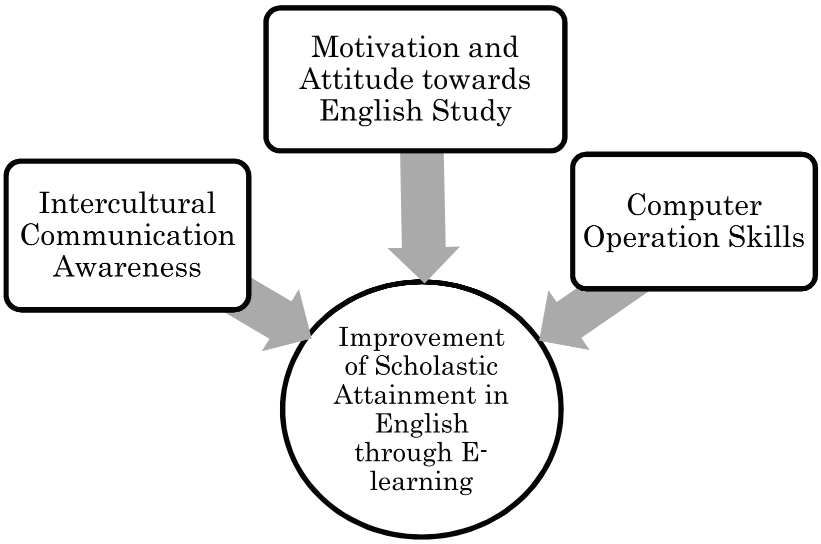


Fig. 2.1 Factors Affecting the Outcome of English Study through Web-based Learning

In summary, pre-course test scores *(PRTS)* and post-course score changes *(PSC)* are two important indexes for student classification, because they reveal the basic nature of students' English proficiencies and show the effectiveness of this online program by reflecting the exact changes in their scholastic achievements. The latter is of more crucial significance, because it suggests students' learning progress in the course study. Therefore, both of these two indexes should be taken into account when making decisions on student classification.

English learning within an online environment is such a sophisticated process that more factors are getting involved than in a face-to-face class. Fig. 2.1 shows the three factors which have been proved to be most influential to students' progresses in English study within an E-learning setting [23][24][25][36][37][38][39][40][41][42][43][44][45][46][47][48][49][50]. This study uses these factors in the three questionnaires (Table 2.1, Appendix II, Appendix III, Appendix IV) as predictor attributes for each record. Students were required to mark the closest comment for each of the 65 items (predictor attributes) in the inquiries. The choices designed for each question are ranked in five degrees ---- Answer "1" stands for "I disagree completely"; answer "2" refers to "I disagree partially"; answer "3" means "I am in the middle"; answer "4" implies "I agree partially" and answer "5" is "I agree completely". These 65 predictor attributes make up a set of data, which aims to offer a precise description about the feature of students with different learning progress. These predictor attributes are called independent variables in this study for the statistical analysis section, because they are responsible for explaining the dependent variable, i.e., Post-course Score Change (*PSC)* and making predictions for classification using the testing data.

Table 2.1 Constitution of Three Inquiries

	Number of Questions included	Theme of Questionnaires
Questionnaire 1 (Appendix II)	35	motivation and attitude toward English study
Questionnaire 2 (Appendix III)	20	intercultural communication awareness
Questionnaire 3 (Appendix IV)	10	computer operation skills; understanding about computer-assisted learning

In order to enhance the objective of this study, students' official records were obtained from the university, which include 10 evaluation items about students' personal information and learning activities, such as attendance rate, average score of all subjects, credits attained during the first semester study and scholarship received. Digital records of students' access to the web-based learning system were provided by the contracted company. 5 evaluation items about their performance in the remedial course are included, such as progressing rate of the total learning assignments, access frequency and total time spent on the program. These 15 items are added to the subset of predictor attributes for the research and are considered as independent variables, too. The purpose is to produce a relatively precise prediction model out of these potential attributes that may have significant impacts on students' learning progresses in the remedial program.

2.2 Basic Information about Pre-course Test Score *(PRTS)* and Post-course Score Change *(PSC)*

2.2.1 Pre-course Test Score *(PRTS)*

As explained above, pre-course test score *(PRTS)* implies the total amount of prior knowledge he/she has acquired before the course learning. Prior knowledge is an indicative exponent, because the prior knowledge level affects the speed for learners to acquire new knowledge and establish innovational learning patterns in a learning process [54][55]. The activation of students' prior knowledge helps to improve their absorption of new knowledge and contributes directly to their future scholastic achievement [56][57]. Therefore, in most cases, pre-course test score (*PRTS*) has been used as the major index which provides evidence for teachers to make decisions on student classification. That is, *PRTS* has been considered a reliable criterion for judging whether a student's scholastic ability matches the level of the class he is allocated to. But unfortunately, the matching practice is usually based on human observation which fails to show evidence to prove its efficacy. Therefore, how to deal with a student's Pre-course Test Score *(PRTS)* from a more scrupulous perspective lays the foundation for the discussion of student classification within the framework of remedial education.

Table 2.2 shows the distribution of students' pre-course test scores *(PRTS)* for the training data. 120 sample data are included in the training data. The average pre-course test score is 68 points, with a wide range spanning from 24 to 100 points. The gap between the maximum and the minimum score points in the pre-course testing is 76 points. This huge gap indicates that there is a great variety of learning needs with the students taking the remedial course. It is questionable that all students could be favored if they are put in the class and given the same teaching material. Some of them will not make desirable progress as expected through the current online assignment in that case, because they have different expectations from remedial learning. Obviously, how to help them satisfy their individual needs is a crucial issue related to the effectiveness of the established remedial program. Teaching materials with different levels need to be prepared in order to help all of the students to make use of their time to the largest extent.

Table 2.2 Basic Statistic Information of Pre-course Test Scores *(PRTS)* (point)

Mean	68
Median	72
Mode	72
Standard Deviation	16.8
Kurtosis	-0.30
Skewness	-0.55
Range	76
Minimum	24
Maximum	100
Sample Number	120

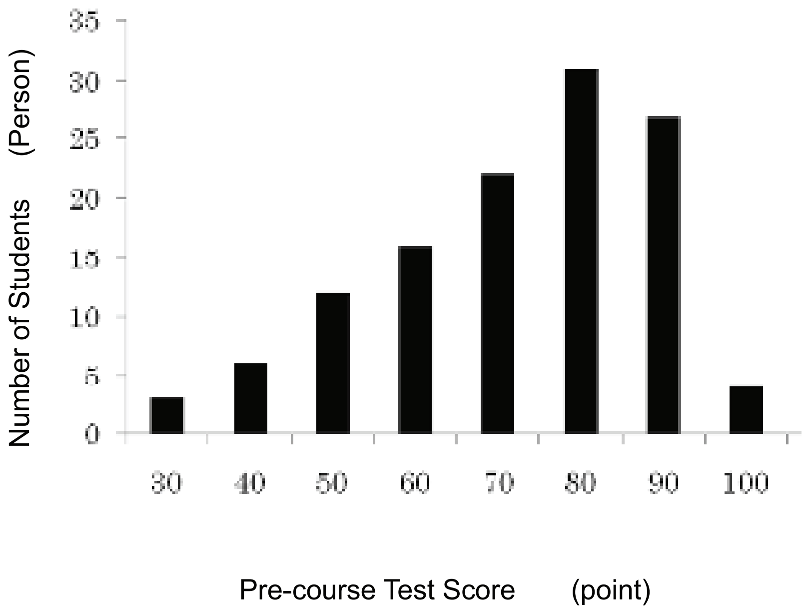


Fig. 2.2 Histograms of Pre-course Test Score *(PRTS)*

The histograms in Fig. 2.2 show the frequencies of students' pre-course test scores with 10 point intervals. The x-axis stands for the score range with the pre-course test scores and the y-axis stands for the total number of students in each score range. 60% (72 students) of the 120 sample data is around 60-90 points. The most noticeable is the 9 students who scored below 40 points. Undoubtedly, they are the main target students of this remedial program and the improvement of their English competences will enormously help to narrow the gap between the top and the bottom students and increase the efficiency of normal curriculum teaching in higher education.

The pre-course test scores for the training data further demonstrate the importance and necessity of classifying students into classes with different academic levels. Apparently, the educational efficiency is largely reduced if students with various scholastic competences are assigned to the same class. Reducing the variations in the academic abilities of the students in the same group is the fundamental solution to this problem, in order that educational resources can be made use of in maximum and desirable educational outcomes are produced.

2.2.2 Post-course Score Change *(PSC)*

Table 2.3 Basic Statistic Information of Post-course Score Change *(PSC)* (point)

Mean	9.5
Median	8
Mode	2
Standard Deviation	11.1
Kurtosis	0.76
Skewness	0.10
Range	64
Minimum	-28
Maximum	36
Sample Number	120

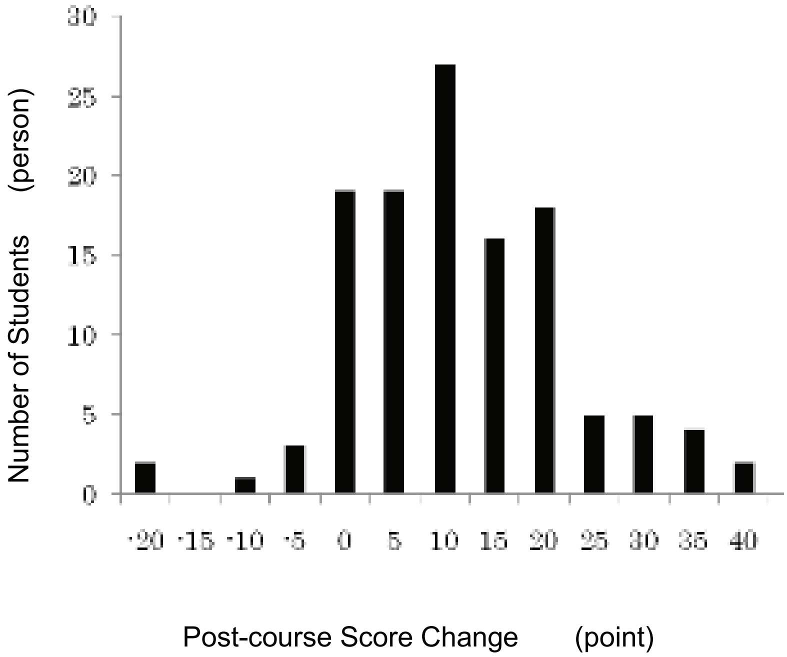


Fig. 2.3 Histogram of Post-course Score Change (*PSC)*

The basic statistic information for post-course score changes (*PSC*) is shown in Table 2.3. The histograms in Fig. 2.3 indicate the wide divergence in students' post-course score changes (PSC). The x-axis stands for the score change range and the y-axis stands for the number of students in each range of post-course score change *(PSC)*. The average post-course score change (*PSC*) is 9.49 points, and 72.5% of the 120 sample data (87 students) center around 0~20 points. The difference between the minimum PSC and maximum PSC reaches as large as 64 points, and 20% of the 120 sample data (24 students) did not show any improvement in their scholastic achievements after the remedial course study. Most of them even show negative figures in their post-course score changes (*PSC*), which means they are not favored by the current course at all.

As mentioned above, the judgment for proper student classification should be based on the criterion whether students have positive gains from the teaching and learning activities within the remedial course. In addition to students' pre-course test scores (*PRTS*), their post-course score changes (*PSC*) are also taken into account as the classification point for student classification. In this study, there are two classification points used for student classification. The first one is pre-course test scores (*PRTS*). Students with their *PRTS* above 20 points are considered candidate students for the established remedial course, which is determined by the observed data. Another classification point is the post-course score changes (*PSC*). Based on the comments by many teachers of English with profound teaching experiences, 10 point is considered a preferable criterion for dividing students. That is, students are divided into two groups (two classes) based on their post-course score changes (*PSC*). If the *PSC* shows a value above 10, then the student is considered to belong to the qualified group, which is labeled as Class 1. On the contrary, if *PSC* shows a value below 10, then the student is considered as unqualified students and is allocated to the unqualified group – Class 0. Because the remedial course offers training for the very basic section about grammatical knowledge and vocabularies for English learners, only those students with potentially admirable progress should be considered as candidate students for the remedial course. Students showing lower progress rates should go to upper level classes where more challenging teaching materials are used.

The criterion proposed above is the basis for all the analyses in the following chapters. The classification points for student classification can be summarized as follows:

Class 1: students who are qualified for the remedial course with $20 \leqq \text{PRTS} \leqq 100$ and $\text{PSC} \geqq 10$;

Class 0: students who are unqualified for the remedial course with $20 \leqq \text{PRTS} \leqq 100$ and $\text{PSC} \leqq 10$.

There is no doubt that the teaching efficiency is largely reduced when students with a broad variety of academic abilities are assigned to the class and working on the same textbook. The current situations show the evidence that student classification is absolutely needed for this remedial course in order to ease the problem caused by the wide dispersion in scholastic competence and learning needs. The wide ranges demonstrated by pre-course test scores (*PRTS)* and post-course score changes (*PSC)* intensify the paradox about English remedial education: What on earth are the most influential predictor attributes affecting students' learning progress? What is the evidence supporting the selection process of qualified candidate students for a prepared remedial course? How to make correct judgment on whether or not a student is favored by a remedial course? What is the difference between students who show desirable progress and who do not.

The effectiveness of an English remedial education program considerably diminishes when the level of didactic materials and exercises does not suit the students perfectly and specifically. It is the teacher's duty to help students to boost their learning outputs and maximize their post-course score changes (*PSC*) regardless of their pre-course academic levels. The core of all issues related to the remedial course lies in the question that to what extent the current remedial course is productive and whether all the students, especially the low-proficient students, are benefited by the established curriculum [33][58]. Student classification is the best solution to the problem caused by the great dispersion in scholastic performances, because it offers leverages for developing students' motivations, learning strategies and academic abilities in any web-based learning system so that educational efficacy is improved [19][33][34][51]. The factors devoting to the students' learning progress should be taken into serious account when selecting students for an established remedial program.

2.3 Correlation Analysis

In order to conduct student classification properly, multiple regression analysis is carried out to discover the influential predictor attributes which are exerting significant influence on students' learning progress. Since the data used in this study are collected by the author, the quality of the data is not so good. Students might give random answers to inquiries without careful thinking and the data contains many noisy predictor attributes. In order to obtain a simple and understandable prediction model through multiple regression analysis, correlation is conducted to prune those noisy data and discover predictor attributes which hold strong correlation with students' post-course score changes (*PSC*).

There are two objectives for using correlation analysis:

(1) Correlation analysis can find out those predictor attributes which hold strong correlation with post-course score change (*PSC*). Because the original database contains 65 predictor attributes, correlation analysis functions as a filter to remove those attributes which don't have much correlation with *PSC*. In other words, correlation analysis shows the predictor attributes with influential impact on *PSC*. It is a basic technique to find useful information from a large amount of data so that simple and understandable prediction model can be obtained through multiple regression analysis later on.

(2) Correlation analysis can find out the multicollinearity problem with the original data. Multicollinearity refers to a phenomenon in which two or more independent variables (predictor attributes) in a multiple regression model are highly correlated. In other words, the two independent variables (predictor attributes) share common parts in the description for the dependent variable, which is called class attribute in this study. Though multicollinearity does not reduce the predictive power or reliability of the prediction model as a whole, it still affects the valid results of the independent variables (predictor attributes) in the prediction model if some redundant factors exist. Therefore, the coefficients for correlation among those strongly related items need to be examined to avoid multicollinearity in the multiple regression analysis to be carried out afterwards.

As mentioned in section 2.1, 65 predictor attributes were collected for each record through three inquiries and they are the independent variables (predictor attributes) for the multiple regression analysis in this study. It is hoped that these independent variables (predictor attributes) help to constitute a meaningful optimized prediction model for students' post-course score changes *(PSC)* and offer clear explanation for the dependent variable (class attribute) *PSC*.

Correlation between the dependent variable (class attribute) *PSC* and the independent variables (predictor attributes) is conducted before multiple regression analysis is carried out. This process filters a large number of independent variables (predictor attributes) by remove insubstantial predictor attributes from the training data and making the outcome of the multiple regression analysis more reflective and useful. In statistics, correlation is a statistical relationship between two or more observed data. It is useful in the case of this study because it suggests a common trend shared by the independent variables (predictor attributes) and the dependent variable (*PSC*) and indicates a predictive relationship among the variables. If certain variables are correlated, it means that they satisfy a mathematical condition of probabilistic dependence. Correlation coefficient r, also known as the Pearson correlation coefficient, after its developer Karl Pearson, is a quantity indicator showing the strength and the direction of linear association between two variables and measures the degree of correlation between them.

The equation for computing the interdependence between two attributes (x, y) is as the follows:

$$r = \frac{\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \overline{x})^2 \sum_{i=1}^{N}(y_i - \overline{y})^2}},$$

$$\{(x_i, y_i): i=1,..,N\} \qquad (2\text{-}2)$$

where,

r is the correlation coefficient,

(x_i, y_i) is the ith pair of (x, y),

N is the number of sample students.

$\bar{x}$ and $\bar{y}$ are the expectation of x_i and y_i, respectively.

The value of coefficient r is within the range of $-1.0< r <+1.0$.

r^2 is called the coefficient of determination. It denotes the strength of the linear association between the two variables and tells us how certain the variance of one variable can be predicted from another variable. It is used to describe how well a regression line fits a certain set of data and indicate the likelihood of the prediction made by the prediction model.

There are many ways for deciding whether the variables are correlated. The compendium method proposed by Ueta [59] is applied to this study because it is the commonly accepted maximum likelihood for the judgment of correlations [60][61]. According to the method, correlation between an independent variable (predictor attribute) and the dependent variable (class attribute) is recognized when the coefficient of determination r^2 meets the following condition (Eq.2.3):

$$r^2 > 4 / (N+2) \qquad (2\text{-}3)$$

where, N stands for the number of the sample students. In the case of this study, the sample number is 120. Therefore, the coefficient of determination r^2 is expected to be larger than $0.0328=4 / (120+2)$ in order to assure the correlation. That is to say, if the correlation coefficient r for post-course score changes *(PSC)* and another independent variable (predictor attribute) is larger than 0.181 or less than -0.181, the correlation between the two variables is considered strong enough to make sense in prediction.

According to the criterion proposed above, 14 independent variables (predictor attributes) including pre-course test score *(PRTS)* demonstrate larger coefficients than 0.181 or smaller coefficients than -0.181, and their strong correlations with the dependent variable (class attribute) *PSC* (Post-course Score Change) are thus proven (Table 2.4). Coefficients of correlations among those 14 independent variables (predictor attributes) are checked, and no correlation coefficient is close to 1. That means, all the 14 independent variables (predictor attributes) are not closely related with each other in the sense of statistics and no multicollinearity is found. Anyway, it is suitable to use those 14 independent variables (predictor attributes) for processing the multiple regression analysis of post-course score changes (*PSC*). Among the 14 relevant independent variables, students' pre-course test scores (*PRTS*) prove to hold the strongest correlation with students' post-course score changes (*PSC*).

Table 2.4 Correlations between Post-course Score Change *(PSC)* and Independent Variables (Predictor Attributes)

Independent Variable (Predictor Attributes)	Correlation (r)
Pre-course test score (*PRTS*)	-0.577
The remedial course is not difficult (*D*).	0.275
I find English learning is more enjoyable than before. (*Y*)	0.258
I am doing extra learning in English. (*L*)	0.241
I don't know what to do when I have a question to ask. (*S*)	0.235
I always review what I learned after the English class. (*W*)	0.233
English is necessary for advanced professional knowledge learning.(*N*)	0.229
I can catch up with the teacher in class. (*C*)	0.225
I am learning for credits. (*A*)	0.222
I always prepare before class.(*P*)	0.221
I like computer games.(*G*)	-0.215
I do not want to lose my current English ability.(*U*)	0.20
One's English proficiency depends on his/her efforts.(*E*)	0.20
Computers are more reliable than human beings. (*R*)	-0.181

2.4 Multiple Regression Analysis

The data is further dissected on the basis of the result of the correlation analysis carried out above. The 14 predictor attributes in Table 2.4 are used as independent variables (predictor attributes) for the multiple regression analysis of post-course score changes *(PSC)* and four independent variables (predictor attributes) remain left in the optimized prediction model for *PSC* finally (Table 2.5). The independent variables (predictor attributes) left in the prediction models for the dependent variable (class attribute) are called identifiers. They function together to give description for the dependent variable (class attribute).

Table 2.5 Identifiers in the Optimized Prediction Model for Post-course Score Change *(PSC)*

	Coefficient
Intercept	37.9
Pre-course test score (*PRTS*)	-0.409**
I can catch up with the teacher in class.(*C*)	2.01
I find English is more enjoyable than before.(*Y*)	1.93
The remedial course is not difficult. (*D*)	-19.6**

**p<0.01

Optimized prediction model for the post-course score change *(PSC)* is expressed by Eq. 2.4:

$$PSC = 37.9 - 0.409 * PRTS + 2.01 * C + 1.93 * Y - 19.6 * D \tag{2-4}$$

where,

PSC is the Post-course Score Change.

PRTS is the Pre-course Test Score.

C denotes “I can catch up with the teacher in class”.

Y denotes “I find English is more enjoyable than before”.

D denotes “the remedial course is not difficult”.

In the statistical hypothesis testing, the *p*-value is the probability of obtaining a test statistic which might exist as an extreme phenomenon, assuming that the null hypothesis is true. A null hypothesis refers to the hypothesis of no change or no effect that may be falsified using a test of the observed data and the *p*-values shows how much evidence we have against a null hypothesis. The *p* in Table 2.5 measures the consistency by calculating the probability of observing the results from the sample data and the general rule is that a small *p*-value is evidence against the null hypothesis, while a large *p*-value means little or no evidence against the null hypothesis. In other words, the smaller the *p*-value is, the greater the inconsistency of the null hypothesis is.

In this study, the *p*-values of Attribute *PRTS* (pre-course test score) and Attribute *D* (The remedial course is not difficult.) are both below 0.01 as shown in Table 2.5, which demonstrates there is one percent chance for the hypothesis to be accidental that the Attribute *PRTS* and Attribute *D* are responsible for the post-course score changes *(PSC)*. In other words, there is one percent possibility that the opposite is true. The result of the multiple regression analysis in Table 2.5 also shows that the *p*-values of Attribute *C* (I can catch up with the teacher in class.) and Attribute *Y* (I find English is more enjoyable than before.) are above 0.05. Likewise, this indicates that there is five percent possibility for the judgment to be accidental. The small *p*-values in Table 2.5 indicate that the prediction model is meaning and consistent from the statistical perspective.

In Eq. 2.4, three identifiers related to students' motivation and attitudes in learning other than *PRTS* do interact with their *PSC* positively. Therefore, it is definite that students' *PSC* can be promoted to a certain extent through the improvement of those identifiers. Students who devote more efforts to their learning activities make more impressive progress. It is notable here that Attribute *D* (The remedial course is not difficult) is more constructive than the other two variables, because its coefficient (19.6) is much larger than the coefficients of the other two predictions (2.01 and 1.93), and its *p*-value $(p<0.01)$ is much smaller than the other two ($p>0.05$). This implies that students who bear more pressures in the remedial course tend to gain more remarkable increase in their post-course test scores. Pressure has turned out to be a significant predictor attribute contributing to the outcome of remedial education.

Because the four identifiers (pre-course test score and other three independent variables) in Eq. 2.4 differ in unit, data need to be processed to achieve their standardized coefficients for the purpose of better scaling and interpreting the dependent variable (class attribute) *PSC* (post-course score change). Standardization, in the field of social and behavioral studies, refers to the practice of redefining regression equations in terms of standard deviation units. It is a change of scale or a linear transformation of the data. It is usually conducted to answer the question on which of the independent variables (predictor attributes) has a greater effect on the dependent variable (class attribute) in a multiple regression analysis when the variables are measured in different units of measurement.

There are many ways to change scales of measurement, but the standardization technique is the one most often adopted by researchers in the social and behavioral study field. The most common practice is to carry out standardization for all the variables (independent and dependent) before fitting the multiple regression equation through Eq. 2.5:

$$\textit{Standardized Data} = (\textit{Observed Data} - \textit{Mean Value}) / \textit{Standard Deviation} \tag{2-5}$$

Standardization is conducted for all the data collected for this study. Multiple regression analysis is carried out for post-course score change *(PSC)* on the basis of the standardized variables. As a result, there is only one identifier, i.e., pre-course test score *(PRTS)*, left in the optimized prediction model Eq. 2.6. Thus, pre-course test score *(PRTS)* is the only identifier responsible for post-course score change *(PSC)* as shown in Eq. 2.6 and Table 2.6:

$$PSC = 37.9 - 0.576 * PRTS\ , \tag{2-6}$$

where,

PSC is the post-course score change.

PRTS is the pre-course test score.

Table 2.6 Identifiers in the Optimized Prediction Model for Post-course Score Change *(PSC)*

	Coefficient
Intercept	37.9
Pre-course test score (*PRTS*)	-0.576**

**p<0.01

According to Eq. 2.6, Pre-course test score *(PRTS)* is the only decisive predictor attribute for the prediction of post-course score change *(PSC)*. It exerts a considerably negative influence on the consequence of post-course score change *(PSC),* which implies the very basic fact that students with higher pre-course test scores tend to gain lower post-course score changes (*PSC*), while those with relatively lower pre-course test scores tend to make more progressive improvements in English proficiencies. When it comes to preciseness in calculation, Eq. 2.6 is viewed as the most acceptable optimized prediction model, because it is the result being processed on the basis of the standardization theory. The interdependence between students *PRTS* and their *PSC* needs to be examined in more detail.

For the sake of accuracy, Eq. 2.6 is used to make prediction of students' progress in learning. The *PSC* in the collected data is called Observed *PSC* (*O-PSC*). Predicted *PSC* (*P-PSC*) can be calculated based on the data from the inquiry. Comparison is made between students' *P-PSC* and *O-PSC*.

Fig. 2.4 shows how to make application of the optimized prediction model (Eq. 2.6) obtained through multiple regression analysis. As explained above, 10 point of post-course score change (*PSC*) is considered admirable Classification Point (CP) for an English remedial course. The oblique line in Fig. 2.4 is created based on Eq. 2.6. It intersects with the 10 point of *PSC* at 48 point of *PRTS*. Therefore, 48 point of pre-course test scores (*PRTS*) is the classification point (CP) for student classification based on Fig. 2.4. That is, students with $20 \leqq PRTS \leqq 48$ are considered as qualified students for the remedial course and they should go to Class 1 because their post-course score changes (*PSC*) are predicted to be higher than 10 points; On the contrary, students with $48 < PRTS \leqq 100$ are considered as unqualified students for the remedial course because their post-course score changes (*PSC*) are predicted to be lower than 10 points and thus they should be labeled as class 0.

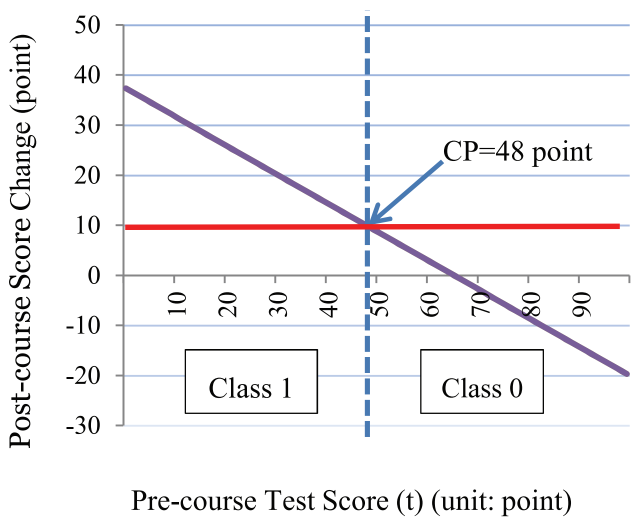


Fig. 2.4　Student Classification using Prediction Model Eq. 2.6

The prediction accuracy is calculated as follows. If the number of students with their $20 \leqq PRTS \leqq 48$ and $PSC \geqq 10$ is defined as *Np*, the number of students with their $48 < PRTS < 100$ and $PSC < 10$ is defined as *Nd*, the total number of students with their $20 \leqq \mathrm{PRTS} \leqq 100$ both in Class 1 and Class 0 is defined as *Ns*, then, the following equation (Eq. 2.7) is used to calculate the prediction accuracy (*Pac*) of the three learning curve models:

$$Pac = (Np + Nd) / Ns \tag{2-7}$$

Consequently, *Pac* demonstrates a ratio of preciseness of 49.2% for the training data and 46.7% for the testing data. Because the prediction accuracy of Eq. 2.7 is lower than 50%, it is not considered a reliable model. The prediction accuracy needs to be improved through further analysis on the relationship between students *PRTS* and *PSC*.

2.5 Summary

The basic statistical analysis of the observed data in this chapter shows that there is a great discrepancy in students' scholastic abilities between the top and the bottom levels. That means, there is a great variety of learning needs with the students taking the remedial course and not all of their needs are being satisfied. Some of the students are not benefited from the remedial course study and educational effectiveness is largely reduced since all of them are being assigned to the same class. Student classification is in urgent need to help students to achieve the most from their learning.

According to the proposed optimized prediction model as in Eq. 2.6 obtained through the multiple regression analysis, students' pre-course test score (*PRTS*) is the decisive factor for the prediction of their post-course score change (*PSC*). But since the prediction accuracy is not high, the interdependence between these two indexes should be further examined in detail through a more intuitive and precise model. Therefore, in Chapter 3, the training data of pre-course test scores (*PRTS*) are fitted to the differential curves of various learning curves and relationship between pre-course test scores (*PRTS*) and post-course score changes (*PSC*) is revealed through learning curves with more flexibility.

Chapter 3 Student Classification based on Logistic Learning Curve Model

According to the result of multiple regression analysis carried out in Chapter 2 (Eq. 2.6), the pre-course test scores *(PRTS)* exerts a decisive impact on students' post-course score changes *(PSC)*. Because student classification should be carried out on the basic idea that students' learning progress should also be taken into account in addition to their pre-course test score *(PRTS)*. Because *PSC* has been proved to be the only identifier for students' scholastic progresses, clarifying the functional relationship between pre-course test scores *(PRTS)* and post-course score changes *(PSC)* is most suggestive for the establishment of an effective student classification model. In sum, the pre-course test score *(PRTS)* should be treated as a prerequisite for the prediction of students' post-course score changes (*PSC*). Therefore, in this chapter, differential curves of various learning curves are used to demonstrate the relationship between pre-course test scores (*PRTS*) and post-course score changes (*PSC*).

Learning curves refer to graphical representation which shows growth and development with the increase of experience or knowledge against practice [62]. It is a term widely used in the research of industrial production and management [63][64][65][66][67], prediction [68][69][70] and other fields because it can be used for the prediction of future patterns. In the logistic learning curve proposed in this study, the x-axis stands for students' *PRTS* and the y-axis stands for their post-course score change rate (*PSCR*=*PSC*/100). As a result, 61 point is proposed as the proper classification point for student classification. That is, students with $20 \leqq PRTS \leqq 61$ are considered as Class1 because their *PSC* are predicted to be higher than 10 points, and students with $61 < PRTS \leqq 100$ are considered as Class 0 because their *PSC* are predicted to be smaller than 10. It is also shown in this chapter that learning curves with S- shape are appropriate in expressing the relationship between students' pre-course test scores (*PRTS*) and their post-course score changes (*PSC*).

3.1 General Description of Learning Curve

Learning curve, also called growth curve, is a graphical representation depicting the progression rate or the progress in the process of mastering a skill or knowledge within a learning circumstance [71]. Typically, it expresses the relationship between experience and efficiency or between investment and efficiency gains in the field of production and management. Workers become more efficient and experienced in carrying out a task after repeated training and the cost for production decreases at a constant rate as time passes.

Although it is a quantitative analysis widely applied in the field of production control and management for cost estimates, setting labor standards and evaluating labor performance [72], as far as the author is concerned, there is no report on the application of learning curves for the research of students' test scores and their score changes. This thesis serves as a preliminary study on the possibility of applying learning curves to the issue of student classification.

Learning curves have been used in the field of education, but the previous studeies are not used for explaining the relationship between students' pre-course academic competence and their post-course achievements. That is, as far as the topic is concerned, there is still no research on revealing the relationship between students' pre-course test scores (*PRTS*) and their post-course score changes (*PSC*) using learning curves. In this study, it is proposed that logistic learning curve can be used as a convenient tool to facilitate the evaluation of students' scholastic achievements and streamline the process of student classification so that the effectiveness of English remedial education can be improved.

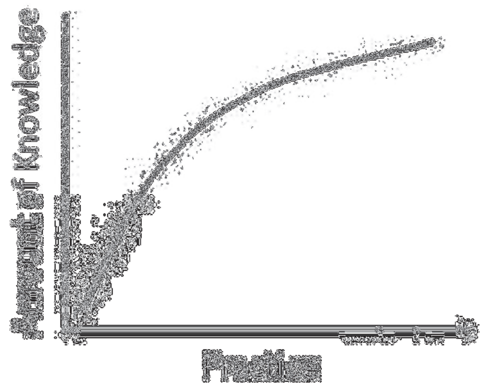


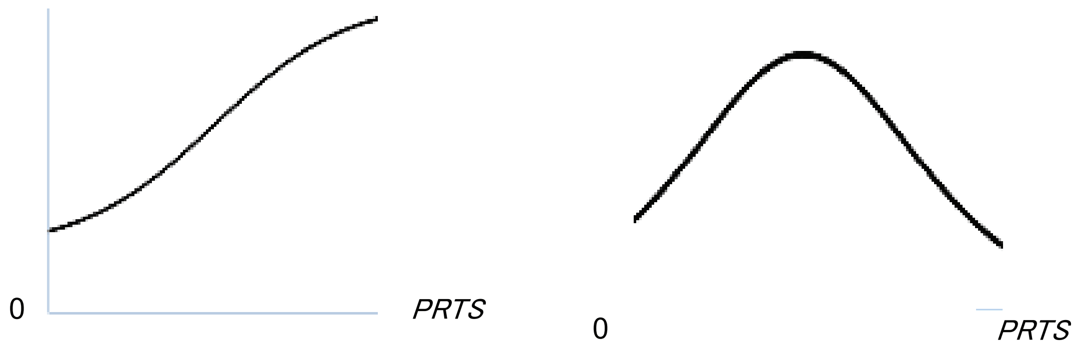
Fig. 3.1　An Example of Learning Curve

According to Thurstone [71], two specific laws need to be abided by formulating a satisfying learning curve for students' total amount of scholastic attainment: (1) The total sum of knowledge students grasp during a continuous learning process is going to increase; (2) "Diminishing returns with practice" is one of the inevitable attributes of any learning process [71]. That is, the amount of newly-acquired knowledge (skills) decreases as the knowledge level rises during the practice process. As shown in Fig. 3.1, the trend for knowledge gaining within the whole process is rising, but as practice increases, the learner's knowledge level rises and the pace for his/her knowledge gaining slows down as it approaches the final stage. These two features give a general description about the learning curve: it keeps a rising trend from a whole, but the rising slope becomes more and more moderate as the practice time accumulates.

The basic idea of the learning curve model proposed in this study is that the education realizes effectiveness because the total amount of changes in students' test scores equals to their newly-acquired academic competence, i.e., the knowledge they are supposed to master through learning. In other words, the accumulation of their test score changes is transferred into the total amount of knowledge they are supposed to master. Therefore, in the learning curve model, a learning curve and its differential curve are used to represent the relationship among students' pre-course test scores (*PRTS*) and post-course score changes (*PSC*). Students' pre-course test scores represent their prior knowledge level. As they learn (practice), their prior knowledge increases. So does their total amount of knowledge. Therefore, the learning curve (integral curve) shows the rising trend of students' knowledge from a whole, while the differential curve shows that the pace of students' knowledge gaining decreases when their prior knowledge level amounts to a certain level. Student classification using the learning curve model focuses on the change of the total amount of knowledge students master and their post-course score changes (*PSC*).

When students make attempts to learn a new knowledge, the increase in their academic competence is sharp at the initial stage. Then, the pace gradually evens out and finally reaches a saturation point. Generally speaking, all learning process presents an incrementally changing process in students' total amount of knowledge, while the speed for their knowledge gaining differs at different levels of their prior knowledge. Therefore, applicable differential cures show ups and downs to show the difference. A sudden steep slope indicates that little effort is needed and task is very easy during that period and the plateau or flat period implies that the acquisition of new knowledge becomes harder as the level of task rises or the input of new knowledge decreases. Creating such a curve means that the improvement in learning is assumed to be consistent theoretically over the successively newly gained knowledge, though the accelerating rate might differ from spot to spot on the prior knowledge scale.

The prerequisite for the learnng curve theory in this study is that the total amount of knowledge students are supposed to master through the course learning equals to the accumulation of the amount of knowledge they acquire at various levels of prior knowledge. That is, the relationship between the total amount of knowledge students should master and the newly-acquired knowledge at each ability level can be explained by an integral curve and its differential curve. Hence, in Fig.3.2 (b), the differential curve stands for the acqusition of new knowledge at the various levels of pre-course test scores (*PRTS*). As a general rule, the efficiency for acquiring new knowledge at the initial stage tends to be higher than at later stages. As the pre-course test score *(PRTS)* increases, the process for acquiring new knowledge tends to slow down because students need to deal with knowledge of higher levels.



(a)Integral Curve (b) Differential Curve

Fig. 3.2 An Example of Learning Curve Model

It is a definite rule for student classification that students with higher pre-course test scores (*PRTS*) should be assigned more challenging tasks, because their returns might be enhanced if the teaching material better matches their English level. The differential curve can be used as a tool for student classification because it expresses the relationship between student' pre-course test scores (*PRTS*) and their learning progress during the course study.

In this study, a learning curve model is proposed based on the theory of growth curve applied to management and production field. Two curves, a learning curve and its differential curve, are created to explain the learning process and the features of students' knowledge gaining. The pre-course test score *(PRTS)* is the x-axis and the total amount of new knowledge they should master in the course study is the y-axis of the integral curve. A new Post-course Score Change Rate *(PSCR)* is used as the label of the y-axis of the differential curve. Post-course change rate *(PSCR)* is defined as the ratio of *PSC* to its maximum value of 100 points *(PSCR= PSC / 100).* Student classification using the differential curve is carried out based on the knowledge they acquire through learning.

3.2 Criterion for Justifying a Learning Curve Model

There are diverse curves which might be used as learning curve models to reveal the growth nature of social phenomena. In this study, three typical non-linear curves are picked up for further examination of model fitting [73], i.e., exponential curve, Gompertz curve and logistic curve.

Model fitting is the judging work for the process of selecting proper parameters for learning curves [64]. Fitting measures should emphasize the generality for the unknown data and the errors are employed to figure out the essential difference between the observed data and the result calculated by the learning curve model [74]. The root mean square method is used as an assessment measure for model fitting in this study, which is carried out by dividing the sum of the squared errors by the total number of sample data and taking the root of it to obtain a mean square statistic [74][75]. This single critical value presenting the data fitting is labeled as Root Mean Square Error. Smaller value of this index indicates a higher preciseness in prediction, and the index suggested by the learning curve model is more convincing and persuasive.

3.3 Fitting Data to Learning Curve Models

The optimized prediction model obtained through the multiple regression analysis (Eq. 2.6) shows that students' pre-course test scores (*PRTS*) are the only determinant responsible for their post-course score changes (*PSC*). An applicable learning curve model should meet all the requirements explained in section 3.1. The learning curve should keep on rising constantly, moreover, it should also show the feature of "diminishing return" as the practice increases. Its differential curve needs show a decreasing trend at a certain prior knowledge level. Learning curves with such features are used for the fitting process below in order to discover the best learning curve which is the most suitable for addressing the relationship between students' knowledge accumulation and the changes in their test scores based on their prior knowledge.

According to the explanation about the learning curve in section 3.1, the learning curve should have a constant rising trend and show a reducing pace in knowledge acquisition. Therefore, in this study, several curves with a rising tendency are used to fit the obtained data. The differential curve of the proposed learning curve implies the relationship between students' pre-course test scores (*PRTS*) and their post-course score change rate (*PSCR*). Because 10-point improvement or more is considered a fruitful performance in the remedial course learning, a criterion using *PRTS* is figured out from the differential curve of the proposed learning curve for student classification.

3.3.1 Exponential Learning Curve Model

Exponential function is a most basic mathematic form to model a phenomenon of change with two attributes (variables). Exponential growth refers to a growing process which starts slow, but then changes faster and faster all the time. It is commonly used to describe the physical processes such as population growth without predators or resource restrictions. Although it does correspond to the situation of the progress in learning knowledge, the attempt is made as the first step for the completion of model fitting.

The x-axis of the exponential function is defined as students' pre-course test scores *(PRTS)* and its y-axis implies the total amount of knowledge mastered through the remedial course study. The equation for the exponential curve is as follows:

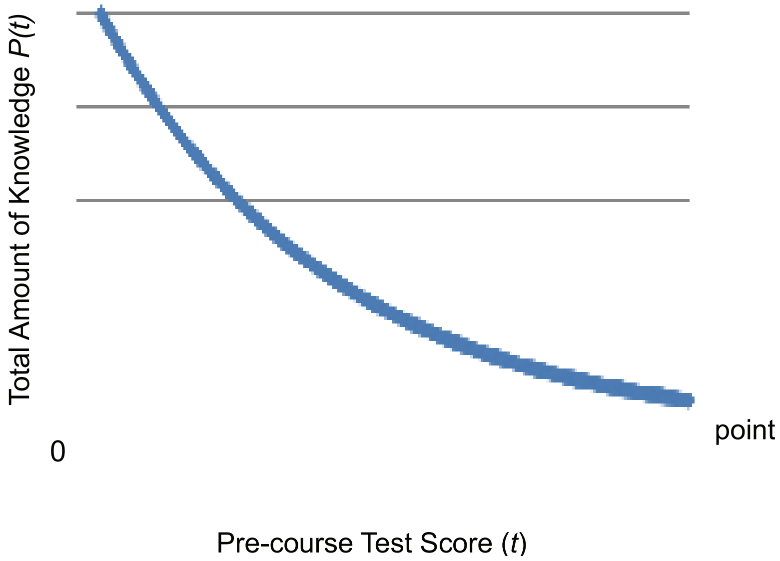
$$P(t) = ab^{t}, \tag{3-1}$$

where,

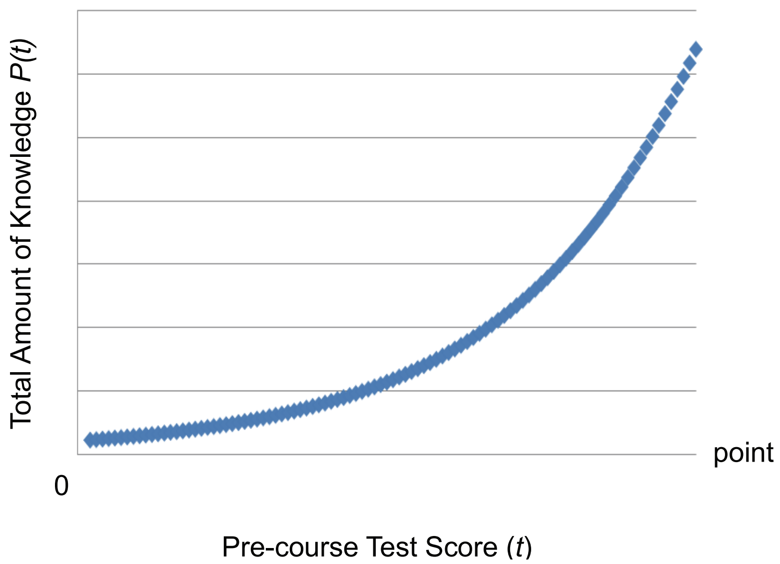
t is the pre-course test score,

a and b are constants.

When $a>0$, $0<b<1$, the curve keeps on descending across the coordinate axes; when $a>0$, $b>1$, the curve shows an ascending trend (Fig. 3.3). The latter case suits the situation of our study because students' total amount of knowledge keeps on increasing as they score higher in the pre-course test.



(1) Exponential Function *(a>0, 0<b<1)*



(2) Exponential Function *(a>0, b>1)*

Fig. 3.3 Exponential Function for *P(t)*

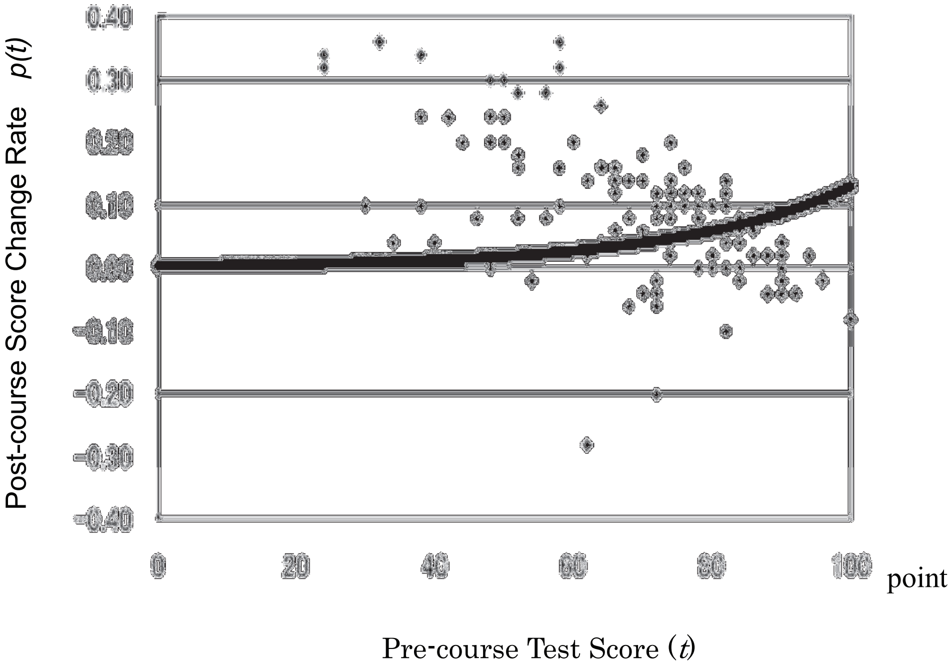


Fig. 3.4 Differential Curve *p(t)* of Exponential Curve
(a>0, b>1)

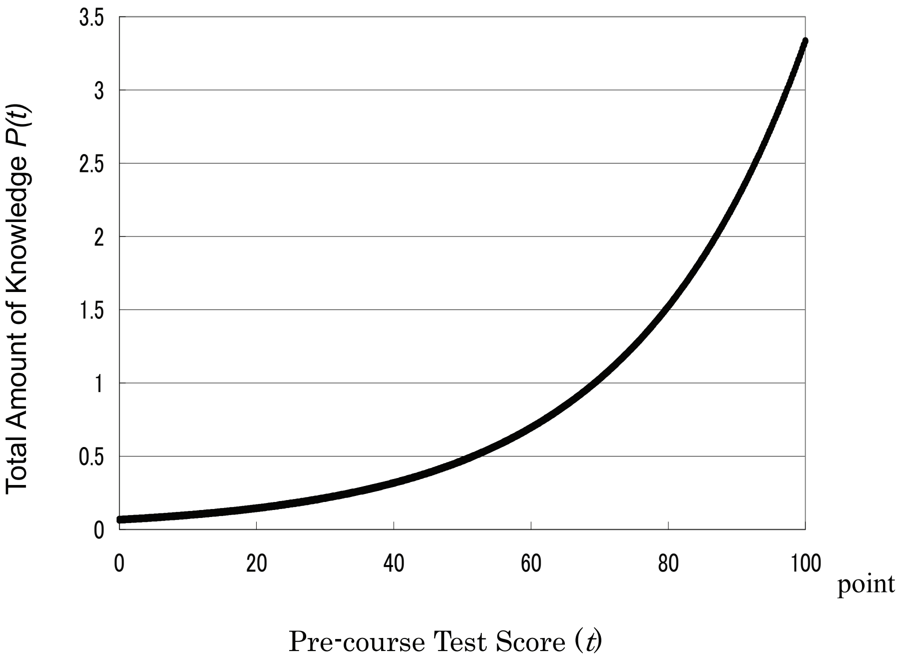


Fig. 3.5 Exponential Curve *P(t)*
(a>0, b>1)

Thus, Eq. 3.2 becomes the equation for the exponential learning curve model in this study. Taking logarithms of both sides of Eq. 3.2 for rearrangement, Eq. 3.3 is derived:

$$P(t) = ab^t \qquad (a>0,\ b>1) \qquad (3\text{-}2)$$

$$\begin{aligned} \ln P(t) &= \ln(ab^t) \\ &= \ln a + t \ln b \end{aligned} \qquad (3\text{-}3)$$

Suppose *p(t)* is the differential curve of the exponential curve of *P(t)*. Its x-axis refers to students' pre-course test scores and y-axis stands for their post-course score change rate *(PSCR)*. The equation for *p(t)* is derived by differentiating Eq.3.3 with *t* :

$$\frac{\frac{d}{dt}P(t)}{P(t)} = 0 + \ln b = \ln b \qquad (3\text{-}4)$$

$$\begin{aligned} p(t) = \frac{d}{dt}P(t) &= P(t)\ln b \\ &= ab^t \ln b \end{aligned} \qquad (3\text{-}5)$$

When *a=0.066* , *b=1.04*, the root mean square error between the observed data and the post-curse socre change rate data obtained through Eq. 3.5 reaches its minimum 2.24. The differential curve of the exponential curve is shown in Fig. 3.4, and the exponential curve is shown in Fig. 3.5. The value of students' total amount of knowledge after learning the remedial course ranges from 0.066 to 3.33 according to the calculation based on Eq. 3.2.

3.3.2 Modified Exponential Learning Curve Model

Modified exponential curve is a developed curve from the exponential curve. Eq. 3.6 for the modified exponential curve is proposed by adding an adjustment constant to the exponential curve, so that the curve can accommodate a larger range along the vertical scale. It exhibits greater flexibility than the exponential curve, and thus the precision of the prediction is largely improved. The modified exponential curve *P(t)* is defined by Eq. 3.6. In this case, *b* is set between 0 and 1.

$$P(t) = k - ab^t \qquad (a>0,\ 0<b<1) \qquad (3\text{-}6)$$

where,

t is the pre-course test score,

k, *a* and *b* are constants.

Suppose *p(t)* is the differential curve of the modified exponential curve *P(t)*. The differential curve *p(t)* refers to students' post-course score change rates (*PSCR*) and the x-axis indicates their corresponding pre-course test scores. Through the following calculation of Eq. 3.6, the equation of *p(t)* is derived from Eq. 3.7:

$$\begin{aligned} p(t) &= \frac{dP(t)}{dt} = -\frac{d}{dt} ab^{t} \\ &= -ab^{t} \ln b \end{aligned} \tag{3-7}$$

We can obtain the following equation by studying the gradient of the differential curve of modified exponential function *p(t)*:

$$\begin{aligned} \frac{d}{dt} p(t) &= -\ln b \frac{d}{dt} ab^{t} \\ &= -\ln b \cdot ab^{t} \ln b \\ &= -ab^{t} (\ln b)^{2} \end{aligned} \tag{3-8}$$

The relationship between constant *a* and constant *b* is discovered by substituting $t=0$ into Eq. 3.7.

$$p(0) = -a \ln b \tag{3-9}$$

According to the definition of the modified exponential curve, values of constant *a* and constant *b* are set within the range of $a>0$, $0<b<1$. It is clear that $lnb<0$ when $0<b<1$. Therefore, a conclusion can be drawn from Eq. 3.9 that $p(0)>0$. Moreover, because the differential curve stands for students' post-course score change rates (*PSCR*), $p(t) \leqq 1$ should be derived from this premise. Thus,

$$-a \ln b \leq 1 \qquad (a>0,\ 0<b<1) \tag{3-10}$$

$$a \leq -\frac{1}{\ln b} \tag{3-11}$$

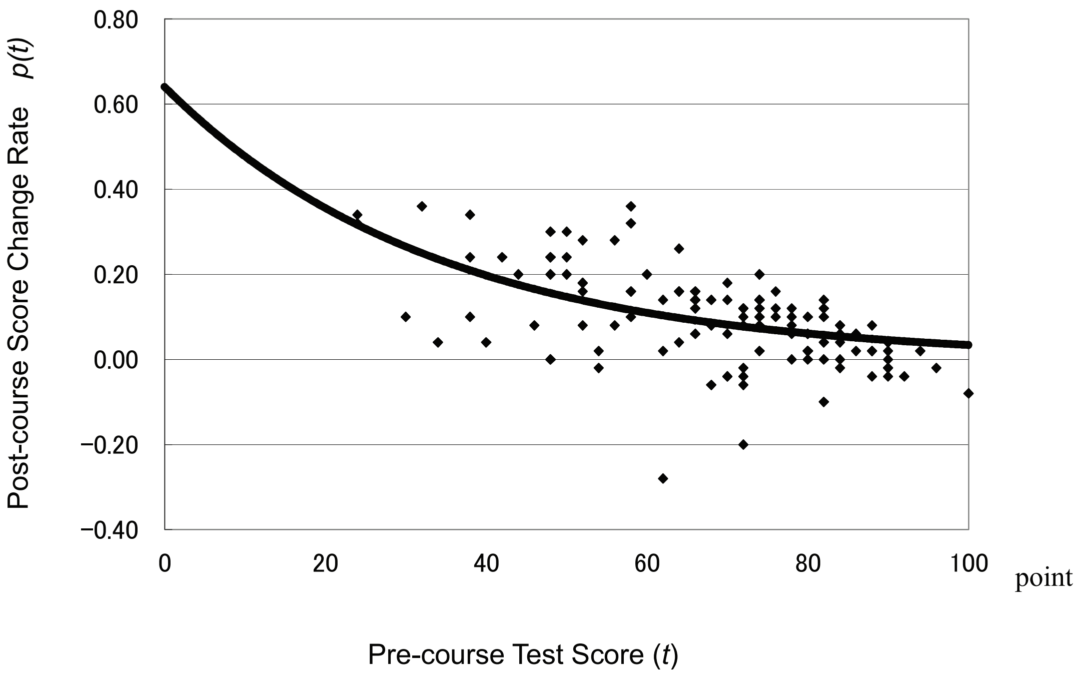


Fig. 3.6 Differential Curve *p(t)* of Modified Exponential Function
(a>0, 0<b<1)

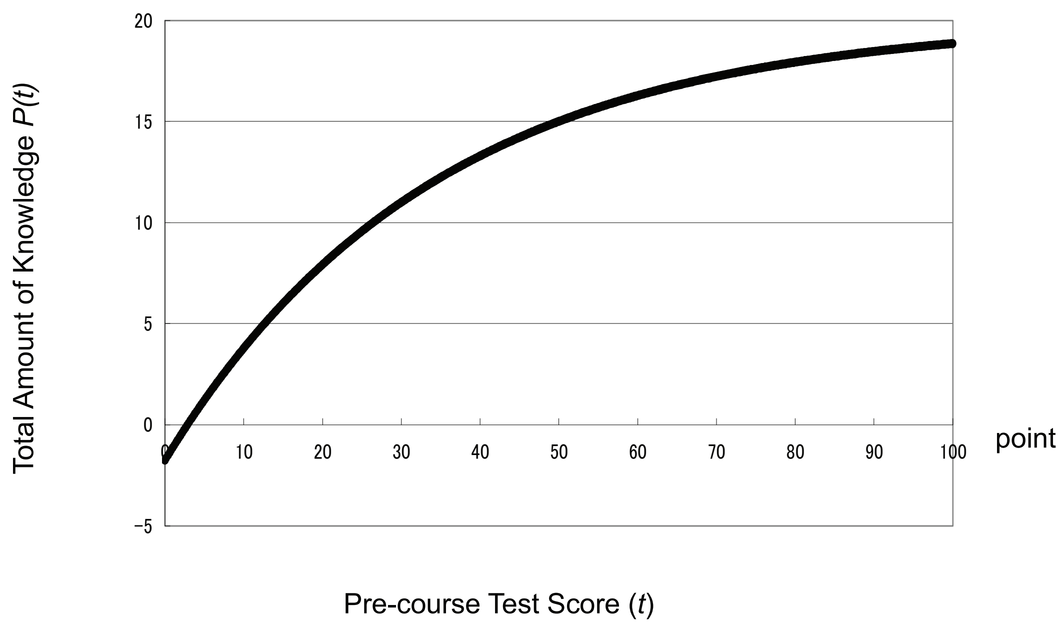


Fig. 3.7 Modified Exponential Function *P(t)*
(a>0, 0<b<1)

The value of the constant a is set within the range of $0 \leqq a \leqq 100$. The value of the constant b is set within the range of $0 \leqq b \leqq 1$. Consequently, when $a=21.76$, $b=0.97$, the root mean square error between the observed data and the post-course score change rate data obtained through Eq.3.9 reaches its minimum 1.04, which is much smaller than that of the differential curve of the exponential learning curve model. The differential curve is shown in Fig. 3.6, and the modified exponential curve is shown in Fig. 3.7.

The fatal disadvantage of this model lies in its indication of students' total amount of knowledge which was mastered during the course study. As clearly suggested in Fig. 3.7, the value of students' total amount of knowledge ranges from -1.77 to 18.85 according to the calculation based on Eq. 3.6. The amount of knowledge should never appear as a negative figure, therefore, the modified exponential learning curve can't be applied for the purpose of explaining students' scholastic achievements in this study.

3.3.3 Gompertz Learning Curve Model

Gompertz curve is another typical form of the sigmoid curves. It presents the growth occurring the most slowly at the start and end of a period, but accelerating sharply when the abscissa is in the middle of the changing process. It is different from most sigmoid curves, because it might not necessarily appear as a symmetrical curve, i.e., one hand of the curve might approach its asymptote much more gradually than the other (An asymptote refers to a line whose distance to a given curve tends to zero).

This non-symmetrical feature helps to create more adaptability for the curve to better reveal the subtle changes hidden in the data. Gompertz curve demonstrates its effectiveness as a tool to show the changes in many natural processes of growth, such as the change in population and birth rate, the uptake of mobile phones, and the modeling of the growth of normal tissues in the medical science field [76][77].

Gompertz curve $P(t)$ is defined by Eq. 3.12:

$$P(t) = k \cdot \exp(-ae^{-bt}), \qquad (k>0,\ a>1,\ 0<b<1) \qquad \text{(3-12)}$$

where,

t is the pre-course test score,

k is a constant,

the upper asymptote of the curve indicating the total amount of knowledge students mastered during the course,

b is the growth rate and refers to the absorption rate of new knowledge in this study,

a is the arbitrary constant.

The differential curve of Gompertz curve *p(t)* indicates the relationship between students' pre-course test scores (*PRTS*) and post-course score changes (*PSC*). The equation of the differential curve *p(t)* of Gompertz curve is derived by transforming Eq. 3.12:

$$\begin{aligned} \ln P(t) &= \ln \{k \cdot \exp(-ae^{-bt})\} \\ &= \ln k + \ln \exp(-ae^{-bt}) \\ &= \ln k + (-ae^{-bt}) \end{aligned} \tag{3-13}$$

By differentiating both sides of Eq. 3.13, Eq. 3.15 for the differential curve of *p(t)* is thus obtained:

$$\frac{\frac{d}{dt}P(t)}{P(t)} = (-a)(-b)e^{-bt} = abe^{-bt} \tag{3-14}$$

$$\begin{aligned} p(t) &= \frac{d}{dt}P(t) = P(t)\,abe^{-bt} \\ &= k \cdot \exp(-ae^{-bt}) \cdot abe^{-bt} \\ &= abk \cdot \exp(-ae^{-bt} - bt) \end{aligned} \tag{3-15}$$

The range of constant *a* is set from 1.01 to 100.0, the range of the constant *b* is set from 0.001 to 0.999 and the range of constant *k* is set from 0.01 to 20.0. The root mean square method is used to determine the values of parameters in this study. Consequently, when *a=2.81, b=0.044,* and *k=16.11*, the minimum value of the root mean square error between the observed data and the post-course score change rate data obtained through Eq. 3.15 reaches 1.02, which is better than both the exponential curve (2.24) and the modified exponential curve (1.04). The differential curve of Gompertz learning curve is demonstrated in Fig. 3.8, and Gompertz learning curve is shown in Fig. 3.9. According to Fig. 3.9, the value of students' total amount of knowledge which was mastered during the course study ranges from 0.97 to 15.6.

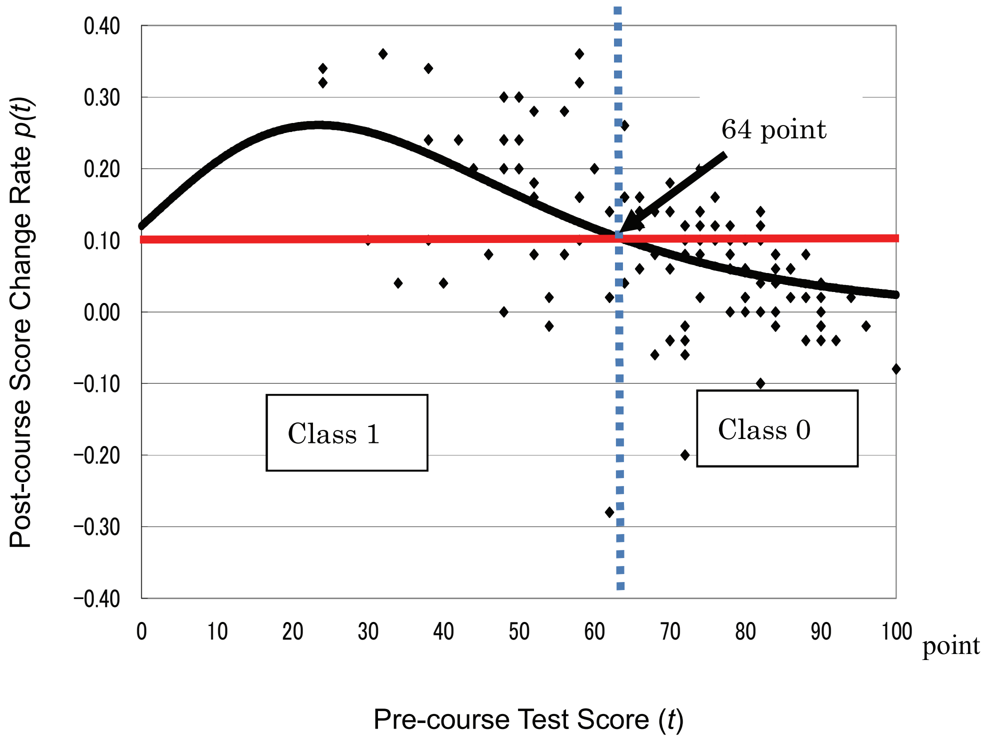


Fig. 3.8　Differential Curve *p(t)*of Gompertz Curve

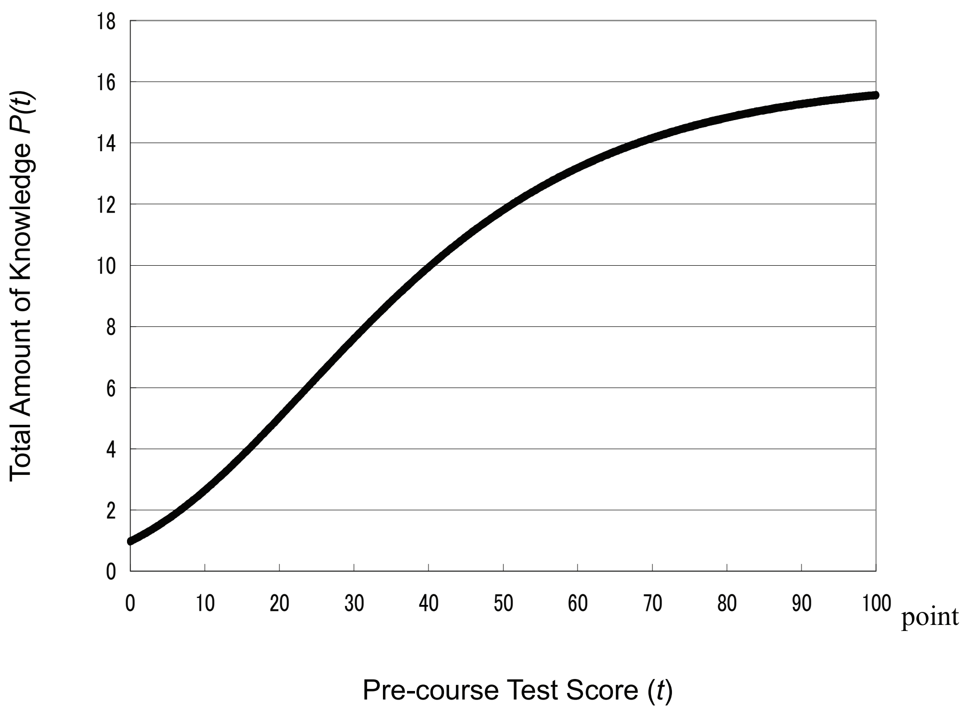


Fig. 3.9　Gompertz Curve *P(t)*

As explained in the previous section, based on the comments of English teachers with profound teaching experiences, 10 point in Post-course Score Changes (*PSC*) (10% change rate in learning progress) or more is considered proper as the educational objective of this remedial course. Therefore, 10-point of *PSC* is used as a classification point to carry out student classification. Students are divided into two classes using this criterion so that they can realize the maximum progress in their remedial course learning.

According to Fig. 3.8, 64 point of pre-course test scores (*PRTS*) can be used as the classification point for student classification when the differential curve of Gompertz learning curve is applied. That is, students with $20 \leqq PRTS \leqq 64$ should be considered the qualified students for the remedial course and be allocated to class 1 because their post-course score changes (*PSC*) are predicted to be higher than 10 points based on the differential curve of Gompertz learning curve. On the contrary, students with $64 < \mathrm{PRTS} \leqq 100$ should be considered unqualified students for the remedial course and should go to class 0 because they tend to show a change rate of less than 10 point in their post-course testing. Students with their pre-course test score (*PRTS*) above 64 points are supposed to be offered more challenging contents since the established remedial education program doesn't meet their English proficiency levels and they are not benefited from the remedial course study. Their English competence can be improved more intensely if proper contexts are offered.

The calculation of the prediction accuracy for student classification using Gompertz learning curve is conducted based on the method proposed in the previous chapter. Consequently, 83 students in the training data show correct classification prediction and the prediction accuracy is 69.2% (=83/120). When Gompertz learning curve is applied to testing data, 76 students show correct prediction results and the prediction accuracy is 63.3% (=76/120). The prediction accuracy has been largely improved than the case of multiple regression analysis.

3.3.4 Logistic Learning Curve Model

3.3.4.1 Defining Logistic Learning Curve

As demonstrated above, Gompertz learning curve turns out to be the model which better fits the observed data than exponential curve and modified exponential curve. The reason for this lies in the shape of Gompertz curve. Its S-shape coincides with changes in students' academic abilities emphasized by Item Response Theory (*IRT*, Appendix V). Students with extremely low academic abilities do not show much difference with each other in the probabilities of giving correct responses to test items. It is the same case with students with extremely high academic abilities. But, students of the medium level make rapid progress during the course study. Thus, the conclusion here is that when pre-course test score is extremely low or extremely high, the post-course score change (*PSC*) shows a slight change and a very small figure; but, when the pre-course test score is around the medium level, the post-course score change (*PSC*) tends to increase more sharply and a big change can be perceived in students' total amount of knowledge.

Logistic curve was used to predict the increasing process of a biological individual organism originally, such as the spreading of infectious disease or the increase of population [77]. It represents the nature of a growing process which can be trailed by an S-shape curve. In contrast to Gompertz curve, logistic curve appears symmetrical, but the difference between Gompertz curve and logistic curve in applications is not clearly indicated. Researchers believe that neither of them shows substantial advantage over the others and their workabiity for a specific case needs to be proved in empirical studies.

Based on the above thinking, logistic learning curve is chosen as the next model for data fitting, since it possesses similar properties as Gompertz. Following the implication of learning curve discussed in the previous section, students' total amount of knowledge relies heavily on their pre-course test scores. It accelerates rapidly as the pre-course test score increases at the beginning stage of the remedial course, but takes the form of plateaus gradually and shows very slight slope. This relationship between students' pre-course test scores and their total amount of knowledge can be expressed by the following differential Eq. 3.16:

$$\frac{d}{dt}P(t) = a\,P(t)\{K - P(t)\} \quad , \qquad (a>0,\ K>0) \qquad (3\text{-}16)$$

where,

t is the pre-course test score which corresponds to the amount of prior knowledge students have before the course study,

a is the rate at which knowledge is absorbed in the course,

K is the asymptote of the logistic function $P(t)$, indicating the total amount of knowledge students need to master during the course study.

In the logistic learning curve, the x-axis *t* stands for the pre-course test score and the y-axis *P(t)* refers to students' total amount of knowledge. *K* is the asymptote whose distance to the given curve *P(t)* tends to be zero and coefficient *a* means the rate at which students absorb new knowledge.

In fact, *a* is not a constant value, because the absorption rate is seriously influenced by the level of the new knowledge and thus keeps on changing. Generally speaking, the absorption rate is high when the level of new knowledge is low. As the level of new knowledge increases, the absorption rate decreases. That is why it takes time for students to understand some difficult questions. But, for the sake of convenience in calculation of this study, the absorption rate *a* is simplified as a fixed value.

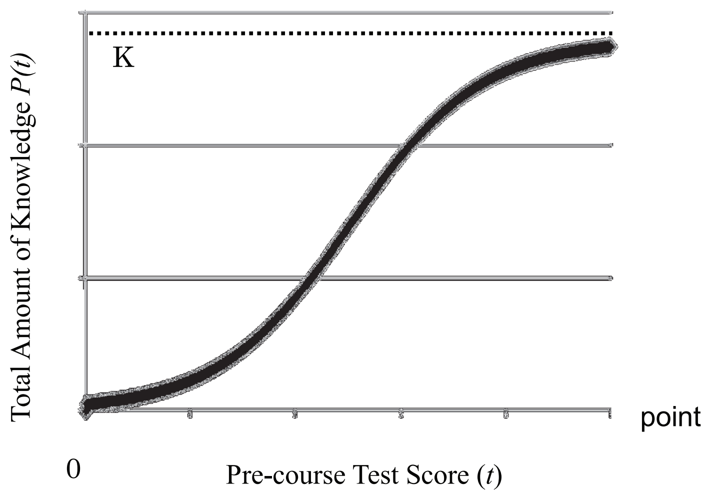


Fig. 3.11 Logistic Learning Curve *P(t)*

Eq. 3.17 is obtained by transforming Eq. 3.16 and integrating in both sides as follows:

$$P(t) = \frac{K}{1 + e^{KC} \cdot e^{-Kat}} \tag{3-17}$$

C is an integral constant. The general view of Eq. 3.17 is shown in Fig. 3.11. The x-axis *t* stands for the pre-course test score and the y-axis of *P(t)* implies students' total amount of knowledge they mastered.

The equation of the differential curve for students' post-course score change rate (*PSCR)* is derived from Eq. 3.18:

$$p(t) = \frac{d}{dt}P(t) = \frac{K^2\, a\, e^{KC}\; e^{-Kat}}{(1 + e^{KC}\; e^{-Kat})^2} \qquad \textit{(a>0, K>0)} \qquad (3\text{-}18)$$

Fig. 3.12 shows the general view of the differential curve *p(t)* of Eq. 3.18.

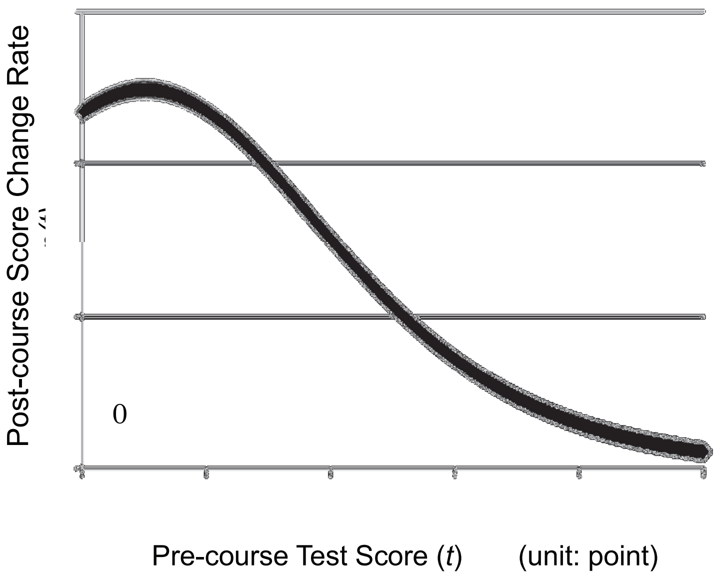


Fig. 3.12 Differential Curve *p(t)* of Logistic Learning Curve

Like in Fig. 3.11, the x-axis in Fig. 3.12 stands for the pre-course test score. The differential function *p(t)* implies their post-course score change rate (*PSCR*). The curve shows the steepness at the starting stage, but after amounting to a maximum point, it tends to decrease until it gets very close to the x-axis. This graph appropriately describes the relationship between students' post-course score change rates (*PSCR*) and the amount of the knowledge they have mastered during the course. Their newly-acquired knowledge is considerably affected by their prior scholastic competence. When they do not have much background knowledge and need to start the learning from scratch, they can obtain new knowledge at a good pace. The tangent point of the differential curve denotes the maximum amount of new knowledge students can obtain through the course learning. The downward slope shows a decline in students' newly-acquired knowledge as the level of their prior knowledge increases. It shows the "diminishing return" feature of the learning curve model explained at the beginning of this chapter.

3.3.4.2 Seeking Applicable Curves for Logistic Learning Curve Model

$P(0)=P_0$ is defined for the case of $t=0$ ($PRTS=0$). The following equation is obtained by substituting $t=0$ in Eq. 3.17 and then making rearrangement:

$$e^{KC} = \frac{K - P_0}{P_0} \tag{3-19}$$

Taking natural logarithms of both sides of Eq. 3.19, the integral constant C is thus fixed:

$$C = \frac{1}{K} \ln \frac{K - P_0}{P_0} \tag{3-20}$$

Suppose that $p(t)$ mounts to its maximum value when $t=t_0$, then, the gradient of the tangential line has to be equal to 0 when $t=t_0$, therefore the differentiation of $p(t)$ can be carried out as follows:

$$\frac{d}{dt} p(t) = -K^3 a^2 e^{KC} e^{-Kat} (1 + e^{KC} e^{-Kat}) \cdot \frac{1 - e^{KC} e^{Kat}}{(1 + e^{KC} e^{-Kat})^4} \tag{3-21}$$

Coefficient a is obtained by Eq. 3.21 on the condition that $dp(t_0)/dt=0$:

$$a = \frac{C}{t_0} \tag{3-22}$$

Assume that the maximum value of $p(t_0)$ is m at the vertex of the curve, then m is derived from Eq. 3.18 as follows:

$$m = \frac{K^2 a e^{KC} e^{-K a t_0}}{\left(1 + e^{KC} e^{-Kat_0}\right)^2} \tag{3-23}$$

In addition, P_0 is available by substituting Eq. 3.20 and Eq. 3.22 for Eq. 3.23:

$$P_0 = \frac{K}{1 + e^{\frac{4 m t_0}{K}}} \tag{3-24}$$

The values of K, m and t_0 are set within the ranges of $0 \leqq K \leqq 100$, $0 \leqq m \leqq 1$ and $0 \leqq t_0 \leqq 50$, respectively. According to the values of K, m and t_0 , the values of C, a and P_0 are determined by Eq. 3.20, Eq. 3.22 and Eq. 3.24, and thus differential function $p(t)$ becomes available.

The root mean square error between the observed data and the post-course score change rate data obtained through Eq. 3.17 turns out to be 1.02, with K=20.2, m=0.25 and t_0 =23.0. When the pre-course test score (*PRTS)* is 23 points, their post-course score change rate (*PSCR*) reaches the maximum value of 0.25. Theoretically, the maximum value of the total amount of knowledge students mastered during the course can reach 20.2 if x-axis extends infinitely. But, the range of the pre-course test score *(PRTS)* is set between 0 ~ 100 as shown in Fig. 3.14, thus, the values of students' total amount of knowledge are settled from 4.9 to 19.8. The value of the constant a is 0.0025.

The differential curve $p(t)$ and the logistic learning curve $P(t)$ fitting the actual data are illustrated in Fig. 3.13 and Fig. 3.14, respectively. According to Fig. 3.13, students with their pre-course test scores (*PRTS*) around 61 points in the pre-course test show a learning progress of 10% in their post-course score change rate. Therefore, it is reasonable to propose that 61 points should be the criterion for student classification using logistic learning curve if 10% of post-course score change rate (*PSCR*) is set as the objective for the remedial course. That is to say, students with $20 \leqq PRTS \leqq 61$ should go to class 1 as qualified students because their *PSC* are predicted to be higher than 10 points based on the differential curve of logistic learning curve; students with $61 < PRTS \leqq 100$ should go to class 0 as unqualified students because their *PSC* are predicted to be lower than 10 points.

The same method proposed in the previous chapters is used for the calculation of prediction accuracy. As a result, 83 students in the training data show correct classification prediction and the prediction accuracy is 69.2% (=83/120). When using the testing data, 77 students show correct prediction results and the prediction accuracy is 64.2%. It is the same as the differential curve of Gompertz learning curve.

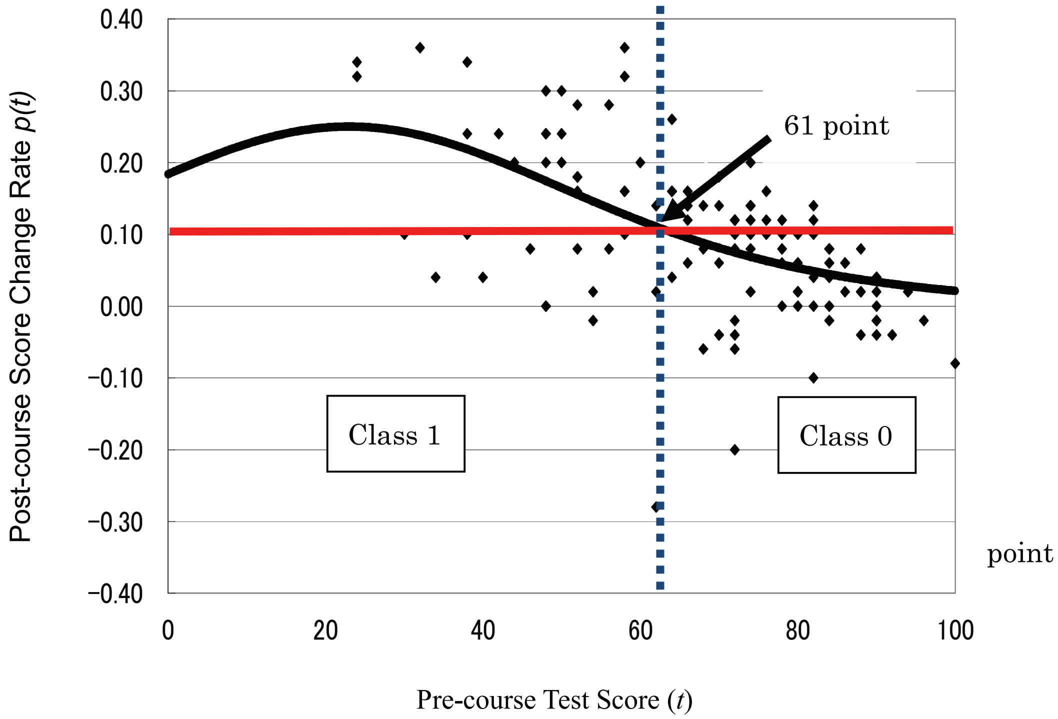


Fig. 3.13 Differential Curve *p(t)* of Logistic Learning Curve *p(t)*

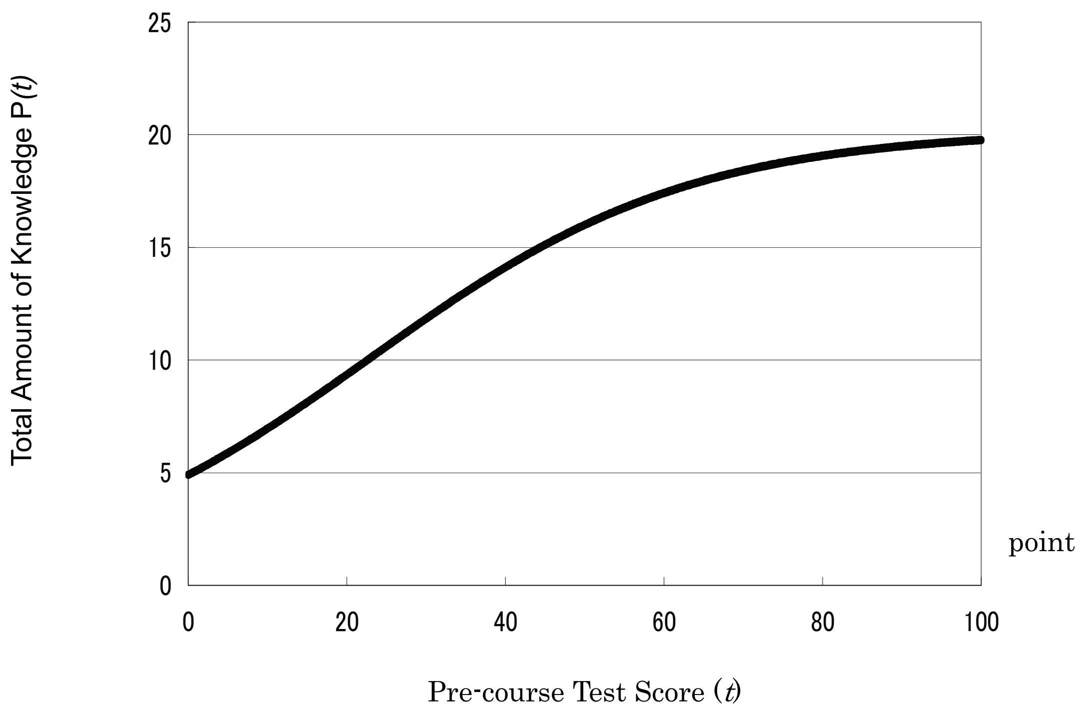


Fig. 3.14 Logistic Learning Curve *P(t)*

3.3.5 Modified Logistic Learning Curve Model

The differential curve of the logistic learning curve in Fig. 3.12 offers a convincing explanation for the relationship between students' pre-course test scores (*PRTS)* and post-course score changes (*PSC)*. It works as a convenient tool to make predictions about students' post-course score changes (*PSC*) according to their pre-course test scores so that student classification can be easily done. But, there is one doubt about the rigorousness of this differential curve. Theoretically, when students' pre-course test score is 100 points, their score change should be exactly zero rather than a positive value illustrated in Fig. 3.12. Judging from Eq. 3.18, the value of the differential function *p(t)* would never equal to zero when *t=100*. Therefore, the modification of the equation is needed in order to meet this newly imposed condition. A constant *L* for the adjustment is added to the equation of the differential function (Eq. 3.18), for the purpose of creating a modified differential curve *p(t)* which might better fit the actual data under the condition that students' post-course score change rate (*PSCR*) equals precisely to zero when their pre-course test score is 100 points (Eq. 3.25):

$$p(t) = \frac{K^2 a e^{KC} e^{-Kat}}{(1 + e^{KC} e^{-Kat})^2} + L \tag{3-25}$$

Accordingly, the modified logistic function appears as follows:

$$P(t) = \frac{K}{1 + e^{KC} \cdot e^{-Kat}} + L \cdot t \tag{3-26}$$

The following *L* is obtained from *p(100)=0*:

$$L = -\frac{K^2 a e^{KC} e^{-Ka\times 100}}{(1 + e^{KC} e^{-Ka\times 100})^2} \tag{3-27}$$

By differentiating the modified differential curve *p(t)* in Eq. 3.25, the gradient of its tangential line can be achieved and the value of coefficient *a* can be determined like Eq. 3.22. The maximum value of *m* ($0 \leqq m \leqq 1$) at $t=t_0$ can be expressed in the following:

$$m = \frac{K^2 a e^{KC} e^{-Kat_0}}{(1 + e^{KC} e^{-Kat_0})^2} + L \tag{3-28}$$

When the values of *K*, P_0 and t_0 are fixed, constant *C* can be determined by Eq. 3.20. By using the value of coefficient *a* of Eq. 3.22 for Eq. 3.27, the value of *L* is obtained. In the same way as the case of the logistic learning curve, the calculation for the optimal parameters of Eq. 3.25 can be done. The ranges for K, P_0 and t_0 are set as $0 \leqq K \leqq 100$, $0 \leqq P_0 < K$, and $0 \leqq t_0 \leqq 50$, respectively.

Consequently, the minimum root mean square error between the observed data and the post-course score change rate data obtained through Eq. 3.25 turns out to be nearly the same as the case of the logistic learning curve (1.02), with K=19.6, P_0=4.51, m=0.23, t_0 =24, and L=-0.02. The values of the two constants are obtained like $C=0.062$, $a=0.0026$. When students' pre-course test score is 24 points, their post-course score change rate (*PSCR*) reaches its maximum value of 0.23. If the x-axis is infinite, the maximum value of the total amount of knowledge is 19.6. When the value of the pre-course test score is set between 0~100 points, the range of the total amount of knowledge is from 4.51 to 17.1.

The modified differential curve and the modified logistic curve fitting the data are illustrated in Fig. 3.16 and Fig. 3.17. According to Fig. 3.16, students with their pre-course test scores (*PRTS*) around 61 points in the pre-course test show a learning progress of 10% in their post-course score change rate. Therefore, it is reasonable to propose that 61 point of pre-course test scores (*PRTS*) should be used as the criterion for student classification when the differential curve of logistic learning curve is applied. That is, students with $20 \leqq PRTS \leqq 61$ should be labeled as class 1 because their post-course score changes (*PSC*) are predicted to be higher than 10 points and they are considered as the qualified students for the remedial course. On the contrary, students with $61 < PRTS \leqq 100$ should be allocated to class 0 because their *PSC* are predicted to be lower than 10 points and they are considered as unqualified students for the remedial course.

The same method proposed in the previous chapters is used for the calculation of the prediction accuracy in this case. 83 students in the training data show correct classification prediction and the prediction accuracy is 70.8% (=85/120). When using the testing data, 78 students show correct prediction results and the prediction accuracy is 65% (=78/120). The prediction accuracy is a little bit higher than the differential curves of Gompertz learning curve and logistic learning curves. But in this study, the simple and most-widely accepted logistic learning curve is adopted as the applicable learning curve model for the observed data. Therefore, 61 point of pre-course (*PRTS*) is considered the proper classification point for student classification and the prediction accuracy is 69.2% for the training data and 64.2% for the testing data.

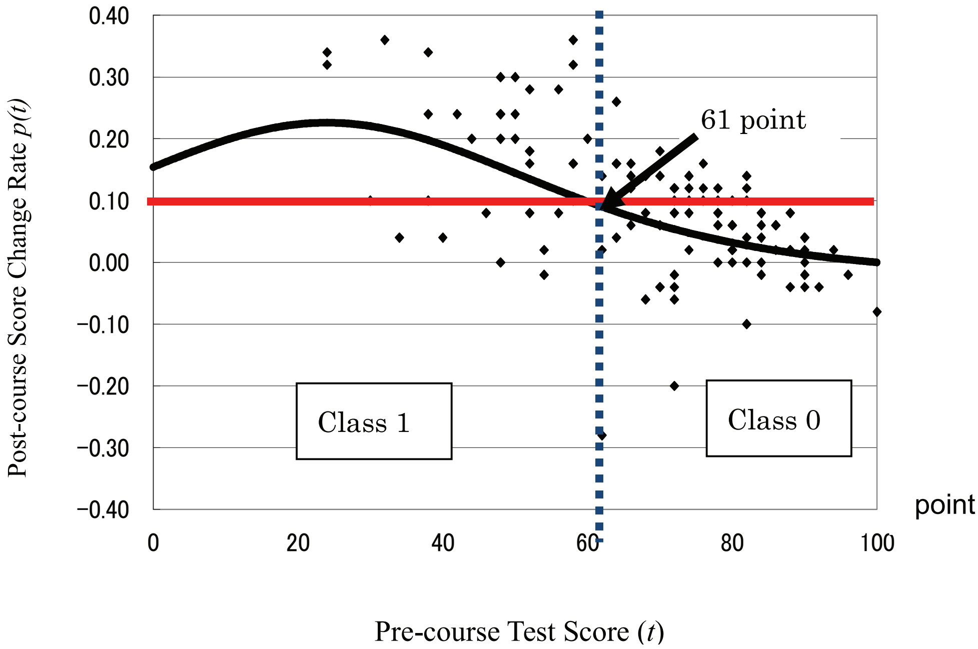


Fig. 3.16 Differential Curve *p(t)* of Modified Logistic Learning Curve

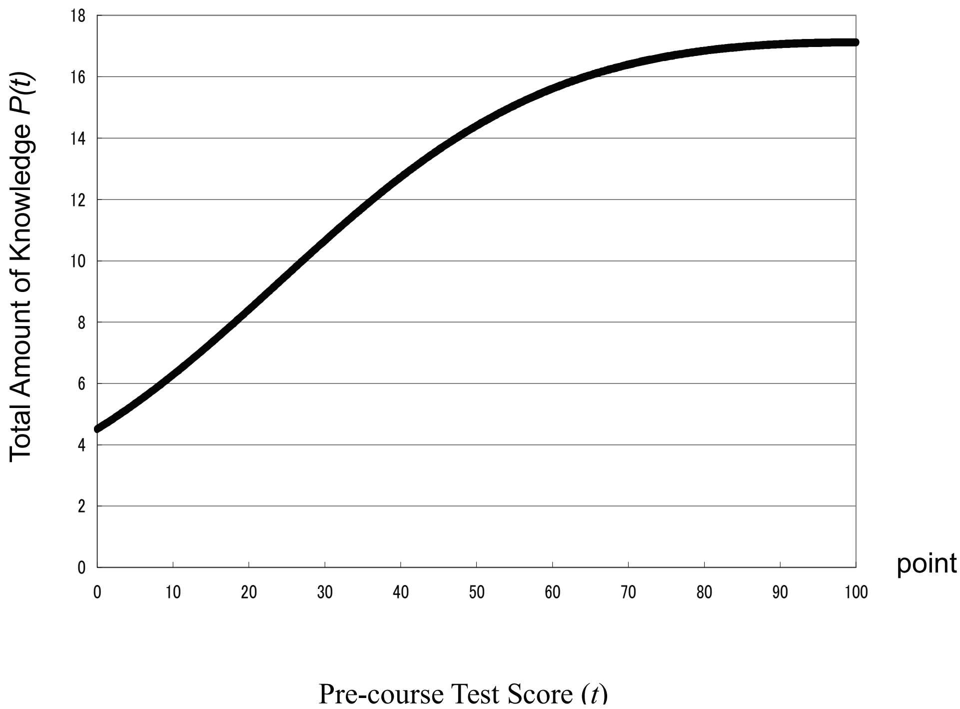


Fig. 3.17 Modified Logistic Learning Curve *P(t)*

Similar results are demonstrated by the results of the differential curves of the logistic learning model and the modified logistic learning curve model. The common point is that students with pre-course test scores *(PRTS)* of around 23~24 points benefit most from the remedial course study, and they realize a maximum post-course score change *(PSC)* of 23~25 points.

3.4 Discussion

Table 3.1 List of Applicable Learning Curve Models for Student Classification

Items		Logistic Learning Curve	Modified Logistic Learning Curve	Gompertz Learning Curve
Total amount of knowledge students have at 0 point in the pre-course test score scale (t=0)		4.9	4.51	0.97
Total amount of knowledge students have mastered at 100 point in the pre-course test score scale (t=100)		19.8	17.1	15.6
Pre-course test scores of those students who demonstrate the highest score change		23	24	23
Highest post-course score change rate (PSCR)		0.25	0.23	0.26
Error		1.02	1.02	1.02
Criterion for student classification when 10% of score change is set as the objective of the course		61	61	64
Prediction Accuracy	Training	69.2%	70.8%	69.2%
	Testing	64.2%	65.0%	63.3%

According to the basic principles of learning curves, students with extremely low academic abilities and those with extremely high academic abilities tend to show unnoticeable improvements in their scholastic achievements, while students of the middle level usually make large progress in learning. The differential exponential learning curve model is inapplicable due to its huge errors comparing with other models. The modified exponential curve is not considered reliable because it fails in offering a correct prediction for students' total amount of knowledge and it shows a negative value when pre-course test scores fall into the range of 0~2.87 points. As a result, the three S-shaped learning curve models (Gompertz learning curve model, logistic learning curve model and the modified logistic learning curve model) are eligible for the assessment of students' academic proficiencies and for the establishment of criterion for proper student classification.

The most outstanding feature of the learning curve model lies in its consideration of students' post-course score changes (*PSC*), rather than the method of using students' pre-course test score (*PRTS*) as the only criterion like in the commonly-accepted method for student classification. Although there are slight differences among the three models in the accuracy of predicting all the major indexes, they demonstrate the identical effectiveness in depicting students' score changes. The following conclusions can be drawn: (1) Students who score around 23~24 points in the pre-course test benefit the most from the remedial course; (2) The maximum score changes in the post-course test are around 24~25 points; (3) As explained in Table 3.1, the standard point for student classification is around 61～64 points if 10% of the score change is set as the objective of the course. Because the scholastic competences of the freshmen in Nishinippon Institute of Technology (Japan) are roughly the same each year, it is meaningful to make a suggestion that 61~64 points should work as a basic criterion for teachers to conduct student classification. According to the criterion for the selection of a learning curve (as explained in Section 3.2), the learning curve with the smallest root mean square error should be used for student classification. In this study, the three S-shaped curves show the same value in their error, because their characteristics in expressing the relationship between pre-course test scores (*PRTS*) and post-course score changes (*PSC*) are the same. Although the modified logistic learning curve shows a little higher prediction accuracy, the most-widely used logistic learning curve is adopted as the tool for student classification. Therefore, 61 point is considered the proper classification point to divide students for the remedial course study. The three applicable learning curve models with their differential curves in S-shape are summarized in Table 3.1.

Learning curve models can be employed as a useful tool to help teachers to make objective assessment about student's academic abilities. Students need to be given exact instructions on what contents they need. If various contents are provided to students on the basis of their performance in learning, students will be able to select courses which offer them the most appropriate contents to fit their individual level, and the effectiveness of English remedial education will be tremendously enhanced. The features of the learning curve model can be summarized as the following:

The differential curves of the S-shaped functions show validity in making prediction for every individual student's scholastic performance during his/her study at the remedial course. Therefore, the criterion suggested by the Gompertz learning curve model, the logistic learning curve model and the modified logistic learning curve model is practically feasible. The logistic learning curve is adopted in this study because it is the most widely-used among the three models.

The current framework for student classification is established on students' pre-course test scores (*PRTS)*, because it is believed to be the key factor reflecting the true nature of respective student categories [41]. Furthermore, this study also focuses on students' post-course changes *(PSC)* rather than only on their pre-course test scores *(PRTS)* and highlights the benefit each student receives from learning the course. Thus, the problem of how to make student classification is solved more effectively because the change in scholastic performance of every individual student is taken into account.

(1) The learning curve model shows more reliability than the commonly-accepted method for student classification. The common practice is to carry out pre-course test before the course study and divide the students into groups based on the information of their pre-course test scores and human observation. This method has been criticized as both time-consuming and low in fidelity [42]. Students are divided according to their scholastic abilities before the course rather than the potential progress they will make through learning. The learning curve method shows the students' current ability levels and their potential changes. The mentioned three learning curve models show slight differences in the final results, which also strengthens the appropriateness of exploiting these methods for student classification. Thus, the consequence is more precious than intuition which props the empirical validity of most classification methods as a vague standard and saves time than human operations.

(2) Techniques for student classification with low-fidelity can be as accurate as live classification carried out by human beings [42]. The S-shaped learning curve model is feasible for describing students' academic abilities and is of more practical convenience due to the simplified procedures in operation. The differential curves of the S-shaped learning curves show their most considerable advantage in student classification because they

provide evidence for the process based on the basic thinking on how much students are benefited from the course. It not only works as a method based on the details of any individual student's performance, but also draws the attention of teachers to the issue of how to make improvements in didactic plan as a whole [39].

(3) In sum, selection bias should be avoided in the process of student classification [78]. Though the pre-course test score is commonly applied as the criterion for student classification, students' progress during the learning process has not been paid enough attention. Student classification based on the logistic learning curve model is a method with scrupulous efficiency to a large extent, because identifiers are proposed to exhibit the relationship between the values of the two significant exponents rather than other less concomitant attributes, i.e., students pre-course test scores (*PRTS)* and their post-course score changes (*PSC)*.

3.5 Practical Application

The essential policy for student classification lies in the point whether students benefit from attending the course study or not. Through the comparison of students' performances in the pre-course test and post-course test, changes in their scholastic abilities are clearly reflected. Therefore, Pre-course Test Score *(PRTS)* is regarded as a reliable and objective reference for student classification. Although the same term is used for student classification, *PRTS* in the logistic learning curve model has a different implication from its common usage. When a student's prospective scholastic progress falls into the range near the maximum value, the arranged course is judged most suitable for his/her academic competence. In other words, the purpose of student classification is to match students with courses, where they can make their score changes the maximum.

The same teaching material may produce diverse outputs. The effectiveness of the teaching material heavily relies on the reciprocal interactions of the various factors involved in the pedagogical environment. Evaluation is one of the most significant challenges that both teachers and researchers have to face [79]. The logistic learning curve hypothesis provides a compendium method for student classification, but human adjustment is needed, while other predictor attributes are taken into account. Subjective factors, such as restrictions of the learning environment and students' individualities might exert impacts on the education production, although the impact differs from case to case [80].

The constructive suggestion of logistic learning curve is to establish a standard evaluation system for all English textbooks and teaching materials on the basis of the logistic learning curve hypothesis. If a logistic learning curve is created for each textbook and teaching material, teachers might be able to make reference for better judgments in textbook selection and student classification. The course is decided by the contents of the textbook and teaching materials, which should refer to the gap in knowledge levels between the starting point and the ending point. The ideal model for the course organization is to have all the candidate students' expected progresses fall within the range of the course. That means the course is suitable for their scholastic competence and they are benefited from the course learning.

The key point about the evaluation system is the standardization of the level assessment for all textbooks and teaching materials. That is, the criterion provided by the differential curve of logistic learning model should be universal so that absolute evaluations of students' academic abilities are carried out. The most convenient way to achieve this goal is to make use of the most widely accepted language proficiency test, such as *TOEIC* (Test of English for International Communication) or *TOEFL* (Test of English as a Foreign Language). They are held by the authoritative test development institution *ETS* (Educational Testing Service in USA), thus the test scores are of less controversy. Test scores of *TOEIC/TOEFL* can be used as pre-course test scores in the analyses, and the differential curve of the logistic learning model indicates the potential post-course score change (*PSC*) for every point along the horizontal axis of pre-course test scores (*PRTS*). Because *TOEIC and TOEFL* are recognized worldwide and their scores are more convincing, the integration of such scores to the logistic learning curve model creates a connection between the logistic learning curve model and the global standard for English communication skill assessment.

In the current study, questions were randomly chosen from the database of the web-based learning system. There is the possibility that students' real abilities are not sufficiently reflected by the scores. If the two testing (pre-course test and post-course test) are substituted by *TOEIC/TOEFL*, the evaluation based on the logistic learning curve hypothesis in predicting students' academic ability progresses will be more persuasive.

Chapter 4 Students' Classification using Enhanced GNP-based Class Association Rule Mining with Attribute Selection through Genetic Algorithm

The differential curve of the logistic learning curve model proposed in Chapter 3 describes the relationship between pre-course test scores (*PRTS)* and post-course changes *(PSC)*. It can function as a convenient tool for student classification by using students' pre-course test scores (*PRTS*) for the remedial course study. In this chapter, attempts will be made to solve the problem of student classification through data mining using GNP-based class association rule mining with consideration of students' pre-course test scores (*PRTS*). Students with desirable learning progress (i.e., $PSC \geqq 10$ points) are defined as Class 1 and students with undesirable learning progress (i.e., $PSC<10$ points) are defined as Class 0. GNP-based (Genetic Network Programming) class association rule mining is applied to the training data to discover rules in both class 1 and class 0. Then, classification is carried out for the testing data using these extracted rules. Advanced GNP-based method with attribute selection through Genetic Algorithm (GA) is applied to this study for the purpose to achieve a higher prediction accuracy. Attribute selection through GA prunes the unrelated or redundant predictor attributes in the original database by forming small data subsets with relevant predictor attributes only. They are giving to the algorithm for class association rule mining instead of the original big database. Therefore, simple but important rules are available for classification. The simulation result shows that a higher prediction accuracy is obtained through the proposed method than the conventional GNP-based method.

4.1 Student Classification through Conventional GNP-based Class Association Rule Mining

In order to realize the goal of conducting student classification without using students' pre-course test scores (*PRTS*), class association rule mining using Genetic Network Programming (GNP) is proposed for student classification in this chapter. Genetic Network Programming (GNP) is the evolutionary optimization technique which can better represent the solutions by using compact directed-graph structures as genes. The purpose of GNP-based class association rule mining is to make further exploration about the collected data by discovering the important attributes which are closely related to students' learning progress. Through the extraction of important rules from the data, features of the students in Class 1 (students with their post-course score changes (*PSC*) of 10 points or more) are discovered for student classification are figured out. The extracted rules offer evidence for teachers' decision-making in student classification and make it possible to conduct prediction for student classification. The extracted rules, which describe the features of the students in Class 1 and Class 0, respectively, are called classifiers in this study.

TID	*Items*	*k*
1	A_2, A_3	1
2	A_1, A_2, A_3	0
3	A_1, A_3,	1

⇨

TID	A_1	A_2	A_3	*K*
1	0	1	1	1
2	1	1	1	0
3	1	0	1	1

Fig. 4.1 Original Database for GNP (1)

TID	*A1*	*A2*	...	*A65*	*K*
01	1	0	…	1	0
02	0	0	…	1	0
...	…	…	…	…	…
N	1	1	…	0	1

Fig. 4.2 Original Database for GNP (2)

Table 4.1 An Example of Extracted Rules

Attribute	Attribute Content
2	I am concerned about what the teacher says in class.
10	A certificate in English will bring me economic benefits.
18	I learn English for credits.
21	One's English proficiency depends on his/her efforts.

Fig. 4.1 and Fig. 4.2 show the explanation for the database used in this study. The total number of predictor attributes for each record is L=66, which include one class attribute k. N=120 records are available from the inquiry to form the original database for this study. *TID* stands for the record number in a database. A_i (i=1,2,3) is predictor attributes in the collected data. k indicates the class attribute. In this study, all students are divided into two groups: students who are qualified candidates for the English remedial course study (Class 1) and students who should be given more challenging contents for learning (Class 0).

Each extracted rule is expressed by the predictor attributes in the data subset. Table 4.1 is an example of the exacted class association rules. The class association rule is expressed by {2, 10, 18, 21} for Class 1 and the predictor attributes of this rule are shown in Table 4.1.

Table 4.2 Parameters of GNP

Parameters	Values
Number of Generations in Evolution	50
Number of Training Records	120
Number of Processing Nodes	10
Number of Judgment Nodes	75
Crossover Rate (*Pc*)	0.2
Mutation Rate (*Pm-1*) for Content	0.2
Mutation Rate (*Pm-2*) for Connection Change	1/3
Sup_{-min}, Con_{-min} and χ^2_{-min}	0.08/0.7/3.84
Maximum Number of Attributes in the Antecedent of a Class Association Rule	8

The detailed description of class association rule mining using GNP can be found as in Appendix V, including the genotype of GNP and the evaluation of the extracted rules and each GNP individuals. The parameters of GNP are listed as in Table 4.2.

The calculation of the fitness value of each GNP individual is based on χ^2 values and the number of attributes in the antecedent part of rules extracted from the individual. Besides, a positive variable *new*(r) is added like Eq. 4.1:

$$F = \sum_{r \in R} \{\chi^2(r) + 10(n(r) - 1) + new(r)\} \tag{4-1}$$

where, R denotes the set of suffixes of important class association rules extracted from the individual, $\chi^2(r)$ is the χ^2 value of rule r; $n(r)$ is the number of attributes in the antecedent part of rule r, and when rule r is a newly extracted rule, $new(r)$ has a constant value, otherwise, it equals 0. The details for the measurement of the extracted rules can be found in Appendix V.

Classification can be conducted using the extracted class association rules. The prediction accuracy is calculated based on the result of the classification. Details for the calculation of the prediction accuracy can be found in Appendix V. Fig. 4.3 shows the average accuracy for student classification through conventional GNP-based class association rule mining using the training data is 56.5 %, and the classification accuracy is 51.2% using the testing data.

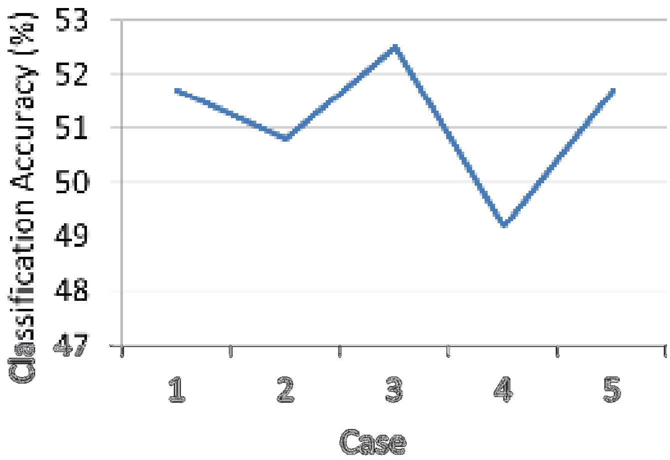


Fig. 4.3 Student Classification Accuracy using Conventional GNP-based Class Association Rule Mining (testing)

The conventional GNP-based class association rule mining has the following features:

*GNP consists of a directed graph and has three kinds of nodes. This simple structure saves both time and space for calculation, especially when it deals with a large database. Therefore, GNP-based class association rule mining works as a more convenient tool for student classification because the class association rule extraction can be finished within a relatively short period of time even when the database is large.

*The purpose of GNP-based rule extraction is to generate a sufficient number of rules for accurate student classification rather than extracting all the association rules hidden behind the huge amount of data. The maximum number for the extracted rules from each GNP individual is predetermined by the user. Therefore, GNP-based rule extraction is much less time-consuming than other methods. GNP-based rule extraction functions as a more efficient method because it executes the operation of student classification based on reliable association rules within a relatively short period of time.

*By using GNP-based rule extraction, the maximum number of predictor attributes in each extracted rule can be determined flexibly by users. Adjustment can be carried out according to the needs of the users. When dealing with a big database, the maximum number of predictor attributes is usually a relatively large figure; On the contrary, when the database gets smaller, the maximum number of predictor attributes tends to be smaller to an acceptable level, which allows the algorithm to exact a sufficient number of associate rules for student classification. In all, the number of predictor attributes is adjusted according to the size of the original database and the convenience of users.

*Through class association rule mining, the identifier for student classification is established on the basis of the acquired information about students in Class 1 (students with desirable post-course score changes (*PSC*), $PSC \geqq 10$ points) and Class 0 (students with undesirable post-course score changes, $PSC<10$ points). The class association rules not only include predictor attributes which are significant in the prediction of students in Class 1 and Class 0, but also show the important combinations of predictor attributes which contribute to students' preferable progress in learning. Therefore, the criterion for student classification becomes more explicit and gives more meaningful suggestions to the improvement of pedagogical plans.

4.2 Framework of Genetic Algorithm (GA) for Attribute Selections

The database collected for this study contains a large number of predictor attributes and some redundant and irrelevant predictor attributes are included in it. Therefore, less powerful identifier tends to be obtained from the extracted rules. If such noisy items are eliminated from the original database beforehand, class association rule mining can be carried out more effectively and important rules with higher classification accuracy can be acquired. This is also one of the features of educational data. Because many factors get involved in the teaching the learning activities, inquiries need to cover all factors which may have potential influence on the educational outcome. As a result, less qualified predictor attributes tend to be extracted. Due to this feature of educational data, attribute selection has become one of the challenging tasks with educational data nowadays to deal with the noisy predictor attributes. Effective computational skills are needed especially when dealing with a large quantity of data.

Educational databases can become more effective through proper attribute selection. It is the prerequisite for the discovery of high-quality knowledge especially for classification, because the preprocessing of data increases the possibility that predictor attributes given to the mining algorithm become relevant to the class attribute. There are two ways to find out predictor attributes with more potential generalization power: wrapper approach and filter approach. Wrapper approach focuses on the attribute selection process which aims at realizing the optimization for a given classification algorithm, while filter approach prefers to preserve the relevant information from the original big database as much as possible [81][82]. Therefore, when classification is taken into account, wrapper approach is preferable compared to filter approach for the attribute selection process because it puts an emphasis on how to figure out qualified identifiers for classification. In this chapter, the wrapper approach is used by using Genetic Algorithm (GA) and GNP-based class association rule extraction to fulfill the task of optimal student classification.

Class association rule mining aims at discovering some correlated relationships between predictor attributes and class attribute in the form of "If X then Y". Usually, redundant and irrelevant predictor attributes are included in a large database and less powerful classifier tends to be obtained due to the irrelevant data. If such noisy attributes are eliminated from the original database beforehand, class association rule mining can be carried out more effectively and important rules with higher classification accuracy can be acquired.

GA techniques have been widely used for the research of attribute selection. Decision tree [83][84], euclidean decision table[85], nearest neighbor[86][87] and neural networks[88][89] are mentioned as the techniques for the selection algorithm in the previous related studies, but as far as we are concerned, there are still no research that the attribute selection is conducted on the basis of class association rule mining. For this reason, the class association rule mining in this study is conducted using Genetic Network Programming (GNP), rather than the conventional methods, such as GA and *a priori* data mining method. GNP demonstrates a superior ability to handle a large amount of information through evolution by using a compact directed-graph structure as genes [90][91]. Instead of a large number of candidate rules, GNP generates a sufficient number of important rules for classification and never causes the bloat [92]. As indicated in many previous studies, GNP is more effective than GP for controlling a large amount of information [90][91][92][93][94][95] and has been widely used for the research on class association rule mining [96][97][98][99][100]. GNP-based class association rule mining has been applied to business and management fields with dynamic and complex databases [100][101][102][103]. In this paper, a preliminary attempt is made to discover the optimal data subset with a small number of predictor attributes from the original database with a large number of predictor attributes and to obtain the higher accuracy for the classification through GNP-based class association rule mining.

The purpose of the following section is to discover the optimal data subset through the reorganization of the original database in order to improve the accuracy for student classification. As mentioned above, the selection of the optimal data subset is of crucial significance for the field of educational data mining. Most educational databases are featured by a large number of uncertain predictor attributes. This is because a variety of factors may get involved in educational activities and influence the educational outcome, which makes it difficult to make precise judgments on the use of predictor attributes. In addition, the major concern about educational data mining is how to classify students properly according to their performances [104][105]. Therefore, redundant and irrelevant predictor attributes have to be removed from the original database so that accurate student classification is obtained and proper pedagogical instructions are given. This study proposes a system called Optimal Data subset through Genetic Algorithm (OASGA), which is a brand-new way for student classification by discovering the optimal data subset with a small number of predictor attributes and realizing higher accuracy for student classification through class association rule mining.

OASGA has the following features comparing to the conventional GNP-based class association rule mining for student classification:

*The proposed method describes a selection process of qualified attributes for student classification. Irrelevant information is pruned for each record in the original database and only predictor attributes which reflect the substantial features of students in the respective class (Class 1 and Class 0) are left to form the new data subsets. Thus, GA individuals (data subsets) with condensed information are available and the efficacy for the association rule extraction is largely improved.

*Simple but important class association rules are discovered for student classification using the newly-established small data subsets. Because the attribute selection process has removed the less substantial predictor attributes from the original database, the extracted association rules are made up of predictor attributes which are less complicated and more understandable to users. These association rules help users to capture the features of students in Class 1 and Class 0, respectively and thus better serve the purpose of student classification.

*Higher accuracy for student classification can be achieved using class association rules extracted from the small data subsets, because the selection of predictor attributes improves their relevance to the class attribute (dependent variable). Irrelevant predictor attributes are pruned from the original database by attribute selection and only predictor attributes with more relevance to the class attribute are included in the extracted class association rules. Therefore, student classification can be carried out with a higher prediction accuracy than the case of the original database.

4.3 Optimal Data subsets through Genetic Algorithm (OASGA)

4.3.1 Basic Structure of OASGA

A newly proposed system with GNP incorporated carries out attribute selection to form data subsets through GA algorithm and makes an outstanding improvement in prediction accuracy. The original database may include less qualified predictor attributes which fail to show the substantial influence on the class attribute (dependent variable) and demonstrate weaker prediction abilities. The basic idea about this method is to establish new small databases (data subsets) consisting of relatively effective attributes by pruning away those less qualified predictor attributes. Class association rule mining is executed using the newly-established data subsets which include attributes with strong prediction abilities only. The possibility to obtain more qualified association rules is largely increased comparing with the original database, and the extracted association rules can better reflect the features of students in Class 1 (students with their post-course score changes (*PSC*) $\geqq$10 points) and Class 0 (students with their *PSC*<10 points) and clarify the reasons for their large progresses.

In order to conduct effective and precise student classification for the collected data, Optimal

Attribute Subset through Genetic Algorithm (OASGA) is established as an extensional system for conventional association rule mining function. Two additional functions are integrated into OASGA to fulfill the task of student classification: (i) OASGA contains a program for obtaining the optimal data subset through attribute selection; (ii) Student classification is carried out based on the GNP-based class association rule mining.

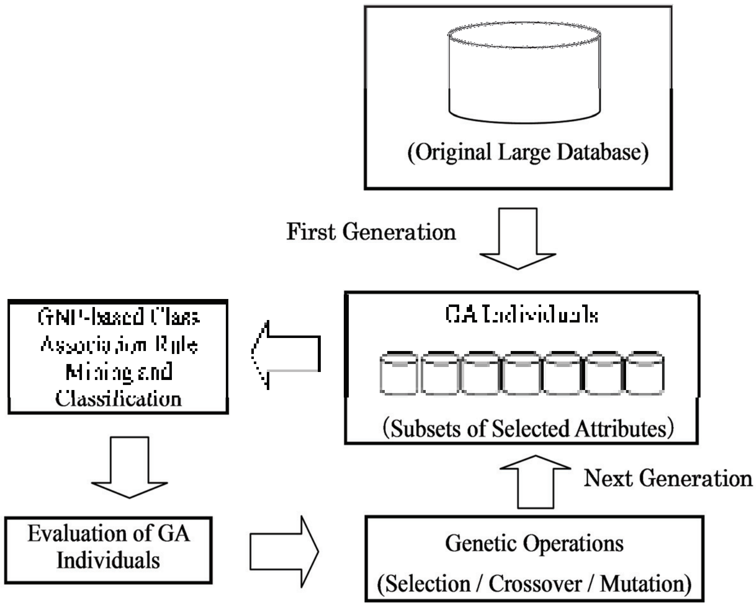


Fig. 4.4 Flowchart of OASGA

Fig. 4.4 shows the general scheme of OASGA. The basic algorithm of OASGA is to carry out a class association rule mining and classification based on the optimal data subset. In order to realize this objective, the following four steps are included in the system: (1) Initialize GA individuals (data subsets); (2) Extract class association rules using GNP; (3) Make evaluation of GA individuals by calculating their fitness value; (4) Generate new data subsets through genetic operations of GA until the optimal data subset is obtained. OASGA is a combination of hierarchical two evolutionary methods, i.e., the upper level GA and the lower level GNP.

The predictor attributes are randomly selected from the database to form the data subsets (i.e., the individuals of GA) in the first generation. Then, the following steps are executed for the selected individuals:

(i) Class association rule mining for both Class 1 (k=1) and Class 0 (k=0);

(ii) Classification based on the extracted rules;

(iii) Evaluation of GA individuals by classification.

After all GA individuals in the population are evaluated by their classification performance, GA repeats its genetic operations to generate the individuals for the next generation.

Compared with the original attributes, relevant attributes are accumulated in the newly-generated data subsets, which lead to the higher accuracy for classification. Besides, a sufficient number of qualified class association rules rather than all the class association rules are extracted through GNP for classification. GNP saves both time and space at the rule extraction stage in comparison with other methods such as *a priori*-based method. Thus, OASGA, which combines GA-based data subset selection and the GNP-based class association rule mining, becomes possible and effective in dealing with classification for large databases.

There are previous studies which attempt to combine GA with GNP for association rule mining with a large and dense database, and the paper by Gonzales et al. is the most typical [106]. Clear differences can be seen in some aspects between the proposed method and paper [106]:

(1) **Purpose**: The purpose of paper [106] is to extract as many association rules as possible and the reorganization of the database is to facilitate the association rule mining process. The purpose of the proposed method is to find the optimal data subset to realize the highest accuracy for classification;

(2) **Genetic operations of GA**: Paper [106] aims at finding out a huge number of association rules. Therefore, GA evolution is designed to convey and transmit the acquired information to the next generation for association rule extraction. That is, the genetic operation of GA is performed with the intention to increase the possibility of extracting new rules. In contrast, the aim of the proposed research is to obtain the optimal data subset which produces the highest accuracy for classification. The role of GA is to find new data subsets. Therefore, genetic operations focus on accumulating relevant data subsets rather than discovering new association rules.

(3) **Output**: In Paper [106], the final output of GA is a set of association rules stored in the rule pool, because its aim to extract as many rules as possible. While in this study, class association rule extraction works as a tool for classification and the final goal is to find the optimal data subset with

the highest accuracy. Therefore, the final output in this case is a collection of data subsets which can produce higher accuracy than the original database.

4.3.2 Attribute Selection by GA

4.3.2.1 Genotypes

As mentioned above, the attribute selection is a process for the rearrangement of data subsets generated from the original attributes. The population in the first generation consists of the predictor data subsets randomly selected from the original attributes. These subsets evolve through genetic operations and new GA individuals are generated for the next generation. GA individuals constitute of *m* attributes, where *m* is a fix parameter predetermined by users.

<table>
<tr><td>Gene</td><td colspan="5">Contents</td></tr>
<tr><td>IND</td><td>NDi</td><td></td><td>…</td><td>NDj</td><td></td></tr>
</table>

Fig. 4.5 Genotype Structure of GA

Fig. 4.5 shows the genotype of GA. *IND* stands for the GA individuals. *NDi~NDj* stand for the different attributes to be included in GA individuals. This indicates that a GA individual is a combination of predictor attributes. When more qualified attributes are included in the GA individuals, more important class association rules can be extracted. Each GA genotype (individual) encodes a potential solution to the problem of finding the optimal subset with the highest classification accuracy. As GA evolves, the less competent individuals in the population are replaced by newly created individuals through crossover and mutation and the new population contains more qualified data subset.

The more attributes are included in the database, the more rules can be extracted. But this does not imply that higher accuracy can be obtained with a larger number of extracted rules. The assumption for the class association rule extraction is that if an attribute appearing frequently in the elite individual of the last GA generation, it is considered as a vital factor in deciding the class of a record.

Fig. 4.6 shows an example of the random selection process of predictor attributes of GA individuals in the initial generation. There are *L* attributes in the original database and *m* attributes are selected randomly to form each data subset. Subsets of predictor attributes are the genes of GA, i.e., there are *n* individuals in the population. The same attribute may appear more than once in different data subsets. Class association rule mining, classification and fitness evaluation continue until the optimal data subset is obtained. After attribute selection, the original large database becomes a small database consisting of *n* data subsets with more qualified predictor attributes.

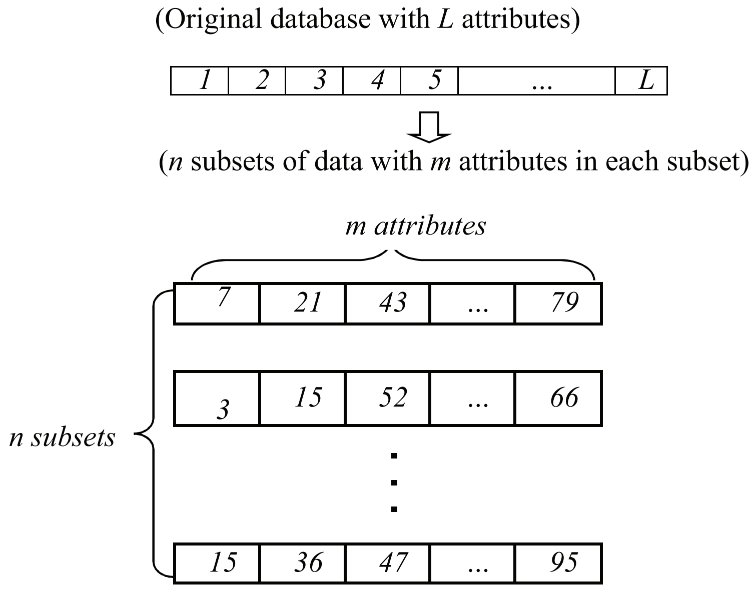


Fig. 4.6 Random Attribute Selection of GA Individuals in Initial Generation

4.3.2.2 Fitness Value

The testing data with 120 records are used in this study to evaluate the classification performance of the proposed system (OASGA) and examine the validity of the identifier. GA individuals are ranked according to the fitness value, i.e., classification accuracy. Prediction accuracy is calculated based on the result of the classification using the extracted rules. The details for calculating the fitness value of GA individuals can be found in Appendix V. The higher the fitness value is, the higher the chance is for the individual to be chosen for the next population.

4.3.2.3 Genetic Operations for GA

The gene of GA encodes each predictor data subset (Fig. 4.6) and GA genetic operations are conducted to produce more qualified predictor data subsets. Candidates of GA individuals are ranked according to their fitness values. Individuals who win the tournament selection are given the chance to be selected for crossover and mutation. The number of GA individuals and the number of selected attributes in each individual are fixed. The population in each generation is produced as follows:

I) All individuals of the current generation are ranked according to their fitness values, i.e., classification accuracy. The individual on the top of the ranking produces more important rules than individuals listed behind them. Thus, it is considered as the elite individual and moved to the next generation.

II) n/2 individuals are generated through crossover. Crossover means the exchange of predictor attributes between two parent individuals and two offspring are obtained. Four candidate GA individuals are randomly chosen out of n individuals for the tournament selection. They are ranked according to their fitness values and crossover is carried out only between the top two individuals.

The frequency for such attribute exchange is labeled as crossover rate (Cr). Eq. 4.2 is used for the calculation of the number of predictor attributes Nc to be exchanged between the parent individuals:

$$Nc = Cr \cdot m, \tag{4-2}$$

where, m stands for the number of predictor attributes in each data subset. Cr heavily depends on the total number of predictor attributes in each data subset and is determined by users. Attributes in the parent individuals are randomly selected for crossover, but the same attribute should not appear in the offspring for more than once.

Fig. 4.7 is an example explaining crossover between GA individuals. Attributes *A2* and A18 in data subset (a) are exchanged with attributes *A13* and *A62* in data subset (b). Data subsets (c) and (d) are the offspring newly generated for the next generation.

TID	*A2*	*A3*	*A18*	*A36*
1	0	1	0	1
2	0	0	1	1
3	1	1	1	1
…	…	…	…	…
N	1	0	1	0

(a) Database A

TID	*A1*	*A13*	*A34*	*A62*
1	1	0	1	1
2	0	0	1	0
3	1	1	1	1
…	…	…	…	…
N	0	0	0	1

(b) Database B

TID	*A3*	*A13*	*A36*	*A62*
1	1	0	1	1
2	0	0	1	0
3	1	1	1	1
…	…	…	…	…
N	0	0	0	1

(c) Database A (2)

TID	*A1*	*A2*	*A18*	*A34*
1	1	0	0	1
2	0	0	1	1
3	1	1	1	1
…	…	…	…	…
N	0	1	1	0

(d) Database B (2)

Fig. 4.7 An Example of GA Crossover

III) The rest individuals are generated through mutation. The number of GA individuals to be produced by mutation is ($n/2-1$). Mutation means to change certain predictor attributes in the parent individual into different ones. Similar to the case of crossover, four candidate individuals are randomly chosen out of the n individuals for the tournament selection. Mutation is conducted to the elite individual among the four. Its attributes are changed at a frequency of Mr, which is called mutation rate. Eq. 4.3 is the expression for calculating the number of predictor attributes Nm to be changed:

$$Nm = Mr \cdot m. \tag{4-3}$$

Similar to Cr, Mr is a fixed parameter determined by the user. It is closely related to the total number of predictor attributes in each data subset. Predictor attributes in the parent individual are randomly selected for mutation, but the same attribute should not appear for more than once in the offspring. Fig. 4.8 shows an example of GA mutation. Attribute *A13* in data subset (e) is changed to Attribute *A62* in the offspring data subset (f).

TID	*A2*	*A13*	*A36*	*A51*
1	0	0	1	1
2	0	0	1	0
3	1	1	1	0
…	…	…	…	…
N	1	0	0	1

(e) Database C

TID	*A2*	*A36*	*A51*	*A62*
1	0	1	1	1
2	0	1	0	0
3	1	1	0	1
…	…	…	…	…
N	1	0	1	1

(f) Database C

Fig. 4.8 An Example of GA Mutation

The features of GA genetic operations can be summarized as follows:

* GA genetic operations prune the redundant attributes, which produce relatively low fitness values, from the candidate individuals so that more effective rule-exaction can be realized.

* GA genetic operations neglect the inferior predictor attributes and strengthen the power of the qualified ones. That is, GA genetic operations functions as a process to condense the original database into a collection of substantial predictor attributes which can better explain the characteristics of each class.

* In order to avoid the diversity loss of predictor attributes in the extracted rules, the conventional method (tournament selection) [107] is used for GA genetic operations in this study.

* The individual on the top of the ranking list of the final GA generation is considered the most influential and powerful for classification because it contains more qualified predictor attributes which can produce more important class association rules.

4.4 Simulation

4.4.1 Parameters

Table 4.3 Parameters of GA

Parameters	Values
Number of Generations in Evolution	100
Number of Individuals	60
Crossover Rate (*Cr*)	0.2
Mutation Rate (*Mr*)	0.1
Number of Attributes in Each GA Individual	10/20/50/100

The setting for GA is indicated in Table 4.3. In order to prove the effectiveness of the predictor data subsets for classification, the number of attributes included in each individual data subset is set at 10, 20, 50, 100 and 130, respectively. The case of 130 attributes refers to the classification process based on the GNP-based class association rule extraction without GA-based predictor attribute selection (the conventional GNP-based class association rule mining as shown in Fig. 4.3). Because the crossover rate Cr is 0.2, the number of predictor attributes Nc for crossover is set at 2, 4, 10, 20 and 26 for the above numbers of attributes. Likewise, the mutation rate Mr is set as 0.1 and the number of attributes Nm for mutation is 1, 2, 5, 10 and 13, respectively.

4.4.2 Experimental Results

The objective of the simulation is to show the effectiveness of OASGA and make the comparison among the data subsets with 5, 10, 20, 50 and 65 attributes. The data subset with 65 attributes does not include the predictor attribute selection process for GA individuals. That is, the data subset with 65 attributes refers to the conventional class association rule mining as shown in Fig. 4.3.

The experiment is carried out using the data subsets with 5, 10, 20, 50 and 65 attributes, respectively. 5 cases of experiments are conducted for each data subset. Fig. 4.9 describes the classification accuracy obtained for the data subsets with 5, 10, 20, 50 and 65 attributes, respectively. It is clear from Fig. 4.9 that the data subset with 20 attributes shows higher accuracy than the data subsets with 5, 10, 50 and 65 attributes. Besides, the data subset with 20 attributes demonstrates a more stable trend in classification than the data subsets with 5, 10 and 65 attributes.

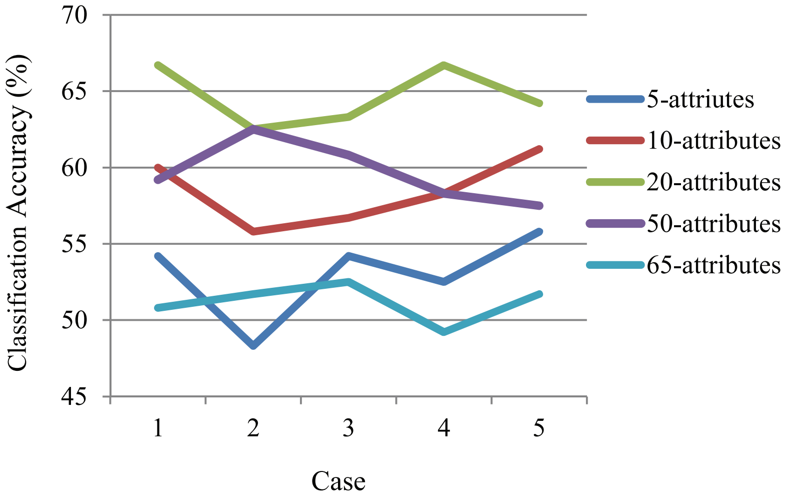


Fig. 4.9 Average Fitness Values using Data subsets of Various Sizes (testing)

Table 4.4 Average Classification Accuracy for Data subsets of Various Sizes

	Training (%)	Testing (%)
5-attributes	59.7	53.0
10-attributes	67.8	58.4
20-attributes	71.9	64.7
50-attributes	66.8	59.7
65-attributes	56.5	51.2

The experiment of the 65 attributes without GA selection shows the worst classification accuracy in Fig. 4.9. This is caused by the local optimal problem of the conventional GNP-based class association rule mining. The conventional GNP-based class association rule mining tends to fall into a local minima when it is applied to a big database with many attributes [106]. In this case, the classification ability is degraded and its accuracy may not be desirable. Our proposed method (OASGA) solves this problem through GA attribute selection because some attributes are selected to form small subsets. This gives the power to discover more important association rules. Therefore, OASGA performs better than the conventional method and is more powerful in dealing with complex educational data.

Table 4.4 describes the average of the classification accuracy using data subsets of various sizes for both training data and testing data. The average of the classification accuracy over 5 cases for the data subsets with 5, 10, 20, 50 and 65 attributes is shown, respectively. The average classification accuracy is 56.5% for the training data and 51.2% for the testing data, when the data subset with 65 attributes is used for the rule extraction and classification without GA selection of predictor attributes. The prediction accuracy is improved to 71.9% for the training data and 64.7% for the testing data when the small data subset with 20 attributes is used. Because the data subset with 20 attributes contains substantially important information, it realizes higher accuracy for student classification. Table 4.4 also indicates that the data subset with 20 attributes shows the highest classification accuracy among the data subsets of different sizes.

Thus, the conclusion for Simulation I is that the data subset with 20 attributes shows higher classification accuracy than the data subsets with 5, 10, 50 and 65 attributes. Actually, it is difficult to obtain a sufficient number of important rules from the data subsets with 5 and 10 attributes. No effective classification can be made with the data subsets with 5 and 10 attributes. While in the case of data subsets with 50 and 65 attributes, more redundant attributes are included than the data subset with 20 attributes. Therefore, the elite individual of the data subset with 20 attributes is the most preferable as the optimal data subset because it produces more important class association rules than the cases of 5, 10, 50 and 65 attributes. OASGA is effective in classification based on class association rule mining, because the optimal data subset in the last GA generation is considered the solution to the problem of student classification.

4.5 Discussion

Attribute selection is an effective method to deal with dense databases, because excellent GA individuals (data subsets) with a small number of predictor attributes can better represent the class attribute and realize more precise classification. OASGA is proposed using GA with GNP-based class association rule mining. The simulation results show that the classification made by the optimal data subset with 20 attributes is more accurate than the conventional method with the original database. The prediction accuracy has been improved from 56.5% to 71.9% for the training data and 51.2% to 64.7% for the testing data.

Attribute selection using OASGA throws the light on the field of educational data mining. Educational database has complicated relations between a large number of predictor attributes and the class attribute. Therefore, class association rules extracted for classification are always too complicated and perplexing for practical purpose. But, after the preprocessing of attribute selection, the original database becomes more revealing and indicative. As GA evolves, the optimal data subset with a small number of attributes is generated and concise class association rules are obtained for classification. Then, teachers can make pedagogical plans for instructions according to the classification results.

As mentioned above, the conventional GNP-based class association rule mining is one of the widely–accepted state-of-art technologies in the field of class association rule mining. There are other class association mining approaches, but they lack feasibility and practicality in the case of our study. Neural networks and SVM can't provide the detailed information on the crucial attributes due to the black box problem. Fuzzy clustering is useful when there are no apparent clear groupings in the database and one sample might belong to several overlapping clusters to some degree, which is obviously not the case in our study. *A priori* methods need to find the frequent itemsets. Therefore, the conventional GNP-based class association rule mining has advantages over other methods for our study because it can find the crucial attributes we need. Our study achieves its basic goal that the effectiveness of the conventional GNP-based class association rule mining is enhanced after GA is integrated into the system. OASGA is more effective than the conventional GNP-based class association rule mining because it shows better performance.

When examining the extracted rules, diversity loss is found when a small number of predictor attributes is included in the data subset. This problem can be solved by increasing the number of predictor attributes in the data subset. But, since the purpose of the proposed method is to obtain higher prediction accuracy, some irrelevant and redundant attributes have to be removed during the evolution because the prediction accuracy is low when many irrelevant predictor attributes are included in the data subsets. The predictor attribute selection through GA prunes those irrelevant predictor attributes and finds out the optimal data subset which leads to the highest classification prediction accuracy.

Chapter 5 Students' Classification Using GNP-based Class Association Rule Mining using Negative Rules

Advanced GNP-based class association rule mining is conducted for student classification through attribute selection using GA. In this chapter, class association rule mining is carried out while taking into account of the negative class association rules. In order to obtain negative class association rules, the original database is amplified when the negative aspect of all predictor attributes are added for each record. Higher prediction accuracy for student classification can be achieved when negative class association rules are taken into account because they offer a more profound and accurate explanation for the features of the students in the respective classes. Thus, more important class association rules are extracted and the classification accuracy can be improved.

5.1 Negative Class Association Rules

Class association rule mining is to discover association rules which associate attribute values with the observed class labels. Negative association rules refer to the rules which associate negations of attribute values to the observed classes [108]. Class association rule mining becomes more encouraging when negative rules are taken into account for classification. Negative association rule mining is often used when the domain contains many factors. It is more effective for classification than the conventional association rule mining because it works in a reverse manner and checks the important relation between the attribute values and the class labels of exceptional instances [109][110][111].

There are many researches which have proved the effectiveness of negative association rule mining because the prediction accuracy is more enhanced than the traditional method. Negative association rule mining has been integrated into different approaches to establish more effective classifiers, such as Apriori [109][110][112], FR Tree [113][114][115] and Genetic Algorithm (GA) [116][117]. But, as far as the author is concerned, there is still no report on a detailed examination of negative association rule mining using GNP. In this research, a preliminary study is conducted to investigate the effectiveness of the established GNP system in Chapter 4 when negative association rule mining is included.

5.2 Simulation

TID	*A1*	*A2*	...	*A65*	*α1*	*α2*	...	*α65*	*K*
01	1	0	…	1	0	1	…	0	0
02	0	0	…	1	1	1	…	0	0
...	…	…	…	…	…	…	…	…	…
N	1	1	…	0	0	0	…	1	1

Fig. 5.1 Amplified Database for GNP

The database to be used for the GNP-based class association rule mining with consideration of the negative rules is shown in Fig. 5.1. In order to discover both positive and negative rules for classification [81], the original database is enhanced by adding the reverse aspect of all the records as in Fig. 5.1. Positive rules take the form of "$X \rightarrow Y$" and indicate the positive association between predictor attributes and the class attribute. Negative rules take the forms of "$\neg X \rightarrow Y$", "$X \rightarrow \neg Y$" and "$\neg X \rightarrow \neg Y$" and indicate the negative association between the predictor attributes and the class attribute. Both positive rules and negative rules are important in decision-making [110][111][113][114][117]. The rearrangement of the original database makes the class association rules better reflect the features of both classes [118][119][120].

Simulation:

The objective of the simulation is to show the effectiveness of optimal data subset with consideration of the negative class association rules. The simulation shows that the efficiency of OASGA has been enhanced when the original database is amplified and the negative class association rules are taken into account. Therefore, Simulation aims at a comparison among the data subsets with 10, 20, 50, 100 and 130 attributes using the amplified database and the optimal data subset selection method. The data subset with 130 attributes refers to the conventional method for classification based on class association rule mining using GNP, which does not include the predictor attribute selection process for GA individuals

The experiment is carried out using the data subsets with 10, 20, 50, 100 and 130 attributes, respectively. 5 cases of experiments are conducted for each data subset. Fig. 5.1 describes the classification accuracy obtained for the data subsets with 10, 20, 50, 100 and 130 attributes, respectively, using testing data. It is clear from Fig. 5.1 that the data subsets with 20 attributes show the highest classification accuracy than the data subsets with 10, 50, 100 and 130 attributes.

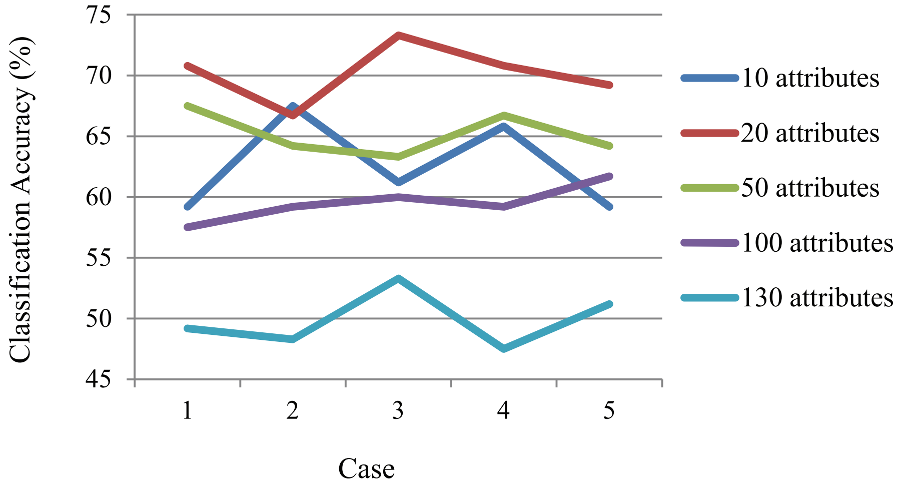


Fig. 5.1 Average Fitness Values using Data subsets of Various Sizes (testing)

Table 5.1 Average Classification Accuracy for Data subsets of Various Sizes

	Training (%)	Testing (%)
10 attributes	73.9	62.6
20 attributes	74.5	70.2
50 attributes	70.6	65.2
100 attributes	64.3	59.5
130 attributes	52.7	50.0

Table 5.1 describes the average of the classification accuracy over 5 cases of the data subsets with 10, 20, 50, 100 and 130 attributes, respectively. It is shown from Table 5.1 that the average classification accuracy is 52.7% for the training data, 50.0% for the testing data when the data subset with 130 attributes is used for the rule extraction and classification without GA selection of predictor attributes, while the accuracy is improved to 74.5% for the training data, 70.2% for the testing data when the small data subset with 20 attributes is used. Because the data subset with 20

attributes contains substantially important information, it realizes the highest prediction accuracy for classification.

Table 5.2 Comparison of Average Classification Accuracy of Various Cases

	Original Database (65 attributes)		Amplified Database (130 attributes)	
	Training (%)	Testing (%)	Training (%)	Testing (%)
10 attributes	67.8	58.4	73.9	62.6
20 attributes	71.9	64.7	74.5	70.2
50 attributes	66.8	59.7	70.6	65.2

Table 5.2 is a comparison of the experimental results between the cases when the original database with 65 attributes and the amplified database with 130 attributes are used, respectively. The amplified database refers to the case when negative association rules are taken into account. As shown in Table 5.2, the average accuracy is improved when the negative aspect of the attributes is added to the original database. The effectiveness of the negative association rules has been varified by the results of the three cases listed in Table 5.2.

5.3 Educational Implications

One of the important advantages of GNP-based class association is that it can find combinations of attributes which influence the prediction significantly. As explained in the previous chapter, GNP-based class association is used for this study because it is different from other blackbox structured technologies. It makes it possible to obtain explicit and understandable class association rules for student classification. By making detailed analysis of the extracted rules, attributes and attribute combinations which appear frequently in the extracted rules can be discovered. They offer hints to teachers of English so that they can pay special attention when giving instructions to students and thus teaching effectiveness can be improved.

As indicated in Appendix V, confidence refers to the strength of the extracted rules and χ^2 refers to the correlation of the extracted rules in classification. The extracted rules with a confidence value of 1.0000 are considered to hold the maximum strength and reliability among all rules. Based on the simulation result above, detailed analysis is conducted for the extracted rules using data subsets with 20 attributes with negative class association rules taken into account. Table 5.4 shows the important rules extracted from the data subsets with 20 attributes using amplified database for both Class 1 and Class 0. Altogether, there are 3597 important rules are available. Among them, 3154 rules are used for describing the features of Class 1 and 443 are used for Class 0. Table 5.5 and Table 5.6 show the lists of attributes which frequently appear in the important rules (with their confidence *Con.*=1.0000) from the data subsets with 20 attributes using the amplified database for Class 1 and Class 0, respectively. The appearance ratio in Table 5.5 and Table 5.6 equals to Appearance Times / Total Number of Extracted Rules.

Table 5.4 Number of Important Class Association Rules with Confidence (*Con.*) = 1.0000

	Class 1	Class 0	Total Number of Important Rules
20-attributes (with Negative Rules)	3154	443	3597

Table 5.5 Attributes Appearing in Important Class Association Rules for Class 1 using 20-attributes (Appearance Ratio >30%)

Attribute Number	Appearance Times in Exacted Rules	Total Number of Rules in Class 1	Appearance Ratio
1	2533	3154	80.3%
18	2149	3154	68.1%
14	1989	3154	63.0%
17	1501	3154	47.6%
19	1285	3154	40.7%
8	1258	3154	39.9%
15	1126	3154	35.7%
3	1047	3154	33.2%

Table 5.6 Attributes Appearing in Class Association Rules for Class 0 using 20 attributes (Appearance Ratio > 30%)

Attribute Number	Appearance Times in Exacted Rules	Total Number of Rules for Class 1	Appearance Ratio
13	288	443	65.0%
11	183	443	41.3%
9	181	443	40.9%
20	170	443	38.4%
5	136	443	30.7%

Table 5.7 and Table 5.8 are detailed analyses of the important attribute combinations which appear frequently in the important class association rules of Class 1 and Class 0 with their confidence *Con.*=1.0000, respectively. The same as in Table 5.5 and Table 5,6, the appearance ratio in Table 5.7 and Table 5.8 equals to Appearance Times / Total Number of Extracted Rules.

Fig. 5.7 Frequently Appearing Combinations in Class 1 Association Rules with *Con.*=1.0000

Attribute Combinations	Appearance Times	Total Number of Rules	Appearance Ratio
$14 \wedge 18$	2174	3154	68.9%
$14 \wedge 19$	2015	3154	63.9%
$1 \wedge 19$	1578	3154	50%
$1 \wedge 16$	1502	3154	47.6%
$17 \wedge 18$	1218	3154	38.6%
$18 \wedge 20$	1164	3154	36.9%
$15 \wedge 18$	1096	3154	34.7%
$14 \wedge 17$	1037	3154	32.9%
$14 \wedge 15$	1001	3154	31.7%

Table 5.8 Frequently Appearing Combinations in Class 0 Association Rules with *Con.*=1.0000

Attribute Combinations	Appearance times	Total Number of Important Rules	Appearance Ratio
13∧17∧20	124	443	28%
11∧13∧14∧11	115	443	26%
5∧9∧16	114	443	25.7%
13∧11	76	443	17.2%

Table 5.9 Explanations of Attribute Numbers

Attribute Number	Contents
1	English learning is very interesting.
3	English will be useful in my future life.
5	I should learn Japanese well first rather than English.
8	I don’t want to lose my current English ability.
9	English class always makes me upset.
11	People need go abroad if they want to learn English well.
12	I will have trouble in my work if I am not good at English.
13	Some people have an inborn ability to learn English well.
14	English is necessary for advanced professional knowledge learning.
15	It will be too late if I don’t learn English from now on.
16	I keep a high attendance rate in all classes.
17	People around me are working hard at English.
18	I learn English for credits.
19	English is the one of the basic skills one should have.
20	I love watching foreign movies.

Table 5.9 offers a detailed explanation of attribute numbers. The important class association rules and the frequently appearing attribute combinations in the important class association rules reveal the features of students who achieve score changes of more than 10 points or less than 10 points. Therefore, instructions concerning them need to be strengthened in the teaching activities in order to improve educational efficacy.

It is not surprising that some of the attributes have not appeared in the extracted rules even once. This simply proves that they do not classify students with different English proficiencies and should not be used as criteria for classification. The features of the proposed method for student classification by using OASGA based on associate rule extraction through GNP are as the follows:

(1) Attribute selection is an effective method to deal with dense databases, because excellent individuals with a small data subset can realize more precise classification.

OASGA is proposed using GA with GNP-based class association rule mining. GNP has a superior power in controlling both space and time preventing the bloating. GNP makes it possible to perform the attribute selection for class association rule mining using GA. The simulation result shows that the classification made by the optimal data subset is more accurate than the conventional classification method with the original database.

(2) Attribute selection using OASGA throws light on the field of educational data mining because as far as the author is concerned, it is the very first attempt to conduct association rule extraction using educational data.

Educational database has complicated relations between a large number of predictor attributes and the class attribute. Therefore, class association rules extracted for classification are always too complicated and perplexing for practical purpose. But, after the preprocessing through the attribute selection, the original database becomes more revealing and indicative. As GA evolves, the optimal data subset with a small data subset is generated and concise class association rules are obtained for classification.

(3) The results of this research have very significant implications on the improvement of educational effectiveness.

Consequently, the attribute combinations in Table 5.7 and Table 5.8, especially Table 5.7, are the attributes contributing most to students' progress in remedial course study. They can be considered as the core of the original inquiries with all of redundant and irrelevant attributes being pruned. These questions are more effective identifiers for student classification than a larger number of questions in the original inquiries. Those who give more positive answers to these questions tend to make more rapid progress in learning and realize an admirable post-course score changes (PSC) of ±10 points. Teachers can not only make pedagogical plans according to this classification result and organize the class in a more effective way, but also put an emphasis on the enhancement of students' motivation and attitudes in their daily teaching in order to achieve desired learning objectives.

Chapter 6 Conclusions

6.1 Summary of this Study

The binocular distribution of students' academic achievements has become the main factor impeding the implementation of effective English remedial education. Mitigation of the conspicuous gap in learning performances between the high-proficient students and low-proficient students is of crucial significance to the enhancement of teaching activities within English remedial education environments. Effective pedagogical schemes are carried out on the condition that qualified candidate students are selected for the prepared course so that they can achieve admirable learning progress during the course study. Therefore, student classification is considered to be the best solution to the problem of notable discrepancy in the scholastic ability of students between the top and bottom levels. Definitely, student classification is the prerequisite for the successful adoption of any curriculum plan for English remedial education.

This study aims at the improvement of effectiveness of English remedial education through the analyses of data collected from a web-based remedial program in the Design Faculty of Nishinippon Institute of Technology, Japan. Correlation analysis is applied to help filter the data and multiple regression analysis is carried out to build the prediction model for the post-course score change (*PSC)*. Parameters used as independent variables (predictor attributes) are restricted to those which hold strong correlations with the dependent variable (class attribute), i.e., post-course score change (*PSC)*. The results suggest that the pre-course test score *(PRTS)* is the most influential classifier for the prognosis of post-course score changes (*PSC)*. Based on the advice of experienced teachers in English education, 10 points in post-course score changes in considered a desirable criterion for student classification. Therefore, 48 point of PRTS is considered as the classification criterion for student classification. That is, students with $20 \leqq \text{PRTS} \leqq 48$ are considered as qualified students because their PSC are predicted to be higher than 10 points according to the optimized prediction model; students with $48 < \text{PRTS} \leqq 100$ are unqualified students because their PSC are predicted to be less than 10 points. The prediction accuracy is 49.2% for the training data and 46.7% for the testing data.

In order to further investigate the relationship between students' pre-course test scores (*PRTS*) and their post-course score changes (*PSC*), the differential curve of the logistic learning curve model is proposed to demonstrate the interdependent relationship between students' pre-course test scores *(PRTS)* and post-course score change rate *(PSCR)*, while the logistic learning curve is applied to express the relationship between students' pre-course test scores *(PRTS)* and the total amount of knowledge they mastered after learning the course. When the root mean square error reaches its minimum value, the differential curve is viewed as most applicable for describing students' score changes and the parameters are thus determined. The differential curve of the logistic learning curve can work as a convenient tool for student classification.

Student classification using the logistic learning curve model exhibits superiority over other methods because it is much simpler and more understandable. The impact of the determinant *PRTS* (pre-course test scores) on *PSC* (post-course score changes) is clearly indicated through the differential curve of the logistic learning curve. It demonstrates empirical validity by overcoming the time-consuming and labor-costing weaknesses of the conventional method for student classification where more intuition and observation are involved. The differential curve of the logistic learning curve provides a perspective of anticipating the situation from the need of every individual student rather than making judgments about student categorization from a whole. After all, the application of the differential curve of the logistic learning curve model tells about the tendency in students' learning progress and thus, enables student classification to be carried out on the basis of the very important policy that students benefit the most out of the established remedial course. Based on the same criterion proposed in Chapter 2, 61 point of *PRTS* is the classification point for student classification. That is, students with $20 \leqq \text{PRTS} \leqq 61$ are qualified for the remedial course and students with $61 < \text{PRTS} \leqq 100$ are unqualified. The prediction accuracy is 69.2% for the training data and 64.2% for the testing data.

The pre-requisition for the application of the differential curve of the logistic learning curve is that students' pre-course test scores (*PRTS*) are available before the course training. When there is no pre-course test scores (*PRTS*) functioning, it does not function for student classification. In order to provide an option for the operation of student classification, GNP-based class association rule mining is applied for the discovery of the important attributes and their combinations using inquiries. Student classification is conducted based on the discovered features of students in the respective class, that is, students with desirable learning progress and students with less desirable learning progress.

Genetic Network Programming (GNP) is applied to the class association rule mining using the training data. GNP has a superior power in controlling both space and time preventing the bloating.

GNP reuses a fixed number of nodes which constitute a directed graph, therefore it is effective in dealing with dense databases because less space and time are required for rule mining. On the basis of the conventional GNP-based class association rule mining approach, attribute selection through GA is integrated into the system so that a much higher prediction accuracy is available. Attribute selection is an effective data mining technique to prune less relevant information and discover high-quality knowledge. It is especially useful for the classification of a dense database, because the preprocessing of data increases the possibility that predictor attributes given to the mining algorithm become relevant to the class attribute. Excellent individuals with a small data subset can better represent the class attribute and realize more precise classification.

GNP makes it possible to perform the attribute selection for class association rule mining using GA. In this study, Genetic Algorithm (GA) called Optimal Data subset through Genetic Algorithm (OASGA) is proposed to find out the optimal data subset (GA individual), which can compensate for the weakness of logistic learning curve and leads to the highest accuracy for classification. Class association rule mining is conducted with small data subsets rather than the original big database, thus, simple but important rules are obtained for classification. Simulation results with data collected in this study shows that the classification made by the optimal data subset is more accurate than the conventional method with the original database. The classification accuracy is largely improved from 56.5% to 71.9% for the training data, 51.2% to 64.7% for the testing data when classification is made with the optimal data subset with 20 attributes through attribute selection.

In order to improve the classification accuracy of OASGA, efforts are made by adding the negative aspects of the predictor attributes to the original database. After the amplification of the database, the extraction of negative class association rules becomes possible. The extracted class association rules can enhance the function of OASGA by giving a more precise description about the features of the students with desirable learning progress. Simulation results show that the classification accuracy is improved to 74.5% for the training data and 70.2% for the testing data when the negative class association rules are taken into account.

Attribute selection using OASGA (the system proposed in this study) throws light on the field of educational data mining. Educational database embodies complicated relations among a large number of predictor attributes and the class attribute. Therefore, class association rules extracted for classification are always too complicated and perplexing for practical purpose. But, after the preprocessing through the attribute selection, the original database becomes more revealing and indicative. As GA evolves, the optimal data subset with a small data subset is generated and concise class association rules are obtained for classification. Then, teachers can make pedagogical plans according to the classification results.

6.2 Limitation of this Study and Future Work

Limitations of this study set obstacles to the improvement of the preciseness in predicting test scores, the reinforcement of the logistic learning curve hypothesis and the enhancement of the algorithm of OASGA. Further studies on the following topics are needed to develop the rigorousness of the hypothesis both theoretically and practically:

(1) The algorithm of the logistic learning curve can be improved through the refinement of detailed analyses.

The current calculation for the logistic learning curve is based on the assumption that students absorb new knowledge at a fixed rate and the arbitrary constant a does not change with the changes of students' academic abilities. In fact, the absorption rate tends to decline as students' knowledge level increases, thus the absorption rate keeps on fluctuating as the ability level changes. Certainly, the fluctuation of constant a is caused by many factors involved in the learning process. The logistic learning curve may be quite unsteady with irregular ups and downs of post-course score changes (*PSC*) when the sample data is inadequate. Large quantities of data shelter those abnormal movements in prediction and leads to conclusions with universal validity. The effectiveness of the logistic learning curve hypothesis will be largely reinforced if the fluctuation of the absorption rate a is taken into consideration in the analytical process.

This is also the similar case with OASGA based on GNP-based class association rule mining. The parameters, such as the crossover rate and mutation rate in GA (*Cr* and *Mr*), crossover rate and mutation rates in GNP (*Pc* and *Pm*) need to be tested through repeated experiments so that more reasonable values can be figured out to help extract more important association rules.

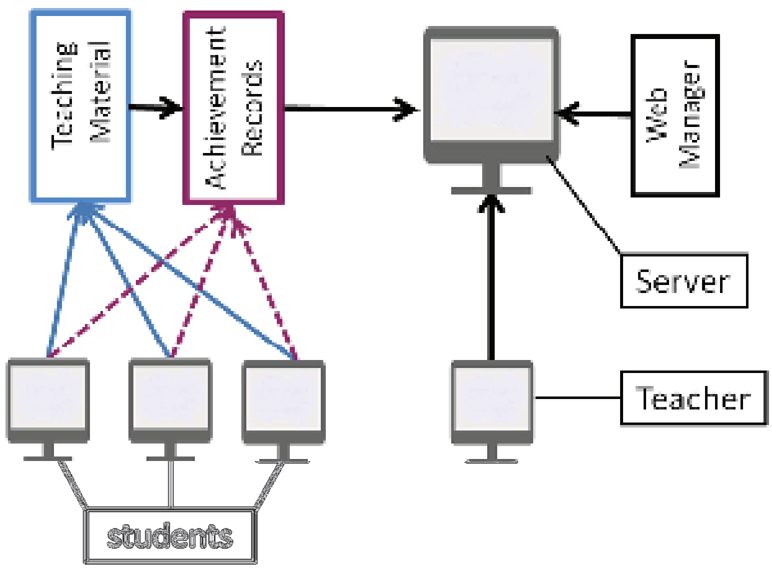
(2) Future works are expected to focus on the improvement of OASGA, such as the method to decrease the time for calculation.

The time-consuming of the system might prevent the method from being applied to more educational data. If the time for calculation can be shortened to a more desirable level, this method will be employed more frequently for practical application.

In summary, educational data mining requires more analytical tools to meet its need for the management of students' performances and progresses. Careful consideration should be given to the logistic learning curve model and the simulation results of classification through GNP in this study if scrutiny in education management is to be realized.

Appendix I Framework of ASP E-learning System

In order to help the new-coming students compensate for their deficiency in English learning, a remedial education program was established in Nishinippon Institute of Technology, Japan in the academic year of 2006. A web-based learning tool called Active Server Pages (ASP) (App. Fig. 1) was adopted and freshman students were supposed to accomplish the assignments for their English study online within a certain period of time. The e-learning system was created by Wao-net Corporation, who also administrates the server as a web-manager. Contents are created by the company, and both the teachers and students are users of the system. Data collected with the students in Design Faculty is used for the quantitative analysis of this study.



App. Fig. 1 Framework of ASP System

Before the start of the online learning, a user list including personal information of all student users, such as name, student number, major, department, is sent to the web manager and registration is carried out on the basis of these information. After the registration is processed, IDs and passwords are allotted to students so that they can access the contents by logging in from a fixed site and carry out their learning activities at any time. Teachers are supposed to log in from a specifically offered website to make confirmation about the contents and students' achievements. Student management function is available with the ASP system and teachers can search for an individual student according to any item of their personal information and check their learning progresses.

App. Fig. 2　Index Page of Grammar Items

The content of the system is very well designed for English remedial education to the first-year students in universities/colleges. It aims at the improvement of students' basic knowledge about English language which they should have before receiving higher education. The contents consist of all fundamental grammatical items, and both a tutoring section and an exercise section are included in each item. The tutoring sections focus on the explanation of grammatical rules, and the exercise sections strengthen the understanding of these rules through exercises. Students can search for the necessary information about a grammatical item from the index page and retrieve the details of both its interpretation section and the exercise section (App. Fig. 2).

Web-based learning systems demonstrate outstanding advantages in teaching and learning activities, and thus are viewed as most effective for English remedial education at the present stage. ASP system is one of the most popular systems widely used in the field of English teaching, because contents are offered by software developers and the burdens of teachers are largely alleviated. Students' progresses are clearly indicated on the student management page through a bar graph and a percentage figure. Teachers can also make confirmation of the average progressing situation of the whole class so that they can make adjustments in instructions (App. Fig. 3). Moreover, web-based learning is welcomed by educational institutions, because server management is conducted by the company and there is no extra work for universities/colleges.

App. Fig. 3　　Layout of Students' Progressing Situation

The main feature of ASP system for English remedial education is that it provides students with a student-friendly learning environment. Students with low-proficiency usually lack interests and good learning habits in study. It is also a part of the objective of the English remedial education to cultivate their abilities of self-control and self-management in learning. Web-based learning makes it possible to fulfill this goal, because students are responsible for their own learning activities and the system is open to access anywhere, anytime only if the internet is available. That is, students can schedule their learning plan and get involved in the study whenever they do not feel compelled. Furthermore, students can work on different sections at the same time even though they might be sitting in the same classroom. Consequently, students' various needs in learning are perfectly satisfied at the same time, which is unimaginable with the case of face-to-face instructions.

Needless to say, the effectiveness of English remedial education through ASP system has been largely improved. The web-based learning system facilitates students with no restrictions in access to the contents and partially assure their persistence in the self-study [1][2]. But, there is one problem with the current English remedial education program. Since the content is created for the purpose of English remedial education, it includes very basic grammatical rules which are taught to beginners. It is difficult to judge whether the content in the system matches the level of students' English proficiencies and whether all of them benefited from learning the course. No reference is available for teachers to give proper instructions to students on their learning assignment.

References

[1] K. Cho, Y. Okuda, S. Takeno and A. Otsu, "A Preliminary Study on Dropout Factors in e-Learning of English – Learners' motivation, preference of learning activities, and language-learning strategies", *LET Kyushu-Okinawa Bulletin*, No.6, P.33-P.47, 2006.

[2] H. Yamagata and R. Zhang, "The Influence of Computer Anxiety and Motivation on Learning Progress in a Web-based English Remedial Education Program", *Memoirs of Nishinippon Institute of Technolgy*, Vol. 39, pp.111-117, 2009.

Appendix II (Questionnaire 1)

Student ID Number: Name:

This is a questionnaire about students' motivation, attitude and strategies in English learning. Respond to each question and circle the corresponding number. Number 1~5 are deciphered as the following in this survey:

1 ……………………… Completely disagree

2 ……………………… Partially disagree

3 ……………………… Not sure

4 ……………………… Partially agree

5 ……………………… Completely agree

1. English learning is very interesting.	1•••••2•••••3•••••4•••••5
2. I am concerned about what the teacher says in class.	1•••••2•••••3•••••4•••••5
3. I always prepare before class.	1•••••2•••••3•••••4•••••5
4. I can do better in English study than my classmates.	1•••••2•••••3•••••4•••••5
5. I should learn Japanese well first rather than English.	1•••••2•••••3•••••4•••••5
6. I can catch up with the teacher in class.	1•••••2•••••3•••••4•••••5
7. I always review what I learned after the English class.	1•••••2•••••3•••••4•••••5
8. I don't want to lose my current English ability.	1•••••2•••••3•••••4•••••5
9. English class always makes me upset.	1•••••2•••••3•••••4•••••5
10. A certificate in English will bring me economic benefit.	1•••••2•••••3•••••4•••••5
11. People need go abroad if they want to learn English well.	1•••••2•••••3•••••4•••••5
12. I will have trouble in my work if I am not good at English.	1•••••2•••••3•••••4•••••5
13. Some people have an inborn ability to learn English well.	1•••••2•••••3•••••4•••••5
14. English is necessary for advanced professional knowledge learning.	1•••••2•••••3•••••4•••••5
15. It will be too late if I don't learn English from now on.	1•••••2•••••3•••••4•••••5
16. I keep a high attendance rate in all classes.	1•••••2•••••3•••••4•••••5

17. People around me are working hard at English.	1·····2·····3·····4·····5
18. I learn English for credits.	1·····2·····3·····4·····5
19. English is the one of the basic skills one should have.	1·····2·····3·····4·····5
20. I love watching foreign movies.	1·····2·····3·····4·····5
21. One's English proficiency depends on his/her efforts.	1·····2·····3·····4·····5
22. I often look up new words into the dictionary.	1·····2·····3·····4·····5
23. I know how to learn English effectively.	1·····2·····3·····4·····5
24. I find English is more enjoyable than before.	1·····2·····3·····4·····5
25. I dislike English learning though I know it is necessary.	1·····2·····3·····4·····5
26. I am doing extra learning in English.	1·····2·····3·····4·····5
27. I will gain respect from people if I speak English well.	1·····2·····3·····4·····5
28. I want to find a job in which English is involved in the future.	1·····2·····3·····4·····5
29. I do the e-learning exercise every day.	1·····2·····3·····4·····5
30. I started learning English in my childhood.	1·····2·····3·····4·····5
31. I feel frustrated when I can't do well in English learning.	1·····2·····3·····4·····5
32. I enjoy learning through e-learning.	1·····2·····3·····4·····5
33. I feel excited when I am praised.	1·····2·····3·····4·····5
34. I don't know what to do when I have a question to ask.	1·····2·····3·····4·····5
35. I have lost confidence in English learning.	1·····2·····3·····4·····5

Appendix III (Questionnaire 2)

Student ID Number: Name:

This is a questionnaire about students' awareness and understanding about intercultural communication. Respond to each question and circle the corresponding number. Number 1~5 are deciphered as the following in this survey:

1 ……………………… Completely disagree
2 ……………………… Partially disagree
3 ……………………… Not sure
4 ……………………… Partially agree
5 ……………………… Completely agree

1. I want to know people from all about the world.	1•••••2•••••3•••••4•••••5
2. I want to travel abroad.	1•••••2•••••3•••••4•••••5
3. I prefer traditional sense of value to new ideas.	1•••••2•••••3•••••4•••••5
4. It is interesting to communicate with foreigners.	1•••••2•••••3•••••4•••••5
5. I will have chances to work with foreigners in the future.	1•••••2•••••3•••••4•••••5
6. English is useful in communication.	1•••••2•••••3•••••4•••••5
7. I love to know about foreign cultures.	1•••••2•••••3•••••4•••••5
8. The common knowledge in Japan should be accepted by foreigners too.	1•••••2•••••3•••••4•••••5
9. I feel ashamed if I can't speak perfect English.	1•••••2•••••3•••••4•••••5
10. I don't care if I share the same table with a foreigner at a restaurant.	1•••••2•••••3•••••4•••••5
11. I have friends from other countries.	1•••••2•••••3•••••4•••••5
12. I think Japan needs to cooperate with other countries for further development.	1•••••2•••••3•••••4•••••5
13. I like to taste the different dishes from other countries.	1•••••2•••••3•••••4•••••5
14. I admire people who are good at English.	1•••••2•••••3•••••4•••••5
15. It is natural that people are different in thinking and behavior.	1•••••2•••••3•••••4•••••5
16. I often make comparisons between Japan and other countries.	1•••••2•••••3•••••4•••••5
17. I am interested in learning about happenings in foreign countries.	1•••••2•••••3•••••4•••••5
18. Japanese life style is the best.	1•••••2•••••3•••••4•••••5
19. I have the experience of travelling abroad.	1•••••2•••••3•••••4•••••5
20. Intercultural communication means to help foreigners to learn about Japan.	1•••••2•••••3•••••4•••••5

Appendix IV (Questionnaire 3)

Student ID Number: Name:

This is a questionnaire about students' awareness and understanding about intercultural communication. Respond to each question and circle the corresponding number. Number 1~5 are deciphered as the following in this survey:

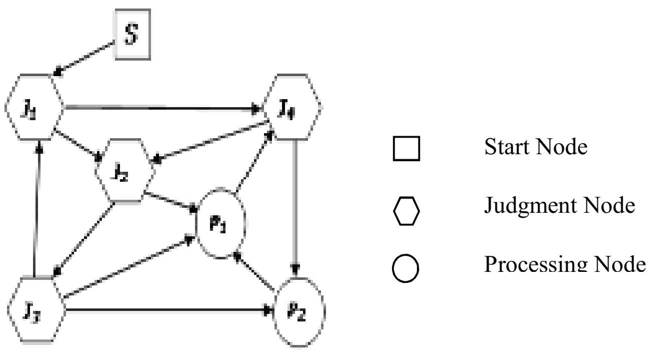
1 ……………………… Completely disagree
2 ……………………… Partially disagree
3 ……………………… Not sure
4 ……………………… Partially agree
5 ……………………… Completely agree

1. The remedial course is not difficult.	1•••••2•••••3•••••4•••••5
2. Computers are helpful with my learning.	1•••••2•••••3•••••4•••••5
3. I always ask for help when I have difficulty with the remedial course.	1•••••2•••••3•••••4•••••5
4. The e-learning program is interesting.	1•••••2•••••3•••••4•••••5
5. The English remedial course is necessary.	1•••••2•••••3•••••4•••••5
6. E-learning makes me feel more interested in English study.	1•••••2•••••3•••••4•••••5
7. I feel lonely when I sit in front of a computer.	1•••••2•••••3•••••4•••••5
8. I like computer games.	1•••••2•••••3•••••4•••••5
9. Computers are more reliable than human beings.	1•••••2•••••3•••••4•••••5
10. E-learning is more effective in English study.	1•••••2•••••3•••••4•••••5

Appendix V Class Association Rule Mining Using Genetic Network Programming (GNP)

1. GNP for Class Association Rule Mining

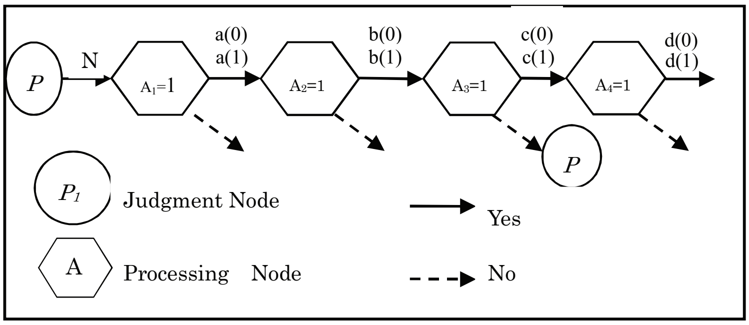
Genetic Network Programming (GNP) is the evolutionary optimization technique which can better represent the program by using compact directed-graph structures as genes. It consists of three kinds of nodes indicated as a start node, judgment nodes and processing nodes. Its basic structure is shown in App. Fig. 4. GNP consists of three kinds of nodes indicated as start node, judgment node and processing node. Start node *S* does not have any functions or conditional branches. The only role of the start node is to boot up the algorithm. Judgment nodes refer to the set of $J_1, J_2, \ldots J_m$, which does the decision-making function based on the *if-then* conditions of each node. Processing nodes are the set of $P_1, P_2, \ldots P_n$, suggesting the various actions or processing required by the various situations in the task. The attributes of each record in the database correspond to judgment nodes and association rules are extracted by the transitions of nodes in GNP. Details of GNP for class association rule mining are described in paper [1].



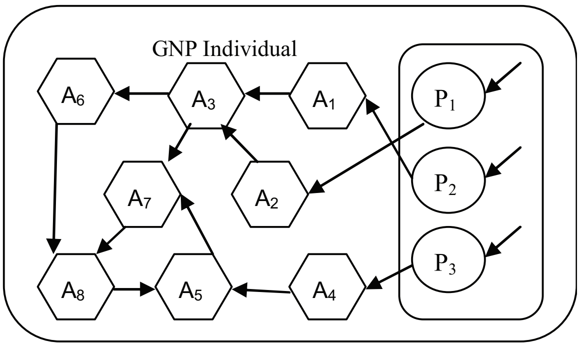
App. Fig. 4 Basic Structure of GNP

Unlike *a priori*-like methods, the objective of GNP is to generate a sufficient number of interesting rules and store them in the rule pool. Therefore, there is no need to identify frequent item sets [1]. GNP has two remarkable features during evolution: re-usage of the fixed number of nodes and realizing partially observable Markov decision processes. It is effective in dealing with a dense database because it takes the advantages of its compact structure and the bloating problem is prevented [2][3][4][5].

The database used in this study is a set of records with attribute values of 1 or 0 belonging to two classes. Class association rule mining aims at discovering some dependent relationships between predictor attributes and class attribute in the form of "If X then Y". Association means correlations or dependent relationships among a set of attributes in a database. Association rules are expressed in the form of " $X \rightarrow Y$", where X is called antecedent and Y is called consequent. It can be interpreted as "If X then Y". It means that if a database record satisfies the antecedent X, then it is likely to satisfy the consequent Y. When consequent Y is confined to the class attribute k ($k \in \{0,1\}$), association rules are labeled as class association rules and can function as identifiers for classification [6].



App. Fig. 5 Connections of Nodes in GNP for Class Association Rule Mining



App. Fig. 6 Basic Structure of GNP for Class Association Rule Mining

Content Gene			Connection Gene				
NTi	*IDi*		*Cil*		...	*Cij*	

App. Fig. 7 Gene Structure of GNP (node *i*)

The purpose of class association rule mining is finally to discover efficient rules so that relatively accurate classification can be made. Class association rule mining is conducted through judgment nodes and processing nodes [1]. N stands for the number of records for class association rule extraction. (1) and (0) stand for Class 1 and Class 0, respectively, while *a, b*, *c* and *d* in App. Fig. 5 refer to the number of records moving to the yes-side of each judgment node from the processing node. For example, the number of records satisfying $(A_1=1)\wedge(A_2=1)\wedge(A_3=1)\wedge(A_4=1) \rightarrow (Class\ 1)$ is $d(1)$ (7). App. Fig. 6 and App. Fig. 7 show the structure of GNP as its gene.

2. Measurements

Support and confidence are used as measurements to assess the quality of the extracted rules. Constraints such as minimum support and minimum confidence are set as conditions for extracting class association rules [6]. Support measures the frequency and confidence measures the strength of an association rule. Let the number of records in the database equal N. If the number of transactions containing X is x, the number of transactions containing *Class k* is y and the number of transactions containing X and *Class k* is z, then $support(X)=x/N=x'$, $support(Class\ k)=y/N=y'$ and $support\ (X \rightarrow Class\ k)=z/N=z'$. The confidence of class association rule "$X \rightarrow Class\ k$" is defined as the follows:

$$Confidence\ (X \rightarrow Class\ k)=support\ (X \rightarrow Class\ k)\ /\ support\ (X)= z'/x' \quad \text{(App. Eq. (1))}$$

In the case of class association rule mining, chi-squared value χ^2 is also used for measurements [3]. χ^2 value of the rule "$X \rightarrow Class\ k$" is given as follows:

$$\chi^2(X \rightarrow Class\ \ k) = \frac{N(z'-x'y')^2}{x'y'(1-x')(1-y')} \quad \text{(App. Eq. (2))}$$

GNP shows advantages in class association rule mining over other conventional methods, such as *a priori* method. The *a priori* method aims at finding all the rules satisfying the user-specific constraints and large computational overheads are caused by the processing of frequent attributes [6]. GNP-based class association rule mining is to exact a sufficient number of candidate rules and to build a classifier for classification on the basis of the extracted rules [3][4]. GNP can prevent the bloating effectively, therefore, it is more powerful in handling dense databases.

Measurements are used for class association rule mining, where the minimum values for the support, confidence and χ^2 are used to check the importance of each class association rule. The condition for the importance of class association rules can be expressed as follows:

$$Support\ (X \rightarrow Class\ k)) \geqq Sup_{-min,} \quad \text{(App. Eq. (3))}$$

$$Confidence(X \rightarrow Class\ k) \geqq Con_{-min,} \quad \text{(App. Eq. (4))}$$

$$\chi^2(X \rightarrow Class\ k) \geqq \chi^2_{-min,} \quad \text{(App. Eq. (5))}$$

where, X is the antecedent of an association rule, while k ($k \in \{0,1\}$) is the class attribute. Sup_{-min}, Con_{-min}, andχ^2_{-min} stand for the minimum values of the support, confidence and χ^2. If the class association rule satisfies App. Eq. (3), App. Eq. (4) and App. Eq. (5), it is regarded as an important candidate of class association rules.

Accordingly, individuals with relatively high fitness values can discover important class association rules for the next generation. The calculation of the fitness value of each individual is based on χ^2 values and the number of attributes in the antecedent part of rules extracted from the individual. Furthermore, a positive variable $new(r)$ is added like App. Eq. (6) [2]:

$$F = \sum_{r \in R} \left\{ x^2(r) + 10\big(n(r) - 1\big) + new(r) \right\} \quad \text{(App. Eq. (6))}$$

where, R denotes the set of important class association rules extracted from the individual, $n(r)$ is the number of attributes in the antecedent part of rule r, and when rule r is a newly extracted rule, $new(r)$ has a constant value, otherwise, it equals 0.

3. Crossover and Mutations

Crossover and mutation are genetic operators used in GNP to generate new offspring from selected individuals with excellent performance in the previous generation. The top 40 individuals with relatively high fitness values in each generation are chosen as the basis for GNP evolution. They are reproduced 3 times to create 120 individuals for genetic operations. Crossover and mutation have their own individuals to produce. After the genetic operations, all newly produced offspring are evaluated by their fitness values and the top 40 individuals are selected to repeat the genetic operations in the next generation.

Crossover and mutation are conducted within the gene of judgment nodes and processing nodes, respectively. Crossover means swapping the contents and connections of judgment nodes between two parent individuals at crossover rate of Pc, while mutation means the change of node connections or contents at mutation rate of *Pm*. The parameters and operations of crossover and mutation in GNP are the same as described in paper [1].

4. Classification by Association Rules

Classification is made by using the class association rules extracted from the training data and the class label of the testing data is determined. The testing data and class association rules are the inputs and the class label is the output in classifications [7]. The method for determining the class of the testing data is as follows:

(1) Suppose that the total number of class association rules in *Class k* extracted from the training data is R_k.

(2) If the antecedent attributes of extracted rules in *Class k* are satisfied by the testing data, the rules are viewed as satisfied. The total number of satisfied rules in *Class k* is T_k.

(3) Score (k) is calculated by App. Eq. (7).

$$Score\ (k) = \frac{T_k}{R_k}. \qquad \text{(App. Eq. (7))}$$

(4) Make the comparison between the two available scores – Score (0) and Score (1). The testing data is supposed to belong to the class with the higher score.

Suppose that the testing is conducted N times and the number of testing data showing correct classification is D. Then, the classification accuracy Pa is obtained as follows:

$$Pa=D/N. \qquad \text{(App. Eq. (8))}$$

Reference

[1] K. Shimada, K. Hirasawa and J. Hu, "Class Association rule mining with chi-squired test using Genetic Network Programming", *in Proceedings of the 2006 International Conference on Systems, Man, and Cybernetics (SMC'06)*, pp.5338-5344, 2006.

[2] Y. Yang, S. Mabu, K. Shimada and K. Hirasawa, "Intertransaction Class Association Rule Mining Based on Genetic Network Programming and its Application to Stock Market Prediction", *SICE Journal of Control, Measurement and System Integration*, Vol. 3, No. 1, pp.50-58, 2010.

[3] K. Shimada, K. Hirasawa and T. Furuzuki, "Data Mining using Genetic Network Programming with the Use of Acquired Information", *Journal of the Information Processing Society of Japan*, Vol.46, No.10, pp.2576-2586, 2005.

[4] C. Yue, S. Mabu, Y. Wang and K. Hirasawa, "Multiple-round English Auction Agent Based on

Genetic Network Programming", *IEEJ Transactions on Electrical and Electronic Engineering*, Vol. 5, Issue 4, pp.450-458, 2010.

[5] H. Zhou, S. Mabu, W. Wei, K. Shimada and K. Hirasawa, "Time Related Class Association Rule Mining and Its Application to Traffic Prediction", *IEEJ Trans.*, Vol. 125, No.1, pp.1-13, 2005.

[6] B. Liu, W. Hsu and Y. Ma, "Integrating Classification and Association Rule Mining", *in Proceeding of the ACM International Conference on Knowledge Discovery and Data Mining*, pp.80-86, 1998.

[7] E. Gonzales, K. Taboada, S. Mabu, K. Shimada and K. Hirasawa, "Combination of Two Evolutionary Methods for Mining Association Rules in Large and Dense Databases", *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.13, No.5, pp.1-12, 2009.

Acknowledgements

It is a pleasure to thank those who made this thesis possible. First and foremost, my deepest thank goes to my ex-supervisor, Prof. Kotaro Hirasawa, whose encouragement, supervision and support at the concluding level enabled me to better understand the subject and make a better improvement about this study. I highly appreciate his patience in giving me constructive comments and invaluable guidance towards the completion of this thesis, and his willingness to share his time with me to answer my questions thoroughly sometimes even to the very depths of the details.

I am heartily thankful to Prof. Fukuya Ishino, my ex-supervisor, who offered me the opportunity to study at the graduate school of Waseda University and encouraged me to keep on making challenges in my research. The experience at the doctoral course of Waseda University helped me to get access to advanced researches in the field of engineering education and to obtain very useful ideas and suggestions to improve my previous studies about English language education from the perspective of practical application.

I would like to extend my gratitude to Prof. Jinglu Hu, Prof. Osamu Yoshie and Prof. Shigeru Fujimura, who agreed to participate in the examining committee and offered me invaluable advice on the revision of this thesis at the completion stage. Their encouragement and structural support were crucial during the process of the degree application.

Special thanks also go to Dr. Kaoru Shimada, who offered me professional help to solve the practical teaching problems in my research. Without his support, the completion of this project would never have been possible.

I want to thank my family, without whom I would never have been able to achieve so much. I cordially thank my husband, whose love is the endless motivation for me to enjoy more challenges in life. I owe a debt to my son, who never complains about my absence in his daily life and has made efforts to comfort me by taking good care of himself while I am far away.

Lastly, I offer my regards and blessings to all of those, my friends, my colleagues and my teachers who supported me in any respect during the completion of the project.

Bibliography

(1) B. G. Moss and W. H. Yeaton, “Shaping Policies Related to Developmental Education: An Evaluation Using the Regression-discontinuity Design”, *Education Evaluation and policy Analysis*, Vol. 28, No. 3, pp.215-229, 2006.

(2) Eastern Kentucky University Homepage, Developmental Education Program Review. http://www.developmentaleducation.eku.edu/description/, available on June 6, 2008.

(3) National Center for Education Statistics (NCES), Remedial Education at Higher Education Institutions in Fall 1910, Washington, DC: U.S. Department of Education, 1996.

(4) National Center for Education Statistics (NCES), Remedial Education at Degree-granting Postsecondary Institutions in Fall 2000, Washington, DC: U.S. Department of Education, 2003.

(5) National Center for Education Statistic (NCES), Digest of Education Statistics. http://nces.ed.gov/programs/digest/d07/tables/dt07_317.asp, available on April 20, 2008.

(6) J. E. Hoyt, “Remedial Education and Student Attrition”, Community College Review, Vol. 27, Issue 2, pp. 51~63, 1999.

(7) V. K. McMillan, S. J. Parke and C. A. Lanning, “Remedial/Developmental Education Approaches for the Current Community College Environment”, Implementing Effective Policies for Remedial and Developmental Education, San Francisco, USA: Jossey-bass Publishers, 1997.

(8) M. Kataoka, C. E. Van Kraayenoord and J. Elkins, “Principals and Teachers’ Perceptions of Learning Disabilities: A Study from Nara Prefecture, Japan”, Learning Disability Quarterly, Vol. 27, Issue 3, pp.161~178, 2004.

(9) H. Ono, “Basic Academic Ability Construction of Japanese Universities Students ---- Observation from Placement Tests”, A Development of Placement Test and E-learning System for Japanese University Students, pp.1-6. National Institute of Multimedia Education, Japan, 2005.

(10) T. Bers, “Evaluating Remedial Education Programs”, AIR Professional File, Number 29, Spring 1987, http://eric.ed.gov/PDFS/ED282492.pdf, available on Jan. 30, 2012.

(11) Ministry of Education, Culture, Sports, Science and Technology (MEXT), Special Support Education for Learning Disabilities: The Situation of National Model Projects. Tokyo: Gyosei, 2002.

(12) P. Parmer and J. Cutler, "Easing the Transition: Building Better Bridges between Developmental and College-level Math, Office of Institutional Planning and Research", http://www.sinclair.edu/, available on Sept. 17, 2008.

(13) M. S. Torrubia, G. M. Lopez and E. C. Holgado, "Detailed and Global Analysis of a Remedial Course's Impact on Incoming Students' Marks", WSEAS Transactions on Advances in Engineering Education, Issue 1, Vol. 5, pp. 43-51, 2008.

(14) R. D. Al-Jarf, "The Effects of Online Grammar Instruction on Low Proficiency EFL College Students' Achievement", Asian EFL Journal. Vol. 7, Issue 4, 10th Article. 2005.

(15) R. Zhang, Y. Nakamura, Y. Zhou and F. Ishino, "A Practical Study on ASP E-Learning Model for Developmental Education", Human Factors in Telecommunication: User Experience of ICTs, pp. 255-262. New York: Prentice Hall, 2008.

(16) Benesse News on Education, http://benesse.jp/blog/20051201/p1.html/, available on Dec.22, 2007.

(17) Benesse Research & Development Center on Education. http://benesse.jp/berd/center/open/dai/between/2001/0708/bet17616.html, available on Aug. 22, 2009.

(18) R. S. Al-Jarf, "The Effects of Online Grammar Instruction on Low Proficiency EFL College Students' Achievement", Asian EFL Journal, Vol. 7, Issue 4, 10th Article, 2005.

(19) K. Beatty, Teaching and Researching Computer-assisted Language Learning. Beijing: Foreign Language Teaching and Research press, 2005.

(20) A. Saruwatari, S. Youshiharu and C. Suzuki, "Development of a Computer Assisted Flashcard System, LET Kyushu-Okinawa Bulletin, No. 6, pp.13-22, 2006.

(21) M. Warschauer, H. Shetzer and C. Meloni, *Internet for English Teaching*, USA: Teachers of English to Speakers of Other Languages, Inc, 2000.

(22) H. Ono, "Utilization of ICT for Developmental Education in Japan and the US: The Present and the Future", Journal of Multimedia Aided Education Research, Vol. 5, No.

1, pp.1-10. Tokyo: National Institute of Multimedia Education, 2008.

(23) R. Zhang, "Instructions of Students at Low-proficiency Level in E-learning", *Report of JSET Conferences*, JSET 07-5, 2007, pp.125-130, 2007.

(24) R. Zhang and K. Hirasawa, "Student Classification for English Remedial Education", Journal of the Japan Association for Developmental Education, Vol.5, No.1, pp. 60-66, 2010.

(25) R. Zhang and K. Hirasawa, "A Logistic Learning Curve Applied as a Compendium Method for Student Classification", International Journal of Continuing Engineering Education and Life-long learning, Vol. 20, No. 6, pp.517-531, UK: Inderscience, 2010.

(26) M. Will, Educating Students with Learning Problems – a Shared Responsibility. Washington, DC: Office of Special Education and Rehabilitation Services, U.S. Department of Education, 1986.

(27) J. Jenkins, C. Pious and D. Peterson, "Categorical Programs for Remedial and Handicapped Students: Issues of Validity", Exceptional Children, Vol.55, Issue 2, Council for Exceptional Children, 1988.

(28) L. M. McDonnell, M. J. McLaughlin and P. Morison, Education One & All: Students with Disabilities and Standards-based Reform, Washington, D.C.: National Academy Press, 1997.

(29) D. K. Richey, J. Mathern, C. S. O'Shea and S. J. Pierce, "Community College/High School Feedback and Collaboration: Preventive Measures", Implementing Effective Policies for Remedial and Developmental Education, San Francisco, USA: Jossey-bass Publishers, 1997.

(30) C. Danilowiez and E. Kukla, "The Application of Adaptive Students' Classification to the Determination of a Learning Strategy in an E-learning Environment", *World Transaction on Engineering and Technology Education* , Vol. 2, No.3, pp. 310-398, 2003.

(31) J. Abedi, "English Language Learners with Disabilities: Classification, Assessment and Accommodation Issue", accessed on Jan. 30, 2013. http://www.testpublishers.org/assets/documents/Special%20issue%20article%202.pdf.

(32) D. M. Wood, "Discrepancy Formulas and Classification and Identification Issues that Affect Diagnoses of Learning Disabilities", Psychology in the Schools, Volume 28,

Issue 3, pp.219-225, 2006.

(33) R. Baker, A. T. Corbett and A. Z. Wagner, “Human Classification of Low-fidelity Replays of Student Actions”, International Working Group on Educational Data Mining, http://www.educationaldatamining.org/, available on Sept. 15, 2009.

(34) D. M. Berger, “Mandatory Assessment and Placement: The View from an English Department”, Implementing Effective Policies for Remedial and Developmental Education, San Francisco, USA: Jossey-bass Publishers, 1997.

(35) T. Fujii, H. Hyase, M. Fellner, A. Bowman, D. Angove, T. Tsuji, K. Nagamine, J. Kuge and S. Hoyashita, “An Effective Method of English Education through E-learning”, *Journal of the Japan Association for Developmental Education*, Vol. 3, No. 1, pp.57-62, 2008.

(36) J. Weissman, J. Bulakowski and M. K. Jumisko, “Using Research to Evaluate Developmental Programs and Policies”, New Directions for Community Colleges, Vol.100, pp.73-80, 1997.

(37) Y. Hiroyuki and T. Hiroaki, “Combination of Qualitative and Quantitative Researches in English Language Education Study: From Hypothesis-testing to Hypothesis-succeeding, Language Education and Technology”, Language Education and Technology, No. 40, pp.161-173, 2003.

(38) R. Zhang, “International or self-conscious? A Slip of Japanese Government in Policymaking”, Studies in Contrastive Linguistics, edited by Figueroa and Gárate, pp.1091-1098, Spain: Universidade de Santiago de Compostela Publicasións, 2006.

(39) N. Kubo, “Orientation-Appraisal Model for Japanese University Students’ Learning of English：Relation among Learning Motive, Cognitive Appraisal, Learning Behavior, and Performance”, The Japanese Journal of Ecuational Psychology, Vol.47, No.4, pp.511-520, 1999.

(40) A. Nakayama, “Testing a Hypothesized Model of English Language Learning: Japanese University Students’ Goal Orientation, Beliefs, and Learning Strategies”, The Japanese Journal of Educational Psychology, Vol. 53, No. 3, pp.320-330, 2005.

(41) R. Zhang and J. Mok, “Globalization, Immigrants and Multilingual Education”, *Journal of Japan Anthropological Linguistic Association*, Vol.5, pp.120-140, 2009.

(42) R. Zhang, "A Literature Study of Repetition in Oral Discourse", Journal of Linguistic and Cultural Studies, Tokyo, Japan: The Japan Association of Language and Culture, Vol.26, pp.53-68, 2006.

(43) Cho, Y. Okuda, S. Takeno and A. Otsu, "A Preliminary Study on Dropout Factors in E-Learning of English – Learners' Motivation, Preference of Learning Activities, and Language-learning Strategies", LET Kyushu-Okinawa Bulletin, No.6, P.33-P.47, 2006.

(44) Yamagata and R. Zhang, "The Influence of Computer Anxiety and Motivation on Learning Progress in a Web-based English Remedial Education Program", Memoirs of Nishinippon Institute of Technolgy, Vol. 39, pp.111-117, 2009.

(45) D. S. Fike and R. Fike, "Predictors of First-Year Student Retention in the Community College", Community College Review, Vol. 36, pp.68-88, 2008.

(46) M. Ikeda, "Placement Tests for Non-English Major Students", Tezukayama Academic Review, Vol. 15, pp.1-6, 2008.

(47) J. Okuda, "Analysis of English Test Scores for Dividing Students into Different Classes", *Annual Reports of Shijonawate Gakuen Junior College*, Vol. 45, pp.80-83, 2012.

(48) S. M. Snyder, "An advocacy for New Placement Methods", Bulletin of Miyazaki Women's Junior College, Vol. 24, pp. 121-130, 1998.

(49) M. Nagai, D. Shimizu, T. Okuda and E. Yamaguchi, "Examination about the Ability Classification in the Information Literacy Class by Questionnaire Survey", Japan Journal of Educational Technologies, Vol. 29, pp. 225-228, 2006.

(50) R. Zhang and F. Ishino, "Key to Effective English Remedial Education: Intimation Derived from Multiple Regression", IAENG Transactions on Engineering Technologies (Volume II), pp. 98-120, New York: American Institute of Physics, 2009.

(51) J. M. Ignash, "Who Should Provide Postsecondary Remedial/ Developmental Education?" Implementing Effective Policies for Remedial and Developmental Education, San Francisco, USA: Jossey-bass Publishers, 1997.

(52) J. Weissman, C. Bulakowski and M. K. Jumisko, "Using Research to Evaluate Developmental Education Programs and Policies", Implementing Effective Policies for Remedial and Developmental Education. San Francisco, USA: Jossey-bass Publishers, 1997.

(53) K. G. Brempong and A. O. Gyapong, “Characteristics of Education Production Functions: An Application of Canonical Regression Analysis”, Economics of Education Review, Vol. 10, No. 1, 1991, pp. 7-17, 1991.

(54) M. Tsai and L. Tsai, “The Critical Success Factors and Impact of Prior Knowledge to Nursing Students when Transferring Nursing Knowledge during Nursing Clinical Practice”, Journal of Nursing Management, Vol. 13, Issue 6, pp.459-466, 2005.

(55) S. Liyanage and R. Barnard, “Valuing of Firms’ Prior Knowledge: a Measure of Knowledge Distance”, Knowledge and Process Management, Vol. 10, Issue 2, pp. 85-98, 2003.

(56) J. Gurlitt and A. Renkl, “Are High-coherent Concept Maps Better for Prior Knowledge Activation? Differential Effects of Concept Mapping Tasks on High School vs. University Students”, Journal of Computer Assisted Learning, Vol.24, Issue 5, pp.407-419, 2008.

(57) V. Woloshyn, E. Wood and T. Willoughby, “Considering Prior Knowledge When Using Elaborative Interrogation”, Applied Cognitive Psychology, Vol.8, Issue 1, pp.25-36, 1994.

(58) R. Zhang, Y. Zhou and F. Ishino, “Extensive Application of Multiple Regression: A Critical Insight into English Remedial Education through Web-based learning”, Lecture Notes in Engineering and Computer Science (WCECS, 2008), USA: IAENG, pp.477-482, 2008.

(59) T. Ueta, “Compendium Method of Judging Correlation”, Operations Research as a management science research, Vol. 42, No. 7, pp.493-496, The Operations Research Society of Japan, 1987.

(60) Y. Shirota, “Recommendation System by Using Consumer’s Kaisei for Soccer Uniform Design”, DBSJ Journal, Vol. 1, No. 1, pp.1-4. The Database Society of Japan, 2003.

(61) T. Shimbo, H. Kimura and T. Oyabu, “Prediction of Plant Bioelectric Potential based on Environmental Information in House Space”, Journal of the Society of Environmental Instrumentation Control and Automation, Vol. 13, No. 1, pp.27-33, 2008.

(62) L. Yelle, “The Learning Curve: Historical Review and Comprehensive Survey”, Decision Sciences, Vol. 10, Issue 2, pp. 302-328, 1979.

(63) H. Tsuchiya and O. Kobayashi, “Mass Production Cost of PEM Fuel Cell by Learning Curve”, International Journal of Hydrogen Energy, Vol. 29, pp. 985-990, 2004.

(64) P. Adler and K. Clark, “Behind the Learning Curve: A Sketch of the Learning Process”, Management Science, Vo. 37, No. 3, pp. 267-281, 1991.

(65) M. Spence, “The Learning Curve and Competition”, The Bell Journal of Economics, Vol. 12, No. 1, pp. 49-70, 1981.

(66) M. Lieberman, “The Learning Curve, Diffusion and Competitive Strategy”, Strategic Management Journal, Vol. 8, Issue 5, pp. 441-452, 1987.

(67) J. Cunningham, “Management: Using the Learning Curve as a Management Tool: The Learning Curve can Help in Preparing Cost Reduction Programs, Pricing Forecasts and Product Development Goals”, IEEE Spectrum, Vol. 17, Issue 6, pp. 45-48, 1980.

(68) J. Duane, “Learning Curve Approach to Reliability Monitoring”, IEEE Transactions on Aerospace, Vol. 2, Issue 2, pp. 563-566, 1964.

(69) G. Riddington, “Forecasting Ski Demand: Comparing Learning Curve and Varying Parameter Coefficient Approaches”, Journal of Forecasting, Vol. 18, pp. 205-214, 1999.

(70) C. Hommes, “Adaptive Learning and Roads to Chaos: The Case of the Cobweb”, Economics Letters, Vol. 36, Issue 2, pp. 127-132, 1991.

(71) L. L. Thurstone, “The Learning Curve Equation”, Psychological Monographs, Vol. XXVI, Whole No. 114. Princeton, N.J.: Psychological Review Company, 1919.

(72) A. Belkaoui, The Learning Curve; A Management Accounting Tool. CA, USA: Quorum Books, 1986.

(73) K. Hasegawa, A Collection of Excel Forms for Statistic Analysis, Tokyo: Kyouritsu, 1997.

(74) G. N. Master and B. D. Wright, “The Partial Credit Model”, in W.J.van der Linden and R. K. Hambleton (Eds.), Handbook of modern item response theory, pp. 101-121. New York: Springer, 1997.

(75) B. D. Wright and G. N. Master, Rating scale analysis. Chicago: MESA, 1982.

(76) S. G. Dastidar, “Gompertz: A Scilab Program for Estimating Gompertz Curve Using Gauss-Newton Method of Lest Squares, Journal of Statistical Software”, Vol.15, Issue 12, pp.1-12, 2006.

(77) Y. Iwasa, Suli Seibutsugaku Nyumon. Tokyo: HBJ Publishment, 1991.

(78) C. P. Winsor, “The Gompertz Curve as a Growth Curve”, Proceedings of the National Academy of Sciences”, Vol.18, No.1, pp.1-8, 1932.

(79) G. A. Haile and A. N. Nguyen, “Determinants of Academic Attainment in the United States: A Quantile Regression Analysis of Test Scores”, Education Economics, Vol, 16, No. 1, pp. 29-57, 2008.

(80) B. Clarke, S. Baker, K. Smolkowski and D. J. Chard, “An Analysis of Early Numeracy Curriculum-Based Measurement: Examining the Role of Growth in Student Outcomes”, Remedial and Special Education, Vol. 29, Issue 1, pp.46~58, 2008.

(81) I. Guyon and A. Elisseeff, “An Introduction to Variable and Feature Selection”, Journal of Machine Learning Research, Vol. 3, pp. 1157-1182, 2003.

(82) A. A. Freitas, Data Mining and Knowledge Discovery with Evolutionary Algorithms, Springer-Verlag, 2002.

(83) J. Bala, K. De Jong, J. Huang, H. Vafaie and H. Wechsler, “Using Learning to Facilitate the Evolution of Features for Recognizing Visual Concepts”, Evolutionary Computation, Vol.4, Issue 3, pp.297-311, 1996.

(84) H. Vafaie and K. DeJong, “Feature Space Transformation Using Genetic Algorithms”, IEEE Intelligent Systems, Vol. 13, No.2, pp. 57-65, 1998.

(85) C. Guerra-Salcedo, S. Chen, D. Whitley and S. Smith, “Fast and Accurate Feature Selection Using Hybrid Genetic Strategies”, in Proceedings of the Congress on Evolutionary Computation, pp.177-184, 1999.

(86) A. Moser and N. Murty, “On the Scalability of Genetic Algorithms to Very Large-Scale Feature Selection”, in Proceedings of the Real-World Applications of Evolutionary Computing: EvoWorkshops, pp.77-86, 2000.

(87) H. Ishibuchi and T. Nakashima, “Pattern and Feature Selection by Genetic Algorithms in Nearest Neighbor Classification”, Journal. of Advanced Computational Intelligence, Vol. 4, No. 2, pp.138-145, 2000.

(88) J. Yang and V. Honavar, “Feature Subset Selection Using a Genetic Algorithm”, IEEE Intelligent Systems, Vol.13, No.2, pp.44-49, 1998.

(89) C. Emmanouilidis, A. Hunter and J. MacIntyre, “A Multiobjective Evolutionary Setting for Feature Selection and a Commonality-Based Crossover Operator”, in Proceedings of the 2000 Congress on Evolutionary Computation. CEC00, pp.16-19,

2000.

(90) S. Mabu, K. Hirasawa and J. Hu, “A Graph-Based Evolutionary Algorithm: Genetic Network Programming (GNP) and Its Extension Using Reinforcement Learning”, Evolutionary Computation, Vol.15, Issue 3, pp.369-398, 2007.

(91) S. Mabu, Y. Chen and K. Hirasawa, “Generating Stock Trading Rules Using Genetic Network Programming with Flag Nodes and Adjustment of Important Indexes”, Electronics and Communications in Japan, Vol. 94, Issue 1, pp.25-33, 2011.

(92) K. Hirasawa, K. Okubo, H. Katagiri, J. Hu and J. Murata, “Comparison between Genetic Network Programming (GNP) and Genetic Programming (GP)”, in Proceedings of the Congress on Evolutionary Computation, pp. 1276-1282, 2001.

(93) S. Mabu, K. Hirasawa, Y. Matsuya and J. Hu, “Genetic Network Programing for Automatic Program Generation”, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol. 9, No. 4, pp. 430-435, 2005.

(94) K. Hirasawa, M. Okubo, H. Katagiri, J. Hu and J. Murata, “Comparison between Genetic Network Programming and Genetic Programming using Evolution of Ant’s Behaviors”, Transaction of IEE Japan, Vol. 122-C, No. 6, pp. 1001-1009, 2001.

(95) H. Katagiri, K. Hirasawa, J. Hu and J. Murata, “Network Structure Oriented Evolutionary Model – Genetic Network Programming and Its Comparison with Genetic Programming”, LATERBREAKING PAPERS of the Genetic and Evolutionary Computation Conference, pp. 219-226, 2002.

(96) G. Yang, S. Mabu, K. Shimada and K. Hirasawa, “An Evolutionary Approach to Rank Class Association Rules with Feedback Mechanism”, Expert Systems with applications, Vol. 38, No12, pp. 15040-15048, 2011.

(97) E. Gonzales, S. Mabu, K. Taboada, K. Shimada and K. Hirasawa, “Evolutionary Selection of Interesting Class Association Rules using Genetic Relation Algorithm”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 6, No. 5, pp. 431-440, 2011.

(98) C. Chen, S. Mabu, K. Shimada and K. Hirasawa, “Network Intrusion Detection using Class Association Rule Mining Based on Genetic Network Programming”, IEEJ Transactions on Electrical and Electronic Engineering, pp. 553-559, Vol. 5, No. 5, 2010.

(99) G. Yang, S. Mabu, K. Shimada and K. Hirasawa, “EvoCMAR: A New Evolutionary Method to Directly Mine Association Rules for Classification”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 5, No. 5, pp. 574—585, 2010.

(100) Y. Yang, S. Mabu, K. Shimada and K. Hirasawa, “Intertransaction Class Association Rule Mining Based on Genetic Network Programming and its Application to Stock Market Prediction”, SICE Journal of Control, Measurement and System Integration, Vol. 3, No. 1, pp.50-58, 2010.

(101) K. Shimada, K. Hirasawa and T. Furuzuki, “Data Mining using Genetic Network Programming with the Use of Acquired Information”, Journal of the Information Processing Society of Japan, Vol.46, No.10, pp.2576-2586, 2005.

(102) C. Yue, S. Mabu, Y. Wang and K. Hirasawa, “Multiple-round English Auction Agent Based on Genetic Network Programming”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 5, Issue 4, pp.450-458, 2010.

(103) H. Zhou, S. Mabu, W. Wei, K. Shimada and K. Hirasawa, “Time Related Class Association Rule Mining and Its Application to Traffic Prediction”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 125, No.1, pp.1-13, 2005.

(104) S. P. Deshpande, and V. M. Thakare, “Data Mining System and Applications: A Review”, International Journal of Distributed and Parallel Systems, Vol.1, No.1, pp.32-44, 2010.

(105) J. Luan, “Data Mining and Knowledge Management in Higher Education: Potential Applications”, in Annual Forum for the Association for Institutional Research, Toronto, Canada, 2002.

(106) E. Gonzales, K. Taboada, S. Mabu, K. Shimada and K. Hirasawa, “Combination of Two Evolutionary Methods for Mining Association Rules in Large and Dense Databases”, Journal of Advanced Computational Intelligence and Intelligent Informatics, Vol.13, No.5, pp.1-12, 2009.

(107) B. Miller and D. Goldberg, “Genetic Algorithms, Tournament Selection and the Effects of Noise”, Complex Systems, Vol. 9, pp. 193-212, 1995.

(108) S. Yip and G. Webb, “Discriminant Attribute Finding in Classification Learning”, in Proceedings of 5th Australian Joint Conference on Artificial Intelligence (AI-92), pp. 1-8, 1992.

(109) G. Kundu, M. M. Islam, S. Munir and M. F. Bari, "ACN: An Associative Classifier with Negative Rules", Proceedings of 11th IEEE International Conference on Computational Science and Engineering, 2008, pp. 369-375.

(110) J. Luo and H. Luo, "Algorithm for Classification based on Positive and Negative Class Association Rules", Proceedings of 3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT), 2010, Vol. 3, pp. 536-540.

(111) M. Antonie and O. R. Zaiane, "An Associative Classifier based on Positive and Negative Rules", In Proceedings of the 9th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, 2004, pp. 64-69.

(112) B. Kavsek and N. Lavrac, "Apriori-SD: Adapting Association Rule Learning to Subgroup Discovery", Applied Artificial Intelligence: An International Journal, Vol. 20, Issue 7, pp. 543-583, 2006. Taylor & Francis Online.

(113) R. Shikhariya and N. Shukla, "An Improved Association Rule Mining with FP Tree using Positive and Negative Integration", Journal of Global Research in Computer Science, Vol. 3, No. 10, pp. 46-51.

(114) B. Ramasubbareddy, A. Govardhan and A. Ramamohanreddy, "Classification based on Positive and Negative Association Rules", International Journal of Data Engineering, Vol. 2, Issue, pp. 84-92, 2011.

(115) N. Jain, V. Sharma and M. Malviya, "Reduction of Negative and Positive Association Rule Mining and Maintain Superiority of Rule using Modified Genetic Algorithm", International Journal of Advanced Computer Research, Vol. 2, No. 6, pp. 31-36, 2013.

(116) Anandhavalli M., S. K. Sudhanshu, A. Kumar and Ghose M. K., "Optimized Association Rule Mining using Genetic Algorithm", Advances in Information Mining, Vol. 1, Issue 2, pp. 1-4, 2009.

(117) R. Dewang and J. Agarwal, "A New Method for Generating All Positive and Negative Association Rules", International Journal of Computer Science and Engineering, Vol. 3, No. 4, pp. 1649-1657, 2011.

(118) J. Li and J. Jones, "Classification Using Multiple and Negative Target Rules", Knowledge-based Intelligent Information and Engineering Systems, Lecture notes in Computer Science, Vol. 4251, pp. 212-219, 2006.

(119) J. Li and J. Jones, "Using Multiple and Negative Target Rules to Make Classifiers More Understandable", Knowledge-Based Systems, Vol. 19, Issue 6, pp. 438-444,

2006. Elsevier.

(120) R. Zhang, K. Shimada, S. Mabu and K. Hirasawa, “Acquisition of Optimal Data subset through Genetic Algorithm Using GNP-based Class Association Rule Mining”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 9, No. 4, 2014. (accepted)

Publication List

Books

1 Rong Zhang and Jeffrey Mok, 2007, Celebrating around the World, Kenkyushya, Tokyo.

2 Yoshiro Takanashi, Rong Zhang, Xanthe Smith and Peter Serafin, 2009, Mastering the TOEFL® iBT Reading and Listening, Kenkyusha, Tokyo.

3 Yoshiro Takanashi, Rong Zhang, Xanthe Smith and Peter Serafin, 2009, Mastering the TOEFL® iBT Speaking and Writing, Kenkyusha, Tokyo.

Journal Papers

1 Rong Zhang, Kaoru Shimada, Shingo Mabu and Kotaro Hirasawa, “Acquisition of Optimal Attribute Set through Genetic Algorithm for GNP-based Class Association Rule Exaction”, IEEJ Transactions on Electrical and Electronic Engineering, Vol. 9, No. 4, 2014,　(Accepted)

2 Rong Zhang and Dennis C. McCornac, “The Need for Private Universities in Japan to be Agents of Change”, International Journal of Educational Management, Vol. 27, Issue 6, pp. 562-577, 2013, UK: Emerald..

3 張栄，平澤宏太郎，“英語リメディアル教育のクラス分け方法について”，リメディアル教育研究，第 5 巻，第 1 号，pp. 60-66, 2010, 日本リメディアル教育学会.

4 Rong Zhang and Kotaro Hirasawa, “A Logistic Learning Curve Applied as a Compendium Method for Student Classification”, International Journal of Continuing Engineering Education and Life-long Learning , Vol.20 , No.6 , pp.517-531, 2010, UK: Inderscience.

International Conference Papers

1 Rong Zhang and Dale Steele, "Improving Intercultural Awareness: A Challenging Task for Japan", In Proceedings of the Cyprus International Conference on Educational Research (CY-ICER 2012), Procedia – Social and Behavioral Sciences, Vol. 47, pp. 52-63, 2012, North Cyprus.

2 Rong Zhang, Yue Zhou and Fukuya Ishino, "Extensive Application of Multiple Regression: A Critical Insight into English Remedial Education through Web-based learning", In Proceedings of the WCECS (2008), Lecture Notes in Engineering and Computer Science, pp. 477-482, October 2008, University of California, Berkeley, USA.

3 Rong Zhang, Yue Zhou, Hiromi Yamagata and Fukuya Ishino, Determinants of English Remedial Education through Web-based Learning", In Proceedings of the 5th WSEAS/IASME International Conference on Engineering Education, New Aspects of Engineering Education, pp.343-348, 2008, Heraklion, Crete Island, Greece.

4 Rong Zhang, Yuji Nakamura, Yue Zhou and Fukuya Ishino, "A Practical Study on ASP E-learning Model for Developmental Education", In Proceedings of the 21st International Symposium on Human Factors in Telecommunication (HFT 2008), pp. 255-262, 2008, Kuala Lumpur, Malaysia.

5 Rong Zhang, "International or self-conscious? A Slip of Japanese Government in Policymaking", In Proceedings of the 4th International Contrastive Linguistic Conference, Studies in Contrastive Linguistic, pp.1091-1098, 2006, Santiago de Compostela, Spain.