**Journal of Information Processing Vol.23 No.2 192–201 (Mar. 2015)**

[DOI: 10.2197/ipsjjip.23.192]

**Regular Paper**

# A Predictive Model to Evaluate Student Performance

Shaymaa E. Sorour[1,2,a)] Tsunenori Mine[3,b)] Kazumasa Goda[4,c)] Sachio Hirokawa[5,d)]

Received: June 27, 2014, Accepted: December 3, 2014

**Abstract:** In this paper we propose a new approach based on text mining techniques for predicting student performance using LSA (latent semantic analysis) and K-means clustering methods. The present study uses free-style comments written by students after each lesson. Since the potentials of these comments can reflect student learning attitudes, understanding of subjects and difficulties of the lessons, they enable teachers to grasp the tendencies of student learning activities. To improve our basic approach using LSA and k-means, overlap and similarity measuring methods are proposed. We conducted experiments to validate our proposed methods. The experimental results reported a model of student academic performance predictors by analyzing their comments data as variables of predictors. Our proposed methods achieved an average 66.4% prediction accuracy after applying the k-means clustering method and those were 73.6% and 78.5% by adding the overlap method and the similarity measuring method, respectively.

**Keywords:** comments data mining, latent semantic analysis (LSA), similarity measuring method, overlap method

## 1. Introduction

Recently, many researchers have turned their attention to explaining and predicting learners' performance. They have contributed to the related literature. By and large, researchers in this field manage to advocate novel and smart solutions to improve performance [15]. Thus, learners' performance assessment may not be viewed as being somewhat separate from learning process. It is a continuous and an integral part of learning processes [10]. By revealing what students already know and what they need to learn, it enables teachers to build on existing knowledge and provide appropriate scaffolding [25]. If such information is timely and specific, it can serve as a valuable feedback to both teachers and students so that it will improve student performance.

Yet interpreting assessment in the learning environment remains a challenge for many reasons. Most teachers lack training in the assessment of understanding beyond the established testing culture. Externally designed tests offer limited information due to less varied and frequent assessment, as well as delayed and coarse-grained feedback [27]. The solution to these problems is to grasp all the class members' learning attitudes and tendencies of learning activities. Teachers can give advices by their careful observation, but it is a hard task to grasp all the class members' learning attitudes all over the periods in the semester.

Goda et al. [12], [13] proposed the PCN method to estimate learning situations from comments freely written by students. The PCN method categorizes the comments into three items: P (Previous activity), C (Current activity), and N (Next activity). Item P indicates the learning activity before the class time. Item C shows the understanding and achievements of class subjects during the class time, and item N expresses the learning activity plan until the next class. These comments have vital roles in educational environments. For example, the comments help students to communicate with their teacher indirectly, and provide a lot of clues or hints to the teacher for improving his/her lessons. Each student writes his/her comments after a lesson; the student looks back upon his/her learning behavior and situation; he/she can express about his/her attitudes, difficulties, and any other information that might help a teacher estimate his/her learning activities.

However [12], [13] did not discuss the prediction of a final student grades. In this paper we first propose a basic prediction method of student grades using comment data with item C (C-comment for short) from the PCN method. The basic method uses LSA technique to extract semantic information from student comments by using statistically derived conceptual indices instead of individual words, then classifies the obtained results into 5 groups according to their grades by using a K-means clustering method. The basic method achieves an average 66.4% prediction accuracy of student grades. To improve the prediction accuracy, we additionally propose overlap and similarity measuring methods.

Experiments were conducted to validate our newly proposed methods; the results illustrated that the proposed methods achieved 73.6% and 78.5% prediction accuracy of student grades by the overlap and the similarity measuring methods, respectively. The contributions of our work are the following:

- The LSA technique is adopted to analyze patterns and relationships between the extracted words and latent concepts

[1] Faculty of Specific Education, Kafr Elsheik University, Kafr Elsheikh, Egypt
[2] Graduate School of Information Science and Electrical Engineering, 744 Motooka Nishiku, Fukuoka, Japan
[3] Faculty of Information Science and Electrical Engineering, Kyushu University, 744 Motooka Nishiku, Fukuoka, Japan
[4] Kyushu Institute of Information Science, 6–3–1 Saifu, Dazaifu, Fukuoka, Japan
[5] Research Institute for Information Technology, Kyushu University, 6–10–1 Hakozaki Higashi-ku, Fukuoka, Japan
[a)] shaymaa@ma.ait.kyushu-u.ac.jp
[b)] mine@ait.kyushu-u.ac.jp
[c)] gouda@kiis.ac.jp
[d)] hirokawa@cc.kyushu-u.ac.jp

contained in an unstructured collection of texts (student comment); we classify the results obtained after applying LSA into 5 groups according to student grades by using the K-means clustering method.
- The similarity measuring method is proposed to calculate similarity between a new comment and comments in the nearest cluster, which is created in the training phase.
- The overlap method is introduced for a stable evaluation that allows to accept the adjacent grade of its original grade corresponding to 5-grade categories. To this end, we classify student marks into 9 grades.
- The experiments were conducted to validate the proposed methods: basic, overlap and similarity measuring methods by calculating the $F$-measure and the accuracy for each method in estimating final student grades. The experimental results illustrate the validity of the proposed methods.

The rest of the paper is organized as follows. Section 2 gives an overview of some related literature. Section 3 introduces the overview of our research and the procedures of the proposed methods. Section 4 describes the methodology of our proposed methods. Section 5 discusses some of the highlighted experimental results. Finally, Section 6 concludes the paper and describes our future work.

## 2. Related Work

The ability to predict student performance is very important in educational environments. Increasing students' success in their learning environment is a long-term goal in all academic institutions. In recent years, there is a growing interest in employing educational data mining techniques (EDM) to conduct the automatic analysis and prediction of learner performance [2], [5], [8], [14], [24]. An emerging trend in EDM is the use of text mining which is an extension of data mining to text data [19], [20], [26]. Many researchers have successfully used text mining techniques to analyze large amounts of textual data in business, health science and educational domains [11], [21], [28], [29]. In our research we focus on using text mining in education fields to predict student grades.

### 2.1 Educational Text Mining

Text mining focuses on finding and extracting useful or interesting patterns, models, directions, trends, or rules from an unstructured text such as in text documents, HTML files, chat messages and emails. In addition, the major applications of text mining include automatic classification (clustering), information extraction, text summarization, and link analysis [3]. As an automated technique, text mining can be used to efficiently and systematically identify, extract, manage, integrate, and exploit knowledge for research and education [1].

Currently, there are only several studies about how to use text mining techniques to analyze learning related data. For example, Tane et al. [28] used text mining (text clustering techniques) to group e-learning resources and documents according to their topics and similarities. Antai et al. [4] classified a set of documents according to document topic areas by using CLUTO program with and without LSA. The results showed that the internal cluster similarity with LSA was much higher than that without LSA. In addition, Hung [16] used clustering analysis as an exploratory technique to examine e-learning literature and visualized patterns by grouping sources that shared similar words and attribute values. In addition, Minami et al. [23] analyzed student attitudes towards learning, and investigated how the attitudes affect final student evaluation; they pursued a case study of lecture data analysis in which the correlations exist between student attitudes to learning such as attendance and homework, as effort, and the student examination scores, as achievement; they analyzed the student own evaluation and lectures based on a questionnaire. Through this study, Minami et al. showed that a lecturer could give feedback to students who tended to over-evaluate themselves, and let the students recognize their real positions in the class.

Previous studies show that we need to understand individual students more deeply and recognize student learning status and attitude to give feedback to them. We need to comprehend student characteristics by letting them describe themselves about their educational situations such as understanding of subjects, difficulties to learn, learning activities in the classroom, and their attitude toward the lesson. Researchers have used various classification methods and various data in their studies to predict student academic performance.

Different from the above studies, Goda et al. [13] proposed the PCN method to estimate student learning situations on the basis of their freestyle comments written just after the lesson. The PCN method categorizes their comments into three items: P (Previous), C (Current), and N (Next) so that it can analyze the comments from the points of views of their time-oriented learning situations. Goda et al. [12] also conducted another study on using PCN scores to determine the level of validity of assessment based on student comments and showed strong correlations between the PCN scores and the prediction accuracy of final student grades. They employed multiple regression analysis to calculate PCN scores. Their results indicated that students who wrote comments with high PCN scores were considered as those who described their learning attitude appropriately. In addition, applying machine learning method support vector machine (SVM), they illustrated that as student comments got higher PCN scores, prediction performance of their grades became higher. Goda et al., however, did not discuss prediction performance of final student grades.

The current study is an extension of Goda et al. [12]; we focus on accuracy of prediction of final student grades. Using C-comments from the PCN method, we try to predict their grade in each lesson and discuss the changes in accuracy and $F$-measure over a sequence of lessons.

## 3. Overview of the Prediction Method

### 3.1 PCN Method and Student Grade

To grasp student lesson attitudes and learning situations and to give feedback to each student are educational foundations. Goda et al. [13] proposed the PCN method to estimate learning situations from comments freely written by students. Each student described his/her learning tendency, attitudes, and understanding

Table 1 Viewpoint categories of student comment [13].

| Viewpoint | Meaning |
|---|---|
| **P(Previous)** | The learning activity before the class time such as review of previous class and preparation for the coming class. |
| **Example** | "I read chapter 3 of the textbook." |
| **C(Current)** | The understanding and achievements of class subjects during the class time. |
| **Example** | "I was completely able to understand the subject of this lesson and have the confidence to make other similar to the ones. I learned in this lesson." |
| **N(Next)** | The learning activity plan until the next class. |
| **Example** | "I will make preparation by next class." |
| **O(Other)** | Other descriptions. |

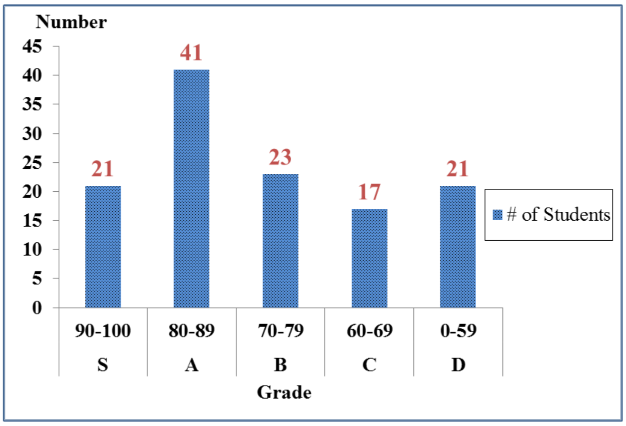



**Fig. 1** The relation between the grades and the range of the marks.

for each lesson according to four items: P, C, N and O. The explanations and the examples of the items are shown in **Table 1**.

Comments data were collected from 123 students in two classes: 60 in Class A and 63 in class B. They took Goda's courses that consisted of 15 lessons. In this research, we use the student comments collected for the last half, from lessons 7 to 15. The main subject and the contents are different from lessons 1 to 6. The main subject from lessons 1 to 6 is computer literacy. From lessons 7 to 15, students begin to learn the basics of programming.

To predict student grades from comment data, 5-grade categories are used to classify student marks. The assessment of each student was done by considering the average mark of three assigned reports, and his/her attendance rate. **Figure 1** displays the number of students in each grade according to the range of the marks. For example, the number of students in grade A is 41 and their marks are between 89 and 80.

### 3.2 Procedures of the Basic Prediction Method

This research aims to predict student performance by analyzing C-comments data. In order to generate a correlation between the comment data and the student grade, **Fig. 2** displays the overall procedures of the proposed method, we call it the Basic Prediction Method. The procedures of the method are based on five phases:
(1) Comment data collection, (2) Data preparation, (3) Training phase, (4) Noisy data detection, (5) Test phase. The details of these phases are as follows:

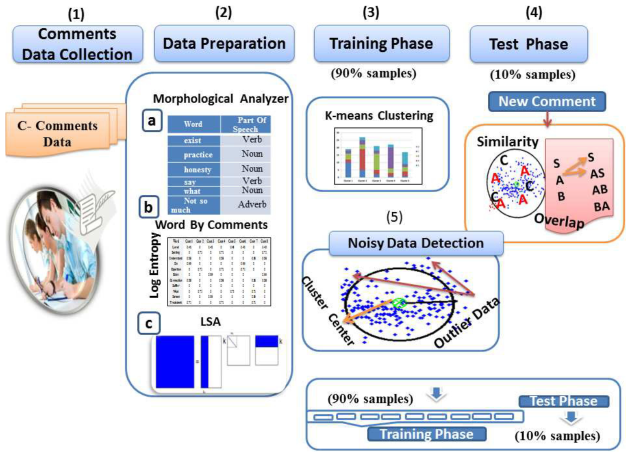



**Fig. 2** Procedures of the proposed method.

Table 2 Number of comments.

| Lesson | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|
| **Number** | 104 | 103 | 107 | 111 | 107 | 109 | 107 | 111 | 121 |

(1) **Comment Data Collection**: This phase focuses on collecting student comments after each lesson. In our research, we collected C-comments from the PCN method. The C-comment indicates students understanding and achievements of class subjects during the class time. In addition, it has a stronger correlation with the prediction accuracy than P- and N-comment [12]. Although we have two class data in each lesson, we combined them to increase the number of comments in each grade; some students didn't submit their comments because they did not write any comments or were absent. **Table 2** displays the real number of comments in each lesson. The number of words appearing in the comments is about 1400 in each lesson. In addition, the number of distinct words in each lesson is over 430 words.

(2) **Data Preparation**: The data preparation phase covers all the activities required to construct the final data set from the initial raw data. This phase includes the following steps:
   (a) Analyze C-comments, extract words and parts of speech with Mecab program *1, which is a Japanese morphological analyzer designed to extract words and identify their part of speech (verb, noun, adjective, and adverb).
   (b) Calculate the occurrence frequencies of words in comments and apply a log entropy term weighting method so as to balance the effect of occurrence frequency of words in all the comments. The detail is discribed in Section 4.1.
   (c) Employ LSA to analyze patterns and relationships between the extracted words and latent concepts contained in unstructured collection of texts (student comment). We call the obtained results LSA results. The details are described in Section 4.2.

(3) **Training Phase**: This phase builds the prediction models of student grades by classifying LSA results into 5 clusters. The model identifies the center of each cluster and the grade

*1 http://sourceforge.net/projects/mecab/

that most frequently appears in the cluster. We call the grade the **dominant grade** in the cluster.

(4) **Test Phase**: The test phase evaluates the performance of prediction models by calculating Accuracy and $F$-measure. This phase first extracts words from a new comment, and transforms an extracted-word vector of the comment to a set of K-dimensional vectors by using LSA. To evaluate the prediction performance, 10-fold cross validation was used. 90% of comments were classified as training data and constructed a model, then the model was applied to the rest 10% of comments as test data, and compared a predicted value with corresponding the original data. The procedure was repeated 10 times and the results were averaged. To improve the prediction accuracy of student grade, we employed the similarity measuring and overlap methods in addition to our basic method. The details of the two methods are described in Sections 4.4 and 4.5, respectively.

(5) **Noisy Data detection**: Outlier analysis can be used to detect data that adversely affect the results. In this paper, we detect outliers in two phases: training phase and test phase. As a threashold to check the outliers in a cluster, we use the standard deviation ($Sd$) of each distance between each member and the center in a cluster in the training phase, and the average of each distance between each member and the center in a cluster, in the test phase. The detail is described in Section 4.3.

## 4. Methodology

This section describes our methodology for predicting student performance from free-style comments.

### 4.1 Term Weighting of Comments

In preparing for LSA, the C-comments are transformed into a standard word-by-comment matrix [6] by extracting words from them. This word-by-comment matrix say $A$, is comprised of $m$ words $w_1, w_2 \ldots, w_m$ in $n$ comments $c_1, c_2, \ldots, c_n$, where the value of each cell $a_{ij}$ indicates the total occurrence frequency of word $w_i$ in comment $c_j$.

To balance the effect of word frequencies in all the comments, a log entropy term weighting method is applied to the original word-by-comment matrix, which is the basis for all subsequent analyses [17]. We apply a global weighting function to each nonzero element of $a_{ij}$ of $A$. The global weighting function transforms each cell $a_{ij}$ of $A$ to a global term weight $g_i$ of $w_i$ for the entire collection of comments.

Here $g_i$ is calculated as follows:

$$g_i = 1 + \sum_{j=1}^{n} (p_{ij} \log(p_{ij}))/\log(n) \quad (1)$$

where $p_{ij} = L_{ij}/gf_i$, $L_{ij} = \log(tf_{ij} + 1)$, $tf_{ij}$ is the number of occurrences of $w_i$ in $c_j$, $gf_i$ is the number of occurrences of word $w_i$ in all comments, and $n$ is the number of all comments.

### 4.2 Latent Semantic Analysis

Latent semantic analysis (LSA) is a computational technique that contains a mathematical representation of language. During the last twenty years its capacity to simulate aspects of human semantics has been widely demonstrated [18]. LSA is based on three fundamental ideas: (1) to begin to simulate human semantics of language, we first obtain an occurrence matrix of terms contained in a comment, (2) the dimensionality of this matrix is reduced using singular value decomposition, a mathematical technique that effectively represents abstract concepts, and (3) any word or text is represented by a vector in this new latent semantic space [7], [18].

#### 4.2.1 Singular Value Decomposition

LSA works through singular value decomposition (SVD), a form of factor analysis. The singular value decomposition of $A$ is defined as:

$$A = USV^T \quad (2)$$

where $U$ and $V$ are the matrices of the term vectors and document vectors. $S = diag(r_1, \ldots, r_n)$ is the diagonal matrix of singular values. To reduce the dimensions, we can simply choose the $k$ largest singular values and the corresponding left and right singular vectors, the best approximation of $A$ with rank-$k$ matrix is given by

$$A_k = U_k S_k V_k^T \quad (3)$$

where $U_k$ is comprised of the first $k$ columns of the matrix $U$ and $V_k^T$ is the first $k$ rows of matrix $V^T$, $S^k = diag(r_1, \ldots, r_k)$ is the first $k$ factors, the matrix $A_k$ captures most of the important underlying structure in the association of terms and documents while ignoring noise due to word choice [30].

When LSA is applied to a new comment, a query, a set of words (like the new comment), is represented as a vector in a $k$-dimensional space. The new comment query can be represented by

$$q' = q^T U_k S_k^{-1} \quad (4)$$

where $q$ and $q'$ are simply the vector of words in a new comment multiplied by the appropriate word weights and the $k$-dimensional vector transformed from $q$, respectively. The sum of these $k$ dimensional word vectors is reflected in the term $q^T U_k$ in the above equation. The right multiplication by $S_k^{-1}$ differentially weights the separate dimensions. Thus the query vector is located at the weighted sum of its constituent term vectors [6].

#### 4.2.2 Feature Selection and Semantic Feature Space

Choosing the number of dimensions $k$ for matrix $A$ is an interesting problem. While a reduction in $k$ can remove much of the noise, keeping to few dimensions or factors may lose important information [7]. In our study, we propose a method which is based on analyzing the first four dimensions of $U$, $S$ and $V$ from comments data. We evaluated the first four columns of $U$ results and confirmed they showed the relation between the meaning of each column and the higher weight words. Therefore, we can predict student performance with more accuracy by employing K-means clustering method [31]. **Tables 3** and **4** show the meaning and the higher weight words of the first four columns after analyzing $U$ results by taking lesson 7 as an example. Words in the first column include the subject of lesson 7 entitled by "An

Table 3 Meaning of dimensions.

| Column | Meaning |
|---|---|
| First | Main subject and learning status |
| Second | Students' learning attitudes |
| Third | Topics in the lesson |
| Fourth | Learning rate and student's behavior |

Table 4 Standard words for lesson 7.

| First | | Second | | Third | | Fourth | |
|---|---|---|---|---|---|---|---|
| Word | Weight | Word | Weight | Word | Weight | Word | Weight |
| Procedure | 0.353 | Be able to | 0.732 | Symbol | 0.504 | Early | 0.448 |
| Language | 0.334 | Make | 0.721 | Meaning | 0.504 | First | 0.441 |
| Symbol | 0.346 | Learning | 0.438 | Various | 0.503 | Training | 0.411 |
| Programming | 0.321 | Procedure | 0.363 | Class | 0.441 | Myself | 0.387 |
| Learning | 0.287 | Myself | 0.346 | Time | 0.373 | Beginning | 0.338 |
| Difficulty | 0.284 | Study | 0.340 | Terminal | 0.37 | Good | 0.332 |
| Use | 0.274 | High | 0.333 | Sufficient | 0.358 | Make | 0.323 |
| Easy | 0.265 | Interest | 0.323 | Compare | 0.357 | End | 0.32 |
| Treatment | 0.248 | Theory | 0.322 | Program | 0.338 | High | 0.303 |
| Knowledge | 0.237 | Good | 0.304 | Save | 0.337 | Finish | 0.209 |

Introduction to C programming language," and learning status such as "understand" or "difficult." In the second column we found words related to student learning attitudes for the lesson take higher weight. In the third column the higher weight words are topics in the lesson, such as "symbol, compare, save or function." In the fourth column, the higher weight words are related to the learning time or rate such as "early, first, full and take time," circumstances, or behaviors performed such as "first time, practical training, or follow." According to the previous analysis, we can conclude the first four dimensions have the strong context to predict student grades with high accuracy.

### 4.3 Noisy Data Detection

Outlier detection discovers data points that are significantly different from the rest of the data [22]. In this paper, detecting outliers are based on two phases: training phase and test phase. We call such outliers **noisy data** from the points of view of grade prediction.

#### 4.3.1 Noisy Data in the *i*-th Cluster in Training Phase

We define noisy data of the $i$-th cluster in the training phase. We calculate $Sd$ to each cluster as a threashold to check a noisy data.

(1) $Sd$ to each cluster

(a) For each cluster, say $i$-th cluster, calculate the centroid $C_i$ of the cluster by finding the average value of K-dimensional vectors (KDV) transformed from comments in the cluster.

$$C_i = \frac{\sum_{k=1}^{n_i} s_{k,i}}{n_i} \tag{5}$$

where $s_{k,i}$ and $n_i$ are the $k$-th singular vectors representing a comment and the number of the comments in the $i$-th cluster, respectively.

(b) Calculate standard deviation $Sd_i$ for the cluster. The higher the $Sd_i$ is, the lower the semantic coherence is [9].

$$Sd_i = \sqrt{\frac{\sum_{k=1}^{n_i} (s_{k_i - C_i})^2}{n_i}} \tag{6}$$

(2) Noisy data detection

Let $s_{k,i}$ be the $k$-th member of the $i$-th cluster; if $|s_{k,i} - C_i| > Sd_i$, then $s_{k,i}$ is a noisy data of the cluster, otherwise $s_{k,i}$ is not a noisy data in the cluster.

#### 4.3.2 Noisy Data Detection in the *i*-th Cluster in Test Phase

In the test phase, instead of the standard deviation, we use the average distance between each comment and a cluster center in a cluster as a threashold to detect noisy data. This is because the prediction accuracy with the average distance became higher than that with the standard deviation from our preliminary experiments.

We define noisy data of the $i$-th cluster in test phase as follows:

- Let $C_i$, $s_{k,i}$ and $d_{i,ave}$ be the centroid of the $i$-th cluster, the $k$-th member in the cluster, and the average distance between members in the cluster and $C_i$, respectively; **if** $|s_{k,i} - C_i| > d_{i,ave}$**, then** $s_{k,i}$ **is a noisy data for the** $i$**-th cluster, otherwise** $s_{k,i}$ **is not a noisy data for the** $i$**-th cluster.**

We separated off about 10% to 15% of comments data as noisy data.

### 4.4 Similarity Measuring Method

The similarity measuring method is proposed to refine the basic prediction method and improve the prediction accuracy of final student grades. We measured the similarity by calculating cosine values between a new comment and each member in the identified cluster by the following equation:

$$Similarity = \frac{S_{new} * S_k}{||S_{new}|| * ||S_k||} = \frac{S_{new} * S_k}{\sqrt{\sum_{i=1}^{k} S_{new}^2} * \sqrt{\sum_{i=1}^{k} S_k^2}} \tag{7}$$

where $S_k$ is the $k$th member in the cluster, and $S_{new}$ is the new comment.

After identifying the nearest cluster center to the new comment, we measure the similarity by calculating cosine values between the new comment $S_{new}$, and each member $S_k$, in the identified cluster, and then return, as an estimated grade of $S_{new}$, the grade of $S_k$ that gets the maximum cosine value among all members in the cluster. This similarity measuring method is used in the Test Phase.

### 4.5 Overlap Method

To predict student grades from comment data, 5-grade categories are used to classify student marks. The method considers prediction is correct only if a grade estimated within the 5-grade categories is the actual grade of a student. We call this method **5-grade prediction method**.

In this paper, in addition to 5-grade categories, we use 9-grade categories so that we can allow the acceptance of a different grade adjacent to the original grade in 5-grade categories of a mark range, i.e., make one mark range correspond to two grades in 5-grade categories. We call this method **overlap method or 9-grade prediction method** for the contrast of 5-grade prediction method. **Tables 5** and **6** show the correspondence relationship between the 5- and 9-grade categories and the range of student marks. For example, we assume a student's mark is 87; the grade of the mark in 5-grade categories is A, and in 9-grade categories

Table 5 5-grades.

| Grade | S | A | B | C | D |
|---|---|---|---|---|---|
| Mark | 90–100 | 80–89 | 70–79 | 60–69 | 0–59 |
| #Student | 21 | 41 | 23 | 17 | 21 |

Table 6 9-grades.

| Grade | S | AS | AB | BA | BC | CB | CD | DC | D |
|---|---|---|---|---|---|---|---|---|---|
| Mark | 90–100 | 85–89 | 80–84 | 75–79 | 70–74 | 65–69 | 60–64 | 55–59 | 0–54 |
| #Student | 21 | 35 | 6 | 10 | 13 | 9 | 8 | 2 | 19 |

is AS; AS corresponds to two grades: A and S, in 5-grade categories.

In the 9-grade prediction method, we consider the prediction is correct if an estimated grade is either A or S. The reasons why we adopt the overlap method are the following: Learning status of students with the upper mark in a grade and others with the lower mark in its one upper grade are not so different from the point of view of the observing teacher. Therefore it is worth noting that handling the two adjacent grades as one grade sometimes helps a teacher to grasp student real learning situations, and to give stable evaluations to students. For example, the mark range of grade AS is from 85 to 89, and that is closer to the lowest mark 90 of grade S than the lowest mark 80 of grade A. The overlap method is used in the Test Phase.

## 5. Experimental Results

### 5.1 Measures of Student Grade Prediction

In our experiment, we evaluated the prediction results by 10-fold cross validation and run evaluation experiments according to 4 values: TP (True Positive), TN (True Negative), FP (False Positive) and FN (False Negative) and calculated Precision, Recall, $F$-measure and Accuracy in each lesson as follows:

Let G be 5-grade categories (S, A, B, C and D), and $X$ be a subset of G; let $obs$ $(s_i, X)$ be a function that returns 1 if the grade of student $s_i$ is included in $X$, 0 otherwise, where $1 \le i \le n$, and $n$ is the number of students; pred$(s_i)$ be a function that returns a set of grade categories only including a predicted grade for student $s_i$; !pred $(s_i)$ returns a complement of pred$(s_i)$.

$$TP = \{s_i | obs(s_i, pred(s_i)) = 1\}$$

$$FP = \{s_i | obs(s_i, pred(s_i)) = 0\}$$

$$TN = \{s_i | obs(s_i, !pred(s_i)) = 1\}$$

$$FN = \{s_i | obs(s_i, !pred(s_i)) = 0\}$$

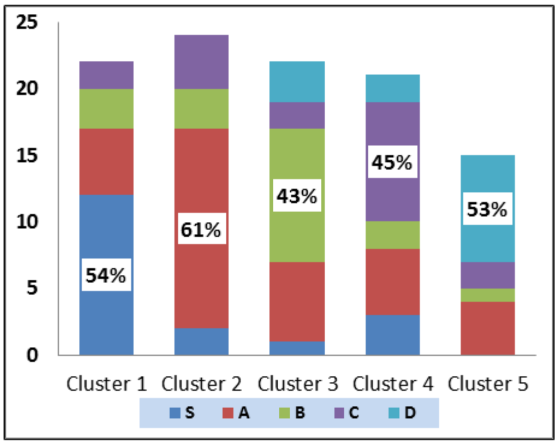
$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + TN}$$

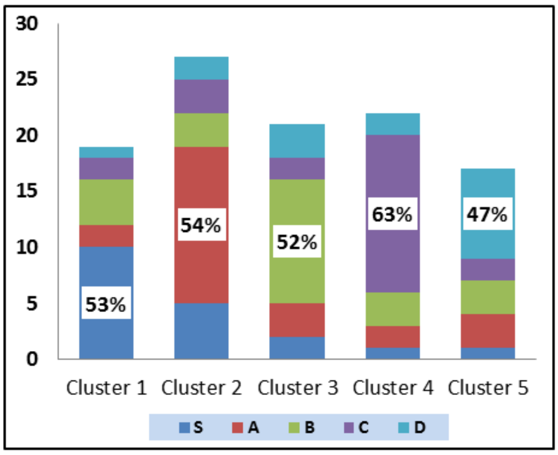
$$F - measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$

$$\text{Accuracy} = \frac{TP + FN}{TP + TN + FP + FN}$$

Actually, $FP$ and $TN$ are important values and affect the prediction results. $FP$ has a strong relation with Precision and $TN$ with Recall. As $FP$ increases, we may more pick up other grade students, say (S) or (A), as a target grade student, say (D). We often want to take care about low level students. At that time, we need to detect all of them. As the value of $TN$ becomes higher, we may more misdetect them. So our study showed the prediction results by calculating Precision, Recall, Accuracy, $F$-measure and standard deviation to the 3 methods: the basic prediction method, the overlap method and the similarity measuring method.




(a) Training phase.




(b) Test phase.

**Fig. 3** Results of classifying students.

### 5.2 Effect of Basic Prediction Method

#### 5.2.1 Training Phase

According to the training phase of the basic prediciton method described in Section 3.2 (3), we built a prediction model.

**Figure 3** (a) displays the results for lesson 7. Grade S accounts for about 54% in Cluster 1; grade A about 61% in Cluster 2; grade B about 43% in Cluster 3; grade C about 45% in Cluster 4; finally, grade D about 53% in Cluster 5. Grade S, A, B, C, and D are dominant grades in Cluster 1, 2, 3, 4, and 5, respectively. We analyzed each lesson from 8 to 15 as well.

#### 5.2.2 Test Phase

We conducted student grade prediction according to the steps described in Section 3.2 (4) and evaluated the prediction performance by 10-fold cross validation.

Figure 3 (b) presents the results of student grade prediction: (Cluster 1, S=53%), (Cluster 2, A= 54%), (Cluster 3, B=52%), (Cluster 4, C=63%), (Cluster 5, D=47%).

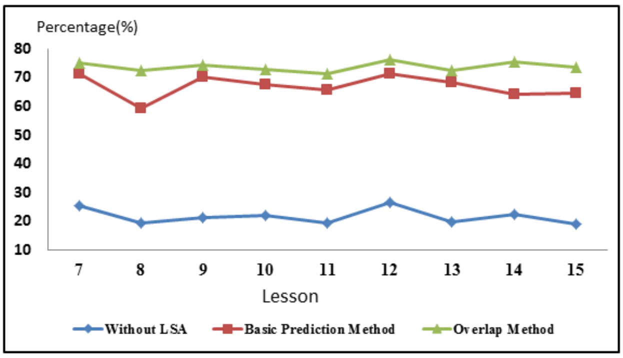
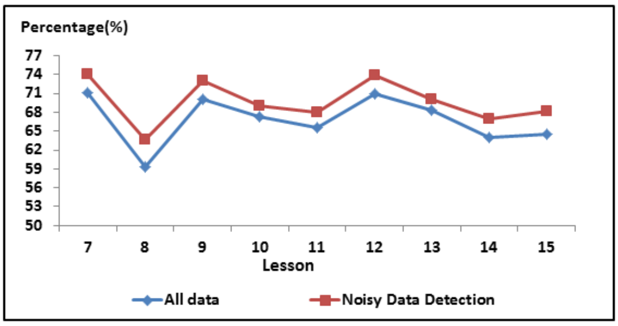
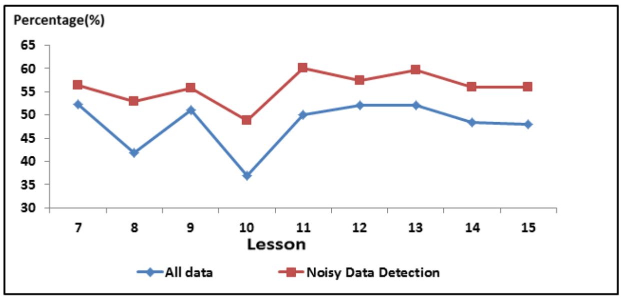



**Fig. 4** The effect of LSA.




(a) Accuracy of prediction results.




(b) $F$-measure of prediction results.

**Fig. 5** Prediction results (with and without noisy data detection) from lessons 7 to 15.

### 5.3 Effect of LSA

We checked the effect of LSA from lessons 7 to 15. We used all comments data. As shown in **Fig. 4**, the average prediction accuracy results of the basic prediction method without LSA were between 19.0% and 26.4%. It is much lower than those with LSA, which were between 59.0% and 71.0%. In addition, adding the overlap method to the basic prediction method with LSA, the average prediction accuracy became between 71.0% and 76.0%.

### 5.4 Effect of Noisy Data Detection

To examine the effect of filtering noisy data, we calculated the average prediction (accuracy and $F$-measure) of student grades before and after detecting noisy data. The results are shown in **Fig. 5**. The prediction accuracy results were between 59.0% and 71.0% for all data, and those became between 63.5% and 74.0% after detecting noisy data as shown in Fig. 5 (a). Also, the average $F$-measure for all lessons was 48.1% and after removing noisy data it became 55.8% as shown in Fig. 5 (b).

The highest accuracy results from the top were obtained in lessons 7 and 12, and the lowest ones from the bottom in lessons 8 and 14.

In addition, **Fig. 6** focuses on the prediction results in each grade after detecting noisy data. We can see that grade A took the highest results of prediction accuracy and $F$-measure, and the grade D was the lowest.

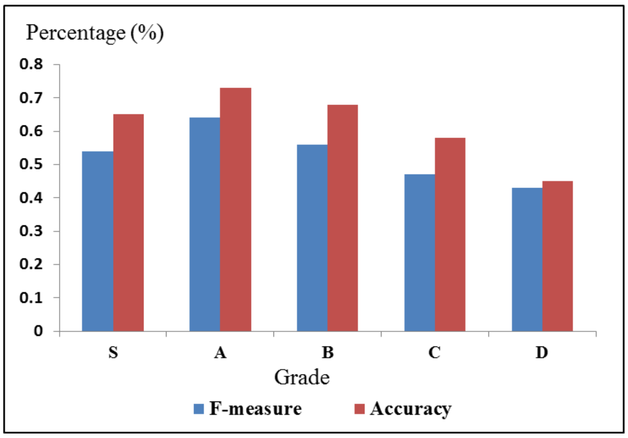



**Fig. 6** The results of accuracy and $F$-measure in each grade after detecting noisy data.

**Table 7** Overall prediction results.

| | Precision | Recall | $F$-measure | Accuracy |
|---|---|---|---|---|
| Method | After(Before) | After(Before) | After(Before) | After(Before) |
| Basic prediction method | 0.530(0.452) | 0.589(0.536) | 0.554(0.480) | 0.696(0.664) |
| Overlap Method | 0.682(0.662) | 0.642(0.545) | 0.622(0.596) | 0.771(0.736) |
| Similarity + 5 grades | 0.645(0.631) | 0.697(0.695) | 0.680(0.661) | 0.822(0.785) |
| Similarity + 9 grades | 0.787(0.765) | 0.735(0.721) | 0.762(0.743) | 0.864(0.842) |

### 5.5 Effect of Overlap and Similarity Measuring Methods

We illustrate the difference between the basic method and two additional methods: overlap and similarity measuring methods. The average overall results of accuracy and $F$-measure are reported in **Table 7**. It shows the effect of noisy data detection by evaluating the prediction results across all the lessons between the basic prediction, similarity measuring, and overlap methods.

Comparing with the basic prediction method, the similarity measuring improved the accuracy results from 66.4% to 78.5% through the analysis of all data, from 69.6% to 82.2% after detecting noisy data. Moreover, the overlap method had a strong effect that increased the prediction accuracy from 66.4% to 73.6% with all data, 69.6% to 77.1% after detecting noisy data. Combining the similarity measuring method and the overlap method, the prediction accuracy increased from 73.6 % to 84.2% with all data, 77.1% to 86.4% after detecting noisy data.

For more clarity, to evaluate the effect of similarity measuring method, **Fig. 7** displays the comparison between the accuracy results of the prediction of student grades with and without the similarity measuring method. We can see the effect of the similarity measuring method, especially when checking the fact that the prediction accuracy results with the similarity measuring method without the overlap method is better than those of the overlap method without the similarity measuring method.

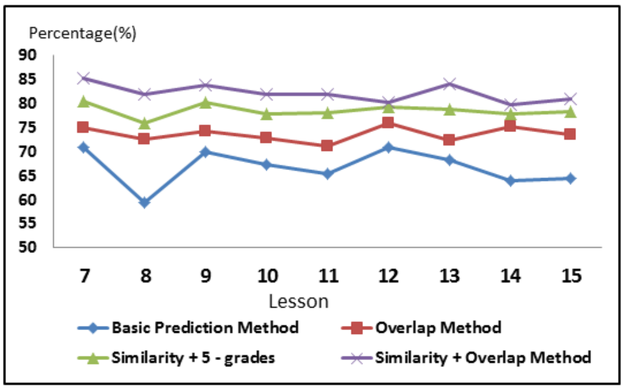



**Fig. 7** The prediction accuracy results with and without similarity measuring method.

**Table 8** $Sd$ of prediction accuracy for the proposed methods.

| Lesson | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|
| **Basic Prediction Method** | 2.76 | 9.19 | 3.73 | 6.02 | 7.89 | 5.19 | 4.57 | 8.07 | 5.63 |
| **Overlap Method** | 2.99 | 6.41 | 3.44 | 5.11 | 6.20 | 4.45 | 6.0 | 5.12 | 4.24 |
| **Similarity Measuring Method** | 2.16 | 5.95 | 4.05 | 4.76 | 7.22 | 4.03 | 4.56 | 6.35 | 4.03 |

**Table 9** Correlation coefficient of $Sd$ and accuracy.

| | |
|---|---|
| **Basic Prediction Method** | −0.89485 |
| **Overlap Method** | −0.5152 |
| **Similarity Measuring Method** | −0.91978 |

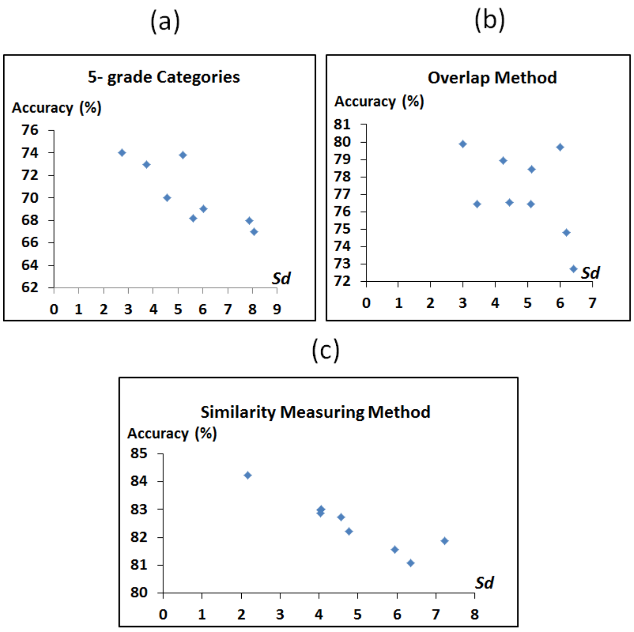



**Fig. 8** The correlation between $Sd$ and prediction accuracy.

### 5.6 Correlation between Standard Deviation and Prediction Accuracy

**Table 8** displays the standard deviation ($Sd$) results of the prediction accuracy from lessons 7 to 15 for the basic prediction method, overlap and similarity measuring methods after detecting noisy data. We can see that higher $Sd$ in lessons: 8, 11, and 14 tend to get lower prediction accuracy and $F$-measure.

**Table 9** shows the correlation coefficients between the $Sd$ and prediction accuracy.

**Figure 8** (a) and (c) show there are strong correlations between the $Sd$ and the prediction accuracy of the basic prediction method and between the $Sd$ and the similarity measuring method. On the other hand, the correlation coefficient between the $Sd$ and that of the overlap method only shows a weak correlation. We believe that this shows a fact that the overlap method gives stable evaluation of student grades.

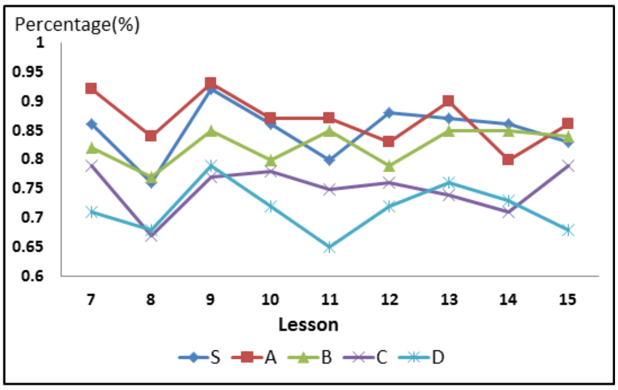



**Fig. 9** The prediction accuracy by grade after applying the similarity measuring method.

### 5.7 Prediction Accuracy Differences between 5 Grades

Finally, **Fig. 9** displays the relationship between C-comments data from lessons 7 to 15 and the prediction accuracy in each grade after applying the similarity measuring method and detecting noisy data. As shown in Fig. 9, the average prediction accuracy results were between 0.65 and 0.93. We can clearly distinguish higher grade group: S, A, and B from lower one: C and D. One of the reasons why prediction accuracy of grades C and D became lower came from the smaller number of comments in those grades.

## 6. Conclusion and Future Work

Learning comments are valuable sources of interpreting student behavior during a class. The present study discussed student grade prediction methods based on their free-style comments. We used C-comments data from the PCN method that presented student attitudes, understanding and difficulties concerning to each lesson.

The main contributions of this work are twofold. First, we discussed the basic prediction method that analyzed C-comments data based on the LSA technique and classified the results obtained using LSA into 5 groups by K-means clustering method. Second, we proposed two new approaches: overlap and similarity measuring methods to improve the basic method and conducted experiments to validate the two approaches. The overlap method allows the acceptance of two grades for one mark to get the correct relation between LSA results and student grades. We made confirmation that the overlap method with 9-grade categories enabled a more stable evaluation than 5-grade categories. The overall results of average prediction accuracy became better than those of classifying student marks to 5-grade categories.

The similarity measuring method calculated similarity between a new comment and comments in the nearest cluster. The results of prediction accuracy with the similarity measuring method became much better than those without the similarity measuring method. By combining with the overlap method (9-grade prediction method), the prediction accuracy became higher.

To sum up, there are still quite a few considerations that would surely add even more value to the results obtained. It is necessary to improve the prediction results of student performance from their comments; other machine learning techniques, such as, neural network and support vector machine, will be candidate for the improvement to compare with the present method (K-means Clustering).

Another interesting issue is expanding the problem to improve student performance by providing advice to them in different classes according to the estimated performance of each student. Measuring motivation after each lesson can help for giving feedback to students and encourage them to improve writing skills; they can describe attitudes and situations, understanding of subjects, difficulties to learn, and learning activities in the classroom. This will help a teacher to give advice and improve their performance.

**Acknowledgments** This work was supported in part by PEARL, enPiT of Project for Establishing a Nationwide Practical Education Network for IT Human Resources Development under the MEXT, Japan, and JSPS KAKENHI Grant Number 24500176, 25350311, 26350357 and 26540183.

## References

[1] Abdousa, M.: A predictive Study of Learner Satisfaction and Outcomes in Face-to-Face, Satellite Broadcast, and Live Video-Streaming Learning Environments, *The Internet and Higher Education*, Vol.13, No.4, pp.248–257 (2010).

[2] Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R. and Honrao, V.: Predicting Students' Using ID3 and C4.5 Classification Algorithms, *International Journal of Data Mining and Knowledge Management Process* (*IJDKP*), Vol.3, No.5, pp.39–52 (2013).

[3] Ananiadou, S.: National Centre for Text Mining: Introduction to Tools for Researchers, (online), available from ⟨http://www.jisc.ac.uk/publications/publications/⟩ (2008).

[4] Antai, R., Fox, C. and Kruschwitz, U.: The Use of Latent Semantic Indexing to Cluster Documents into Their Subject Areas., *Fifth Language Technology Conference. Springer. ASME Design Engineering Technical conferences, DETC2001/DTM* (2011).

[5] Baker, R.S.J.D.: Data Mining for Education, *International Encyclopedia of Education* (*3rd edition*). Oxford, UK: Elsevier (2009).

[6] Berry, M.W., Dumais, S.T. and O'Brien, G.W.: Using Linear Algebra for Intelligent Information Retrieval, *SIAM Review*, Vol.37, No.4, pp.573–595 (1995).

[7] Deerwester, S., Dumais, S.T., Furnas, G.W., Landauer, T.K. and Harshman, R.: Indexing by Latent Semantic Analysis, *Journal of the American Society for Information Science*, Vol.41, No.6, pp.391–407 (1990).

[8] Delavari, N., Phon-Hon-Amnuaisuk, S. and Beikzadeh, M.R.: Data Mining Application in Higher Learning Institutions, *Informatics in Education*, Vol.7, No.1, pp.31–54 (2008).

[9] Dhillon, I.S. and Modha, D.S.: Concept Decompositions for Large Sparse Text Data using Clustering, *Machine Learning*, Springer, Vol.42, No.1-2, pp.143–175 (2001).

[10] Earl, L.: *Assessment of Learning, for Learning, and as Learning*, chapter 3, Thousand Oaks, CA, Corwin Press (2003).

[11] Fuller, C.M., Biros, D.P. and Delen, D.: An Investigation of Data and Text Mining Methods for Real World Deception Detection, *Expert Systems with Applications*, Vol.38, No.7, pp.8392–8398 (2011).

[12] Goda, K., Hirokawa, S. and Mine, T.: Correlation of Grade Prediction Performance and Validity of-Self-Evaluation Comments, *SIGITE 2013, The 14th Annual Conference in Information Technology Education*, pp.35–42 (2013).

[13] Goda, K. and Mine, T.: Analysis of Students' Learning Activities through Quantifying Time-Series Comments, *KES 2011, Part II, LNAI 6882*, Springer-Verlag Berlin Heidelberg, pp.154–164 (2011).

[14] Gorissen, P., van Bruggen, J. and Jochems, W.: Usage Reporting on Recorded Lectures, *International Journal of Learning Technology*, Vol.7, No.1, pp.23–40 (2012).

[15] Hord, S.M.: *Professional Learning Communities: Communities of Continuous Inquiry and Improvement*, Southwest Educational Development Laboratory (1997).

[16] Hung, J.: Trends of E-learning Research from 2000 to 2008: Use of Text Mining and Bibliometrics, *British Journal of Educational Technology*, Vol.43, No.1, pp.5–15 (2012).

[17] Landauer, T.K., Foltz, P.W. and Laham, D.: An Introduction to Latent Semantic Analysis, *Discourse Processes*, Vol.25, pp.259–284 (1998).

[18] León, J., Olmos, R., Escudero, I., Jorge-Botana, G. and Perry, D.: Exploring the Assessment of Summaries: Using Latent Semantic Analysis to Grade Summaries Written by Spanish Students, *Procedia - Social and Behavioral Sciences*, Vol.83, pp.151–155 (2013).

[19] Li, S. and Ho, H.: Predicting Financial Activity with Evolutionary Fuzzy Case Bbased Reasoning, *Expert Systems with Applications*, Vol.36, No.1, pp.411–422 (2009).

[20] Lin, F.-R., Hsieh, L.-S. and Chuang, F.-T.: Discovering genres of online discussion threads via text mining, *Computers and Education*, Vol.52, pp.481–495 (2009).

[21] Liu, B., Cao, S.G. and He, W.: Distributed Data Mining for E-Business, *Information Technology and Management*, Vol.12, No.1, pp.1–13 (2011).

[22] Mansur, M.O. and Md. Sap, M.N.: Outlier Detection Technique in Data Mining: A Research Perspective, *Postgraduate Annual Research Seminar* (2005).

[23] Minami, T. and Ohura, Y.: Lecture Data Analysis towards to Know How the Students' Attitudes Affect to their Evaluations, *International Conference on Information Technology and Applications* (*ICITA*), pp.164–169 (2013).

[24] Parack, S., Zahid, Z. and Merchant, F.: Application of data mining in educational databases for predicting academic trends and patterns, *IEEE International Conference on Technology Enhanced Education* (*ICTEE*) (2012).

[25] Pellegrino, W., Chudowsky, N. and Glaser, R.: *Knowing What Students Know: The Science and Design of Educational Assessment*, Washington. DC: National Academy Press (2001).

[26] Romero, C., Espejo, P.G., Zfra, A., Romero, J.R. and Ventura, S.: Web Usage Mining for Predicting Final Marks of Students That Use Moodle Courses, *Computer Applications in Engineering Education* (2010).

[27] Shepard, L.A.: *The Role of Classroom Assessment in Teaching and Learning*, (CSE Technical Report 517), Los Angeles CA: Center for the Study of Evaluation (2000).

[28] Tane, J., Schmitz, C. and Stumme, G.: Semantic Resource Management for the Web: An e-learning Application, *WWW Conference*, New York, USA, pp.1–10 (2004).

[29] Ur-Rahman, N. and Harding, J.: Textual Data Mining for Industrial Knowledge Management and Text Classification: A business Oriented Approach, *Expert Systems with Applications*, Vol.39, No.5, pp.4729–4739 (2012).

[30] Yu, B., ben Xu, Z. and hua Li, C.: Latent Semantic Analysis for Text Categorization Using Neural Network, *Knowledge-Based Systems*, Vol.21, pp.900–904 (2008).

[31] Zaiane, O.R.: *Introduction to Data Mining*, chapter I, CMPUT690 Principles of Knowledge Discovery in Databases (1999).

**Shaymaa E. Sorour** received her B.S. degree in Education Technology, at Faculty of Specific Education, Kafr Elsheikh University, Egypt and a M.S. degree in Computer Education at Department of Computer Instructor Preparation, Faculty of Specific Education, Mansoura University, Egypt, in 2004 and 2010 respectively. Since 2005 until now, she is working as an assistant lecture at Department of Education Technology, Faculty of Specific Education, Kafr Elsheikh University, Egypt. She is currently a Ph.D. student in Graduate School of Information Science and Electrical Engineering, Department of Advanced Information Technology, Kyushu University, Japan. She is a member of IEEE and IPSJ.

**Tsunenori Mine** received his B.E. degree in Computer Science and Computer Engineering, in 1987, and his M.E. and D.E. degrees in Information Systems, in 1989 and 1993, respectively, all from Kyushu University. He was a lecturer at College of Education, Kyushu University, from 1992 to 1994 and at Department of Physics, Faculty of Science, Kyushu University from 1994 to 1996. He was a visiting researcher at DFKI, Saarbruecken, Germany from 1998 to 1999, and at Language Technology Institutes of CMU, Pittsburgh, PA, USA in 1999. He is currently an Associate Professor at Department of Advanced Information Technology, Faculty of Information Science and Electrical Engineering, Kyushu University. His current research interests include Natural Language Processing, Information Retrieval, Information Extraction, Collaborative Filtering, Personalization and Multi-Agent Systems. He is a member of IPSJ, IEICE, JSAI, NLPSJ and ACM.

**Kazumasa Goda** received his B.E. and M.E. degrees in 1994 and in 1996 from Kyushu University, respectively. He has been an Associate Professor at Kyushu Institute of Information Science since 2008. His research interests include Programming Theory, Programming Education, and Computer Education. He is a member of JSiSE, JAEiS, and IPSJ.

**Sachio Hirokawa** received his B.S. and M.S. degrees in Mathematics and Ph.D. degree in Interdisciplinary Graduate School of Engineering Sciences from Kyushu University in 1977, 1979, and 1992. He founded the search engine company Lafla (http://www.LafLa.co.jp) in 2008 as a university venture company. He has been working as an outside director of the company. He is currently a Professor in the Research Institute for Information Technology of Kyushu University, since 1997. His research interests include Search Engine, Text Mining, and Computational Logic. He is a member of JSAI, IPSJ, IEICE and IEEE. He has been serving as a general chair of international conference AAI (Advanced Applied Informatics) since 2012.