Predicting Students' grades based on free style Comments Data by Artificial Neural Network

Sorour, Shaymaa E.
Faculty of Specific Education, Kafr Elsheik University

Mine, Tsunenori
Faculty of Information Science and Electrical Engineering, Kyushu University

Goda, Kazumasa
Kyushu Institute of Information Science

Hirokawa, Sachio
Research Institute for Information Technology, Kyushu University

出版情報：2014 IEEE Frontiers in Education Conference Proceedings, 2014-10. IEEE
バージョン：
権利関係：

Predicting Students' grades based on free style Comments Data by Artificial Neural Network

Abstract—**Predicting students' academic achievement with high accuracy has an important vital role in many academic disciplines. Most recent studies indicate the important role of the data type selection. They also attempt to understand individual students more deeply by analyzing questionnaire for a particular purpose. The present study uses free-style comments written by students after each lesson, to predict their performance. These comments reflect their learning attitudes to the lesson, understanding of subjects, difficulties to learn, and learning activities in the classroom. To reveal the high accuracy of predicting student's grade, we propose Latent Semantic analysis (LSA) technique to overcome the problems caused by using statistically derived conceptual indices instead of individual words, then apply Artificial Neural Networks (ANN) model. We chose five grades instead of the mark itself as a student's result to predict their grade. The potentials of ANN for approximating extremely complex problems help us to develop an estimation model of student performance. Our proposed method averagely achieves 82.6% and 76.1% prediction accuracy and F-measure, respectively of students' grades.**

Keywords*—*Comments Data, Latent Semantic analysis (LSA), Artificial Neural Networks (ANN), predicting performance.

I. Introduction

Recently, many researchers have drawn their attention to enhance learners' performance and have contributed to the related literature. By and large, researchers in this field manage to advocate novel and smart solutions to improve performance. Thus, learners' performance assessment may not be viewed as being somewhat separate from learning process. It must be a continuous process. On the contrary, performance assessment is an integral part of learning processes and ultimately, should aim to improve the quality of students' learning.

In the classroom, there are very varieties of students. They have wide-ranging performance. Some are assiduous and self-motivated, others have difficulty to understand the lesson or frustrate from the subject; they're waiting instructions from the teacher to follow other students. Teachers can give advices by their careful observation, but it is hard task to grasp all the class members' learning attitudes all over the periods in the semester.

To control students' learning behavior and situations, the previous studies use various regular assessment methods such as e-learning logs, test marks and questionnaires [1, 2]. These methods have difficulty of meaning interpretation and creating good questions. Although, teachers observation has vital role to improve educational situation, they pick up some cases according to their needs mainly based on their experience in the class.

Goda et al. proposed the PCN method to estimate a learning situation from a comment freely written by students [3]. While describing comments, the students can reflect on their learning attitudes or behaviors. Therefore, they call the students' comments as free-style comments with their self-reflection or self-evaluation comments. The PCN method categorizes comments into three items of P (previous), C (current) and N (next). Item P is learning activities for preparation of a lesson and reviewing of previous class. Item C is the understanding of the lesson and learning attitudes to the lesson. Item N is the learning plan and goal until the next lesson. To expose high possibility of comments data for predicting students' performance, we propose a method for predicting students' grades using comments of C item (C-comments in short) from the PCN method. We apply LSA and ANN to the C-comments. In addition, we calculate similarity to the comments in each cluster, after classifying LSA results into 5 clusters. Experimental results averagely achieves 82.5% prediction accuracy of students' grades by ANN, and 78.5% with similarity measuring method. The contributions of our work are the following:

- We apply ANN to students' comments and made it learned the relationships between analyzed comments data by (LSA) and the final grade of the students. It can handle linear and nonlinear problems for text categorization and learn the problem presented.
- We propose a similarity measuring method that calculates similarity between a new comment and comments in the nearest cluster, which is created in the training phase.
- We conduct experiments to validate our proposed methods by calculating F-measure and accuracy for estimating the final grades in each method. The experimental results illustrate the validity of the proposed methods by showing better prediction accuracy and F-measure of students grades.

The rest of the paper is organized as follows. Section 2 discuses related work. Section 3 introduces overview of our research. Section 4 describes the proposed methods to predict the students' final grades. Section 5 introduces the procedures and methodology of our proposed method. Section 6 discusses some of highlighted experimental results. Finally Section 7 concludes the paper and describes our future work.

II. Related Work

The ability to predict students' 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 applying educational data mining technique *(EDM)* to conduct the automatic analysis of learner performance and behavioral data with learning environments. An emerging trend in *EDM* is the use of text mining which is an extension of data mining to text data. Various experiments have been carried out in two areas to predict students' academic performance. Machine learning techniques are overlapped with two methods to distinguish important data classes and predict future data trends.

A. Educational Data Mining

Educational data mining *(EDM)* is defined to be "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in." Furthermore, several leading *EDM* experts [6, 7, 8] classify work in *EDM* into a few categories such as statistics and visualization, prediction (classification, regression, and density estimation), clustering, relationship mining, outlier detections. *EDM* can be applied to assess students learning performance, to improve the learning process and guide students learning, to provide feedback and adapt learning recommendations based on students' learning behaviors, to evaluate learning materials and courseware, to detect abnormal learning behaviors and problems, and to achieve a deeper understanding of educational phenomena [7, 8, 9, 10].

Quite a few *EDM* studies have been found in the most recent literature from 2010 to 2013. For example, Gorissen et al. analyzed the interactions of students with the recorded lectures using educational data mining techniques. The data logged by the lecture capture system (LCS) was used and combined with collected survey data. They found discrepancies as well as similarities between students' verbal reports and actual usage as logged by the recorded lecture servers. The data suggests that students who do this have a significantly higher chance of passing the exams [10]. Jovanovica et al. applied classification models for predicting students' performance, and cluster models for grouping students based on their cognitive styles in e-learning environment. They indicate that the classification models helped teachers, students and business people, for early engaging with students who are likely to become excellent on a selected topic [11]. Parack et al. used multiple data mining algorithms for student profiling and grouping. They found that data mining can be very useful in discovering valuable information which can be used for profiling students based on their academic record such as exam scores, term work grades, attendance and practical exams [12]. Wu He analyzed the online questions and chat messages, that automatically recorded by a live video streaming (LVS) system. He applied data mining and text mining techniques to analyze two different data sets and then conducted correlation analysis for two educational courses with the most online questions and chat messages respectively. The study found the discrepancies as well as similarities in the students' patterns and themes of participation between online questions and online chat messages [13].

B. Educational Text Mining

Text mining is focused on finding and extracting useful or interesting patterns, models, directions, trends, or rules from unstructured text such as 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 [14]. As an automated technique, text mining can be used to efficiently and systematically identify, extract, manage, integrate, and exploit knowledge for research and education [15].

Currently, there are only several studies about how to use text mining techniques to analyze learning related data. For example, Tane et al. used text mining (text clustering techniques) to group e-learning resources and documents according to their topics and similarities [16]. Hung used clustering analysis as an exploratory technique to examine e-learning literature and visualized patterns by grouping sources that share similar words and attribute values [17]. Goda et al. proposed a students' grade prediction model based on the students' comments by using the PCN method. Their experimental results illustrated that as comments of students get higher PCN scores, the prediction performance of the students' grades becomes higher [4]. Sorour et al. proposed a prediction method of students' grade based on comments data, and evaluated the effects of their proposed method by investigating prediction accuracy of their grades in each lesson [5].

The objective of our study is to reveal the high accuracy of predicting student's grade. We proposed two methods(ANN and similarity measuring)with LSA technique. In addition, we improve the method proposed by Sorour et al. and show that our proposed methods achieve higher prediction accuracy of students' grade than those of their method.

III. Background

A. Comments Data

In this research we use C-comment from [3], to predict student's grade. C-comment indicates the understanding and achievements of class subjects during the class time. We illustrate examples of C comments written by students' as follows.

- I was completely able to understand the subject of this lesson and have confidence to make other functions similar to ones I learned in this lesson.
- I didn't finish all exercise, because i can't understand the last two methods and the time is up.

Comments data collected from 123 students in two classes. They took Goda's courses that consisted of 15 lessons. In this research, we use the students' comments collected for the last half, from lesson 7^{th} to 15^{th}. Main subjects in those lessons are introductory C programming. Although, we have 123 students in all lessons, some students didn't submit their comments because they did not write any comments or were absent. Table I displays the real number of comments in each lesson that we analyzed them. The number of words appeared in the comments is about 1400 in each lesson, and the number of words in all the comments without duplication is over 430 in each lesson.

B. Students' Grades

To predict students' grades from their comments, 5-grade categories are used to classify students' marks. We consider

TABLE I: Number of comments

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

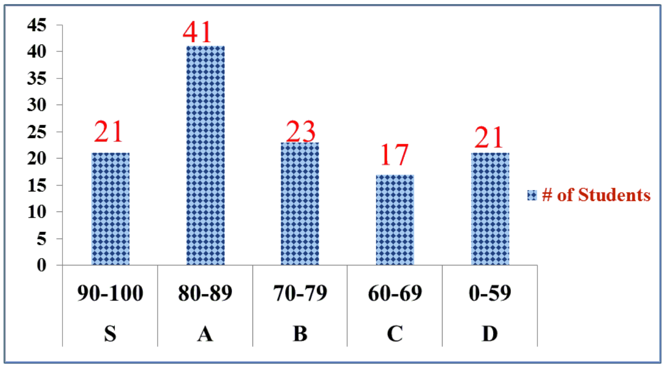


Fig. 1: The correspondence between grades and the range of marks.

prediction is correct only if one estimated grade with 5-grade category is the actual grade of a student. Fig.1 shows the correspondence between the grades and the mark. For example, number of students' in grade A=41 and their marks between 89 and 80.

C. Procedures of the Proposed Method

Fig.2 displays the overall procedures of our proposed method; we have four phases:

1) Comment Data: This phase focuses on collecting and analyzing C- comments by extracting words and part of speech.
2) Data Preparation: The data preparation phase covers all the activities required to construct the final data set from the initial raw data. In our method, we calculate the word frequencies, apply entropy weighting method and LSA technique to reduce the dimensions of a matrix and obtain the most significant vectors.
3) Training Phase: In this phase, we use LSA results from the previous step then build network model for each lesson by applying **artificial neural network (ANN) model**. Also, we classify LSA results into 5 clusters and measure cosine similarity in each cluster. We call this method **similarity measuring method**.
4) Test Phase: This phase revolves on extracting words from a new comment, and transforming an extracted-words vector of the comment to a set of K-dimensional vector (KDV) by using LSA.

We evaluate the prediction performance by 10-fold cross validation. We separated comments data by using 90% of them as training data and constructed a model. Then we applied the model to the rest 10% data and compared a predicted value with corresponding observed data. The procedure is repeated 10 times and the results were averaged.

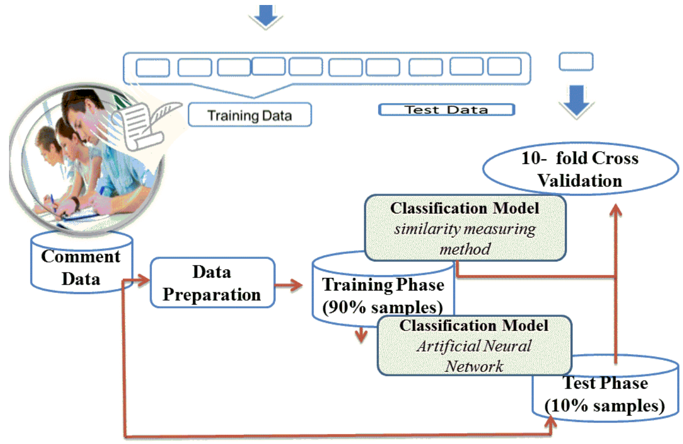


Fig. 2: Procedures of the proposed method

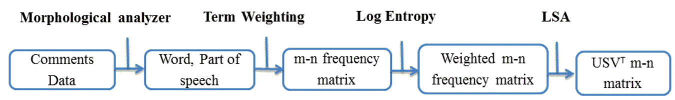


Fig. 3: Procedures to use LSA for measuring similarity between comments.

IV. METHODDOLOGY

This section describes our methodology for predicting student performance from free-style comments data. We analyzed data and applied LSA technique, then classified the obtained results by ANN model and similarity measuring method.

A. Term Weighting to Comments

Fig.3 provides an overview of how we use LSA to process free-style comments data. It illustrates the procedures before applying LSA. First, we analyze comments data with Mecab[1] program, which is a Japanese morphological analyzer to extract words and their part of speech (verb, noun, adjective, and adverb). Next we create a word-by-comment matrix with extracted words. This word-by-comment matrix, say A, is comprised of m words w_1, w_2,. . , w_m in n comments c_1, c_2, ...,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, log entropy term weighting is applied to the original word-by-comment matrix, which is the basis for all subsequent analyses [18].

B. Latent Semantic Analysis

- *Semantic vector space generation*

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. LSA is based on three fundamental ideas: (1) to begin to simulate human

[1] http://sourceforge.net/projects/mecab/

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 makes the tool a latent semantic space, and (3) any word or text is represented by a vector in this new latent semantic space [17, 19].

- *Singular value decomposition*

Latent semantic analysis (LSA) works through singular value decomposition (SVD), a form of factor analysis. From the training comments, we can get the term by comment matrix $A(m*n)$, it means there are m distinct terms in n comments collection . The singular value decomposition of A is defined as

$$A = USV^T \tag{1}$$

where U and V are the matrices of the term vectors and document vectors. $S = diag(r1, ..., r_n)$ is the diagonal matrix of singular values. For reducing 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

$$A_k = U_k S_k V_k^T \tag{2}$$

where U_k is comprised of the first k columns of the matrix U and V_k^T is comprised the first k rows of matrix V^T, $S^k = diag(r_1, ..., r_k)$ 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 [20].

- *Feature selection and semantic feature space*

In this section, we explain why we used only the first four dimensions from comments data. Our objective is to extract a close relationship from the comments data for each lesson to predict student's grade. We evaluate the first four columns of U results in all lessons to determine the meaning to each column. We found the higher weight to each column as the following:

- **First Column**: Main subject of each lesson and the learning status.
- **Second Column**: Students' learning attitudes for the lesson.
- **Third Column**: Topics in the lesson.
- **Forth Column**: Learning rate.

Tabel II displays the standard words for lesson 7 based on the first four columns of U results. The subject of lesson 7 "An Introduction to C programming language," and students' learning status such as "understand" or "difficult." In the second column we found words related to students' 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."

TABLE II: 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

- *LSA for information retrieval*

When LSA is used for information retrieval for new comments, a query is represented as a vector in k-dimensional space. The query is represented by

$$q' = q^T U_k S_k^{-1} \tag{3}$$

Where q' and q are the vector of words in a new comment multiplied by the appropriate word weights and the K-dimensional vector (KDV) transformed from q, respectively. The sum of these k dimensional word vectors is reflected by the term $q^T U_k$ in the above equation. The right multiplication by S_k^{-1} differentially weights the separate dimensions[20].

LSA has become one of the most widely-used computational tools of recent years, and one of the fastest-growing areas of application has been the field of education. Today, LSA is already a reality in some U.S. classrooms and it is gradually finding its way into more and more schools as a means of helping to improve students' writing and comprehension strategies [21, 22]. LSA has been applied to text categorization in many previous works. Yang used SVD for noise reduction so as to improve the computational efficiency in text categorization [23]. Zelikovitz and Hirsh performed LSA expanded term-by-document matrix in conjunction with background knowledge in text categorization [24]. In our research, we analyze comments data based on LSA technique to obtain more similarity between word by comment matrix and detect noisy data by reducing number of dimensions. Our objective is to establish relationship between analyzed comments and students' grades.

C. K-means Clustering Method

K-means clustering algorithm is one of the simplest unsupervised learning algorithms. It is a process in which a set of objects are split into a set of structured sub-classes, bearing a strong similarity to each other, such that they can be safely treated as a group. Such sub-classes are referred to as clusters [25].

Sorour et al. applied LSA technique to the comments data. Then they classified the results into 5 groups by using K-means clustering method. They carried out test data by comparing clustering results with students' grades. Their results proved, each cluster has almost one grade, that most frequently appears in a cluster.

Sorour et al. established the following steps to predict students' grade:

1) Extract words from a new comment.
2) Transform the comment to a K dimensional vector (KDV) using LSA.
3) Identify which cluster center is the nearest to the comment, by measuring the distance between the comment and cluster centers.
4) Return the dominant grade in the cluster to which the identified cluster center belongs to, where the dominant grade in a cluster means the grade that most frequently appears in the cluster.

After performing the above steps, they evaluated the prediction performance by 10-fold cross validation, and presented the results of student grade prediction.

The following an example of the obtained results from lesson 7, (Cluster 1, S_{grade} = 53%) and the remaining for grades A, B, C and D. (Cluster 2, A_{grade}=54%), (Cluster 3, B_{grade}=52%), (Cluster 4, B_{grade}=63%) and (Cluster 5, D_{grade}=47%) [5].

D. Similarity Measuring Method

In our research, We found, although each cluster has almost one grade, it contains other grades. we measured the similarity by calculating cosine values between the new comment and each member in the identified cluster as the follows:

$$Similarity = \frac{S_{new} * S_k}{||S_{new}|| * ||S_k||} = \frac{S_{new} * S_k}{\sqrt{\sum_{i=1}^{k} S_{new}^2} * \sqrt{\sum_{i=1}^{k} S_k^2}} \quad (4)$$

Here S_k be the kth member of the cluster, S_{new} be 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.

E. Artificial Neural Network (ANN)

A supervised ANN has been widely used in areas of prediction. The wide range of applications of ANN in many fields and sectors are due to its power to model behavior to produce an approximation of given output [26, 27]. A three-layered perception have been established in our research to estimate student grade. We constructed the network model to each lesson. The structure of the ANN as shown in Fig.4. Layer 1 of each network, which is the input layer, consists of LSA results that characterize similarity between words. Layer 2 consists of 30 hidden neurons. The number of hidden neurons is chosen heuristically because 30 neurons in the hidden layer showed the least error during the training of the data set after combining data of two classes. Layer 3, the output layer, consists of 1 neuron denoting the student's grade (S, A, B, C, and D).

ANN is trained by Back Propagation (BP). The system is trained based on the principle of gradient descent learning. In our research we use WEKA machine learning program to build our network models. Each network weight is adjusted according to the presented input and error to the network by changing the learning rate parameter of error that is adjusted with the weight of presented input. After performance was examined, training time=10000 epochs showed the most predictive power in generalizing the problem with all lessons. The number of instances used in the training depend on the number of comments in each lesson.

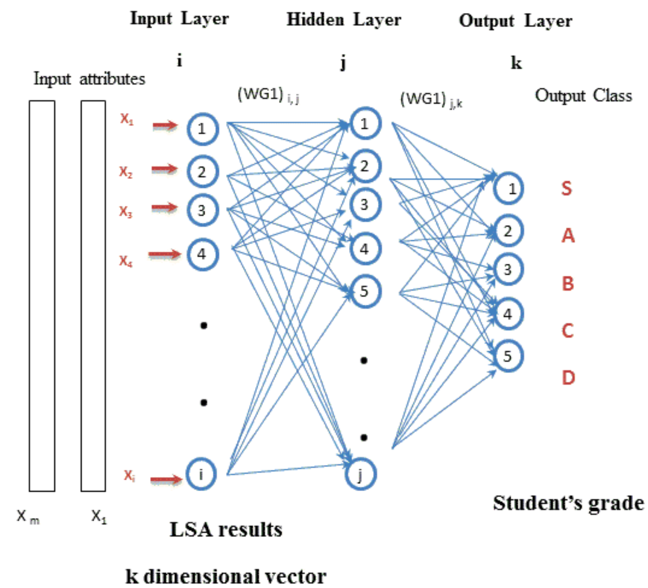


Fig. 4: The structure of the ANN.

The convergence rate between the actual and desired output is achieved by, 30 hidden nodes, 0.1 learning rate, 0.85 momentum coefficient and 10.000 epochs.

V. EXPERIMENTAL RESULTS

In this section, we consider to predict students' final grade from their comments. We evaluate the prediction accuracy by 10-fold cross validation. We run evaluation experiments by calculating Accuracy and F-measure by the following :
Let G be 5 grade category S, A, B, C, 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 $1 \leq i \leq n$, 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 category for student s_i; $!pred(s_i)$ returns a complement of $pred(s_i)$. TP, FP, FN and TN are defined as follows:

TP = $\{s_i | obs(s_i, pred(s_i)) = 1\}$,
FP = $\{s_i | obs(s_i, pred(s_i)) = 0\}$,
FN = $\{s_i | obs(s_i, !pred(s_i)) = 1\}$,
TN = $\{s_i | obs(s_i, !pred(s_i)) = 0\}$.

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

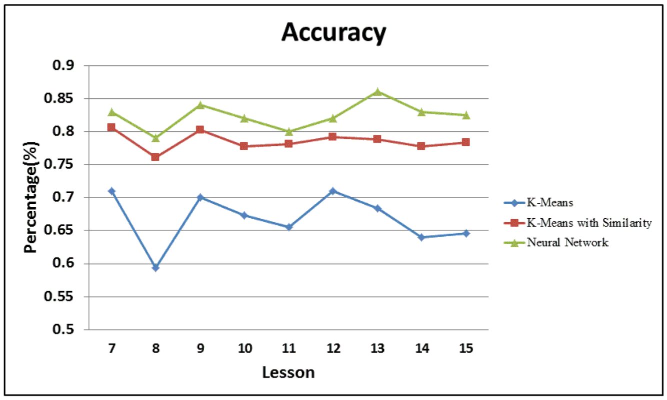
Recall = $\frac{TP}{TP+FN}$

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

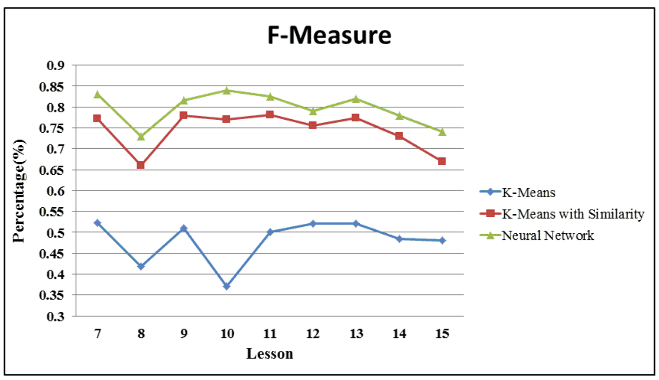
Accuracy = $\frac{TP+TN}{TP+TN+FP+FN}$

TABLE III: Overall prediction results

Models	Precision	Recall	F-Measure	Accuracy
K-Means	0.452	0.536	0.480	0.664
Similarity Measuring	0.631	0.697	0.661	0.785
Neural Network	0.787	0.654	0.761	0.826



(a) Accuracy results from lesson 7^{th} to15^{th}.



(b) F-measure results from lesson 7^{th} to15^{th}.

Fig. 5: Prediction results by lesson.

A. Overall prediction results(Accuracy / F-measure)

We considered to predict each student's results from his/her comments with high accuracy. We applied ANN to comments vectors to predict if the student s_i got the grade S, A, B, C or D based on the comments vectors. In addition, we present the results of similarity measuring method and compare our results with the results obtained from Sorour et al. [5].

With regard to students' grades prediction performance, the results revealed that the accuracy of students' responses in ANN and similarity measuring method was 82.6% and 78.5%, respectively. Table III shows the overall prediction results among ANN, similarity measuring method, and K-means clustering method. Fig.5 tells the details of difference in accuracy and F-measure in each lesson. We see that the results of ANN are the best ones between similarity measuring method and K-means method. According to the accuracy results from Fig.5 (a), we can distinguish the overall accuracy results in each lesson. The prediction accuracy results between 71.0% and 76.0% by the similarity measuring method , 79.0% to 86.0% for ANN model and 59.0% to 71.02% for K-means clustering method.

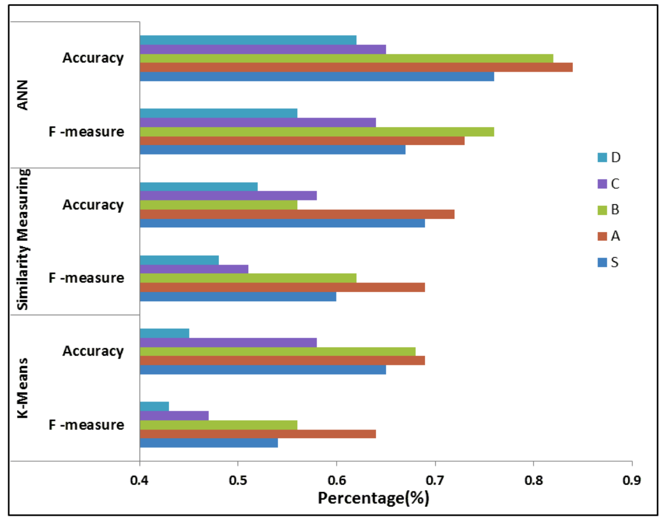


Fig. 6: Prediction of grade by ANN.

By comparing our results with Sorour et al., we can see that our methods have better accuracy to predict students' grades in all lessons than K-means clustering method.
Through the analyzing results, we can see that, we have more similarity between comments data and students' grades specially in lesson 7^{th}, 9^{th} and 13^{th}. Although, the rate is not the same in three methods, we can see the highest level of accuracy and F-measure in the previous lessons.

Fig.5 (b) shows the average prediction F-measure results of 5 grades by comparing between three methods. The ANN model achieves the best results among 3 methods and the K-means clustering method has the lowest results. The prediction F-measure results were averagely 76.1% (for ANN model), 66.1% (by similarity measuring method) and 48.0% (with K-means clustering method).

B. Prediction of Grade by ANN

This section displays the average overall results between C-comments from lesson 7 to lesson 15, and the prediction results (Accuracy / F-measure) in each grade. Fig.6 displays the difference in each grade, we can distinguish higher grade groups A and B from lower ones C and D. In addition the results of grade S not higher or lower than other grades, but performed well with ANN and similarity methods than K-means. Predicting students' grades by ANN has the highest accuracy and F-measure than similarity measuring and K-means methods. Moreover, the accuracy and F-measure results for grade C and D although they have the lowest results, the results become better with ANN model than other methods.

According to Fig.6, ANN model has the highest results for predicting students' grades in all lessons. Although the difference between ANN and similarity measuring method was (4.2%) in average overall accuracy results and the difference between F-measure was (10%), the prediction results in grade B, C and D are better than similarity method.

TABLE IV: Prediction results for class A and class B

Precision	Recall	F-Measure	Accuracy
0.845	0.732	0.823	0.863

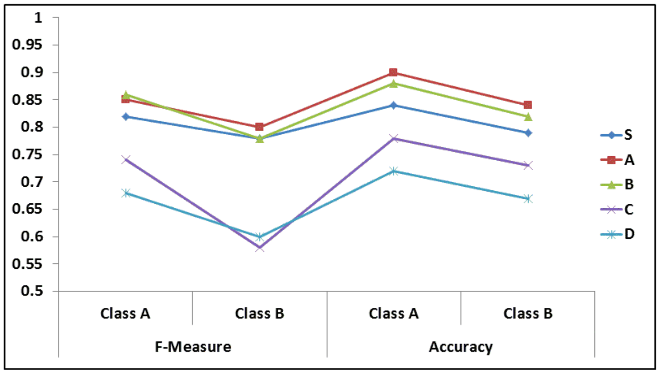


Fig. 7: Prediction results for class A and class B.

C. Class A and Class B

To further clarify, if there is any difference between two class data about their effect on the prediction results, we conducted experiments in each class by using the same results from LSA. We followed the previous approach and created neural network model in each class. We have established network model as we mentioned previously: number of neuron in hidden layer =15 neurons, 0.1 learning rate, 0.85 momentum coefficient and training time= 1.000 epochs.

We compared two class data by calculating the average accuracy and F-measure. Table IV displays the overall prediction results after analyzing each class. We evaluate the prediction performance by 5-fold cross validation in each class data.

Fig.7 displays the difference between two class data. class A has the higher results than class B according to F-measure results, but the rate not higher than accuracy results. In addition, the results of grade S, A and B are higher than grade C and D.

From the previous results, we can say that, there is difference between two class data. Although students learn the same subject in two classes, there exist difference between writing comments data in the two classes.

D. The effect of correctly classified instances

To study the capacity of machine learning technique and LSA. To detect difference between ANN and similarity measuring methods. *t-test* was carried out between number of correctly classified instances in the test data between 9 lessons, from lesson 7 to 15. We compared between ANN and similarity measuring methods. Table V shows the obtained results after applying *t-test*. There's statistically significant difference for ANN, where $P \leq 0.05$. For more clarity, we evaluate the TP (True Positive) rate to each grade. Table VI shows the *t-test* results between two methods (ANN and similarity).There's statistically significant difference $p \leq 0.05$ between grade A

TABLE V: Correctly Classified Instances

Model	Std.Deviation	Mean	S.E Mean	F	t	*p*
ANN	6.30	71.78	2.10	2.20	2.35	0.03
Similarity	5.32	65.33	1.77			

TABLE VI: Correctly classified instances by grade

Grade	ANN Mean SD	sim Mean SD	F	t	*p*
S	78.1 *7.2*	74.6 *7.9*	0.14	1.03	0.32
A	76.0 *8.14*	58.8 *6.72*	0.35	2.28	0.04*
B	61.4 *11.47*	50.2 *7.98*	1.13	2.40	0.03*
C	63.4 *6.29*	57.4 *6.22*	0.27	2.06	0.06
D	54.3 *6.89*	49.8 *5.16*	0.97	1.56	0.14

and grade B for ANN and no statistically significant difference was found between S, C and D grades.

E. Number of dimensions

The main difficulty in the application of LSA to comments classification is the high dimensionality of the input feature space which is typical for textual data. This is because each unique term in the vocabulary represents one dimension in the feature space, so that the size of the input of the neural network depends upon the number of stemmed words [23]. Although we selected 4 dimensions in our experiments and the results of the first four K- dimensions vector (KDV) can strongly predict student's grade, we checked the accuracy prediction results from 2 to 100 dimensions by ANN model, to clarify whether increasing number of dimensions can reflect their prediction performance or not. Fig.10 shows the average accuracy between all lessons with the different number of dimensions of LSA results. According to Fig.8, we can see the

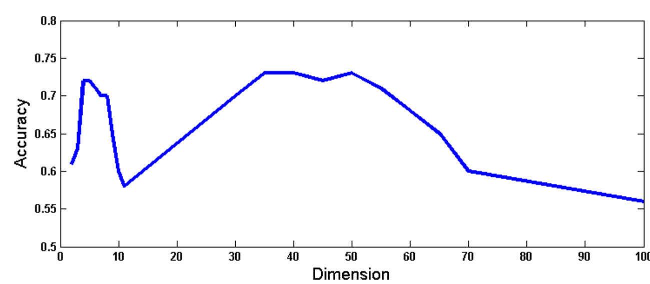


Fig. 8: Number of dimensions.

highest range from the dimension 2 to 10 and from 30 to 40. By scanning the highest range, we found the highest dimensions results exist in 4 and 35. Table V shows the average overall prediction results between all lessons by 4 and 35 dimensions. Although the results become higher with 35 dimensions than 4 dimensions, there's no statistically significant difference was found between two dimensions. Where F (3.285)= 7.34, p>0.05. We compare the correctly classified instances for all lessons by the obtained results from weka model.

TABLE VII: Number of dimensions

K-dimension vector	Precision	Recall	F-Measure	Accuracy
4	0.787	0.654	0.761	0.826
35	0.845	0.753	0.813	0.845

VI. CONCLUSION AND FUTURE WORK

Text classification is regarded as one core technical component of knowledge management systems because it can support us to handle explicit knowledge more systematically.
In this paper, we proposed student's grade prediction methods based on their free-style comments by applying Artificial Neural Network (ANN) and latent semantic analysis (LSA) techniques. The introducing of latent semantic analysis not only reduces the dimensions drastically, but also overcomes the problems existing in commonly used vector space model method for text representation and the categorization performance has been improved [21, 22].

On the other side, ANN has been successfully applied to solve a variety of classification and function approximation problems [27, 28]. We conducted our experiments with LSA technique and ANN model to reveal the high potential of comments data for predicting students' performance. In addition, We propose a new method that calculates similarity between a new comment and comments in each cluster. In near future, we will try to investigate such words that are clues for improving students' performance, or for judging learning problems. Also, we will collect new comments in order to improve students' performance. We believe this will help a teacher give advice to students and improve their performance. Finally this model is a step to improve students' performance. If we can collect their comments with high quality that include important contents and their attitudes in each lesson, we believe their comments would help a teacher to estimate their learning situation more precisely and correctly, to give advice appropriate to them, and to improve their performance as a result. In addition, students' comment evaluation must be continued throughout the semester to give more advice to them.

REFERENCES

[1] Barnett C. W., Matthews, H. 2009. " Teaching evaluation practices in colleges and schools of pharmacy" *Am J Pharm Educ*vol. 73 (6): Article 103.

[2] Luo, L. 2011. " Fusing research into practice: The role of research methods education" *Journal of Library and Information Science Research*,Vol.33, pp. 191-201.

[3] Goda, K., Mine, T. 2011. " Analysis of Students Learning Activities through Quantifying Time-Series Comments." *Proc. KES 2011*, Part II (LNAI 6882).

[4] Goda, K., Hirokawa, S.,Mine, T. 2013. " Correlation of Grade Prediction Performance and Validity of-Self-Evaluation Comments" *Proceedings of the SIGITE 2013 The 14th Annual Conference in Information Technology Education*, pp. 35-42, October 2013.

[5] Sourour, S., Goda, K., Hirokawa, S.,Mine, T. 2014. Ffficiency of LSA and K-means in prediction students' Academic Performance Based on Their Comments Data" *Proceedings of the CSEdu 2014 International Conference on Computer Supported Education*, pp. 63-74, April 2014.

[6] Baker, R., Yacef, K. 2009. "The state of educational data mining in 2009: A review and future visions," *Journal of Education Data Mining*Vol.1, No.1, pp. 3-17.

[7] Baker,R. 2009. "Data Mining for education." In B. Mcgaw, P. Peterson, Baker (Eds.)*International encyclopedia of education (3rd ed.. Oxford, UK: Elsevier)*.

[8] Romero, C., Espejo, P. G., Zafra, A., Romero, J. R.,Ventura, S. 2010. " Web usage mining for predicting final marks of students that use Moodle courses." *Computer Applications in Engineering Education.*

[9] Delavaria, N., Phon-Amnuaisuka, S., Reza Beikzadehb, M. 2008. " Data mining application in higher learning institutions. " *Informatics in Education*,Vol. 7(1),31.

[10] Gorissen, P., Bruggen, J., Jochems, W. 2012. " Usage reporting on recorded lectures using educational data mining" *Usage reporting on recorded lectures using educational data mining,International Journal of Learning Technology*, 7(1), pp. 23-40 .

[11] Jovanovica, M., Vukicevica, M., Milovanovica, M., Minovica, M. 2012. " Using data mining on student behavior and cognitive style data for improving e-learning systems": A case study. *International Journal of Computational Intelligence Systems*,5(3), pp. 597-610.

[12] Parack, S., Zahid, Z. Merchant, F. 2012. " Application of data mining in educational databases for predicting academic trends and patterns": A case study. *In Proceedings of 2012 IEEE international conference on technology enhanced education (ICTEE)* pp. 1-40.

[13] He,W. 2013. " Examining students' online interaction in a live video streaming environment using data mining and text mining" *Journal of Computers in Human Behavior*,Vol. 29, pp. 90-102 .

[14] Ananiadou, S. 2008. National centre for text mining: Introduction to tools for researchers.
http://www.jisc.ac.uk/publications/publications/

[15] Abdous, M., Yen, C., J. 2010. " 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, pp. 248-257.

[16] Tane, J., Schmitz, C., Stumme, G. 2004. " Semantic resource management for the web: An e-learning application. " *In Proceedings of the WWW conference, New York, USA,* 7(1), pp. 1-10.

[17] Hung, J. 2012. " Trends of E-learning research from 2000 to 2008: Use of text mining and bibliometrics " *British Journal of Educational Technology,*,Vol.43 (1), pp. 5-16.

[18] Landauer, T. K., Dumais, S. T. 1997. " A solution to Plato's problem: The Latent Semantic Analysis theory of the acquisition" *induction, and representation of knowledge, Psychological Review*, Vol. 104, pp. 211-240.

[19] Foltz, P. W., Laham, D. Landauer, T. K. 1999. "The Intelligent Essay Assessor: Applications to educational technology." *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*.

[20] Deerwester,S.C. Dumais, S. T. 1990. Landauer,T.K. ,Furnas, G.W. ,Harshman R.A. ," Indexing by latent semantic analysis" *Journal of the American Society of Information Science*, Vol. 41(6) pp. 391-407.

[21] Berry, M. W., Dumais, S. T. OBrien, G. W., Samarin, F. 1995. *Using linear algebra for intelligent information retrieval*, SIAM Review ,Vol. 37(4).

[22] Dikli, S. 2007. "An Overview of Automated Scoring of Essays" *Journal of Technology, Learning, and Assessment,*Vol. 5(1).

[23] Yany, Y. 1995. " Noise reduction in a statistical approach to text categorization" *Proceedings of the ACM 1995 the 18th International Conference on Rexorch ond Development in Information Retrieval*, pp. 256-263.

[24] Zelikovitz,S., Hirsh, H. 2001. " Using LSI for text classification in the presence of background text" *Proceedings of the ACM 2001 the tenth international conference on Information and knowledge management*, pp. 113-118.

[25] Zaiane,O. 1999. " Principles of Knowledge Discovery in databases" *chapter 8. Data Clustering lecturing slides for CmPUT 690*, University of Alberta.

[26] Lippman, R. 1987. *An Introduction to Computing with Neural Nets*,Addison-Wesley, Vol. 4, IEEE Trans. ASSP Magazine.

[27] Andrews, R., Diederich, J., Tickle, a. B. 1995. "Survey and critique of techniques for extracting rules from trained artificial neural networks." *Knowledge Based System*, Vol. 8, pp. 373-389.