Journal of Information Processing Vol.23 No.6 784–794 (Nov. 2015)

[DOI: 10.2197/ipsjjip.23.784]

Regular Paper

Evaluation of Effectiveness of Time-Series Comments by Using Machine Learning Techniques

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

Received: January 9, 2015, Accepted: July 1, 2015

Abstract: Understanding individual students more deeply in the class is the most vital role in educational situations. Using comment data written by students after each lesson helps in the understanding of their learning attitudes and situations. They can be a powerful source of data for all forms of assessment. The PCN method categorizes the comments into three items: P (Previous learning activity), C (Current learning activity), and N (Next learning activity plan). The objective of this paper is to investigate how the three time-series items: P, C, and N, and the difficulty of a subject affect the prediction results of final student grades using two types of machine learning techniques: Support Vector Machine (SVM) and Artificial Neural Network (ANN). The experiment results indicate that the students described their current activities (C-comment) in more detail than previous and next activities (P- and N-comments); this tendency is reflected in prediction accuracy and *F*-measure of their grades.

Keywords: comment data mining, student grade prediction, PCN method, LSA, ANN, SVM

1. Introduction

Knowledge Management (KM) has increased in popularity and credibility as a management tool, as well as a research discipline, over the past decade [5]. KM in education is the monograph that makes eminent sense about a wonderful combination of good intuition, practical know-how, and a feel for what might be best described as a set of emerging theories focusing on the effective management of knowledge in educational institutions. In addition, KM in education supplies a framework for understanding how good assessment practice depends on effective information management [16].

Classroom assessment is the knowledge and skill necessary for compiling data about students' achievement and for effectively utilizing the assessment process and outcomes to develop and improve the quality of instruction of teachers and learning of students [4].

Assessment benefits both teachers and students in a number of ways: 1) it yields data that can be used to improve the appropriateness of teachers' teaching, 2) it enables teachers to monitor students' learning throughout the year and to improve students' learning before year-end assessment, 3) it provides teachers with data to use in selecting teaching methods that are suitable for each group of students, 4) students can use the data from the assessment and feedback to improve their knowledge and understanding, 5) students have the chance to develop or improve their self assessment ability and consider assessment as part of the learning process, and 6) it helps students make decisions about how they can acquire knowledge and skills [23]. In addition, classroom assessment yields important data for teachers regarding students' learning, which leads to further development and improvement of teachers' instruction and revision of curriculum content to better serve the students' needs, enabling them to learn efficiently and effectively [17]. Thus, classroom assessment is an important method for developing the quality of students.

Teachers who have sufficient background knowledge about assessment are able to integrate different assessment methods into learning and to use an instructional format that is suitable for students. On the other hand, using many sources of evidence help teachers accurately interpret what each student really knows and can do.

Using traditional paper-and-pencil tests (e.g., multiple-choice and short-answer) and informal day-to-day measures of student progress such as observation and questioning strategies can help to interpret student performance and give a comprehensive assessment. However, the instructor lacks a comprehensive view of each student in the classroom. In fact, even in classroom courses with a small number of students, there could be thousands of messages and instructions generated in each lesson, the instructor is faced with the difficulty of interpreting and evaluating learning situations. Evaluating students in such a case is very difficult, considering that current learning environments do not provide many indicators or information regarding the structure of interactions between students and teachers [8], [24]. A solution to this problem is the use of quantitative and qualitative evaluation to understand and grasp each student's performance in the classroom over the whole period of the semester.

[1] Faculty of Specific Education, Kafr Elsheik University, Kafr Elsheikh, Egypt
[2] Graduate School of Information Science and Electrical Engineering, Fukuoka 819–0395, Japan
[3] Kyushu Institute of Information Science, Dazaifu, Fukuoka 818–0117, Japan
[4] Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819–0395, Japan
[a)] shaymaa@ma.ait.kyushu-u.ac.jp
[b)] gouda@kiis.ac.jp
[c)] mine@ait.kyushu-u.ac.jp

© 2015 Information Processing Society of Japan

Analyzing free-style comment written by students has some benefits for student assessment, such as understanding students' behaviors, attitudes and situations, reflecting their activities and difficulties of learning in each lesson. Comment data enables student interactions, especially for the students with an introvert character, and helps them feel less threatened about expressing their views or asking questions. In addition, it synchronously allows teachers to develop monitoring of assessment tasks.

To further contribute to the understanding of student learning situations and to enhance individualized feedback to them, this paper presents new methods to predict student grades by comparing their comment data from the point of view of three time-series items: P, C, and N from the PCN method [9], [10]. The current study aims to estimate and assess the unknown value of student performance through predicting their final grades using comment data mining methods.

In this study, we use Latent Semantic Analysis (LSA) to grasp student learning attitudes and situations. LSA constructs a conceptual vector space in which each comment is represented as a vector in the space. It not only greatly reduces the dimensions, but also uncovers the important associative relationship between comments. We create prediction models based on comments analyzed by LSA using ANN and SVM models.

The experiments are conducted using data obtained from 15 lessons in two classes. The difficulty of the subject in each lesson affects student attitudes to expressing their behavior and sometimes does not give the students leeway to write comments. Therefore predicting student grades using their comments is a challenging problem.

1.1 Research Questions

The major research question in this study is to reveal the high prediction results from comment data. The results are measured by recall, precision, F-measure and accuracy. Many parameters will affect the prediction results. This paper reports those that largely impact the analysis of comment data. The following are the research questions investigated in this paper.

- **Question 1**: Are there any differences between lessons for predicting student grades from their comments with the three viewpoints: P (Previous learning activity), C (Current achievement activity) and N (Next activity plan) ?
- **Question 2**: Which machine learning technique can obtain better prediction results, ANN or SVM ?
- **Question 3**: Are there any differences between higher grade students (S, A or B) and lower ones (C or D) in predicting their grades ? If so, what causes the differences ?
- **Question 4**: Are there any clues to explain the prediction results obtained in each lesson?
- **Question 5**: Are there any relationships between the difficulty of a subject and prediction accuracy of student grades?
- **Question 6**: Are there any differences between two class data (Class A and Class B) in predicting student grades?

The rest of the paper is organized as follows: Section 2 discusses some related work. Sections 3 and 4 describe the procedure and the methodology of our proposed methods. Sections 5 and 6 display and discuss some of the highlighted experiment results. Finally, Section 7 concludes the paper and describes our future work.

2. Related Work

Predicting student performance is one of the most useful applications of Educational Data Mining (EDM) and its goal is to estimate student performance, knowledge, score or mark from other information, aspects or behavior of those students [19]. This is a difficult problem to solve due to the large number of factors or characteristics that can influence student performance, such as demographic, cultural, social, or family factors, socioeconomic status, psychological profile, previous schooling, prior academic performance, interactions between students and the faculty, etc. [2]. Predicting student performance has been studied with different techniques: classification (when the predicted variable is a categorical value), regression (when the predicted variable is a continuous value) or density estimation (when the predicted value is a probability density function) [13]. It is also important to notice that most of the current research on the application of EDM for predicting student performance has been applied primarily to the specific data and there are only a few studies about how to use text mining techniques to analyze learning related data [13], [20]. For example, Minami et al. [15] analyzed student attitudes toward learning, and investigated how they affect their final 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 students' own evaluation and lectures based on a questionnaire. They 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. Also, Rodrigues et al. [18] proposed a system for assessment of free-text answers. Their main goal is to design a system to work as a formative assessment tool for students and to help teachers creating and assessing exams as well as monitoring student progress. The system automatically created training exams for students based on questions from previous exams and assisted teachers in the creation of evaluation exams with various kinds of information about student performance. The system automatically assessed training exams to give automatic feedback to students. The correction of free-text answers was calculated based on the syntactic and semantic similarity between the student answers and various reference answers defined by the teacher concerning parts of the answer or its sub goals. The results indicated that there was a good correlation between the evaluation of the instructors and the evaluation performed by the proposed system. Dringus et al. [7] demonstrated a strategy for embedding data /text mining techniques to extract temporal information from a threaded discussion forum. They provided a strategy for assessing discussion forums in a manageable way, developed an assessment tool set that could be embedded in a threaded discussion forum, and pointed out the complexity and inconsistence inherent in a natural language text. In addition, Romero et al. [20] proposed the use of different data mining approaches for improving prediction of final student performance, starting from participation indicators in both quantita-

© 2015 Information Processing Society of Japan

Table 1 Examples of comments written by students.

Viewpoint	Comment data from Lessons 1 to 6
P	- I have logged in Web CT at home, and prepared lessons before class according to the materials.
C	- The way of input is quite difficult; I feel that I could hardly follow all the steps.
N	- I would like to practice again, because I'm not so good in designing slides.
	Comment data from lessons 7 to 15
P	- I read the text of programming for the next lesson, but I didn't understand.
C	- I have tried to learn and practice programming language, and I'm very glad that I managed to follow the lesson.
N	- I recognized that I should do exercise not only in mind, but also by hand.

tive, qualitative and social network forums. Their objective was to determine how the selection of instances and attributes, the use of different classification algorithms and the data gathered affect the accuracy and comprehensibility of the prediction. A new Moodle's module for gathering forum indicators was developed and different executions were carried out. The results indicated the suitability of performing both a final prediction at the end of the course and an early prediction before the end of the course, of applying clustering plus class association rule mining instead of traditional classification for obtaining highly interpretable student performance models, and of using a subset of attributes instead of all available attributes, and not all forum messages but only student messages with content related to the subject of the course for improving classification accuracy.

Previous studies show that we need to understand individual students more deeply, and recognize students' characteristics and attitudes to give feedback to them. Also, we need to comprehend students' characteristics by letting them describe themselves, their learning situations, such as understanding of subjects, difficulties of learning, learning activities in the classroom, and their attitudes toward the lesson.

Different from the above studies, Goda et al. [9], [10], 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. Goda et al. [9], [10] collected comment data from students to analyze and estimate their learning situation. While describing comments, the students can reflect on their learning attitudes or behaviors. Therefore, they call the student comments as free-style comments with their self-reflection or self-evaluation comments.

However Goda et al. [9], [10] did not discuss the prediction of final student grades. Sorour et al. [22] proposed a method based on the C-comment from the PCN method. They used the LSA technique and the K-means clustering method. They conducted their experiments from lessons 7 to 15 by combining comment data from the two classes. To improve the prediction accuracy results, they proposed similarity measuring and overlap methods based on their previous method.

In this paper we propose new methods to improve the prediction results of final student grades; the new methods use two machine learning techniques: ANN and SVM, and analyze comment data from the three viewpoints: P, C and N items.

Table 2 Student grade.

grade	S	A	B	C	D
Mark	100-90	89-80	79-70	69-60	59-0
#students	21	41	23	17	21

3. Overview of the Prediction Method

3.1 Subject of the Study

In this study, we used the same comment data as Sorour et al. [22]. They were collected from Goda's courses consisting of 15 lessons. The main subject from lessons 1 to 6 of the course is computer literacy, giving information on how to use some IT tools. From lessons 7 to 15, students learn the basics of programming. The main subject in those lessons is introductory C-programming [10].

In the classroom, the teacher had 90 minutes in each lesson. He organized the lesson time as follows:

- The first 45 to 60 minutes: The teacher taught the lesson subject.
- The last 30 to 45 minutes: He gave the students some questions to answer or practical exercises like writing a program or solving a problem that was related to lesson objectives.

Table 1 shows examples of the PCN comments written by students from lessons 1 to 6 and lessons 7 to 15, where original ones were described in Japanese.

The assessment of each student was done by considering the average mark of three assigned reports, and his/her attendance rate. In this research, we chose five grades instead of the mark itself as a student result to predict his performance from his/her comments. **Table 2** shows the correspondence between each grade and the range of the marks.

Although we have two class data in each lesson, we combined them to increase the number of comments in each grade [*1]; some students didn't submit their comments because they did not write any comments or were absent. **Table 3** displays the real number of comments in each lesson that we analyzed. 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. **Figure 1** displays the number of students who did not submit their comments from lessons 1 to 15 with the three viewpoints: P, C and N. It can be seen that there are differences between the number of P, C and N comments written by students. Also, we can show that the grade D has the greatest number of

[*1] Although more practical setting is to use the two class data separately, we in this paper focus on evaluating the improvement of our proposed methods by comparing with our previous results based on the two class data combined.

© 2015 Information Processing Society of Japan

Table 3 Number of comments from lessons 1 to 15.

Lesson	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
P	108	121	118	115	123	116	104	103	107	113	110	109	107	110	114
C	100	121	118	115	123	116	104	103	107	111	107	109	107	111	121
N	109	121	118	115	123	116	104	103	107	113	110	109	107	104	110

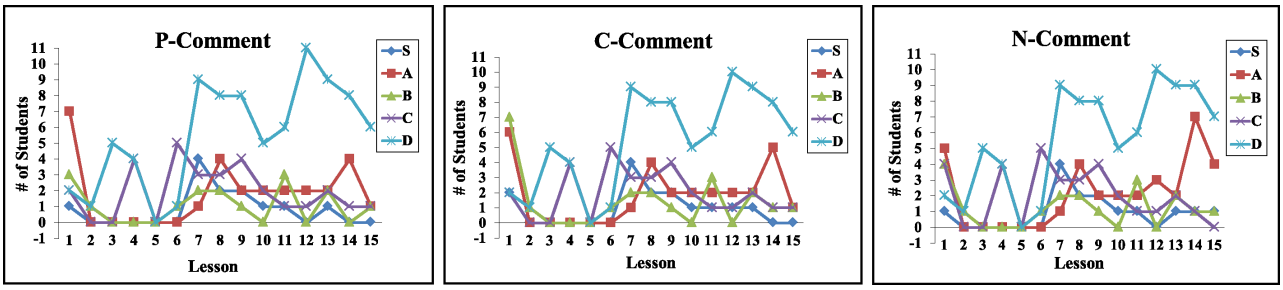


Fig. 1 The relation between the number of students who did not submit comments and their grades.

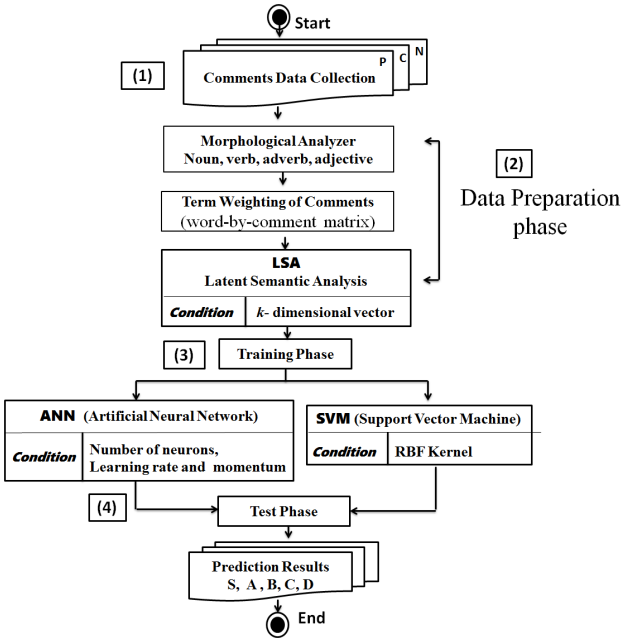


Fig. 2 Procedure of the proposed method.

students who didn't submit the comments. On the other hand, the grade S has averagely the smallest number of such students.

3.2 Procedure of the Prediction Method

Figure 2 displays the overall procedure of our proposed method; we have four phases:

(1) Comment Data Collection: This phase focuses on collecting student comments after each lesson. Comment data were collected from 123 students in two classes: (Class A = 60 students) and (Class B = 63 students), who took the introductory information processing course consisting of 15 lessons (weeks). Students write their comments using a form of triple: P, C and N.

(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 P, C and N comments, extract words and parts of speech with Mecab program [*2], which is a Japanese morphological analyzer designed to extract words and identify their parts of speech (verb, noun, adjective, and adverb).

(b) Calculate the occurrence frequencies of words in comments. We create a word-by-comment matrix with extracted words. This word-by-comment matrix, say A, is comprised of m words $w_1, w_2, \ldots, w_m$ 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, log entropy term weighting is applied to the original word-by-comment matrix, which is the basis for all subsequent analyses [14].

(c) Apply LSA to the word-by-comment matrix 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 Sections 4.1 and 4.2, respectively.

(3) Training Phase: This phase builds prediction models of student grades based on LSA results using the ANN and the SVM models. We chose the ANN and the SVM models because they are popular strategies for supervised machine learning and classification, and it's not clear which method is better for a particular problem. Although we tested the decision tree (C4.5) algorithm in the preliminary stage of our experiment to predict final student grades, the results were worse than SVM and ANN models. The details about the ANN and the SVM models are described in Sections 4.3 and 4.4, respectively.

(4) Test Phase: This phase evaluates the performance of prediction models by calculating the accuracy and the F-measure in each lesson.

To evaluate the prediction performance, each evaluation has been done for each lesson separately. Thus, we did not merge any comments that appeared in different lessons. 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 remaining 10% of comments as test data, and compared a predicted value corresponding with the original data. The procedure

[*2] http://sourceforge.net/projects/mecab/

© 2015 Information Processing Society of Japan

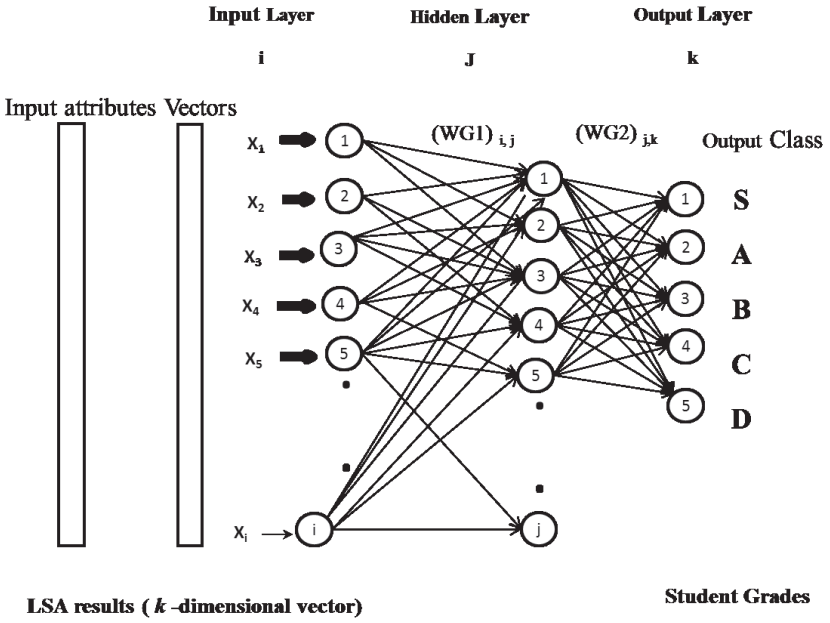


Fig. 3 The structure of the ANN.

was repeated 10 times and the results were averaged. The detail is described in Section 5.

4. Proposed Methodology

4.1 Semantic Vector Space Generation

Latent semantic analysis is a technique that projects the original high dimensional document vectors into a space with "latent" semantic dimensions. Once a term-by-document matrix is constructed, LSA requires the singular value decomposition (SVD) of this matrix to construct a semantic vector space which can be used to represent conceptual term-document associations, reduce the dimensions drastically and overcome the problems of lexical matching [6].

4.2 Latent Semantic Analysis for Our Methods

LSA is originally proposed as an information retrieval method. Nowadays, it is also widely used in text categorization [3], [6]. In our research, we use LSA to analyze patterns and relationships between the extracted words and latent concepts contained in an unstructured collection of texts (student comments) and detect noisy data that adversely affects the results by reducing the number of dimensions. Our objective is to establish a strong relationship between analyzed comments and student grades in each lesson.

4.3 Model Estimation by ANN

Supervised ANNs have been widely used in areas of prediction. The wide range of applications of the ANN in many fields and sectors is due to its power to model behavior to produce an approximation of given output [1].

A three-layered perceptron was established in our research to estimate student grades. We constructed a network model for each lesson. The structure of the ANN is shown in **Fig. 3**, layer 1 of each network, which is the input layer, consists of a k-dimensional vector of LSA results that characterize similarity between words. Layer 2 consists of one hidden layer; the number of neurons in the hidden layer is chosen heuristically because they showed the least error during the training of the data set with lessons. The number of the neurons established for all lessons by using the LSA method was between 30 and 40. The output layer, consists of 5 neurons denoting student grades: S, A, B, C and D. The total output y_k produced by the neuron can be summarized by

$$y_k = g'(a_k) \tag{1}$$

where g' is the activation function of output units and a_k is the total weight from the previous layer. The ANN was trained by back propagation (BP) [21] which was based on the principle of gradient descent learning. Each network weight will be adjusted according to the presented input and the error to the network as shown in Eq.(2):

$$w_{ij}(n+1)=w_{ij}(n)+\eta.e_i(n)x_j(n)+\alpha(w_{ij}(n)-w_{ij}(n-1)) \tag{2}$$

where η is a learning rate parameter of error e_i that is adjusted with the weight of the presented input x_i. The adjustment of the weight w_{ij} is the weight between processing elements (i) and (j) at iteration (n), and $x_j(n)$ is the presented value of the hidden layer at processing element (j). α is a momentum parameter. For the purpose of training data, we set the weight randomly for all input parameters. The weight values for updating η and α were 0.3 and 0.65, respectively for all lessons. Each network was trained with more than 10,000 iterations to determine the predictive power.

4.4 Model Estimation by SVM

SVM is a powerful solution to the classification problems. The main advantages of SVM used as a classifier are its extremely powerful learning procedure and its ability to lead to the global minimum of the defined error function [11]. In our research, we employed the SVM method with a radial basis function (RBF) kernel to generate models from lessons 1 to 15 and to predict a

© 2015 Information Processing Society of Japan

student grade as one of five grades: S, A, B, C, and D based on the results obtained by the LSA model. We used the MATLAB LibSVM tool [*3] as a library of SVM.

5. Experiment Results

This section will report prediction results of final student grade from lessons 1 to 15. Section 5.3 demonstrates the difference between lessons from the three viewpoints: P, C and N in predicting student grades using the SVM and the ANN models, and answers the research **Questions 1** and **2** described in Section 1.1. Section 5.4 answers research **Question 3** that explains the relationships between comment data and grade prediction results compared with higher and lower student grades. Section 5.5 displays the correlation between standard deviation (Sd) of prediction F-measure results and the prediction F-measure from lessons 1 to 15 with the three viewpoints: P, C and N comments; the correlation answers the research **Questions 3** and **4**. Sections 5.3 and 5.4 reveal if the difficulty of a subject affects the prediction results of final student grades; the results answer the research **Question 5**. Finally, Section 5.6 shows if there are any differences between Class A and Class B data from lessons 1 to 15; the results answer the research **Question 6**.

5.1 Evaluation Methods

A 10-fold cross-validation [12] approach is used to predict student grades. We 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 $1 \leq i \leq n$, 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 for student s_i; !pred (s_i) returns a complement of pred(s_i).

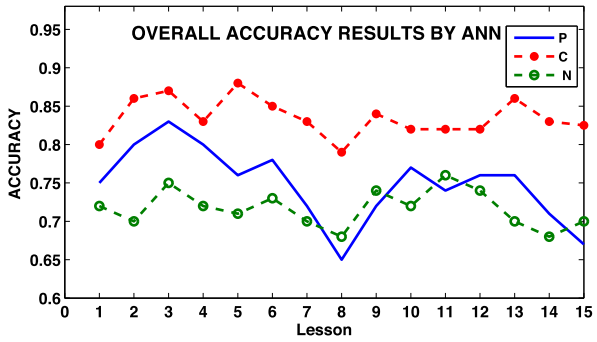
$$TP = \{s_i | obs(s_i, pred(s_i)) = 1\}$$
$$FP = \{s_i | obs(s_i, pred(s_i)) = 0\}$$
$$TN = \{s_i | obs(s_i, !pred(s_i)) = 1\}$$
$$FN = \{s_i | obs(s_i, !pred(s_i)) = 0\}$$
$$\text{Precision} = \frac{TP}{TP + FP}$$
$$\text{Recall} = \frac{TP}{TP + TN}$$
$$F\text{-}measure = 2 * \frac{(Precision * Recall)}{(Precision + Recall)}$$
$$\text{Accuracy} = \frac{TP + FN}{TP + TN + FP + FN}$$

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 pick up more 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 misdetect them more. So our study shows the prediction results by calculating Precision, Recall, Accuracy and F-measure using the ANN and the SVM models.

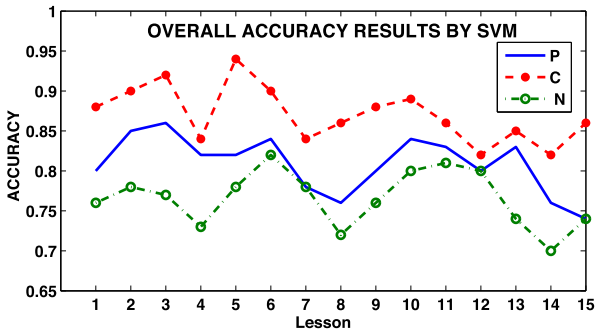
[*3] http://www.csie.ntu.edu.tw/cjlin/libsvm/index.html

Table 4 Number of dimensions.

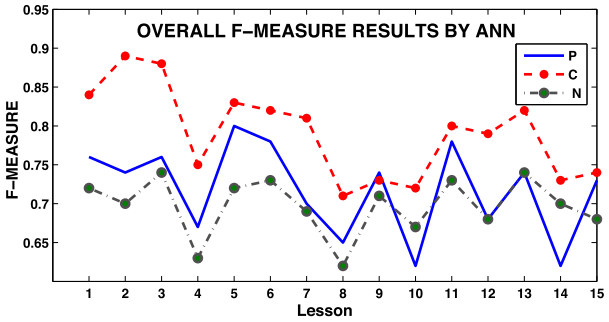
Viewpiont	ANN	SVM
P	8	12
C	4	8
N	8	10



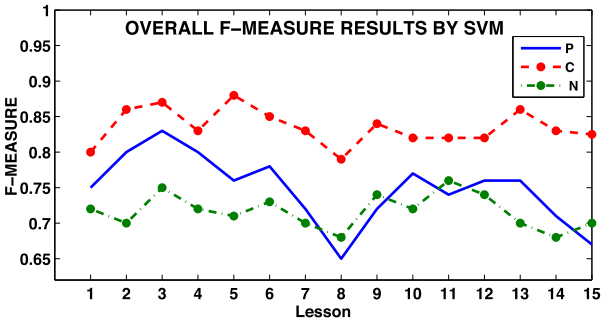
(a) Overall accuracy results after applying the ANN model.



(b) Overall accuracy results after applying the SVM model.



(c) Overall F-measure results after applying the ANN model.



(d) Overall F-measure results after applying the SVM model.

Fig. 4 Overall accuracy and F-measure results with the three viewpoints: P, C and N comments.

5.2 Number of Dimensions

The main difficulty of our application of using the LSA technique is to choose the number of dimensions k for the matrix A so as to predict student grades with high accuracy. In our research,

© 2015 Information Processing Society of Japan

Table 5 Overall prediction results.

Models		Precision	Recall	*F-measure*	Accuracy
ANN	P	0.668	0.774	0.717	0.749
	C	0.750	0.833	**0.788**	**0.830**
	N	0.654	0.763	0.702	0.726
SVM	P	0.762	0.788	0.772	0.802
	C	0.802	0.888	**0.842**	**0.868**
	N	0.732	0.762	0.743	0.765

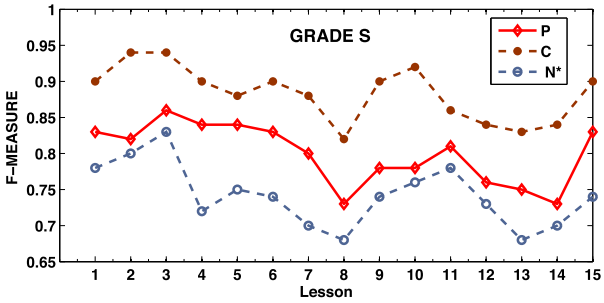
we checked the *F*-measure and the accuracy of prediction results from 2 to 50 dimensions using ANN and SVM models. We chose the highest *F*-measure prediction results as reducing the size of dimensions. **Table 4** shows the number of dimensions that have been chosen for P, C and N comments.

5.3 Overall Prediction Results (Accuracy / *F*-measure)

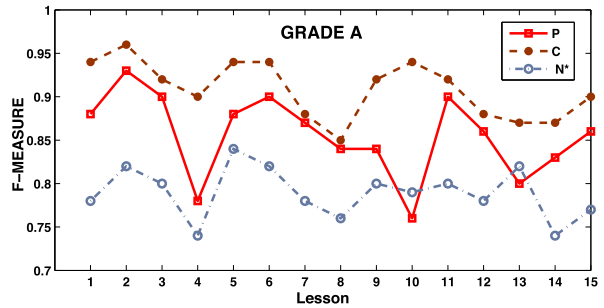
This section discusses the overall prediction results of final student grades with the three viewpoints: P, C and N comments. The objectives in this section are to answer the research Questions 1 and 2 in Section 1.1, by finding which item and machine learning technique (ANN or SVM) gets the best prediction results, and to discover whether the prediction accuracy and *F*-measure results will outperform the previous research [22]. **Figure 4** and **Table 5** display the average prediction *F*-measure and accuracy results using the ANN and the SVM models. We applied LSA to comment data from lessons 1 to 15 with three items.

The prediction accuracy results after employing ANN were between 65.0% and 82.4% for the P-comment, from 79.2% to 88.4% and from 68.5% to 76.3% for the C- and the N-comments, respectively. On the other hand, the accuracy and the *F*-measure of the prediction results increased using the SVM model. The accuracy results achieved from 75.0% to 86.2% from 82.3% to 93.7%, and from 69.7% to 81.0% for the P-, the C-, and the N-comments, respectively. From Fig. 4 we can show the differences among the P, the C-, and the N-comments. The C-comments had the highest prediction results; students described their current activities better than previous and next activities. Also, the N-comments had the lowest results using ANN and SVM; students didn't describe well their plans concerning the next lesson. In addition, the SVM model had higher prediction results than the ANN model in all lessons. (See Table 5). Also, we can see that the average overall prediction results from lessons 1 to 6 were higher than from lessons 7 to 15 in the most of the lessons. The highest accuracy/F-measure results from the top were obtained in lessons 3 and 5, and the lowest ones from the bottom in lessons 8 and 14. Lesson 4 has the lowest results from lessons 1 to 6.

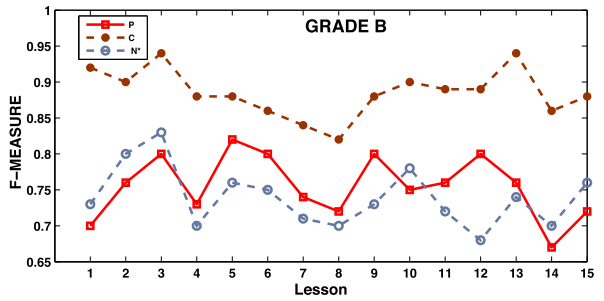
By comparing our results with Sorour et al. [22], we find that their method achieved an average 66.4% prediction accuracy using the k-means based methods with C-comment from the PCN method. They conducted their experiments only from lessons 7 to 15. Although the scores were increased to 73.6% and 78.5% by adding the overlap method and the similarity measuring method, respectively, our methods outperformed their methods as shown in Table 5. The average prediction accuracy results of final student grades for C-comments were 83.0% and 86.8% using the ANN and the SVM models, respectively.



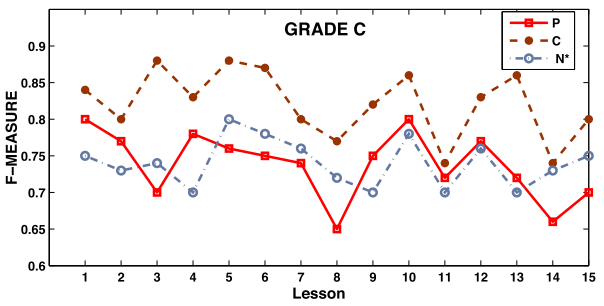
(a) Prediction results for grade S



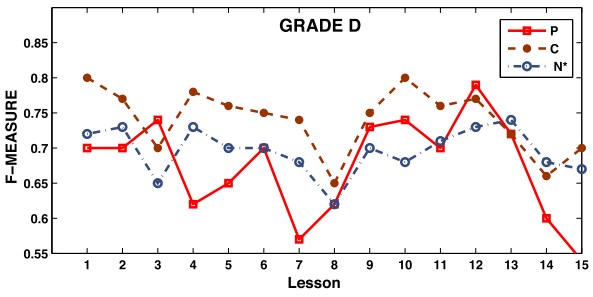
(b) Prediction results for grade A



(c) Prediction results for grade B



(d) Prediction results for grade C



(e) Prediction results for grade D

Fig. 5 Overall prediction results (*F*-measure) by grade using the SVM model.

© 2015 Information Processing Society of Japan

Table 6 Overall prediction results by grade.

		P				C				N			
Grade	Method	Precision	Recall	*F*-Measure	Accuracy	Precision	Recall	*F*-Measure	Accuracy	Precision	Recall	*F*-Measure	Accuracy
S	ANN	0.725	0.822	0.767	0.827	0.825	0.892	**0.857**	**0.876**	0.703	0.833	0.748	0.743
	SVM	0.733	0.892	0.802	0.864	0.852	0.924	**0.884**	**0.934**	0.785	0.726	0.749	0.806
A	ANN	0.675	0.873	0.757	0.843	0.861	0.932	**0.895**	**0.909**	0.761	0.826	0.789	0.825
	SVM	0.824	0.934	0.872	0.865	0.853	0.965	**0.902**	**0.925**	0.784	0.804	0.789	0.827
B	ANN	0.661	0.734	0.679	0.763	0.744	0.854	**0.795**	**0.823**	0.576	0.723	0.645	0.736
	SVM	0.791	0.943	0.758	0.885	0.795	0.734	**0.859**	**0.879**	0.790	0.671	0.725	0.822
C	ANN	0.671	0.786	0.721	0.765	0.671	0.751	0.709	**0.786**	0.672	0.743	0.703	0.691
	SVM	0.824	0.754	0.747	0.763	0.854	0.842	**0.819**	**0.864**	0.686	0.847	0.752	0.745
D	ANN	0.624	0.743	0.658	0.613	0.653	0.721	**0.684**	**0.699**	0.563	0.722	0.635	0.656
	SVM	0.675	0.693	0.679	0.725	0.723	0.784	**0.748**	**0.776**	0.630	0.784	0.697	0.652

Table 7 Sd of prediction F-measure for the ANN and the SVM methods.

Model	Lesson	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
ANN	P	3.34	2.45	3.56	6.54	3.03	4.56	5.43	4.54	3.45	4.67	2.54	3.67	3.03	6.65	4.34
	C	2.13	2.41	2.44	5.11	3.20	2.05	3.72	5.12	5.87	4.91	2.76	4.76	2.65	6.76	3.43
	N	4.56	5.43	3.45	5.55	4.56	3.66	5.65	8.98	7.67	4.56	4.65	5.67	3.54	5.67	3.43
SVM	P	4.53	2.55	4.06	5.76	4.22	2.03	6.56	6.35	4.03	2.17	5.96	3.06	4.76	5.22	4.03
	C	1.34	2.01	1.44	3.11	2.20	1.45	3.02	4.12	2.81	1.65	2.02	3.60	5.61	5.52	3.04
	N	3.84	5.29	2.11	6.75	3.79	2.82	6.81	6.71	7.73	2.72	5.82	2.79	3.82	6.737	5.24

5.4 Correlation of PCN Comments with Grade Prediction Performance

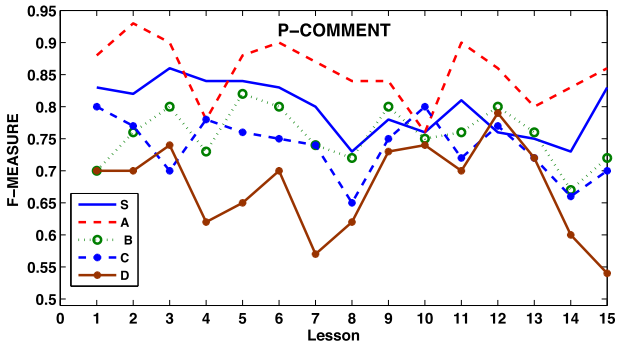
This section explains the relationships between comment data and grade prediction results and answers the research Question 3 in Section 1.1. Whether there are any differences between higher grades and lower grades on their prediction results with the P, C and N-comments, using the ANN and SVM models. **Figure 5** shows there are differences between P-, C- and N-comments on prediction F-measure of final student grades and the C-comments had the best prediction F-measure among the three types of comments. **Figure 6** displays the results from the point of view of five grades: S, A, B, C and D and shows there are clear differences between higher grades: S, A and B from lower ones: C and D. Grade A had the highest results and grade D had the lowest results with the P-, the C- and the N-comments. The details of the results are shown in **Table 6**.

5.5 Correlation between Standard Deviation and Prediction F-measure

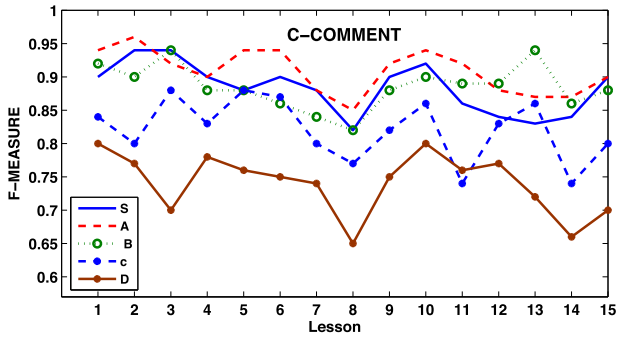
The aims in this Section are to discover whether there are any correlations between Sd of prediction F-measure and prediction F-measure results, and if there are any influences on the correlations from the differences between lesson subjects and between the prediction models built by the SVM and the ANN models. These answer the research Questions 4 and 5 in Section 1.1, respectively. **Table 7** displays the Sd results of the prediction F-measure from lessons 1 to 15, using the ANN and SVM models with the three viewpoints: P, C and N comments. We calculated the Sd to the students from lessons 1 to 15 as follows:
we use 5 values (0, 1, 2, 3 and 4) instead of grade symbols: S, A, B, C and D to compute the prediction error to each student.

We define the prediction error as the absolute difference value between an estimated student grade and an actual student grade. For example, if the actual grade for student (St.1) is S and his predicted grade is A, then the prediction error =1. If the predicted grade is S, then the prediction error to St.1 = 0.

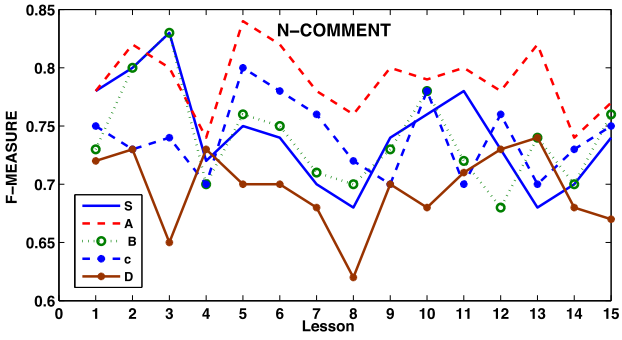
From Table 7, we can see that lessons with greater Sd such as



(a) Prediction results for P-comment



(b) Prediction results for C-comment



(c) Prediction results for N-comment

Fig. 6 Predicting student grades using SVM model, for P, C and N comments.

© 2015 Information Processing Society of Japan

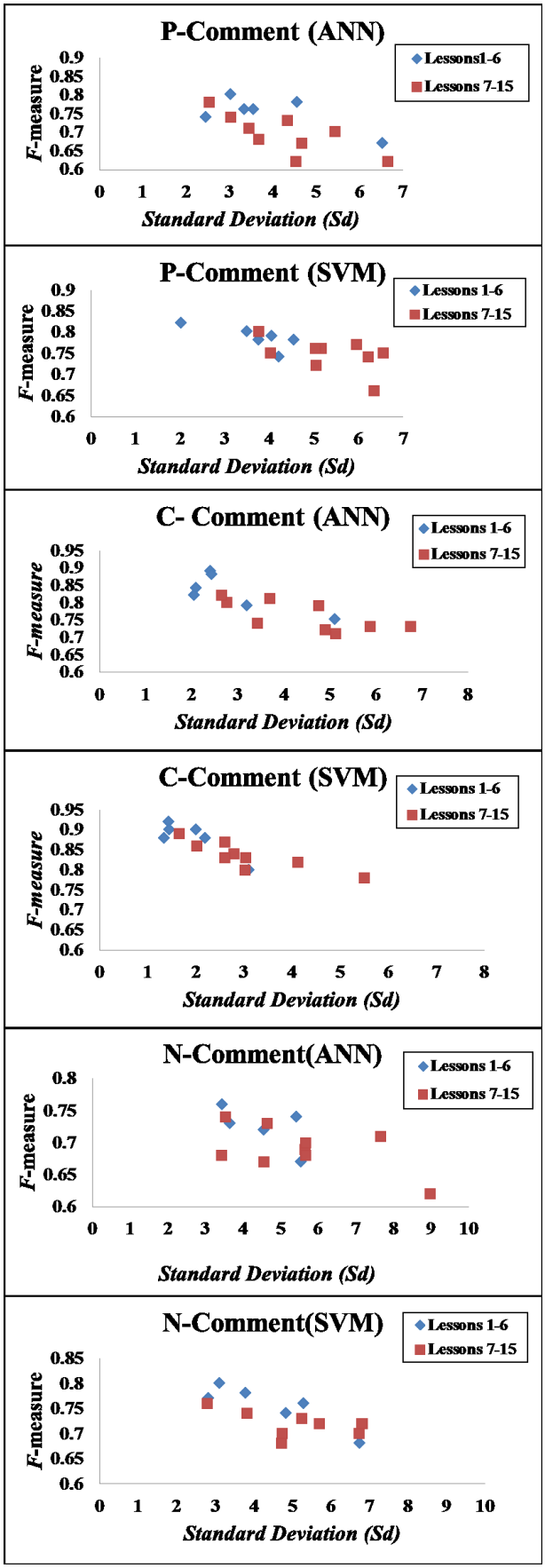


Fig. 7 The correlation between Sd and prediction F-measure results for P, C and N comments.

Table 8 Average correlation coefficient of Sd and F-measure.

Model		Overall	Lessons 1-6	Lessons 7-15
ANN	P	−0.654	−0.699	−0.664
	C	−0.749	−0.740	−0.715
	N	−0.613	−0.650	−0.566
SVM	P	−0.723	−0.740	−0.700
	C	**−0.855**	**−0.859**	**−0.844**
	N	−0.622	−0.786	−0.568

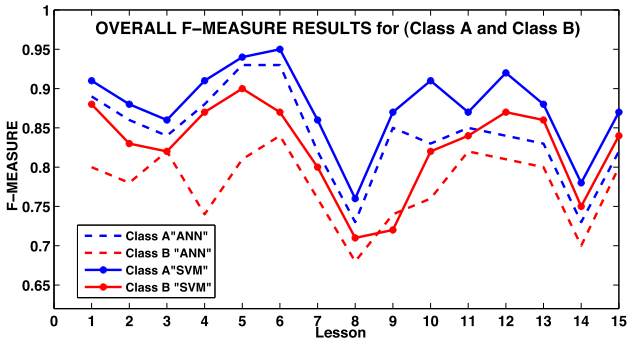


Fig. 8 Overall F-measure results for C-comments in class A and class B.

lessons: 4, 7, 8 and 14 tend to get a lower prediction F-measure and accuracy. We assume that student comment descriptions became drastically changed by some causes such as lesson subjects or questions. Actually from lesson 7, C-programming lesson got started. **Table 8** shows the correlation coefficients between the Sd and the prediction F-measure results. The SVM model with the C-comments had a stronger correlation than the P- and N-comments. On the other hand, **Fig. 7** displays the correlations between the Sd and the prediction F-measure results of the P-, the C- and the N-comments using the ANN and the SVM models.

The correlation coefficients between the Sd and that for the N-comment show a weak correlation. Also, the correlation coefficient from lessons 1-6 were higher than those from lessons 7-15 with viewpoints: P- and C-comments. In addition, C-comment shows a stronger correlation than P- and N-comments.

5.6 Class A and Class B

After conducting the procedures of mining all the comment data in each lesson separately, we decided to narrow down the analysis and select each class data to further clarify if there are any differences between two class data and their effects on the prediction results. This section discusses research Question 6 in Section 1.1. We conducted experiments in each class using the same LSA results. We followed the previous approach and created the ANN and the SVM models in each class. We established a network model of ANN as we mentioned previously: the number of neurons in hidden layer =15 neurons, 0.3 learning rate, 0.65 momentum coefficient and training time= 1,000 iterations. Also, we applied the SVM model to each class data. The proposed method compared the two class data by calculating the average F-measure results as shown in **Table 9**. We evaluated the prediction performance by 5-fold cross validation in each class data using the ANN and SVM models.

Figure 8 shows the average prediction results (F-measure) in class A and class B for C-comments using the ANN and the SVM models; the results in class A were higher than that in class B.

© 2015 Information Processing Society of Japan

Table 9 Overall F-measure results for class A and class B.

		ANN			SVM		
Viewpoint		**Overall**	**Lessons 1-6**	**Lessons 7-15**	**Overall**	**Lessons 1-6**	**Lessons 7-15**
P	**Class A**	0.795	0.823	0.776	0.863	0.883	0.844
	Class B	0.749	0.758	0.743	0.821	0.853	0.788
C	**Class A**	0.853	0.887	0.830	**0.876**	**0.902**	**0.852**
	Class B	0.781	0.798	0.770	0.814	0.864	0.783
N	**Class A**	0.769	0.781	0.767	0.823	0.863	0.783
	Class B	0.765	0.765	0.770	0.803	0.845	0.755

The results of the SVM model were higher than those of the ANN model in class A and class B.

6. Discussion

In this section, we discuss the answers of the six research questions described in Section 1.1.

- **Question 1**
  The results shown in Section 5.3 and Fig. 4 answer the research Question 1 that there were differences in the prediction results (accuracy and F-measure) from lessons 1 to 15, with the three viewpoints: P-, C-and N-comments. The prediction results for the C-comments were higher than those for the P- and the N-comments. Students described the current activity more clearly which distinguished their grades; this tendency was reflected in the prediction accuracy and F-measure of their grades. On the other hand, the prediction results using the P-comments were higher than those using the N-comments in most of the lessons; the previous action included better clues to estimate student learning situations than the next activity plan.
- **Questions 2 and 3**
  All the previous results confirm that the SVM model performed better than the ANN model in predicting student grades. The research Question 3 investigated whether the prediction results of higher grade students were better than those of lower grade students. From Fig. 6, we can distinguish the prediction results for students with higher grades: S, A and B from those for lower ones: C and D. Also, we can see that the prediction results for grade A had the best ones among the five grades for the following reasons: The number of comments of grade A students in all lessons was greater than that of the other grade students. On the other hand, we had the worst prediction results for grade D students, because the number of their comments was smaller than the other grade students in most of the lessons.
- **Question 4**
  The results displayed in Table 7 and Fig. 7 illustrate the strong correlation between the standard deviation (Sd) of the prediction F-measure and the F-measure using the SVM model with C-comments. N-comments had the weaker correlation than the P-comments. In addition, the correlation from lessons 1-6 were higher than those from lessons 7-15.
- **Question 5**
  Research Question 5 concerns the relationship between the difficulty of a subject and prediction results of student grades. From Figs. 4 and Fig. 5, we assumed that the difficulty of the subject had influenced the quality of the written comments; students wrote their learning situations precisely in their comments during lessons about Computer Literacy compared to lessons about C-programming. In addition, the prediction results with C- and P-comments from lessons 1 to 6 were higher than those in lessons 7 to 15.
- **Question 6**
  The results shown in Section 5.6 and Fig. 8 answer the research Question 6 that there were differences between the comments although students in each class took the same course given by the same instructor; we assume each class data has its own characteristics and unique features.

7. Conclusions and Future Work

In this paper, we proposed student grade prediction methods based on their free-style comments with the three viewpoints: P, C and N items. First, comment data was analyzed using the LSA technique; we calculated similarity between words using a comment matrix and detected noisy data by reducing the number of dimensions. Second, two types of machine learning technique: ANN and SVM were employed to build prediction models of student grades based on LSA results.

From the previous results, we can conclude that the difficulty of the subject in the lesson affected student attitudes to expressing their behavior and sometimes did not give students leeway to write comments; they wrote better comments while learning Computer Literacy from lessons 1 to 6 than while learning C-programming from lessons 7 to 15. We can assume that the dropping of prediction results from lesson 7 due to the nature of the comments; students started coding and the comments included additional noise, i.e., programming/technical content, while Computer Literacy education was compulsory throughout senior high schools in Japan, with only a few differences in the details of course contents.

Also, students described their current activities (C-comment) better than previous and next activities (P- and N-comments); this tendency is reflected in the (accuracy/F-measure) of their predicted grades. On the other hand, the SVM model performed better to predict student grades than the ANN model in all lessons.

In future research, the effort can be spent in the following directions to further improve student performance. First, measuring motivation after each lesson can help to provide advice to students and improve their tendency and attitudes in each lesson. Second, the next task is to further explore the relation between student grades and their comment data to extract some clues according to their grades and to give automatic feedback so that we can improve their performance.

Acknowledgments This work was supported in part by JSPS

© 2015 Information Processing Society of Japan

KAKENHI Grant No.25350311 and 26540183.

References

[1] Andrews, R., Diederich, J. and Tickle, A.B.: Survey and critique of techniques for extracting rules from trained artificial neural networks, *Knowledge-Based Systems*, Vol.8, No.6, pp.373–389 (1995).

[2] Araque, F., Roldan, C. and Salgueroa, A.: Factors influencing university drop out rates, *Computer and Education*, Vol.53, pp.563–574 (2009).

[3] 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).

[4] Chappuis, J., Stiggins, R., Chappuis, S. and Arter, J.: *Classroom assessment for student learning: Doing it right-using it well.*, Assessment Training Institute (2012).

[5] Cranfield, D.J. and Taylor, J.: Knowledge Management and Higher Education: A UK Case Study, *The Electronic Journal of Knowledge Management*, Vol.6, No.2, pp.85–100 (online), available from ⟨www.ejkm.com⟩ (2008).

[6] 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).

[7] Dringus, L.P. and Ellis, T.: Using data mining as a strategy for assessing asynchronous discussion forums, *Computers & Education*, Vol.45, pp.141–160 (2005).

[8] Florez, M.T. and Sammons, P.: *Assessment for learning: Effects and impact*, Oxford University Department of Education (2013).

[9] 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).

[10] 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).

[11] Hsu, C.-W., Chang, C.-C. and Lin, C.-J.: A Practical Guide to Support Vector Classification, *National Taiwan University, Taipei 106, Taiwan*, pp.1–16 (online), available from ⟨http://www.csie.ntu.edu.tw/ cjlin/ papers/guide/guide.pdf⟩ (2010).

[12] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, *14th International Joint Conference on Artificial Intelligence*, Vol.2, No.1137-1145 (1995).

[13] Kotsiantis, S., Patriarcheas, K. and Xenos, M.: A combinational incremental ensemble of classifiers as a technique for predicting students performance in distance education, *Knowledge-Based System*, Vol.23, No.6, pp.529–535 (2010).

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

[15] 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).

[16] Petrides, L.A. and Nodine, T.R.: *Knowledge Management in Education: Defining the Landscape*, The Institute for the Study of Knowledge Management in Education (2003).

[17] Qualters, D.M.: *Using classroom assessment data to improve student learning*, Northeastern University (2001).

[18] Rodrigues, F. and Oliveira, P.: A system for formative assessment and monitoring of students' progress, *Computers & Education*, Vol.76, p.3041 (2014).

[19] Romero, C. and Ventura, S.: Data mining in education, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol.3, No.1, pp.12–27 (2013).

[20] Romero, C., Lopez, M., Luna, J. and Ventura, S.: Predicting students' final performance from participation in on-line discussion forums, *Computers & Education*, Vol.68, pp.458–472 (2013).

[21] Rumelhart, D.E., Hinton, G.E. and Williams, R.J.: *Learning internal representations by error propagation*, Vol.1, chapter 8, pp.318–361 (1986).

[22] Sorour, S.E., Mine, T., Goda, K. and Hirokawa, S.: Predictive Model to Evaluate Student Performance, *Information Processing Society of Japan* (*IPSJ*), Vol.23, No.2, pp.192–201 (2015).

[23] Thomas, J., Allman, C. and Beech, M.: *Assessment for the diverse classroom: A handbook for teachers*, Bureau of Exceptional Education and Student Services Florida Department of Education (2004).

[24] Yamtim, V. and Wongwanich, S.: A study of classroom assessment literacy of primary school teachers, *Procedia - Social and Behavioral Sciences*, Vol.116, pp.2998–3004 (2014).

Shaymaa E. Sorour received her B.S. degree in Education Technology, from the Faculty of Specific Education, Kafr Elsheikh University, Egypt and her M.S. degree in Computer Education from the Faculty of Specific Education, Mansoura University, Egypt in 2004 and 2010, respectively. From 2005 until the present time, she has been working as a demonstrator and an assistant lecturer, respectively at the Department of Education Technology, Faculty of Specific Education, Kafr Elsheikh University, Egypt. She is currently a Ph.D. student in the Graduate School of Information Science and Electrical Engineering, Department of Advanced Information Technology, Kyushu University, Japan. She is a member of IEEE and IPSJ.

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.

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 an Assistant Professor at the College of Education, Kyushu University, from 1992 to 1994 and at the 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 the Language Technology Institutes of CMU, Pittsburgh, PA, USA in 1999. He is currently an Associate Professor at the Department of Advanced Information Technology, Faculty of Information Science and Electrical Engineering, Kyushu University. He received the Best Paper Award in 1993, and the Social Activity Service Award in 2015, both from the IPSJ. His current research interests include Natural Language Processing, Information Retrieval, Information Extraction, Information Recommendation, Personalization and Multi-Agent Systems. He is a member of IPSJ, IEICE, JSAI, NLPSJ, IEEE and ACM.

© 2015 Information Processing Society of Japan