# Predicting Student Grade based on Free-style Comments using Word2Vec and ANN by Considering Prediction Results Obtained in Consecutive Lessons

Jingyi Luo
Graduate School of Information Science and Electrical Engineering,Kyushu University
Fukuoka,Japan

Shaymaa E.Sorour
Graduate School of Information Science and Electrical Engineering,Kyushu University
Fukuoka,Japan

Kazumasa Goda
Kyushu Institute of Information Science
Dazaifu,Japan

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

## ABSTRACT

Continuously tracking students during a whole semester plays a vital role to enable a teacher to grasp their learning situation, attitude and motivation. It also helps to give correct assessment and useful feedback to them. To this end, we ask students to write their comments just after each lesson, because student comments reflect their learning attitude towards the lesson, understanding of course contents, and difficulties of learning. In this paper, we propose a new method to predict final student grades. The method employs Word2Vec and Artificial Neural Network (ANN) to predict student grade in each lesson based on their comments freely written just after the lesson. In addition, we apply a window function to the predicted results obtained in consecutive lessons to keep track of each student's learning situation. The experiment results show that the prediction correct rate reached 80% by considering the predicted student grades from six consecutive lessons, and a final rate became 94% from all 15 lessons. The results illustrate that our proposed method continuously tracked student learning situation and improved prediction performance of final student grades as the lessons go by.

## Keywords

PCN Method, Word2Vec, ANN, Comment Mining, Grade Prediction

## 1. INTRODUCTION

Learner performance assessment is a continuous and an integral part of the learning process [4]. During studying, exams are used to help teachers know how good students are learning, as well as to help them find out the difficulties with the course. However preparing a good exam is a laborious and resource demanding work, so it's still hard to obtain assessment by exams over all periods of a semester.

Thus, in the past four decades, researchers have been working on predicting individual or group performance in courses for getting assessments. By accurate predictions, we can detect students who have difficulties with the courses early, and help them improve [1].

To control students' learning behavior and situations, previous studies have used various regular assessment methods, such as e-learning logs, test marks and questionnaires. The current study proposes a new method to predict student grades. Our method is based on students' free-style comments collected after each lesson.

K.Goda, S.Hirokawa, and T.Mine [3] [2]proposed the PCN method to estimate student learning situations from free-style comments written by the students. The PCN method categorizes the comments into three items: P (Previous activity), C (Current activity), and N (Next activity).

In this paper, we apply the Word2Vec method to the comments data to get a vector representation of each comment. Then we use an artificial neural network (ANN) model to predict student grades based on the vectors. The experiments were conducted to validate the proposed methods by calculating the F-measure and accuracy for each lesson. After acquiring a prediction result for each lesson, we applied a window function and a majority vote method to get a final prediction result based on multiple lessons. The experiment results illustrate that the prediction correct rate reached 80% by considering the predicted student grades obtained from six lessons, and the final rate became 94% from all 15 lessons.

Contributions of this paper are threefold. First, we propose a new method to predict final student grades by using Word2Vec and ANN. Second, we improve the prediction performance by considering the results obtained in consecutive lessons. We show as the size of the lessons increases, the prediction performance becomes better. Third, we conduct experiments to illustrate the effectiveness of the proposed methods. The experiment results show the validity of the proposed methods.

## 2. RELATED WORK

Extensive literature reviews of the Educational Data Mining (EDM) research field are mainly focused on retention of students, improving institutional effectiveness, enrollment management and alumni management. In the past four decades, a considerable amount of research has gone into predicting individual or group success in exams and courses.

Schoor and Bannert [7] studied sequences of social regulatory processes (i.e. individual and collaborative activities of analyzing, planning...aspects) during collaborative sessions and their relationship to group performance. They used process mining to identify process patterns for high versus low group performance dyads. The result models showed that there were clear parallels between high and low achieving dyads in a double loop of working on the task, monitoring, and coordinating.

Liu and Xing [5] aimed to develop a predictive model of student behavior by an ensemble approach composed of creation of sampled sets, generation of base models, and selection of base models to be aggregated for obtaining the final ensemble model. The solution required less computation resource, had satisfying prediction performance and produced prediction models with good capability of generalization.

Different from the above studies, Goda et al. [3] proposed the PCN method to estimate students' learning situations with their free-style comments written just after a lesson. They applied Support Vector Machine (SVM) to the comments for predicting final student results in 5 grades. The experiment results illustrate that as student comments get higher PCN scores, prediction performance of student grades becomes better. Sorour et al.[8] applied machine learning technique: artificial neural network (ANN) and made it learn the relationships between comments data analyzed by Latent semantic analysis(LSA) and the final student grades. They constructed a network model to each lesson. The average prediction accuracy of student final grades was 82.6%. In this study, as an extension of Sorour et al. [8], we focused on using different text mining method Word2Vec combined with the ANN model to get prediction on each lesson, and obtain prediction results based on consecutive multiple lessons. Our method outperformed the method of Sorour et al.[8].

# 3. METHODOLOGY

## 3.1 Collecting Comments

In this research, we used the same comment data as Sorour et al.[8]. The comments were collected after each lesson in a course including 15 lessons. 123 students attended this course. They were asked to fill in three simple questionnaire items about their learning status. Goda et al. [3] called the three items, P (Previous), C (Current) and N (Next) items. In this paper, we mainly focus on the C (Current) comments. Table 1 displays the real number of comments in each lesson that we analyzed. On average there is 111.13 comments in each lesson.

Table 1: Number of comments for each lesson

| Lesson | Num | Lesson | Num | Lesson | Num |
|---|---|---|---|---|---|
| 1 | 100 | 6 | 116 | 11 | 107 |
| 2 | 121 | 7 | 104 | 12 | 109 |
| 3 | 118 | 8 | 103 | 13 | 107 |
| 4 | 115 | 9 | 107 | 14 | 111 |
| 5 | 123 | 10 | 111 | 15 | 121 |

## 3.2 Comments Data Preparation

### *3.2.1 Comments Data Preprocessing*

This step covers all the preparations required for constructing the final dataset from the initial data. Our method used a Japanese morphological analyzer Mecab[1] to analyze C comments, extract words and part of speech. In this experiment, we only used noun, verb, adjective and adverb. 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.

### *3.2.2 Word2Vec*

Word2vec is a popular neural network based approach to learning distributed vector representations for words released by Google in 2013. This tool adopts two main model architectures, Continuous Bag-of-Words (CBOW) and Skip-Gram[6].

## 3.3 Training Phase

After the previous step and before we applied ANN to train the data, we have some pretreatments for preparing training data for ANN.

We have got a list of vocabularies and their corresponding vectors after the previous step. Now we need to find out all the words one student have used in his/ her comment which existed in the vocabulary list, and add the vectors indicating these words up to get a final vector for that student.

After obtaining a list of vectors for each student, we need to proceed the training phase with the list. In this research, we used a three-layered Artificial Neural Network to estimate student grades. In our work, we used FANN Libraries[2] to build our network model. We took the results from the former step and put them into the input layer of ANN. For all the lessons, we applied the same model with 0.1 learning rate and 0.3 momentum.

## 3.4 Test Phase

To predict student grades, we used 5 grade categories instead of real marks to classify final student marks.

Table 2: 5 Grades Categories

| Real Marks | Grades | Num of Students |
|---|---|---|
| $\geq 90$ | S | 21 |
| 80-89 | A | 41 |
| 70-79 | B | 23 |
| 60-69 | C | 17 |
| $\leq 60$ | D | 21 |

Since in each lesson, there exist some students who did not fill in questionnaires, we can't predict their grade. In these cases, we treat them as grade D instead.

After training the ANN model, we proceed the test phase to get prediction results of final student grades in each lesson. In the test phase, we evaluated prediction performance (Accuracy, F-measure) by 10-fold cross validation. We separated comments data by using 90% as training data and the rest 10% as test data. The procedure was repeated 10 times and the results were averaged. Afterwards, we apply an window function and the majority vote method to obtain a continuous prediction. The details of the window function and the majority vote method will be described in Section 4.1.

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

[2]http://leenissen.dk/fann/wp/

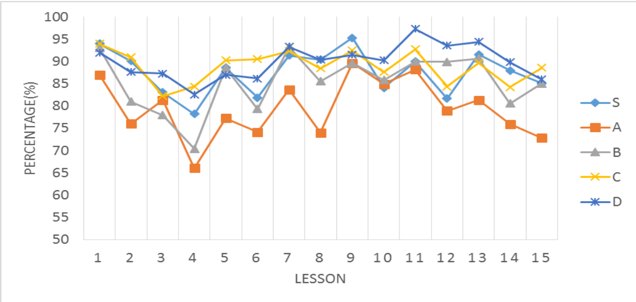


Figure 1: Accuracy for different grades

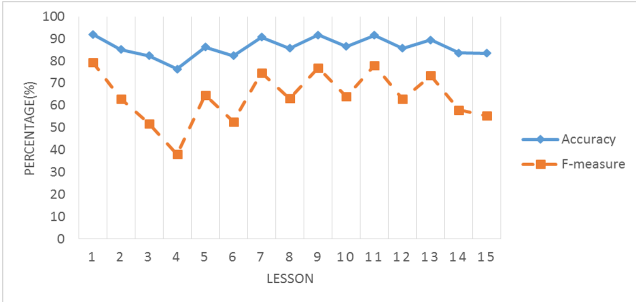


Figure 2: Average accuracy and F-measure of all the grades in each lesson

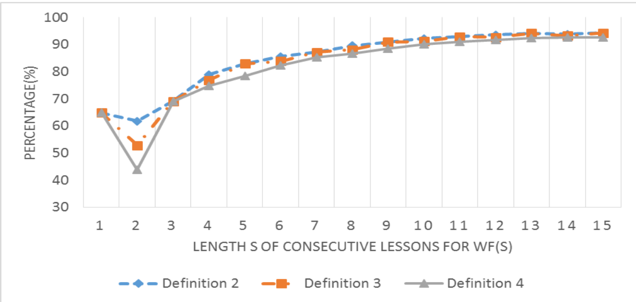


Figure 3: Average TP rate based on different definitions

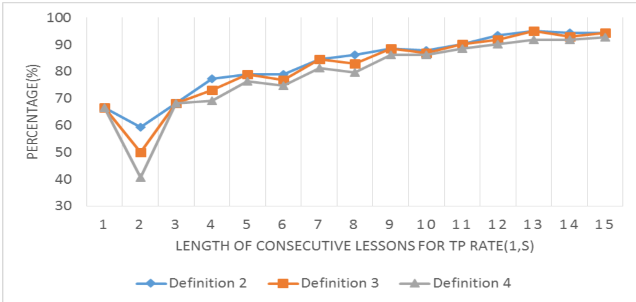


Figure 4: TP rate for different length of consecutive lessons from lesson 1

# 4. PREDICTION PERFORMANCE

## 4.1 Measure of Prediction Performance

We define the majority vote method and the window function as follows:
Let G be a set of grades $\{g_0, g_1, g_2, g_3, g_4\}$; each element of G corresponds to each grade, i.e., $g_0$, $g_1$, $g_2$, $g_3$, and $g_4$ correspond to S, A, B, C, and D, respectively. Let $MV_k(m,n)$ be the function of Majority Vote of student $k$ from lessons $m$ to $n$ . $MV_k(m,n)$ returns a set of predicted student $k$'s grades whose occurrence frequency from lessons $m$ to $n$ became the greatest. We define $MV_k(m,n)$ in Definition 1.

Definition 1. $MV_k(m,n)$

$$MV_k(m,n) = \underset{g_i \in G}{\operatorname{argmax}}\ f(k,g_i)(m,n)$$

where $f(k,g_i)(m,n)$ returns the occurrence frequency of predicted grade $g_i$ of student $k$ from lessons $m$ to $n$.

For example, if the predicted grades of student 1 from lessons 1 to 3 are respectively S ($=g_0$), A ($=g_1$), and S ($=g_0$), then $f(1,g_0)(1,3) = 2$ and $f(1,g_1)(1,3) = 1$ . So, $MV_1(1,3)$ returns $\{g_0\}$. If the predicted grades of student 1 from lessons 1 to 3 are respectively S($=g_0$), A($=g_1$), B($=g_2$), then $f(1,g_0)(1,3) = 1$, $f(1,g_1)(1,3) = 1$, and $f(1,g_2)(1,3) = 1$. So, $MV_1(1,3)$ returns $\{g_0, g_1, g_2\}$.

Function $\delta$ returns a score according to the results returned by a Majority Vote function $MV(m,n)$ defined in Definition 1. Three $\delta$ functions: $\delta_1$, $\delta_2$, and $\delta_3$, are defined in Definitions 2, 3, and 4. Here we use the notation $|.|$ that denotes the cardinality of a set. For example, if $MV_1(1,3)$ returns $\{g_0, g_1, g_2\}$, then $|MV_1(1,3)| = 3$.

Definition 2. $\delta_1$

$\delta_1(MV_k(m,n))$ returns 1 if $g_k$ is the actual grade of student $k$, $g_k \in MV_k(m,n)$ and $g_l \notin MV_k(m,n)$ such that $|l-k| > 1$, 0 otherwise.

For example, we assume that the actual grade of student $k$ is $g_0$, if $MV_k(m,n) = \{g_0, g_1\}$, then $\delta_1(MV_k(m,n)) = 1$. If $MV_k(m,n) = \{g_0, g_2\}$ then $\delta_1(MV_k(m,n)) = 0$, because $|2-0| > 1$.

Definition 3. $\delta_2$

$\delta_2(MV_k(m,n))$ returns $\frac{1}{|MV_k(m,n)|}$ if $g_k \in MV_k(m,n)$ where $g_k$ is the actual grade of student $k$, 0 otherwise.

Definition 4. $\delta_3$

$\delta_3(MV_k(m,n))$ returns 1 if $g_k \in MV_k(m,n)$ and $|MV_k(m,n)| = 1$, 0 otherwise.

Next, we define $TP(m,n)$ that returns True Positive (TP) rate from lessons $m$ to $n$ in Defition 5.

Definition 5. $TP(m,n)$

$$TP(m,n) = \frac{\sum_{k=1}^{N_s} \delta(MV_k(m,n))}{N_s}$$

where $N_s$ is the number of students.

Now we define function $WF(s)$, which returns the average TP rate in $s$ consecutive lessons, in Defition 6. Here $s$ denotes the length of consecutive lessons, i.e. the number of lessons.

Definition 6. $WF(s)$

$$WF(s) = \frac{\sum_{k=1}^{N-s+1} TP(k,k+s-1)}{N-s+1}$$

where $N$ is the number of all lessons in a course, 15 in this research.

For example, when $N = 15$, $WF(1)$ to $WF(15)$ are computed as follows:

$$WF(1) = \frac{TP(1,1) + TP(2,2) + .. + TP(15,15)}{15}$$

$$WF(2) = \frac{TP(1,2) + TP(2,3) + .. + TP(14,15)}{14}$$

$$WF(3) = \frac{TP(1,3) + TP(2,4) + .. + TP(13,15)}{13}$$

$$\cdots$$

$$WF(14) = \frac{TP(1,14) + TP(2,15)}{2}$$

$$WF(15) = \frac{TP(1,15)}{1} = TP(1,15)$$

## 4.2 Results in Each Lesson

We examined the same model on all the students with different final grades. Results are shown in Figures 1 and 2. Figure 1 displays the plot of accuracy results of students with different grades in each lesson. Table 3 shows the average overall prediction accuracy and F-measure for the different grades. As for accuracy, the result of grade D is the highest, which scores 89.5%, and the lowest average is grade A, which scores 79.1%. Also, according to Figure 2, lesson 1 has the highest accuracy and F-measure, while lesson 4 has the lowest results.

Table 3: Average accuracy and F-measure for different grades

| Grades | Accuracy | F-measure |
|---|---|---|
| S | 87.3 | 65.6 |
| A | 79.1 | **71.3** |
| B | 85.0 | 62.6 |
| C | 88.5 | 57.2 |
| D | **89.5** | 62.3 |
| Average of all grades | 85.9 | 63.8 |

## 4.3 Results after Using Window Function and Majority Vote

Before we apply the window function to all the consecutive lessons, we first treat all the students who did not describe comments as Grade D. After this step, it also ensures that for each lesson, every student has one predicted grade. After we get the prediction result in each lesson, we apply the window function and the majority vote method to get a continuous track of student performance.

Here, we only consider TP rates. First we investigated the effect of size $s$ of $WF(s)$ by varying the value of $s$ from 1 to 15. As we can see, in Figure 3, the TP rate was increased as the value of $s$ increased. As an example of the results, even though the strictest way of counting the correct case by Definition 4, the correct rate still raised over 80% after considering more than six lessons. In addition, with all the lessons, the correct rates all reached over 90%. And with Definition 2 and 3, they both reached 94%. The results by Definition 4 reached 92.7%.

Figure 4 shows the result of TP rate from TP(1, 1) ,TP(1, 2),TP(1, 3) to TP(1, 15) with three different definitions. With the growing of window function size, the TP rate raised over 80% with more than 7 lessons, which is slightly lower than the average.

Considering the results of Figures 3 and 4, we can say the both results took similar tendency that the TP rates became greater as the size of lessons increased.

# 5. CONCLUSIONS AND FUTURE WORK

In this paper, we discussed the prediction method of student grade based on the C comments data from Goda et al. [3]. We applied the Word2Vec and ANN methods to the comments to obtain prediction of their grades in each lesson. Then we used the window function and the majority vote method to improve the prediction results based on consecutive multiple lessons. The experiment results illustrate the validity of the proposed method.

This study expressed the correlation between self-evaluation descriptive sentences written by students and their academic performance by predicting their grade. Especially when using prediction results obtained in consecutive lessons, the prediction result has quite high credibility. This could help giving feedback to students during the semester to help students achieve higher motivation and know their learning conditions better.

However, there still remain some room for improving prediction results in each lesson. In the future, we will try to apply better models to achieve higher accuracy in predicting student grades.

# 6. ACKNOWLEDGMENTS

This work was partially supported by JSPS KAKENHI Grant Numbers 25350311 and 26540183.

# 7. REFERENCES

[1] M. Fire, G. Katz, Y. Elovici, B. Shapira, and L. Rokach. Predicting student exam's scores by analyzing social network data. Lecture Notes in Computer Science Volume 7669, pp 584-595, ,2012.

[2] K.Goda, S.Hirokawa, and T.Mine. Correlation of grade prediction performance and validity of self evaluation comments. In Proc. of the 14th annual ACM SIGITE conference on Information technology education, Florida, USA, 2013.

[3] K.Goda and T.Mine. Analysis of student' learning activities through qualifying time-series comments. Proc. of the KES 2011, Part 2, LNAI 6882, Springer-Verlag Berlin Heidelberg, pp.154-164, 2011.

[4] L.Earl. Assessment of Learning, for Learning, and as Learning. Thouand Oaks,CA,Corwin Press, 2003.

[5] K. Liu and Y. Xing. A lightweight solution to the educational data mining challenge. In Proceedings of the KDD 2010 cup 2010 workshop knowledge discovery in educational data(pp. 76-82), 2010.

[6] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. Advances in Neural Information Processing Systems (NIPS), 2013.

[7] C. Schoor and M. Bannert. Exploring regulatory processes during a computer-supported collaborative learning task using process mining. Computers in Human Behavior, 2012.

[8] S. E. Sorour, T. Mine, K. G., and S. Hirokawa. Predictive model to evaluate student performance. In Journal of Information Processing (JIP), Vol. 23, 2015.