# The Department Score Prediction and Analysis of the College Entrance Examination *

Chun-Chieh Hsu[a], Chang-Biau Yang[a]†, and Yung-Hsing Peng[b]

[a]Department of Computer Science and Engineering
National Sun Yat-sen University, Kaohsiung, Taiwan

[b]Innovative DigiTech-Enabled Applications & Services Institute
Institute for Information Industry, Kaohsiung, Taiwan

**Abstract**

Prediction systems for *college entrance examination* (CEE) are used to give some recommendations for students to choose their ideal department. In Taiwan, more than 40% students of senior high schools attend the CEE to enter a university/college. In this paper, we first introduce four CEE prediction methods, and then propose an effective method based on the forward method. We analyze the differences between predicted scores and real scores to propose a factor to improve the performance. In addition, we compare these four prediction methods with ours, and explain how to apply our method to the prediction system. In the experiments, the CEE data from 2004 to 2014 are considered. We use three indicators based on the *root mean square error* (RMSE) and two indicators recall and precision to evaluate the performance of various methods. The experimental results show that our method outperforms the previous methods. This new method has been implemented on our website (http://uexam.nsysu.edu.tw/) to provide free CEE prediction service.

**Keywords:** college entrance examination, minimum required score, prediction

## 1 Introduction

For each senior high school student in Taiwan, to enter a university/college is a very important choice in his (her) life. There are three ways to enter universities/colleges, including the *star recommendation*, the application for admission to a university/college, and the *College Entrance Advanced Subjects Test* (CEAST). The CEAST or simply the *Advanced Subjects Test* (AST), whose predecessor was called the *Joint College Entrance Examination* (JCEE), is organized by the *College Entrance Examination Center* (CEEC). The CEAST covers more than 40% of the university/college entrance quota, which is a very important way for students to enter universities/colleges. The CEAST contains ten subjects, including Chinese, English, Mathematics A, Biology, Physics, Chemistry, Mathematics B, History, Geography, and Social Science. According to their own expertise, the examinees can freely choose to take some or all of the subjects, which offers a better focus on professional subjects. The least score to enter a department is called the *minimum required test score* (MRTS) of the department. After the CEAST, students have to fill in their application lists based on their CEAST scores and preference to departments. For this situation, a *college entrance examination* (CEE) prediction system can help students evaluate whether they can be admitted by their preferred departments.

In this paper, our goal is to construct a prediction system with historical data to predict the MRTS of a department, whose accuracy is measured by RMSE, recall and precision. The organization of this paper is as follows. In Section 2, we first introduce some prerequisite knowledge, including the adopted tools and the previous work. In Section 3, we present a detailed explanation of our method. The experimental results and conclusions are given in Section 4 and Section 5, respectively.

*

†Corresponding author. E-mail: cbyang@cse.nsysu.edu.tw (Chang-Biau Yang).

# 2 Preliminaries

## 2.1 The Forward Method

The forward method [4] is a direct prediction for the department's MRTSs of the CEE. There are two assumptions that the preferences of students to schools are not changed in two consecutive years, and that the rank of a department is unique, regardless to subjects. With these assumptions, the prediction indices of the department's MRTSs could be obtained from historical information. With the *cumulative score table* (CST) released by the CEEC [10], a score can be transformed into its corresponding percentage and vice versa. To use the CST, we first define some variables. Let $sco_{(i,j)}$ be a student's score of the $j$th subject in year $i$, $w_{(s,d,i,j)}$ be the weight put on the $j$th subject in year $i$ by department $d$ in school $s$, $MRTS_{(s,d,i)}$ be the MRTS of department $d$ in school $s$ in year $i$, and $pts_{(i,j)}(\ )$ be the function that transforms a high-to-low cumulative percentage of the $j$th subject into its corresponding score in year $i$. Note that $w_{(s,d,i,j)}$ is required because each department can have its emphasis on different subjects. Assume that $per_{(s,d,i)}$ is the rank in percentage for department $d$ of school $s$ when considering subject $j$ in year $i$, where top departments are of less percentages. Let $MRTS_{(s,d,i)}$ be the minimum required test score for department $d$ of school $s$ in year $i$, whose equation is given as follows.

$$MRTS_{(s,d,i)} = \sum_{j=1}^{|SUB|} pts_{(i,j)}(per_{(s,d,i)}) \times w_{(s,d,i,j)}, \tag{1}$$

where $SUB$ denotes the set of required subjects, and $|SUB|$ means the number of subjects in $SUB$. By Equation 1, a binary search can be applied to obtain suitable percentage (rank) $per_{(s,d,i)}$ for all departments. Based on the assumption that this rank is the same in two consecutive years, we can predict the department's MRTSs in year $i$ as $MRT\hat{S}_{(s,d,i)}$ with the information $per_{(s,d,i-1)}$ in year $i-1$. The equation of predicting a department's MRTS is given as follows:

$$MRT\hat{S}_{(s,d,i)} = \sum_{j=1}^{|SUB|} pts_{(i,j)}(per_{(s,d,i-1)}) \times w_{(s,d,i,j)}. \tag{2}$$

## 2.2 Forward Method with Enrolment Quota

Let the function $cpn_{(i,j)}(\ )$ transform a cumulative percentage into its corresponding cumulative number of students in CST. Let $pn_{(s,d,i,j)}$ be the number (quota) of students who can enroll for department $d$ in school $s$ in year $i$ if the $j$th subject is required for this department. In Taiwan, $pn_{(s,d,i,j)}$ is released by CEEC every year. If the $j$th subject is not necessary for department $d$, we set $pn_{(s,d,i,j)}$=0. Note that all departments can be ranked by their $per_{(s,d,i)}$ in year $i$, which forms a ranking list of all departments. Let $dpn_{(s,d,i,j)}(\ )$ be a function to transform a percentage $per_{(s,d,i)}$ into its corresponding cumulative number of students based on $pn_{(s,d,i,j)}$, whose equation is given as follows.

$$dpn_{(s,d,i,j)}(per_{(s,d,i)}) = \sum_{per_{(s',d',i)} \le per_{(s,d,i)}} pn_{(s',d',i,j)}. \tag{3}$$

In other words, $dpn_{(s,d,i,j)}$ is the cumulative number of students by considering the departments that require the $j$th subject, which is more meaningful than $cpn_{(i,j)}$. Therefore, we can compute $per'_{(s,d,i,j)}$ for the $j$th subject in the department $d$ in school $s$ by the equation as follows:

$$cpn_{(i,j)}(per'_{(s,d,i,j)}) = \frac{cpn_{(i-1,j)}(per_{(s,d,i-1)})}{dpn_{(s,d,i-1,j)}(per_{(s,d,i-1)})} \times dpn_{(s,d,i,j)}(per_{(s,d,i-1)}). \tag{4}$$

By equation 4, we can obtain an estimated cumulative percentage $per'_{(s,d,i,j)}$. By rewriting Equation 2, a department's MRTS can be predicted as follows.

$$MRT\hat{S}_{(s,d,i)} = \sum_{j=1}^{|SUB|} pts_{(i,j)}(per'_{(s,d,i,j)}) \times w_{(s,d,i,j)}. \tag{5}$$

## 2.3 Lin's Method

In 2008, Lin [5] constructed four kinds of predictors and determined the suitable one for each department. Lin's prediction models require the means and standard deviations of all students in all examination subjects, in which some of them are real ones and some other are predicted. The real means $\mu$ and standard deviations $\sigma$ are released by CEEC [10]. The predicted means and standard deviations are denoted as $\hat{\mu}$ and $\hat{\sigma}$, respectively. Let $Z_i^{(n)}$ denote the $n$th prediction index of one department in year $i$, $w_{ij}$ denote the weight of the $j$th subject in year $i$, $|SUB|$ denote the number of all examination subjects, and $X_{ij}$ denote one department's MRTS of the $j$th subject in year $i$. The four prediction models by Lin are given as follows.

$$Z_i^{(1)} = \frac{\sum_{j=1}^{|SUB|} w_{ij}(X_{ij} - \mu_{ij}^{(1)})}{\sum_{j=1}^{|SUB|} w_{ij}}, \tag{6}$$

where $\mu_{ij}^{(1)}$ is a real median score of the $j$th subject.

$$Z_i^{(2)} = \frac{\sum_{j=1}^{|SUB|} w_{ij}(X_{ij} - \hat{\mu}_{ij})}{\sum_{j=1}^{|SUB|} w_{ij}}, \quad (7)$$

where $\hat{\mu}_{ij}$ is the predicted mean of the $j$th subject in year $i$.

$$Z_i^{(3)} = \frac{\sum_{j=1}^{|SUB|} w_{ij}(X_{ij} - \hat{\mu}_{ij})}{\sqrt{\sum_{j=1}^{|SUB|} w_{ij}\hat{\sigma}_{ij}^2}}, \quad (8)$$

where $\hat{\sigma}_{ij}$ denotes the standard deviation of all students' scores of the $j$th subject in year $i$.

$$Z_i^{(4)} = \frac{\sum_{j=1}^{|SUB|} w_{ij}(X_{ij} - \mu_{ij})}{\sqrt{\sum_{j=1}^{|SUB|} w_{ij}\sigma_{ij}^2}}, \quad (9)$$

where $\mu_{ij}$ and $\sigma_{ij}$ are denoted as the real mean and standard deviation of all students' scores of the $j$th subject in year $i$, respectively. In order to verify the assumption that a student has same preference for departments in the two consecutive years, Lin applied the simple linear regression to compute the correlation coefficient between $Z_{i-1}^n$ and $Z_i^n$. As results, the correlation coefficient is very high, which implies that the assumption is true. By comparing the four prediction indices, the numerical analysis of accuracies [5] indicates that Equation 8 outperforms others. Therefore, the prediction model $Z_i^{(3)}$=$Z_{i-1}^{(3)}$ is used to predict each department's MRTS.

### 2.4 Prediction with Fuzzy Time Series

The concept of fuzzy set was first introduced by Zadeh [11] in 1965. Since then, the fuzzy set has been massively applied in several areas, including industry, finance, meteorology, etc. Based on definitions given by Song and Chissom [8], time-invariant and time-variant models for fuzzy time series [7, 9] can be constructed. Chen and Hsu [3] proposed a new method to forecast enrolments with fuzzy time series. In 2007, Sah [6] constructed an enrolment prediction model based on the first-order fuzzy time series. Haneen *et. al.* [1] forecasted the enrolments of students by using genetic algorithms and fuzzy time series. Through the above study of fuzzy time series, a CEE prediction model with fuzzy time series can be designed.

In this paper, the data points of the time series are variations of cumulative percentages (variations of ranks) between two consecutive years. For conciseness, in the following we give a brief example for our fuzzy model. To predict the MRTS with fuzzy time series, we assume that the variations of *per* between two consecutive years are expressed within the range [-0.6,0.6]. Let $U$ be the universe with range [-0.6,0.6], which is further partitioned into five intervals $u_1$=[-0.6,-0.36], $u_2$=[-0.36,-0.12], $u_3$=[-0.12,0.12], $u_4$=[0.12,0.36], and $u_5$=[0.36,0.6]. The linguistic variable of variations can be expressed as two fuzzy sets $A_1$ (decreasing) and $A_2$ (increasing), whose rule is to give the membership degree 1 to $u_j$ if the given number is in the interval $u_j$, and to give 0.5 if it falls in the neighboring intervals of $u_j$. The following examples illustrate the results for two given numbers -0.3 and 0.5, where the number preceding the slash symbol "/" denotes the membership degree of $u_j$ to $A_1$ or $A_2$.

- -0.3 $\in \{0.5/u_1, 1/u_2, 0.5/u_3, 0/u_4, 0/u_5\}$.
- 0.5 $\in \{0/u_1, 0/u_2, 0/u_3, 0.5/u_4, 1/u_5\}$.

For predicting the variations, the techniques of finding fuzzy relationships are involved [1]. We omit the detailed descriptions because it is beyond the scope of this paper. However, it is clear that the predicted variations can be used to estimate the rank $per_{(s,d,i)}$ of a department, and then its MRTS can be further predicted by Equation 1.

## 3 An Improved Prediction Method for Entrance Scores

### 3.1 The Admission Rate

According to our historical experiments, the forward method [4] is more accurate than other previous methods for predicting MRTSs. Therefore, we are inspired to analyze the predicted MRTSs by the forward method. By observing the difference between the predicted MRTS and real MRTS of each department, the predicted MRTS could be *over-estimated* (OE) or *under-estimated* (UE), if the predicted MRTS is higher or lower than the real MRTS, respectively. The percentages of OE and UE for the total departments in each year are summarized in Table 1. In Table 1, we find that the percentages in each year are seriously biased. In other words, the over-estimation and under-estimation are not balanced. If we can balance the over-estimation and under-estimation in every year, the prediction would be more precise. For solving this issue, we examine the historical admission rates. The admission rate is calculated as the total number (quota) of admitted

Table 1: The percentages of over-estimation (OE) and under-estimation (UE).

| Year | 2005 | 2006 | 2007 | 2008 | 2009 |
|---|---|---|---|---|---|
| OE | 72.02% | 78.62% | 88.20% | 87.74% | 77.40% |
| UE | 27.98% | 21.38% | 11.80% | 12.25% | 22.60% |
| Year | 2010 | 2011 | 2012 | 2013 | 2014 |
| OE | 32.60% | 16.86% | 30.00% | 85.69% | 75.47% |
| UE | 67.40% | 83.14% | 70.00% | 14.31% | 24.53% |

Table 2: The admission rates and the biased prediction results.

| Year | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|
| OE / UE | – | OE | OE | OE |
| Admission Rate | 75.60% | 76.29% | 80.82% | 85.86% |
| Year | 2008 | 2009 | 2010 | 2011 |
| OE / UE | OE | OE | UE | UE |
| Admission Rate | 91.02% | 94.17% | 86.51% | 80.29% |
| Year | 2012 | 2013 | 2014 | – |
| OE / UE | UE | OE | OE | – |
| Admission Rate | 77.49% | 82.95% | 84.21% | – |

students divided by the total number of applicants for CEAST. In every year, we give the mark OE if the percentage of OE is greater than that of UE; otherwise we give the mark UE. The relation between OE, UE, and admission rates are illustrated in Table 2. We can see that if the admission rate rises (falls), then the department's predicted scores will be higher (lower) than the real MRTSs. Accordingly, we can assume that the variation of the admission rates in two consecutive years is an important factor for the biased prediction scores.

### 3.2 Our Prediction Method

From the previous section, we find that the admission rates in two consecutive years are deeply correlated to the prediction accuracy of MRTS. Note that if the admission rate of this year is higher than that of last year, it means the threshold of entering the university becomes lower. Therefore, a department's real MRTS should also become lower, which could be surpassed by our predicted score. That is, to consider the admission rate will be an effective way to improve our prediction accuracy. Here we propose the term *ratio of admission rates* (RAR) as the ratio between consecutive two years. The RAR of year $i$ is defined as follows.

$$RAR_i = \frac{AR_i}{AR_{i-1}}, \tag{10}$$

Table 3: RAR with H and L.

| Year | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|
| OE / UE | – | OE | OE | OE |
| Admission Rate | 75.60% | 76.29% | 80.82% | 85.86% |
| $RAR_i$ | – | 1.0091 | 1.0595 | 1.0623 |
| Year | 2008 | 2009 | 2010 | 2011 |
| OE / UE | OE | OE | UE | UE |
| Admission Rate | 91.02% | 94.01% | 86.51% | 80.29% |
| $RAR_i$ | 1.0601 | 1.0325 | 0.9187 | 0.9281 |
| Year | 2012 | 2013 | 2014 | – |
| OE / UE | UE | OE | OE | – |
| Admission Rate | 77.49% | 82.95% | 84.21% | – |
| $RAR_i$ | 0.9652 | 1.0705 | 1.0152 | – |

where $AR_i$ denotes the admission rate of year $i$. The values of RAR, OE and UE are summarized in Table 3. Our new approach is to adjust the prediction index $per_{(s,d,i)}$ by multiplying it with RAR. This method is called the *forward method with admission rate* (FMAR). The equation of FMAR is given as follows.

$$MRTS^*_{(s,d,i)} = \sum_{j=1}^{|SUB|} pts_{(i,j)}(per_{(s,d,i-1)} \times RAR_i) \times w_{(s,d,i,j)}. \tag{11}$$

Next, we explain the detailed steps of our prediction method.

- Step1: Predict each department's MRTS of year $i$ by Equation 11.
- Step2: Obtain the subject scores input by the student using our website of prediction service.
- Step3: Pick out the valid departments that use these subjects as required scoring criteria.
- Step4: Calculate the application scores of the student for each valid department according to each department's weights on the required subjects.
- Step5: Calculate the predicted admission level ($PAL$ in Equation 12) between each department's predicted MTRS and the student's application score.

$$PAL = \frac{aps_{(s,d,i)} - MRTS^*_{(s,d,i)}}{MRTS^*_{(s,d,i)}} \times 100\%, \tag{12}$$

where $aps_{(s,d,i)}$ denotes the application score of a student to department $d$ of school $s$ in year $i$, and $MRTS^*_{(s,d,i)}$ denotes the predicted MRTS for the department in year $i$.

- Step6: Output different messages to the student according to $PAL$ as follows.

1. If $PAL > 6\%$, then the student can definitely be admitted by this department.
2. If $2\% < PAL \leq 6\%$, then the student has high probability to be admitted by this department.
3. If $-2\% < PAL \leq 2\%$, then the student is recommended to apply for this department as a good choice.
4. If $-6\% < PAL \leq -2\%$, then the student has low probability to be admitted by this department.
5. If $PAL \leq -6\%$, then it is almost impossible for the student to be admitted by this department.

### 3.3 Performance Evaluation

We adopt several indicators for evaluating the performance. The first indicator RMSE (root mean square error) is calculated by the difference between the predicted MRTS and the real MRTS. The second indicator $PRMSE1$ gives an error in percentage between the predicted MRTS and the real MRTS. Because the difference between predicted MRTS and real MRTS may be much greater than the real MRTS, we propose the third indicator $PRMSE2$ to reduce the negative influence. The indicator $PRMSE2$ is obtained by dividing $PRMSE1$ with the maximum score between predicted MRTS and real MRTS. In addition to RMSE and PRMSE, we compute the recall and the precision [12] by considering the difference between the predicted admission level ($PAL$) and the real admission level ($RAL$), defined in Equation 12 and 13, respectively.

$$RAL = \frac{aps_{(s,d,i)} - MRTS_{(s,d,i)}}{MRTS_{(s,d,i)}} \times 100\%, \quad (13)$$

To perform the experiments, the students' application scores are generated by normal distribution with mean $x_i$ and standard deviation 5%, where $x_i$ denotes a cumulative percentage, for $x_1$=10%, $x_2$=20%,..., $x_9$=90%. The random variables $rv_j$ for the $j$th subject's cumulative percentage with mean equal to $x_i$ and standard deviation equal to 5% are generated by Box and Muller's method [2]. We also use the function $pts_{(i,j)}(\ )$ to transform a cumulative percentage $rv_j$ to its corresponding score. In our experiments, we compare $PAL$ and $RAL$ with five specified intervals [-2%,2%], [-4%,4%], [-6%,6%], [0%,∞], and [-∞,0%). In the following, we explain some fundamental terms used in the performance measurements.

1. True positive ($TP$): the number of departments that both $PAL$ and $RAL$ fall in the same specified interval.
2. False positive ($FP$): the number of departments that $PAL$ is within the specified interval while $RAL$ is not.
3. True negative ($TN$): the number of departments that both $PAL$ and $RAL$ are outside the specified interval.
4. False negative ($FN$): the number of departments that $RAL$ is within the specified interval while $PAL$ is not.

The formulas for recall and precision are expressed as follows.

$$recall = \frac{TP}{TP + FN}. \quad (14)$$

$$precision = \frac{TP}{TP + FP}. \quad (15)$$

## 4 Experimental Results

In this section, we show the performance of our forward method with admission rate (FMAR), and compare it with the previous methods, including the forward method [4], the forward method with enrolment quota, Lin's method [5] and the fuzzy method. The five performance indicators used in our experiments are RMSE, $PRMSE1$, $PRMSE2$, recall, and precision.

The averages of the five performance measurements from 2005 to 2014 are shown in Table 4, except that the results of fuzzy method are from 2008 to 2014. In this table, $P1$ denotes $PRMSE1$, $P2$ denotes $PRMSE2$, FMEQ denotes forward method with enrolment quota, *rec* denotes recall, and *prec* denotes precision. In Table 4, we can see that FMAR predicts the departments' MRTSs more accurately than others in almost all performance measurements. The detailed comparison of recall and precision in the interval [-2%,2%] for year 2013 is shown in Table 5. Each of the cumulative percentages from 10% to 90% is the mean for generating 10 subjects' scores by normal distribution. And each result is the average of 1000 generated students' scores. If the $RAL$ or $PAL$

Table 4: The performance comparison of FMAR and the previous methods from 2005 to 2014, where $P1$ denotes $PRMSE1$, $P2$ denotes $PRMSE2$, and FMEQ denotes forward method with enrolment quota.

| | | | | [-2%,2%] | | [-4%,4%] | | [-6%,6%] | | [0%,∞] | | [-∞,0%) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | $P1$ | $P2$ | rec | prec | rec | prec | rec | prec | rec | prec | rec | prec |
| Forward [4] | 20.62 | 14.35 | 8.48 | 35.48% | 35.14% | 51.55% | 51.28% | 61.08% | 60.80% | 85.86% | 91.45% | 97.71% | 95.63% |
| FMEQ | 25.57 | 19.29 | 10.07 | 32.23% | 31.19% | 46.55% | 44.79% | 55.18% | 53.61% | 76.11% | 91.63% | 98.65% | 93.69% |
| Lin [5] | 21.30 | 15.20 | 8.83 | 34.71% | 34.99% | 51.70% | 52.40% | 61.02% | 61.55% | 86.01% | 91.58% | 97.45% | 94.94% |
| Fuzzy | 24.10 | 13.79 | 9.53 | 35.56% | 36.28% | 51.16% | 52.46% | 60.34% | 62.23% | 88.75% | 89.96% | 96.64% | 95.89% |
| FMAR | 17.49 | 12.85 | 7.92 | 44.41% | 44.15% | 61.07% | 60.58% | 69.76% | 69.18% | 89.73% | 92.16% | 97.83% | 96.59% |

Table 5: The detailed performance comparison of FMAR and the previous methods by recall (rec) and precision (prec) in the interval [-2%,2%] for year 2013.

| | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---|---|---|---|---|---|---|---|---|---|
| Forward [4] | | | | | | | | | |
| rec | 87.08% | 71.48% | 46.03% | 33.8% | 24.96% | 15.78% | 14.93% | 8.58% | 2.88% |
| prec | 80.6% | 63.45% | 37.45% | 31.62% | 24.52% | 14.42% | 18.75% | 7.46% | 7.56% |
| FMEQ | | | | | | | | | |
| rec | 78.64% | 78.95% | 63.29% | 46.65% | 35.27% | 21.67% | 9.35% | 1.17% | 0.88% |
| prec | 78.45% | 72.96% | 55.30% | 43.80% | 31.77% | 18.48% | 10.82% | 0.88% | 1.00% |
| Lin [5] | | | | | | | | | |
| rec | 82.81% | 79.16% | 51.74% | 39.84% | 29.86% | 18.01% | 15.60% | 7.37% | 2.49% |
| prec | 78.71% | 69.01% | 42.25% | 37.17% | 28.71% | 16.80% | 19.11% | 6.69% | 6.86% |
| Fuzzy | | | | | | | | | |
| rec | 84.26% | 70.35% | 43.23% | 32.80% | 18.98% | 11.47% | 5.96% | 0.47% | 10.61% |
| prec | 77.66% | 58.42% | 35.48% | 31.43% | 18.24% | 9.51% | 8.15% | 1.28% | 17.17% |
| FMAR | | | | | | | | | |
| rec | 81.05% | 78.8% | 66.66% | 49.84% | 50.60% | 49.45% | 42.55% | 34.39% | 10.02% |
| prec | 81.82% | 75.21% | 61.47% | 51.63% | 51.63% | 52.89% | 48.22% | 32.56% | 11.65% |

Table 6: The detailed recall (rec) and precision (prec) comparison of FMAR and the previous methods in the interval [-2%,2%] for each year, where each performance measure is obtained by the average of the students' scores generated by normal distributions with mean from top 10% to 50%.

| Year | 2005 | | 2006 | | 2007 | | 2008 | | 2009 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rec | prec | rec | prec | rec | prec | rec | prec | rec | prec |
| Forward [4] | 45.52% | 44.10% | 47.17 | 43.49% | 51.28% | 47.57% | 46.54% | 44.02% | 48.15% | 46.10% |
| FMEQ | 32.67% | 29.80% | 39.52% | 35.16% | 46.93% | 43.91% | 20.29% | 18.15% | 47.89% | 44.50% |
| Lin [5] | 43.78% | 43.90% | 47.94% | 45.74% | 50.29% | 46.12% | 43.33% | 44.52% | 43.92% | 43.09% |
| Fuzzy | – | – | – | – | – | – | 52.44% | 50.48% | 51.39% | 50.32% |
| FMAR | 48.08% | 46.98% | 58.54% | 57.82% | 57.86% | 58.28% | 62.08% | 62.32% | 53.29% | 52.43% |

| Year | 2010 | | 2011 | | 2012 | | 2013 | | 2014 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rec | prec | rec | prec | rec | prec | rec | prec | rec | prec |
| Forward [4] | 59.18% | 60.89% | 53.90% | 57.78% | 76.39% | 77.36% | 52.67% | 47.53% | 66.09% | 65.43% |
| FMEQ | 60.17% | 58.88% | 63.06% | 59.80% | 69.89% | 67.83% | 60.56% | 56.46% | 66.25% | 73.36% |
| Lin [5] | 50.72% | 54.69% | 54.97% | 58.69% | 69.34% | 73.04% | 56.68% | 51.17% | 60.09% | 60.72% |
| Fuzzy | 48.06% | 50.66% | 48.79% | 52.46% | 67.39% | 67.36% | 51.41% | 45.92% | 60.00% | 65.13% |
| FMAR | 48.70% | 46.26% | 70.95% | 70.75% | 72.74% | 74.41% | 65.39% | 64.35% | 69.38% | 69.89% |

Table 7: The detailed recall (rec) and precision (prec) comparison of FMAR and the previous methods in the interval [-2%,2%] for each year, where each performance measure is obtained by the average of the students' scores generated by normal distributions with mean from 60% to 90%.

| Year | 2005 | | 2006 | | 2007 | | 2008 | | 2009 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | rec | prec | rec | prec | rec | prec | rec | prec | rec | prec |
| Forward [4] | 14.11% | 15.37% | 5.80% | 6.70% | 2.39% | 2.42% | 3.77% | 5.30% | 7.89% | 7.73% |
| FMEQ | 0.92% | 0.81% | 0.00% | 0.00% | 7.08% | 6.65% | 0.37% | 0.42% | 3.86% | 3.27 |
| Lin [5] | 11.14%% | 12.69% | 12.72% | 12.71% | 2.58% | 2.52% | 6.03% | 7.41% | 13.78% | 11.46% |
| Fuzzy | – | – | – | – | – | – | 7.95% | 9.10% | 14.63% | 14.29% |
| FMAR | 13.61% | 14.06% | 29.98% | 32.11% | 21.06% | 19.64% | 22.41% | 21.80% | 17.22% | 15.35% |
| Year | 2010 | | 2011 | | 2012 | | 2013 | | 2014 | |
| | rec | prec | rec | prec | rec | prec | rec | prec | rec | prec |
| Forward [4] | 23.17% | 22.92% | 8.78% | 8.80% | 15.90% | 19.26% | 10.54% | 12.05% | 21.53% | 21.56% |
| FMEQ | 11.99% | 10.62% | 8.46% | 8.21% | 22.28% | 21.52% | 8.27% | 7.80% | 27.47% | 31.88% |
| Lin [5] | 20.02% | 19.19% | 12.11% | 12.60% | 15.07% | 17.49% | 10.87% | 12.37% | 20.59% | 20.09% |
| Fuzzy | 12.53% | 11.73% | 6.94% | 6.88% | 19.01% | 22.43% | 7.64% | 8.20% | 18.85% | 21.72% |
| FMAR | 6.03% | 5.70% | 38.30% | 36.43% | 26.84% | 31.55% | 34.10% | 36.33% | 28.10% | 27.73% |

Table 8: The RMSE comparison of FMAR and the previous methods.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Forward [4] | 12.46 | 12.44 | 24.99 | 25.12 | 26.28 | 23.61 |
| FMEQ | 24.02 | 32.82 | 24.91 | 36.73 | 37.60 | 30.03 |
| Lin [5] | 13.51 | 12.64 | 25.46 | 26.03 | 25.78 | 25.94 |
| Fuzzy | – | – | – | 23.89 | 23.54 | 27.49 |
| FMAR | 11.97 | 8.36 | 17.15 | 19.46 | 26.67 | 34.38 |
| Year | 2011 | 2012 | 2013 | 2014 | – | – |
| Forward [4] | 23.22 | 20.54 | 20.01 | 17.54 | – | – |
| FMEQ | 21.67 | 12.25 | 19.98 | 15.66 | – | – |
| Lin [5] | 24.32 | 20.15 | 20.54 | 18.58 | – | – |
| Fuzzy | 27.56 | 22.81 | 25.03 | 18.38 | – | – |
| FMAR | 14.46 | 13.34 | 12.63 | 16.46 | – | – |

Table 9: The $PRMSE1$ comparison of FMAR and the previous methods.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Forward [4] | 5.73 | 5.77 | 35.93 | 36.49 | 18.35 | 14.41 |
| FMEQ | 13.45 | 15.95 | 35.63 | 50.69 | 26.49 | 26.50 |
| Lin [5] | 6.59 | 5.89 | 35.88 | 39.57 | 18.69 | 17.04 |
| Fuzzy | – | – | – | 34.42 | 15.91 | 15.41 |
| FMAR | 5.41 | 3.35 | 26.80 | 26.04 | 18.22 | 29.42 |
| Year | 2011 | 2012 | 2013 | 2014 | – | – |
| Forward [4] | 7.92 | 6.28 | 6.49 | 6.10 | – | – |
| FMEQ | 8.84 | 3.46 | 6.55 | 5.36 | – | – |
| Lin [5] | 8.60 | 6.10 | 6.67 | 6.99 | – | – |
| Fuzzy | 9.38 | 6.99 | 8.09 | 6.30 | – | – |
| FMAR | 5.80 | 3.89 | 3.95 | 5.61 | – | – |

to a departments falls in the interval [-2%,2%], it would suggest that the department is one of the goal departments for the student. The predicted MRTSs will be close to the real MRTSs if the recall and precision are high. One can see that the predicted MRTSs are much closer to the real MRTSs for high application scores (with cumulative percentage 10%). FMAR illustrates its significant superiority for cumulative percentages between 20% to 80%, and its performance for the cases 10% and 90% remains acceptable. Next, we investigate to see the performance comparison of high and low scores, which are shown in Tables 6 and 7, respectively. In general, the accuracies of high scores are much better than those of low scores. The comparison of RMSE, $p1$, and $p2$ of our FMAR method and the previous methods are shown in Tables 8, 9, and 10. In our experiment, we notice that FMAR fails to give an accurate prediction in year 2010. From Table 11, we can

Table 10: The $PRMSE2$ comparison of FMAR and the previous methods.

| Year | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
|---|---|---|---|---|---|---|
| Forward [4] | 5.24 | 5.15 | 12.23 | 13.26 | 12.45 | 11.15 |
| FMEQ | 10.61 | 11.95 | 12.14 | 16.57 | 15.03 | 13.87 |
| Lin [5] | 5.86 | 5.11 | 12.38 | 13.31 | 11.79 | 12.99 |
| Fuzzy | – | – | – | 13.09 | 11.87 | 12.99 |
| FMAR | 5.00 | 3.19 | 9.58 | 13.66 | 15.45 | 14.90 |
| Year | 2011 | 2012 | 2013 | 2014 | – | – |
| Forward [4] | 7.89 | 6.28 | 5.73 | 5.37 | – | – |
| FMEQ | 6.61 | 3.21 | 5.72 | 4.95 | – | – |
| Lin [5] | 8.54 | 6.09 | 5.86 | 6.32 | – | – |
| Fuzzy | 9.35 | 6.96 | 6.92 | 5.52 | – | – |
| FMAR | 4.68 | 3.87 | 3.75 | 5.12 | – | – |

Table 11: The student quotas of admission and the real number of students getting admission between years 2008 to 2010.

| Year | 2008 | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|---|
| Quota | 85270 | 82264 | 72674 | 66052 | 58841 |
| Real | 81409 | 76434 | 71165 | 66683 | 59738 |

see that the student quota of admission descends sharply in year 2010. Therefore, the threshold for entering a university/college should be higher than year 2009. However, the real number of students getting admission is in fact less than the student quota, which means the competition of entering a university/college is not as strong as we expected. This is the reason why the prediction of FMAR is not so accurate in year 2010. With regard to the overall performance, however, FMAR still achieves the best result.

## 5 Conclusion

In this paper, we first investigate four prediction methods for CEE, including the forward method [4], the forward method with enrolment quota, the Lin's method [5], and the fuzzy method. After that, we propose a new prediction method, called the forward method with admission rate (FMAR), whose performance is better than others. We also adopt statistical measures to test our prediction method. From experimental results, one can see that our prediction method achieves the best accuracy. In Taiwan, more than 40% of senior high school students attend CEAST to enter a university/college, and they need a reliable CEE prediction system for department recommendation. We have implemented the FMAR as a free website service (http://uexam.nsysu.edu.tw/), for helping the senior high students who attend the CEAST. The score data in this website is annually updated, and new CEE prediction methods will further be developed, verified, and supported in the future.

## Acknowledgement

This research was partially supported by the National Science Council of Taiwan under contract NSC 101-2221-E-110-020-MY3. This research was also partially supported by the "Online and Offline Integrated Smart Commerce Platform (2/4)" of the Institute for Information Industry, which is subsidized by the Ministry of Economy Affairs of Taiwan.

## References

[1] H. Al-Wazan, K. Ibraheem, and A.-G. Salim, "A hybrid algorithm to forecast enrolment based on genetic algorithms and fuzzy time series," *International Arab Journal of Information Technology*, Vol. 11, No. 3, 2014.

[2] G. E. P. Box and M. E. Muller, "A note on the generation of random normal deviates," *The Annals of Mathematical Statistics*, Vol. 29, No. 2, pp. 610–611, 06 1958.

[3] S.-M. Chen and C.-C. Hsu, "A new method to forecast enrollments using fuzzy time series," *International Journal of Applied Science and Engineering*, Vol. 3, pp. 234–244, 2004.

[4] Y.-S. Chen, "The prediction of the department score of the college entrance examination in Taiwan," Master's thesis, National Sun Yat-Sen University, Kaohsiung, Taiwan, 2012.

[5] C.-L. Lin, "On the study of forecasting indices after the college entrance examination with real data," Master's thesis, National Cheng Kung University, Tainan, Taiwan, 2008.

[6] M. Sah and K. Y. Degtiarev, "Forecasting enrollment model based on first-order fuzzy time series," *International Journal of Computer, Information Science and Engineering*, Vol. 2, No. 1, 2007.

[7] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series part I," *Fuzzy Sets and Systems*, Vol. 54, No. 1, pp. 1–9, 1993.

[8] Q. Song and B. S. Chissom, "Fuzzy time series and its models," *Fuzzy Sets and Systems*, Vol. 54, No. 3, pp. 269–277, 1993.

[9] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series part II," *Fuzzy Sets and Systems*, Vol. 62, No. 1, pp. 1–8, 1994.

[10] The College Entrance Examination Center, "CEEC." http://www.ceec.edu.tw/, 1989.

[11] L. Zadeh, "Fuzzy sets," *Information and Control*, Vol. 8, No. 3, pp. 338–353, 1965.

[12] H. Zhang and X. Zhang, "Comments on data mining static code attributes to learn defect predictors," *IEEE Transactions on Software Engineering*, Vol. 33, No. 9, p. 635, 2007.