1. Introduction

In recent years, educational institutions have started managing students’ data electronically. For example, confirmation of their attendance is achieved via recognition of an IC chip embedded in their ID cards, and their attendance records are maintained on a server. Similarly, the scores of exercises are also maintained on the server. These data have been used only as an index of the grades given by a teacher. However, there is no feedback to the students, which means that some of the data are not being fully exploited.

To correct this underutilization, we propose the idea that each student’s future academic status is forecasted from the huge amount of data available for the students. Students are given educational advice and warnings corresponding to their forecasted future academic status. Using this idea, we are able to convert underutilized data into useful information for students.

Moreover, in particular, we focus on “anxious students”; that is, those students who are worried about dropping out. In university education, some students tend to be absent without notice, and when they advance to the next level, they are not able to follow the lectures that build on material studied in prior courses. We want to help these students. We propose to prevent dropping out by giving academic advice to students who are forecasted to be anxious about the future.

There are many studies related to student’s grade forecast and data analysis. Some examples are: a study which analyzes the learning trend from student data about attendance and exercise scores; another which forecasts future grades of students by neural networks from student data about attendance and exercise scores; another one which investigates characteristics of the student by a Bayesian network created from sequences of class questionnaires; and a final one which forecasts a student’s properties by a Bayesian network created from exercise scores. These studies handle data about one class. Reference (2) forecasts the final grade when the student had completed at least eight exercises, and they give advice and new exercises corresponded to the forecast result.

Instead of one course, we propose forecasting the academic status for one year. In this study, using grading data for each course of a certain university, we propose forecasting the grade score of the second year from that of the first year. The grade score is the student’s Grade Point Average (GPA). In the university, there are a lot of basic courses in the first year and specialized courses begin in the second year. We expect that the learning trend can be found for the second year when there are more specialized courses than in the first year. If the result of this study is good, we can say there is a relationship between the basic courses and the specialized courses.
The reason why we choose the future GPA of the second year for forecasting is because just before starting the second year is a good time to provide academic advice. For the second semester of the first year, there is not sufficient information for making a good forecast. For the third year, the course of study is almost over. Thus such advice then is meaningless. By giving academic advice to students who are forecast to be “anxious students” about their future just before the start of the second year, we can reduce the number of students who cannot keep up in course work in the future.

There are many forecast methods that require large amounts of data to be processed by calculation: regression analysis, multivariate statistics, neural networks, genetic programming, and Bayesian networks can be listed as examples. For our purpose, we suggest a forecasting method using a Bayesian network. A Bayesian network is a probabilistic graphical model that represents a set of random variables and their conditional dependencies. It is used in probability theory as a forecast method. In this study, we handle GPA scores as variables of a Bayesian network and forecast a student’s academic status from the network result.

One of the merits of generating forecasts using the Bayesian network is that the output is in the form of a probability value, the meaning of which can be easily understood by users without specialized knowledge. We think we can provide more flexible academic advice by handling probability values as forecast results. For example, we give a strong warning to an “anxious student” given a probability value of 100%. Moreover, we give advice to students given probability values from 30% to 50%. We can change the contents of the advice to make them suit the student’s situation. For this reason, the output of the Bayesian network is user-friendly.

Furthermore, the Bayesian network holds well for non-linear events which is another merit. For example, some good first year student become poor in their second year. A simple forecast method such as regression analysis cannot identify these students. However, we think that the Bayesian network can. To verify that the Bayesian network is suitable in this study, we compare the forecast Bayesian network result accuracy with the regression analysis result.

In order to generate such a forecast using the Bayesian network, it is necessary to construct a Bayesian network model. In brief, the model construction process is as follows. First the random variables are defined. Second, a directed graph is created. Finally, the conditional probability is estimated.

In this study, about 100 variables which are based on GPA scores and some number of grades received during a course (exercises and exam scores) are defined previously. However, it is necessary to select some variables for generation of models because all variables are not useful. As the method of selecting variables, we use correlation-based feature selection (CFS), which is one data mining technique. In addition, variables of the Bayesian network must be nominal. Thereby, it is necessary to discretize numeric variables selected by CFS. As the method of discretization, we use a decision tree. A decision tree is a method of classification and it is also useful for discretization. Using these methods, we overcome the difficulty in generation of Bayesian network models.

The rest of the paper is organized as follows. In Section 2, the Bayesian network and the generation method of the Bayesian network models are outlined. In Section 3, models are generated which are suited to giving academic advice. In Sections 4 and 5, evaluation of the method and the results are described.

2. Outline of the Bayesian Network

In this study, we generated Bayesian network models for forecast from variables defined by GPA scores and some number of grades received. First, we define the Bayesian network; then, we describe a forecasting method using the Bayesian network and finally, we describe the process from definition of the variables to generation of the models.

2.1 Definition

The Bayesian network is a graphical model that presents the dependence relationship among some variables in a graph structure. By calculating the probability value using the model, it is possible to make forecasts and to infer unknown events.

The Bayesian network consists of three elements: random variables, a directed acyclic graph, and conditional probabilities. Conditional probabilities are stored in a conditional probability table (CPT). In a model such as that shown in Figure 1, the elements are the variables $X$, $Y$, $Z$ and $W$, the directed graph, and the probabilities $P(X)$, $P(Y|X,Z)$, $P(Z)$, and $P(W|Z)$. 

33
2.2 Forecast method by Bayesian network

The Bayesian network is implemented to generate a forecast by using a model to perform the probability calculation. We suppose that the probability value that we want to forecast is $X$ and the information concerning another event is $e$; by calculating $P(X|e)$, the expectation value and confidence factor of one event are obtained. From this calculation, we can get the probability of being an “anxious student” in the future.

2.3 How to generate a model

First, we explain the process of generating the models described below.

**STEP 1.** About 100 variables are defined from course grade data like shown in Table 3.

**STEP 2.** Variables which have strong causality with the purpose variable are selected by the CFS technique. Specifically, we apply the best one selected by the CFS technique as the index of selection. Therefore, we are able to carefully select about 10 variables from about 100 variables.

**STEP 3.** A decision tree which has the purpose variable as target of classification is generated from selected variables. Then all numeric variables selected by the CFS technique are discretized by using the consequence from generation of the decision tree.

**STEP 4.** A Bayesian network model is generated by the Weka software using the purpose variable and variables which are selected and discretized. As the index of generating the optimal models, we use AIC(6).

Weka Environment for Knowledge Analysis (Weka)$^7$ software is used to create the directed graphs and to estimate conditional probabilities. This free software provides data mining tools and many calculation algorithms.

About 100 variables are defined previously (see STEP 1). If all defined variables are used for generating models, the accuracy of the Bayesian network model becomes worse, so it is necessary to select variables. In addition, because there are numeric variables in defined variables, it is also necessary to convert numeric data into nominal data; that is to discretize them. To solve this problem, we use information gain for selection, whereas a decision tree is used for discretization. We outline the information gain and decision tree next.

2.4 Information gain

Information gain (also called information divergence or relative entropy) is the measure of the difference between two probability distributions(6).

If we suppose that $P$ and $Q$ are probability distributions, the information gain is defined as Equation (1).

$$D(P|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)}$$  \hspace{1cm} (1)

Because of its character, information gain is often expressed as the “distance of probability distributions”.

For constructing the model, we must select variables which are suitable to the forecast model from among many variables. The CFS technique is used as the index of selection. CFS is an algorithm that couples the evaluation formula with an appropriate correlation measure and a heuristic search strategy. Equation (2) defines the quantity CFS, where $k$ denotes the number of variables, $Z$ denotes the response variable, and $Y_i$ denotes the explanatory variable. $Y_i$ which maximizes Equation (2), is selected as the variable for the model.

$$CFS = \frac{\sum_{i=1}^{k} SU(Y_i, Z)}{\sqrt{k + \sum_{i=1}^{k} \sum_{j=1}^{k} SU(Y_i, Y_j)}}$$  \hspace{1cm} (2)

Symmetrical Un-certainty (SU) in Equation (2) is defined as Equation (3). $H(A)$ and $H(B)$ are entropies.
2.5 Decision tree

A decision tree is a decision support tool that uses a tree-like graph or model of decisions and their possible consequences. The graph or model has nodes which describe the rules, and leaves which represent the class labels. In this paper, in particular, we use a binary tree. By following the rules at the nodes, a forecast can be made. If the random variables are expected to yield a response variable, both numeric and nominal types are applicable in the decision tree. Using this feature, we can discretize numeric type random variables. The discretization method by the decision tree is based on information gain. Therefore, this method is suitable for the Bayesian network.

Using imaginary data of height and sex and making a classification tree from these data, we present an explanatory example of a decision tree. In Table 1, M and F denote “Male” and “Female”. The decision tree based on these data is shown in Figure 2. The height dispersion is \([-\infty, 156), (156, 166), (166, 168), (168, +\infty)\].

\[
SU(A, B) = 2^n \frac{D(AB)}{H(A) + H(B)}
\]  

(3)

3. Academic Advice Based on Forecast

Now, we generate Bayesian network forecast models from student data and we evaluate whether it is possible to give academic advice based on their prediction. In addition, we evaluate the generated models and discuss their usefulness.

3.1 Original data and their normalization

To generate forecast models, we use original (raw) data related to 171 students in a university. These data consist of about 8,000 recorded scores representing each student’s grade for various courses. These data include both first and second year information. In the first year, students take basic courses. In the second year, they take specialized courses. Therefore, when the first year’s courses are different from those of the second year, the models generated in this study are still suitable. Moreover, we do not consider students who intentionally take first year courses in the second year because we did not find such cases in the university data. In order to make forecast models, it is necessary to convert the original data into a variable and we use Grade Point Average (GPA) as the variable. The GPA is the score that evaluates a student’s performance. It is calculated by dividing the number of grade points earned in a given time period by the total number of credit points of the courses taken.

In addition, it is possible to calculate the GPA score for a certain course, such as “math GPA”, or “foreign language GPA”, or for a certain time period, such as “first year GPA”. Table 2 shows the number of courses by subject field and Table 3 shows some examples of variable definitions. We do not consider defining variables by a method such as factor analysis; like \(A = 0.7B + 0.3C\). In addition, to reflect the significance of the data, the standard deviation of the scores is calculated. Therefore, because the grade scores are converted to a relative one, we think influence in the difference of difficulty among each subject field’s grade distributions disappears.

When creating our Bayesian network forecast models, all variables are selected and discretized using the CFS technique and the decision tree.
Table 2. Number of Courses by Subject Field.

<table>
<thead>
<tr>
<th>Subject field</th>
<th>Number of courses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First year</td>
</tr>
<tr>
<td></td>
<td>First semester</td>
</tr>
<tr>
<td>Foreign language</td>
<td>2</td>
</tr>
<tr>
<td>Foundation of science (1)</td>
<td>6</td>
</tr>
<tr>
<td>Specialized (1)</td>
<td>4</td>
</tr>
<tr>
<td>Humanities and culture (2)</td>
<td>1</td>
</tr>
</tbody>
</table>

(1) – Including subjects for selection from the second year
(2) – Subjects for selection

Table 3. Examples of Variable Definitions.

<table>
<thead>
<tr>
<th>Variable number</th>
<th>Variable content – Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>GPA of first year</td>
</tr>
<tr>
<td>2</td>
<td>GPA of first year first semester</td>
</tr>
<tr>
<td>3</td>
<td>GPA of first year second semester</td>
</tr>
<tr>
<td>17</td>
<td>Math GPA of first year first semester</td>
</tr>
<tr>
<td>28</td>
<td>Number of “A” grades in the first year</td>
</tr>
<tr>
<td>88</td>
<td>(GPA of the foreign language course in the first year)–(GPA of the specialized courses in the first year)</td>
</tr>
</tbody>
</table>

Table 4. The Purpose Variable of Model A.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-P</td>
<td>The variable that gives YES if the deviation value of the GPA score for the second year is less than 40.</td>
</tr>
</tbody>
</table>

3.2 About forecasting

The results data of 171 students who completed the course for two years are used for construction of the models. After building the forecasting models, at the end of a time period, the data for one year are input into the model, and the student’s future academic status is forecast.

The posterior probability value of an objective variable is obtained using the forecasting model. We attempt to generate a decision as to whether giving academic counseling to a student would be useful, based on the probability value.

3.3 Target groups of academic counseling

Before making a model to forecast the benefit of giving a student academic advice, we should discuss how the decision to give such counseling is reached. In short, target groups for academic counseling must be discussed. Two targets, Model A and Model B, are assumed here.

3.3.1 Model A

Our basic aim is to give useful academic counseling to students who are worried about their academic performance. To achieve this, we first have to define “the student who will be anxious about his studies from now on”, hereafter called the “anxious student”.

In this model, an “anxious student” is defined as one who will have a low GPA score in the second year. When the first year was completed, we forecast whether students are an “anxious student”, and if so, we give them academic advice. The criterion that defines an “anxious student” is a GPA score deviation value in the second year of less than 40.

The purpose variable is defined in order to predict which students are the targets of academic guidance. The content of the purpose variable is given in Table 4.

3.3.2 Model B

In Model B, a “student who will be worried” is defined as one who will have a low GPA score in the second year specialized courses because the second year
specialized courses are more difficult than those in the first year which disconcerts many students.

If students are classified as a “student who will be worried”, we give them academic counseling. By giving the appropriate advice, we can support the students in their studies.

The criterion that defines a “student who will be worried” was a GPA deviation value of less than 45 in the second year specialized courses.

The purpose variable is defined in order to identify students who should be the target of academic guidance. The content of the purpose variable is described in Table 5.

### Table 5. The Purpose Variable of Model B.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-P</td>
<td>The variable that shows YES if the deviation value of the GPA score for second year specialized courses is less than 45.</td>
</tr>
</tbody>
</table>

The variables selected in model A are listed in Table 6. The generated decision tree is shown in Figure 3. Finally, the forecasting model that is generated is shown in Figure 6.

Moreover, the probability value that denotes “anxious student” is calculated using the forecasting model. The results are listed in Tables 8 and 9. In the range of a certain probability value, the number of students who are actually “anxious students” is counted. The coincidence rate of “anxious student” in each probability value is examined, and we verify the probability value that serves as the range within which academic advice would be effective.

### Table 6. Variable List of Model A.

<table>
<thead>
<tr>
<th>Notation</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-1</td>
<td>GPA of the foreign language courses in the second semester of the first year</td>
</tr>
<tr>
<td>A-2</td>
<td>GPA of the science foundation courses in the first year</td>
</tr>
<tr>
<td>A-3</td>
<td>GPA of the specialized courses in the second semester of the first year</td>
</tr>
<tr>
<td>A-4</td>
<td>The number of “S” grades got in the first semester of the first year</td>
</tr>
<tr>
<td>A-5</td>
<td>The number of “B” grades got in the first year</td>
</tr>
<tr>
<td>A-6</td>
<td>The number of “B” grades got in the second semester of the first year</td>
</tr>
<tr>
<td>A-7</td>
<td>(GPA of the foreign language courses in the first year) - (GPA of the specialized courses in the first year)</td>
</tr>
</tbody>
</table>

### Figure 3. Decision Tree of Model A.

### Figure 4. Forecasting Model of Model A.
4. Evaluation and Validation of the Models

To evaluate models of this study, we compare them with another forecast method. As the other method, we use one which makes a linear relationship between each purpose variable and variable groups. Concretely, it is given as Equation (4).

\[ Y = a_0 + \sum_{i=1}^{n} a_i \times X_i \quad (4) \]
Y is the purpose variable. n is the number of selected variables. $X_i$ is the variable selected by the CFS technique. $a_i$ is a parameter. Optimal parameters are determined by least-squares method.

At first, we compare accuracy of the Bayesian network models with the linear relationship. These results are listed in Tables 10 and 11. In the forecast by Bayesian network models, a high probability value is the forecast result. For example, if the probability value of “YES” is 52, it is “YES”. If the value is 48, it is “NO”.

From these results, it seems that the linear relationship is better than the proposed method. However, we expect many “anxious students” are helped by our approach. Then, we investigate the percentage of the students who actually became so in the students who were predicted as “YES” at each model. These results are listed in Tables 12 and 13.

From these results, the prediction accuracy as a whole is lower, for the goal we want to identify students whose academic status may go down, it can be seen that the better prediction is got in the Bayesian network; it is effective in predicting student performance. This means that there is no simple linear relationship between the one-year and two-year results. In fact, because learning of the first year is mainly about the basic courses and more specialized courses start from the second year, the relationship cannot be expressed as linear simply. We consider, since the Bayesian network can treat the non-linear object without difficulty, the Bayesian network is suitable in this study.

We also should pay attention to the results of Tables 7 and 9. We know that students with the higher probability value went down actually. Thus, we can consider the probability value output of “risk of going down”. For example, we give them the strong warning for the students with the “YES” probability is close to 100% and give them the weak warning for the students from around 30% to 50%. This is said to be flexible teaching. Examples of academic advice are shown in Table 14. Because this idea is difficult to achieve using other methods, the Bayesian network model makes sense for our research proposal.

5. Conclusion

In this paper, we proposed a Bayesian network for identifying students who would benefit from academic
We demonstrated that the decision about whether broad and flexible academic advice should be given to a student should not be made by using a simple prediction classification, but by using the posterior probability value of a Bayesian network forecasting model that identifies an “anxious student”. Moreover, since the computational cost of using a Bayesian network is low, two or more models can be run simultaneously. The academic advice that is given can be adapted according to the results of the forecasting model. We showed that the Bayesian network is useful as a technique for predicting whether academic advice should be given to a student.

However, not every candidate for educational guidance is actually identified. In Model A, appropriate counseling by an academic advisor was not given to two of the 24 students who would have benefited from it. Moreover, in order to build a more general-purpose model, it is necessary to use the student’s data for another academic year as study and test data. For this purpose, it is necessary to gain a better understanding of the Bayesian network.

In addition, further discussion on model building and validation is needed. For example, for the definitions of variables, it may be necessary to also consider and adopt the application of techniques such as factor analysis or data of a different format. In addition, we did not study conditional probability so much. In a Bayesian network, since the conditional probability is one of the important elements, we need to look at it in future research.

Moreover, the concept of a concrete and overall model for practical use is also required. We plan to investigate these issues in the future.

In addition, although we use only grade data, we want to improve the accuracy by adding attendance data from the IC card roll call system.

### References

Hirotaka Itoh received B.Eng., M.Eng. and Ph.D. degrees from Nagoya Institute of Technology in 1994, 1996 and 1999, respectively. He joined the Educational Center for Information Processing, Nagoya Institute of Technology in 1999. He is currently Assistant Professor of the Information Technology Center, Nagoya Institute of Technology. His research interests include e-Learning, educational technology and data mining. He is a member of the Institute of Electrical Engineers of Japan (IEEJ) and the Japan e-Learning Association (JeLA).

Keisuke Itoh received B.Eng. degree from Nagoya Institute of Technology in 2012. He is currently a master student in Nagoya Institute of Technology. His research interests include educational technology and data mining.

Kenji Funahashi received B.Eng. degree from Gifu University in 1993. He received M.Eng. and Ph.D. degrees from Nagoya University in 1995 and 1998, respectively. He joined the Educational Center for Information Processing, Nagoya Institute of Technology in 1998. He was a Visiting Researcher, University of British Columbia in 2003. He is currently Associate Professor of the Information Technology Center, Nagoya Institute of Technology. His research interests include virtual reality and computer graphics. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), the Virtual Reality Society of Japan (VRSJ), the Institute of Electronics, Information and Communication Engineers (IEICE) and the Japanese Society for Medical and Biological Engineering (JSMBE).