A Decision Support System for Evaluating International Student Applications

Nguyen Thi Ngoc Hien and Peter Haddawy
Computer Science and Information Management Program, Asian Institute of Technology
NguyenThiNgoc.Hien@ait.ac.th, Haddawy@ait.ac.th

Abstract - **In today's transnational admission environment, evaluating applicant qualifications is becoming increasingly challenging. While standardized tests can be helpful, studies have shown that they are rather noisy predictors of performance. Predicting educational outcome is a viable alternative in such heterogeneous environments. Performance prediction models can be built by applying data mining techniques to enrollment data. In this paper we present an approach to using Bayesian networks to predict graduating cumulative Grade Point Average based on applicant background at the time of admission. While such prediction models can be helpful, their recommendations may not be followed by departmental faculty members making admission decisions if they are presented as black boxes. We thus present a novel approach to deriving a case-based retrieval mechanism from the Bayesian network prediction model in such a way that the similarity measure used by the case-based system is consistent with the predictive model. The case-based component retrieves the past student most similar to the applicant being evaluated. The Bayesian network model is evaluated using stratified ten-fold cross validation.**

Index Terms - Bayesian networks, case-based retrieval, student evaluation.

INTRODUCTION

The Asian Institute of Technology (AIT) is an autonomous postgraduate institution with a highly international student population. The institute has an enrollment of 2000 students from 45 different countries with the majority nationality comprising only 35% of the total student body. Each year we receive applications from students who have completed previous bachelors or masters degrees at any one of approximately 600 different institutions. This extreme diversity in applicants for admission makes accurate evaluation a highly challenging task. The evaluation of applications has been traditionally performed by faculty members who have some degree of knowledge of each particular country's educational system. Unfortunately, it is not always possible to find a faculty member in each program who has good knowledge of each country's system of higher education. We were thus motivated to produce a decision support system to provide a more methodical approach.

Research on academic performance [3, 4] suggests using student outcome as a good basis to assess applicants' qualifications. A performance prediction model can be built by applying data mining to available admission and graduation grade point average data. Fortunately, AIT has a large database of information on past and current applicants. We decided to make use of this information by applying data mining techniques to predict student performance at AIT based on the information contained in the application.

Decision support systems have been built to help advisors instruct students in choosing suitable courses and appropriate study plans [1, 2]. Previous work on student performance prediction used logistic regression to examine the impact of various factors on student performance [1]. Bekele and Menzel [5] used Bayesian networks to predict mathematics performance of high school students. Their model categorized students into three categories: below satisfactory, satisfactory, and above satisfactory. The work reported in the present paper differs from theirs in the highly international nature of the applicant pool and the more fine grained prediction.

In this paper we present an approach to using Bayesian networks to predict graduating cumulative Grade Point Average based on applicant background at the time of admission. Evaluation by stratified ten-fold cross validation on three years of admissions data shows the model has a mean absolute prediction error of 0.22 grade point for master program applicants and 0.20 for doctoral program applicants. While the Bayesian prediction model can provide valuable information to departmental faculty members in making admission decisions, it functions as a black box and thus faculty members may not completely trust the predictions, despite the published accuracy. They may be more comfortable with the predictive results if the system can show them the past student most similar to the applicant being considered. Thus we present a novel approach to integrating a case-based component with the prediction model. The challenge here is to define similarity of cases (applicants) in a way that is consistent with the prediction model. Our approach is to use the prediction model itself to compute the similarity.

The rest of this paper is organized as follows. Section 2 provides background on Bayesian networks, the methodology for their application to this problem and the approach for integrating case-based component with the Bayesian model in the context of this problem. Section 3 presents empirical evaluation of the prediction model and of the case-based

component. The paper ends with conclusions and directions for future work.

METHODOLOGY

I. Bayesian Networks

A Bayesian network [6] is a graphical representation of a probability distribution. It is a directed acyclic graph in which nodes represent random variables and links represent probabilistic influences between the variables. Probabilistic dependence and independence are expressed by the presence or lack of paths between nodes in the graph. The fact that probabilistic dependence is encoded in the network topology in this way permits probability distributions over large numbers of random variables to be compactly represented and permits calculations to be performed efficiently. Due to the inherent uncertainty of the performance prediction problem, we chose to use Bayesian networks for the modeling task. Using a probabilistic model has the advantage that it can later become a component of a higher level optimization model.

II. Data Pre-Processing

At AIT coursework is generally completed within the first year of study. So we used the grade point average (GPA) accumulated after the first year as the dependent variable. The numeric GPA at AIT ranges from 0 to 4.0, which is also translated into a letter grade with values of A, B+, B, C+ and Fail (C or below). The number of students classified as Fail is low due to an institute policy of continuous review of students and either special tutoring for or dismissal of students with low grades.

Based on research on student performance [3, 5] and available attributes in the admission data, ten attributes were chosen as predictors of performance: age, gender, marital status, nationality, English test score, institute of the previous degree, major of the previous degree, GPA of the previous degree, field of study and degree program (master or doctor) to which the applicant is applying. The attributes major and field of study (FoS) have a large number of values so they were processed by clustering the values into groups of similar majors and similar fields of study.

The attributes nationality and institute of the previous degree have large numbers of values (86 and 1700, respectively) without any intrinsic meaning. We thus decided to transform them into more meaningful values.

The socio-economic environment can play a major role in the performance of students. So we used the World Bank classification of countries according to their Gross National Income (GNI) to group countries into GNI categories of LIC (Low Income), LMC (Low Middle Income), UMC (Upper Middle Income), NOC (High Income, non-OECD), and OEC (High Income, OECD).

The most significant factor concerning the university of the previous degree is the quality of the academic programs there. We gauged this by correlating the GPA of the previous degree with that obtained at AIT. If students consistently enter AIT with somewhat average grades from a particular university but graduate with high grades, we take this difference as evidence of the high quality of that institution and vice versa otherwise. To account for the nonlinear nature of the grading scale at AIT, the GPA distance was calculated by the difference between the incoming GPA and the squared outgoing GPA from AIT. We rated institutions on continuous scale from 1 to 100. With this rating scale, we found that universities offering primarily programs in economics and social sciences were getting ranked quite highly. The reason for this was discovered that graduates from these schools were typically applying to our school of management, where grades tend to be higher than in our other two schools as shown in Table I. So we normalized the AIT grades to account for differences in the average grades given in the institute's three schools. This resulted in a more accurate ranking of universities.

TABLE I
AVERAGE OF CGPA BY SCHOOLS

School	Average of CGPA
AT (School of Engineering and Technology)	3.36
ED (School of Environment, Resources and Development)	3.43
SM (School of Management)	3.47

II. Development of the Model

To train and test the model we used admissions data from 2003 through 2005 consisting of 1386 masters and 212 doctoral students. The distribution of CGPA by degree (doctor and master) and the distribution of age by degree are shown in Figure 1 and Figure 2.

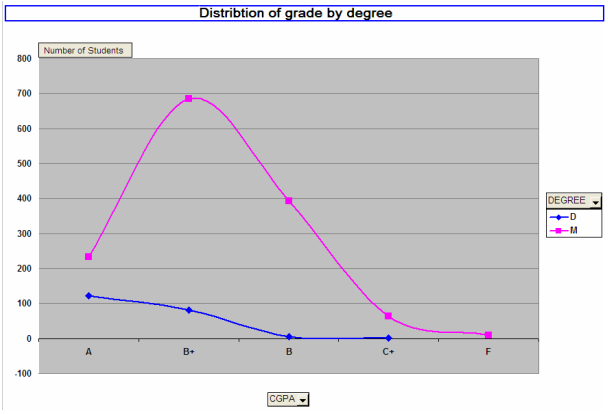

FIGURE 1
THE DISTRIBUTION OF GRADE BY DEGREES

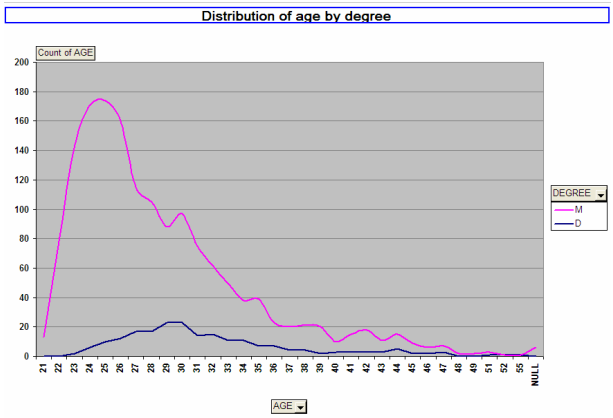

FIGURE 2
THE DISTRIBUTION OF AGE BY DEGREE

As would be expected, grades are distributed quite differently between doctoral and master students, with doctoral students tending to receive higher grades than masters students, as shown in Figure 1. Doctoral students also tend to be older than master students, as shown in Figure 2. This difference in distributions leads us to adopt separate prediction models for master and doctoral students. Initial experimentation with a number of different Bayesian network structures showed that the simple Naïve Bayesian model structure shown in Figure 3 produced the best results when trained/tested on the master student data and on the doctoral student data. Although the structures of the two models were identical, the probabilities in the models were different since they were trained on different data sets.

In the figure, predictor attributes are displayed along the sides and top of the graph. The model contains two nodes for the predicted variable CGPA (cumulative GPA). The one from which the edges to the other attributes emanate is a discrete valued node and the other with the double edge is continuous. The continuous CGPA represents a set of conditional Gaussian distributions. This was done in order to be able to provide a numeric CGPA prediction under the limitation that Bayesian networks do not permit continuous variables to be parents of discrete variables. The model was built using the Hugin Researcher 6.5 software. When the values of the predictor variables are specified the model provides a probability distribution over the predicted letter grades and a numeric CGPA prediction, corresponding to the expected value of the CGPA.

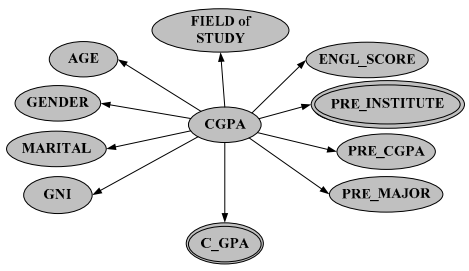

FIGURE 3
NAÏVE BAYESIAN MODEL FOR PERFORMANCE PREDICTION

III. Refinement of the Models

Sensitivity analysis is the study of how uncertainty in the output of a model can be apportioned to different source of uncertainty in the model input [8]. Sensitivity analysis can be used to determine which input parameters contribute the most to output variability and which parameters are insignificant and can be eliminated from a model [7]. There are a number of variance-based methods used in sensitivity analysis such as Correlation Ratios or Importance Measures, Sobol Indices and FAST Indices. The first method measures the main effect contribution of each input on the output variance while the two others are capable of computing sensitivity of the output by taking into account the interaction between inputs. The importance measure of each input parameter is computed by estimating the following quantity [7, 8]:

$$I_X = \frac{Var_X[E(Y|X)]}{Var(Y)} \quad (1)$$

where Y is the output variable or C_GPA in this problem; X is each input variable of the model; and $E(Y|X)$ is the expectation of Y given a possible value of X. The numerator in (1) denotes the variance of $E(Y|X)$ taken over all possible values of X, and the denominator denotes the variance of Y.

Here we use sensitivity analysis computing importance measures to assess the importance of each predictor variable on the CGPA and to judge which predictors can be eliminated to refine the model. Analysis results are displayed for the master and doctoral degree models in Table II and Table III respectively.

TABLE II
SENSITIVITY ANALYSIS FOR MASTER MODEL

Predictor variables	Importance measures (%)
PRE_INSTITUTE	91.82
PRE_CGPA	2.73
GNI	1.70
ENGL_SCORE	1.54
FIELDofSTUDY	0.78
PRE_MAJOR	0.53
MARITAL	0.41
AGE	0.33
GENDER	0.16

TABLE III
SENSITIVITY ANALYSIS FOR DOCTORAL MODEL

Predictor variables	Importance measures (%)
PRE_INSTITUTE	73.71
PRE_CGPA	22.25
GNI	1.05
FIELDofSTUDY	1.03
PRE_MAJOR	0.87
ENGL_SCORE	0.71
AGE	0.35
MARITAL	0.03
GENDER	0.00

Sensitivity analysis reveals the dominance of variables PRE_CGPA, PRE_INSTITUTE over the others, especially in the master model, as well as the insignificance of variables GENDER, MARITAL, AGE. Consequently, we experimented to determine whether eliminating these three variables would

increase the model's accuracy. The experimental results reported in the next section confirmed that eliminating GENDER from the master model and eliminating GENDER and MARITAL from the doctoral model improved each model's accuracy.

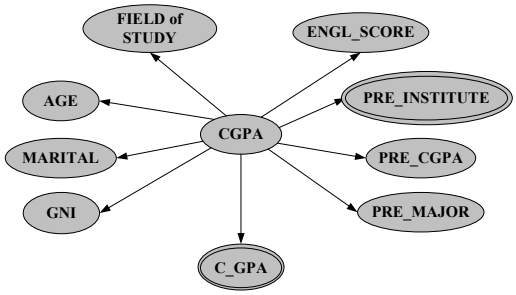

FIGURE 4
FINAL PREDICTION MODEL FOR MASTER STUDENTS

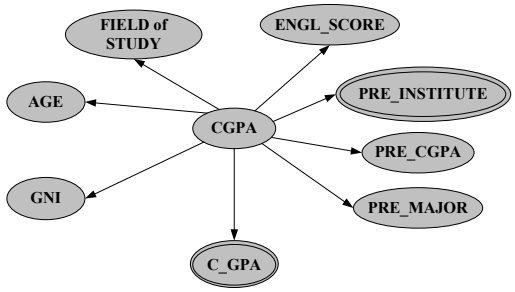

FIGURE 5
FINAL PREDICTION MODEL FOR DOCTOR STUDENTS

IV. Case-Based Component

The objective of this component is to retrieve the past student (a case) in the database (case base) who is most similar to the applicant being evaluated (the new case). A case or a student is described by the same ten predictor variables or attributes as used in the Bayesian model. Similarity between two cases is estimated by computing the local similarity of the pair of values within each attribute and then summing the local similarities over all attributes:

$$Sim(C_1, C_2) = \sum_{i=1}^{n} localSim(C_{1i} = a_{ik}, C_{2i} = a_{il}) \qquad (2)$$

where $Sim(C_1,C_2)$ is the similarity metric between case C_1 and case C_2; C_{1i} and C_{2i} denote the attribute i^{th} of C_1 and C_2; a_{ik} and a_{il} denote possible values of the attribute i^{th}; n is the number of attributes representing a case.

The challenge here is how to calculate local similarities within each attribute or similarities between values pair wise of the same attribute. Our solution takes advantage of the prediction from the Bayesian model. The similarity between two values a_k, a_l of the attribute A is expressed as the inverse of their dissimilarity which is estimated by the difference between two expected CGPAs given each value of A:

$$localSim_A(a_k, a_l) = -localDisSim_A(a_k, a_l) = \frac{\left|E(CGPA \mid A = a_k) - E(CGPA \mid A = a_l)\right|}{\max CGPA - \min CGPA} \qquad (3)$$

where $E(CGPA|A=a_k)$ is the expected CGPA given A getting value of a_k; *maxCGPA* and *minCGPA* are the maximum and the minimum CGPA. The denominator of (3) is a normalization factor to scale $localSim_A$ into the range from 0 to 1.

To illustrate the similarity measure, Table IV shows local dis-similarities of the attribute Pre_CGPA according to prediction model for master students. This attribute has six values A, B+, B, C+, C and F that are labeled in row and column headers. Each cell of the table shows the dis-similarity between the grade specified in the row header and the one in the corresponding column header. Therefore, the table is symmetric. A grade of B+ is more similar to a grade of A than to a grade of B. This is consistent with the grading policy and tradition of the institute.

Since faculty members do not find it useful to retrieve cases from other fields when evaluating an applicant, we treat the field of study attribute as a hard constraint requiring an exact match. This also results to the retrieval space becoming smaller and speeding the retrieval process.

TABLE IV
LOCAL DIS-SIMILARITIES OF THE PRE_CGPA IN MASTER MODEL

	PRE_CGPA					
	A	B+	B	C+	C	F
A	0	0.015	0.138	0.208	0.222	0.300
B+	0.015	0	0.123	0.193	0.207	0.285
B	0.138	0.123	0	0.070	0.084	0.162
C+	0.208	0.193	0.070	0	0.014	0.092
C	0.222	0.207	0.084	0.014	0	0.078
F	0.300	0.285	0.162	0.092	0.078	0

EVALUATION AND RESULTS

To train and test the model we used admissions data from 2003 through 2005 consisting of 1386 masters and 212 doctoral students. Because variables PRE_INSTITUTE and PRE_CGPA showed their dominant importance through sensitivity analysis, cases missing values of these two variables were eliminated from the training and test set. Therefore, there remain 962 masters and 170 doctors. In the data, the master students' CGPA had mean value of 3.380 and variance of 0.158; the doctoral students' CGPA had mean value of 3.731 and variance of 0.068.

Stratified ten-fold cross-validation was used to evaluate the model. The data was separated into folds so that each fold had the same distribution of grades as the entire data set. To predict the letter grade, we used the grade with the maximum probability. The confusion matrices showing the predicted versus actual results are shown in Table V and Table VI for master and doctoral models respectively. We evaluated the

numeric CGPA prediction using the mean absolute error (MAE), which is defined as

$$MAE = \frac{\sum_{i=1}^{n} | a_i - p_i |}{n} \quad (4)$$

where a_i is the actual CGPA; p_i is the predicted CGPA; and n is the number of data points. The average mean absolute error over the ten folds was **0.21 for the master model** and **0.16 for the doctoral model**.

TABLE V
CONFUSION MATRIX OF MASTER MODEL

Actual Class	Predicted Class					
	A	B+	B	C+	Fail	
A	68	99	8	0	0	175
B+	53	345	56	2	0	456
B	4	117	151	13	0	285
C+	0	3	27	11	0	41
Fail	0	2	1	2	0	5
	125	566	243	28	0	962

TABLE VI
CONFUSION MATRIX OF DOCTOR MODEL

Actual Class	Predicted Class					
	A	B+	B	C+	Fail	
A	73	26	0	0	0	99
B+	32	36	0	0	0	68
B	0	1	0	0	0	1
C+	0	2	0	0	0	2
Fail	0	0	0	0	0	0
	105	65	0	0	0	170

To compare performance between experimental models, a relative measure is more appropriate than an absolute one. We used the relative squared error which is defined in (5) for this purpose.

$$RSE = \frac{\sum_{i=1}^{n} (p_i - a_i)^2}{\sum_{i=1}^{n} (a_i - \bar{a})^2} \quad \text{where } \bar{a} = \frac{\sum_{i=1}^{n} a_i}{n} \quad (5)$$

Table VII shows the little improvement in performance of the final master model in Figure 4 and the final doctoral model in Figure 5 (after eliminating insignificant attributes) compared to ones of the model proposed in Figure 3 (without eliminating insignificant attributes).

TABLE VII
COMPARISON MODEL PERFORMANCES BEFORE AND AFTER ELIMINATING INSIGNIFICANT ATTRIBUTES

	MAE		RSE	
	Before	After	Before	After
Master model	0.215	0.214	0.603	0.597
Doctoral model	0.163	0.163	0.787	0.786

Performance of the case-based component was evaluated on the pool of master degree applications for the year 2006. Some cases from different field of study were randomly selected for assessment. Some of case retrieval results are displayed in Table VIII. Highlighted rows represent test cases followed by the three most similar past cases with the dissimilarity metric shown in the column labeled METRIC. The CGPAs shown for test cases are predicted CGPAs, while those for retrieved cases are actual CGPAs. For test cases that happen to be enrolled students, the actual CGPA is shown in italics beside the predicted one.

TABLE VIII
CASE RETRIEVAL

AGE	GENDER	MARITAL	GNI	INSTITUTION RANK	PREVIOUS MAJOR	PREVIOUS CGPA	ENGLISH	FOS	CGPA	METRIC
30	M	S	LMC	70	AgrForMari	C+	N	Civil	**3.35219**	
29	M	S	LMC	70	AgrForMari	C+	N	Civil	3.12909	0
28	M	S	LMC	70	ConsTrans	B	N	Civil	3.79909	0.13568
29	M	S	LMC	72	ConsTrans	C+	Y	Civil	3.03	0.1675
28	M	S	LIC	56	AgrForMari	B	Y	Development	**3.14591** ***(3.35)***	
29	M	S	LIC	61	Economics	B	Y	Development	3.04444	0.0549
31	F	S	LMC	54	ArtSolLawPol	B	Y	Development	3.14	0.16464
27	M	S	LIC	44	ConsTrans	B	Y	Development	2.62778	0.18644
32	M	S	LMC	70	ISE	B+	N	Energy&Env	**3.39531** ***(3.42)***	
30	M	S	LMC	70	ISE	B+	N	Energy&Env	3.09438	0.01376
25	M	S	LMC	70	ISE	B+	N	Energy&Env	3.445	0.05336
29	M	S	LMC	64	ISE	B+	N	Energy&Env	3.32818	0.07376
26	M	S	LMC	48	IT	B	NULL	ICT	**2.99109**	
26	M	S	LMC	47	IT	B	NULL	ICT	2.94571	0.01
25	M	S	LMC	52	IT	B	E1	ICT	3.218	0.04
25	F	S	LIC	46	IT	B	Y	ICT	2.69667	0.07704
30	M	S	LIC	54	ISE	B	E1	ISE	**3.11735** ***(2.94)***	
30	F	S	LIC	54	ISE	B	E1	ISE	3.035	0
30	M	S	LMC	55	ISE	B	C	ISE	3.12889	0.07554
28	M	S	LMC	52	ISE	B	NULL	ISE	2.876	0.07704
27	M	S	LMC	70	EnvEnergy	B+	E4	Resources	**3.65744**	
29	F	S	LMC	70	AgrForMari	B+	Y	Resources	3.72	0.13566
23	F	S	LMC	68	AgrForMari	B+	E4	Resources	3.5	0.16914
23	F	S	LMC	70	AgrForMari	A	E3	Resources	3.72	0.20545
30	F	M	LIC	66	ArtSolLawPol	C+	Y	SOM	**3.2887** ***(3.67)***	

27	F	M	LIC	65	ArtSolLawPol	B	Y	SOM	3.5	0.07924
27	M	S	LIC	65	ISE	C+	Y	SOM	3.38571	0.095
43	M	M	LIC	65	ISE	B	Y	SOM	3.57962	0.0955

* ENGL_SCORE indicates the English proficiency of applicants. Because of the change in coding this information in admission data at AIT over time, ENGL_SCORE can get one of following values:

- N - not proficient
- Y - proficient
- C - national certificate of English
- E1 - IELTS score in 4.5 - 5.4
- E2 - IELTS score in 5.5 - 5.9
- E3 - IELTS score in 6.0 - 6.4
- E4 - IELTS score in 6.5 - 7.4
- E5 - IELTS score in 7.5 - 9.0

CONCLUSION

In today's transnational admissions environment, educational institutions are facing the need for a precise and thoughtful method to evaluate and select the most qualified applicants graduating from various institutions in many countries. Our approach using Bayesian networks combined with case-based reasoning for predicting applicant performance has the potential to meet this need. The technique can be applied at any institution that has a good database of student and applicant information.

Several directions remain open for future work. The most important is to deal with the imbalance in the data. We have relatively large amounts of data for students with high GPA's and little for students at the low end. This causes the model to overestimate the GPA for students with grades B, C+, and Fail. One way to overcome this may be to use data concerning students who have not been admitted, assuming that the judgment not to admit accurately reflects the students' likelihood of not performing well.

REFERENCES

[1] Chowdhury A. A., "Predicting success of a beginning computer course using logistic regression", *ACM conference on Computer Science*, 1987, p449.

[2] Dekhytar A., Goldsmith J., "The Bayesian advisor project", online at http://www.cs.engr.uky.edu/~goldsmit/papers/#BAP.

[3] Hadkkinen I., "Do University entrance exams predict academic achievement?", *Working Paper Series*, Department of Economics, Uppsala University, 2004.

[4] Golding P., Donaldson O., "Predicting academic performance", *Proc. 36th ASEE/IEEE Frontiers in Education Conference*, 2006, 21-26.

[5] Bekele R., Menzel W., "A Bayesian approach to predict performance of a student (BAPPS): A Case with Ethiopian Students", *Proc. IASTED International Conference on Artificial Intelligence and Applications*, 2005.

[6] Jensen F., "Bayesian Networks and Decision Graphs", *Springer-Verlag*, 2002.

[7] Chan K., Saltelli A., Tarantola S., "Sensitivity analysis of model output: Variance-based methods make the difference", *Proc. of the Winter Simulation Conference,* 1997.

[8] Saltelli A., "Sensitivity analysis for importance assessment", *Risk Analysis,* Vol. 22, Issue 3, June 2002, 579-590.