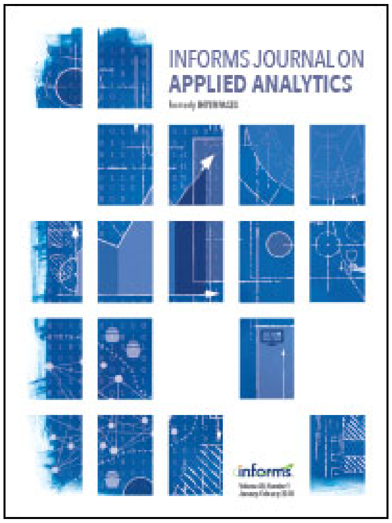


INFORMS Journal on Applied Analytics

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Optimizing the Multistage University Admission Decision Process

Donghyun Kim, Namyong Kim, Junoh Cho, Hayong Shin

informs
http://pubsonline.informs.org/journal/inte
INFORMS JOURNAL ON APPLIED ANALYTICS
Vol. 49, No. 6, November–December 2019, pp. 422–429
ISSN 0092-2102 (print), ISSN 1526-551X (online)

Optimizing the Multistage University Admission Decision Process

Donghyun Kim,[a] Namyong Kim,[a] Junoh Cho,[a] Hayong Shin[a]

[a] Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea

Contact: dhk618@kaist.ac.kr, https://orcid.org/0000-0002-5049-821X (DK); steen@kaist.ac.kr (NK); jocho@kaist.ac.kr (JC); hyshin@kaist.ac.kr (HS)

Received: August 7, 2017
Revised: February 22, 2018; November 29, 2018; April 13, 2019
Accepted: April 17, 2019
Published Online in Articles in Advance: November 1, 2019

https://doi.org/10.1287/inte.2019.1009

Abstract. The admission decision process is an important operational management problem for many universities. Admission control processes may, however, differ among universities. In this paper, we focus on the problem at Korea Advanced Institute of Science and Technology (KAIST). We assume that individual applications are evaluated and ranked based on paper evaluations and (optional) interview results. We use the term "university admission decision" to mean determining the number of admission offers that will meet the target number of enrollments. The major complexity of an admission decision comes from the enrollment uncertainty of admitted applicants. In the method we propose in this paper, we use logistic regression with past data to estimate the enrollment probability of each applicant. We then model the admission decision problem as a Markov decision process from which we formulate optimal decision making. The proposed method outperformed human expert results in meeting the enrollment target for the validation data in 2014 and 2015. KAIST successfully used our method for its admission decisions in academic year 2016.

History: This paper was refereed.
Funding: This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and ICT [Grant 2017R1A2B4006290].

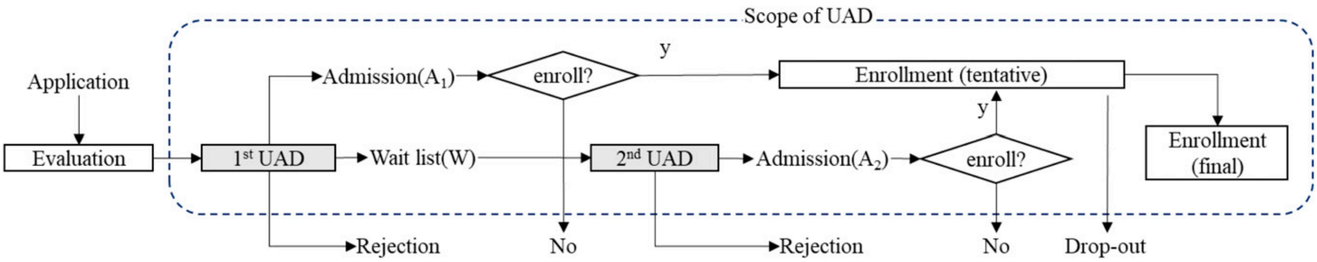
Keywords: university admission process • enrollment probability estimation • Markov decision process • logistic regression

Making a decision about admitting applicants to a university is an important operational management decision problem at a university (Chade et al. 2014). Because admission control processes may differ among universities, we present the problem that Korea Advanced Institute of Science and Technology (KAIST) faced. Based on the capacity of facilities and academic programs, KAIST's target is to enroll about 750 freshmen each year. The university evaluates individual applications and ranks them based on the applicants' documents and interviews. In this paper, the term "university admission decision" (UAD) refers to determining the number of admission offers a university should make to meet its target number of enrollments. The main complexity of UAD comes from the enrollment uncertainty of the admitted applicants. Unlike many other universities in Korea that meet their enrollment target after many rounds of offering admission to candidates on a waiting list, KAIST uses only two rounds of admission offers (Figure 1). KAIST uses only two rounds as the result of a strategic decision that goes beyond the scope of this paper. When the gap between the target and the actual enrollment reached about 10% of the target in 2014, the KAIST admissions office and the authors performed a study of the UAD problem based on past data. We present that study and its results in this paper.

Below, we explain the KAIST admissions and enrollment process (Figure 1).

- Each applicant receives an evaluation score based on a review of the documents that applicant has submitted and an optional interview.
- Applicant admissions are determined based on the evaluation scores in descending order. The number of admissions offers generated from the UAD process determines the admission of individual applicants.
- UAD comprises two stages. In the first UAD stage, we determine the number of admissions (A_1) and the size of the waiting list (W). Applicants ranked lower than (A_1+W) are rejected. Those who receive an admissions offer in the first UAD (called the A_1 group hereafter) may choose to enroll (or not), because they usually have applied to multiple universities, and their preferences differ. However, decisions to enroll are not final. Later, they can drop out of the tentative enrollment list by canceling their registration at a small penalty cost.
- After the initial tentative enrollment (also called registration), KAIST conducts the second UAD stage, which determines the number of admissions (A_2) from the waiting list.

Figure 1. (Color online) The KAIST Admission and Enrollment Process Flow



Tentatively enrolled applicants may drop out until the cutoff date, which is usually the start of the new academic year. Those who remain enrolled until that date constitute the actual enrollment. The objective is to as closely as possible match the number of actual enrollments to the target. The penalty for shortage and overage may differ. The target number is carefully chosen each year by considering the capacity of the university (e.g., dormitory, laboratory, and faculty size) and some strategic factors (e.g., government policy and long-term policy change plan).

A few related studies on the applicant admissions process are available. Rebbapragada et al. (2010) quantified the value of an applicant to a university using data-mining techniques by predicting the applicant's freshman grade point average. Using these values, they maximized the overall quality of admitted applicants by using revenue-management techniques. Walczak and Sincich (1999) compared the results of a neural network model for applicant enrollment decision modeling with the results of logistic regression analysis to demonstrate improvements that can be obtained by using neural networks. Maltz et al. (2007) implemented an enrollment model using a financial-aid matrix and other predictors to optimize enrollment by using a logistic regression function and a financial objective function. To the best of our knowledge, no study has built a sequential decision-making model and used enrollment probability to solve the optimal decision problem for UADs.

From a preliminary study of past enrollments, we found that the individual enrollment probability varies significantly, and there are some predictable patterns in enrollment probability as a function of applicants' attributes, such as evaluation scores, high school types (e.g., public, private, or special purpose), and home location. Figure 2 compares the conventional UAD of human experts with the overall structure of the UAD support system presented in this paper. Key differences are that the proposed system predicts individual enrollment probabilities rather

Figure 2. (Color online) We Compare the Overall Structure of the KAIST UAD Support System to the Conventional Process

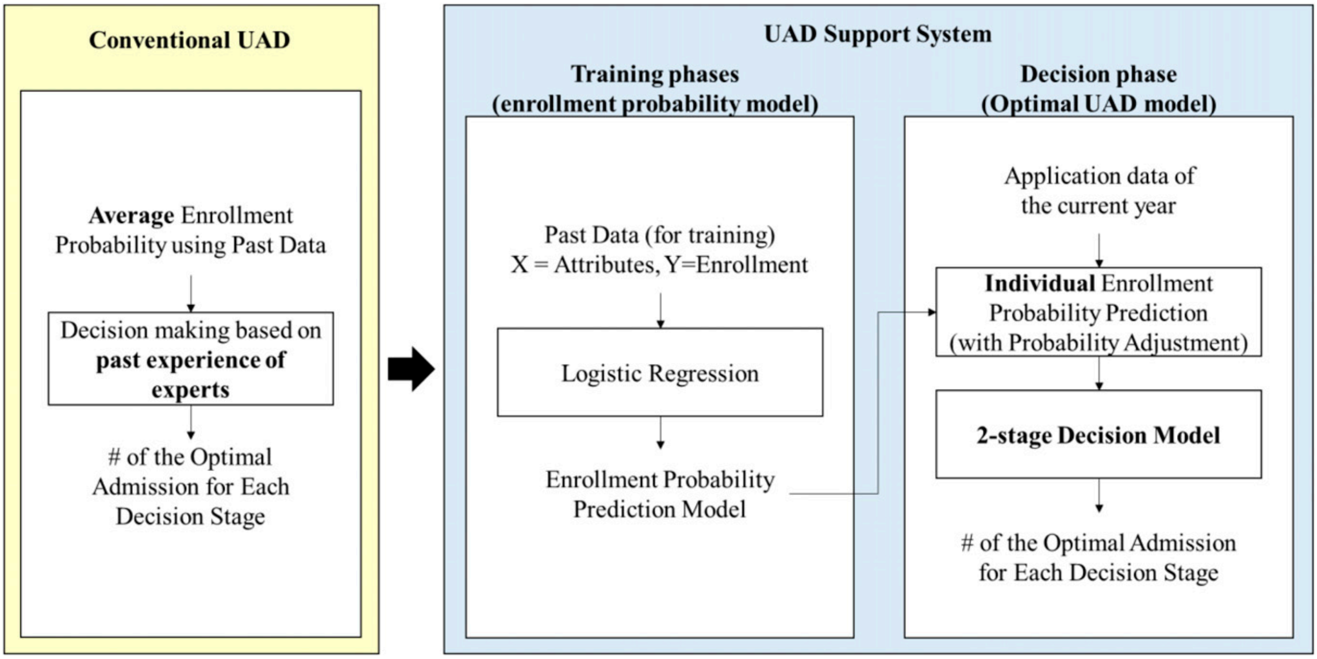
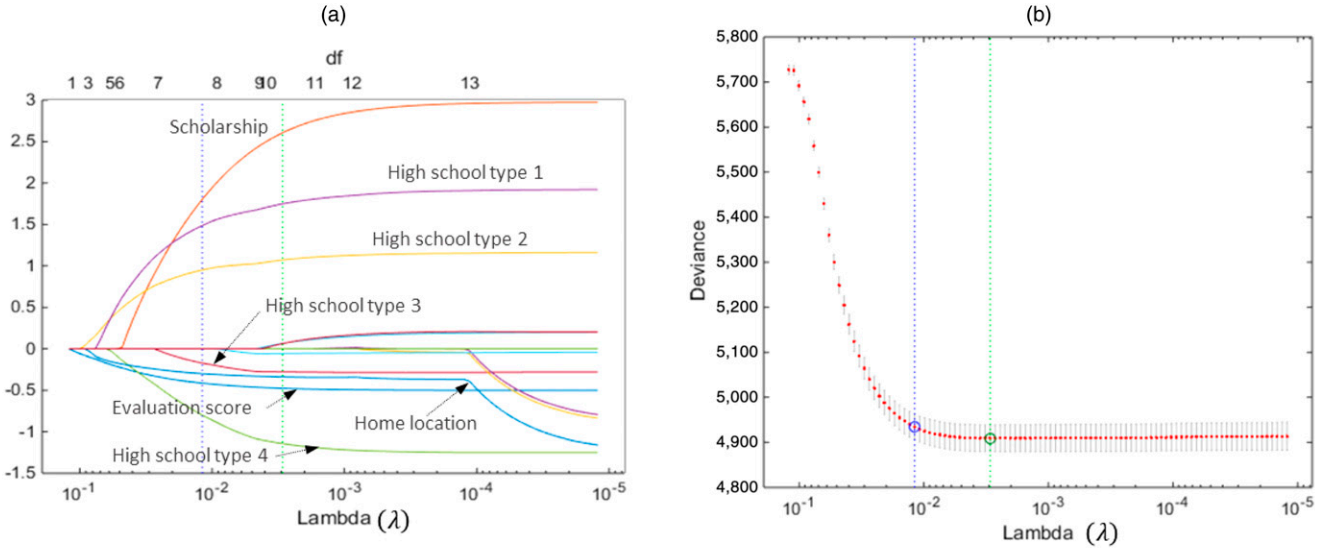


Figure 3. (Color online) The Graphs Illustrate Variable Selection for the Enrollment Probability Model



Notes. (a) Trace plot of coefficients fit by Lasso. (b) Cross-validated deviance of Lasso fit.

than average probability, and it uses them for optimal UAD. The proposed UAD support system comprises two phases. In the training phase, the enrollment probability prediction model learns parameters from past application and enrollment data by using logistic regression. In the decision phase, we estimate the individual enrollment probability for each applicant in the current year. Then, we formulate this phase as a two-stage decision problem and solve the problem to find the optimal number of admissions to minimize the expected loss. In the sections below, we explain each phase in detail and discuss the results.

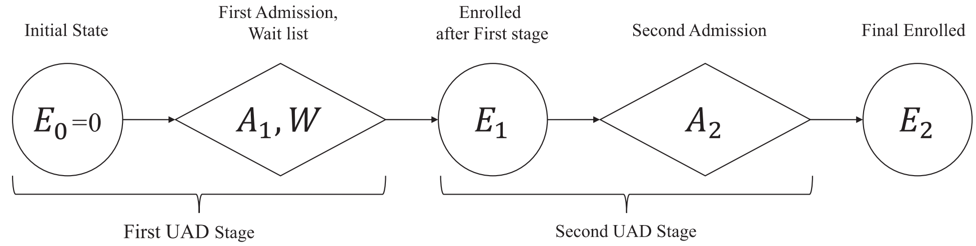
Enrollment Probability Estimation

In this section, we describe the prediction of enrollment probability using logistic regression (Cox 1958). In the KAIST conventional UAD, the overall enrollment probability was computed from the enrollment records of the previous years. Hence, if the overall enrollment probability was 70% and the target number was 600, then the number of admissions was naively determined to be 857 (600/0.7). To increase the prediction accuracy, experts in the admissions office grouped applicants based on their key attributes, such as high school type, and computed the enrollment probability for each group. In our proposed method, we estimate individual enrollment probability to overcome the seemingly obvious limit of the conventional approach.

We collected the application-enrollment data set for 3 years (2013–2015) to train the prediction model. In the original data set, each applicant record has about 20 attributes. To avoid overfitting, we used the Lasso procedure, as explained by Tibshirani (1996). Figure 3(a) shows a trace plot of logistic regression coefficients corresponding to independent variables (i.e., regressors). As the Lasso regularization parameter λ grows (from right to left in the graph), the regression coefficients gradually vanish. Figure 3(b) shows the 10-fold cross-validated deviance statistics over λ. The curve is the average of the deviance for each λ, shown with ± 1 standard deviation. The right circle in the curve is where the deviance is minimum, and the left circle is the closest point to the minimum deviance plus one standard deviation. We select regressors with nonzero coefficients when λ is set to that of the left circle point, as suggested in the *one-standard-error* rule by Krstajic et al. (2014). As a result, four variables are chosen: evaluation score, scholarship, home location, and high school type. High school type is a categorical variable of four types. For confidentiality reasons, we represent each type as a binary dummy variable (i.e., high school types 1–4). So, we select seven variables.

We also examined the interaction effects of independent variables by including product terms (such as multiplication of evaluation score and scholarship). However, adding higher-order terms did not meaningfully improve the prediction accuracy. By reapplying logistic regression using only the seven selected variables, we obtained the model for the enrollment probability of the ith applicant, as shown in Appendix A. By applying the logistic regression model trained with the past 3 years' data to the current applicants, we obtained their predicted enrollment probability. At this stage, human experts in the admissions office can intervene by subjectively adjusting the probability to reflect changes in the

Figure 4. The Flow of the Two-Stage Optimal UAD



business environment. Although the UAD support system provides some functionality to support this expert intervention, we omit the details for brevity.

University Admission Decision Problem

Figure 4 shows the flow of the two-stage optimal UAD decision-making process. The rhombus node in the diagram denotes a decision-making node, and the circle node is the enrollment-state vector representing the enrollment of each applicant at each decision stage. Initial state $\boldsymbol{E_0}$ is a zero vector because no student has yet enrolled. Here, $\boldsymbol{E_1}$ is the state vector after the first-stage decision. Note that the enrollment at this time is only tentative because applicants can cancel their registration at any time. Here, $\boldsymbol{E_2}$ is the enrollment-state vector at the final cutoff date (after the second-stage decision, second enrollment, and cancellations). Let $|\boldsymbol{E_2}|$ denote the final number of enrollments. Our goal is to make $|\boldsymbol{E_2}|$ as close to the *target* as possible.

The decision variables are the number of admissions (A_1) and the size of the waiting list (W) for the first stage and the number of second-stage admissions (A_2) from the waiting list for the second stage. Applicants who are admitted in the first and second stages are called the A_1 group and A_2 group, respectively. The A_2 group should clearly be a subset of the waiting list.

One may think that the first-stage decision can be made easily by setting $A_1 = target$ and a very large W (e.g., $W =$ the rest of the applicants). Setting A_1 smaller than the target is clearly unnecessary. If we set $A_1 = target$, which is the minimum value for A_1, the enrollment will certainly be smaller than the *target*, and we will lose the opportunity to see the enrollment results for some applicants who are ranked lower than A_1. However, if A_1 is too large, we are in danger of overage. Similarly, if we choose a small value for W and the number of enrolled applicants in the first stage is fewer than expected, achieving the target will be impossible. On the other hand, if we choose too large a value for W at the first stage, it may harm the university's reputation factor. Thus, we need some quantitative definition of the penalty function. The total penalty function is the sum of the two types of penalties involved. One is the reputation penalty shown in Figure 5(a), and the other is the gap penalty shown in Figure 5(b). We set a very simple reputation penalty function as the linear function (see Figure 5(a)). From our testing, we found that the result is not sensitive to the slope of the reputation penalty. The gap penalty function captures the penalty related to the gap between $|\boldsymbol{E_2}|$ and the *target*. Because KAIST is a nationally funded university that does not rely substantially on tuition from students, the underage penalty is considered smaller than the overage penalty. As Figure 6(b) shows, we set the overage penalty to be twice as large as the shortage penalty until *target* + α, after which the overage penalty again doubles. (In the *Experiments* section, we used $\alpha = 0.05$**target*.) Note that a smaller W will result in a smaller reputation penalty, but it may cause a larger gap penalty.

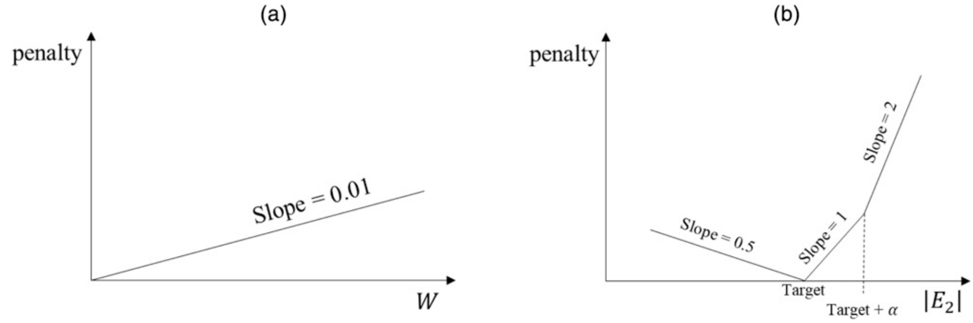
Although we cannot say conclusively that the penalty function in Figure 5 is the best choice, it is simple and captures the essential elements well. Admission professionals generally agree with the choice of a penalty function. Another choice of penalty function would require only a small change in the UAD support system.

To determine the optimal values of decision variables A_1, A_2, and W to minimize the total loss function, we formulate this problem as a two-stage dynamic-programming (DP) problem, which is an effective way to solve such sequential decision-making problems (Bertsekas 1995). A DP problem contains states, actions, a reward (loss) function, and a transition function. In Appendix B, we explain the formulation in detail, and we highly recommend reading this appendix to better interpret our experiment results.

Experiments

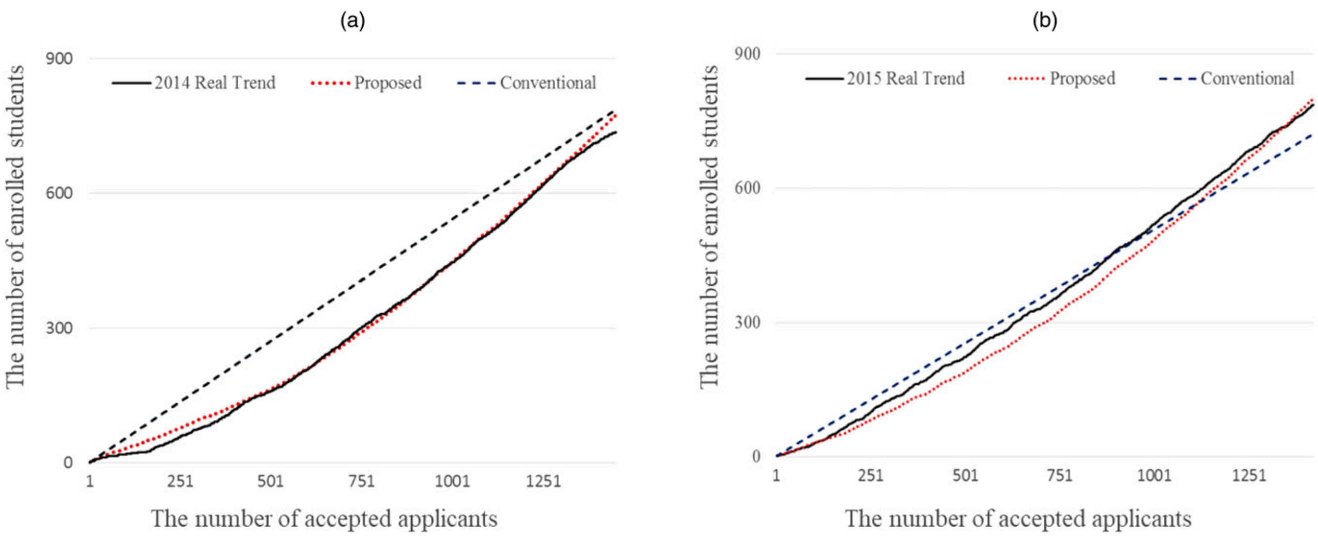
We first validated the model using data from 2013, 2014, and 2015 academic-year admissions; these sets of data contained 1,700; 1,600; and 1,600 samples, respectively. We made a logistic regression model using these applicant data and compared the performance of the conventional and proposed approaches. The conventional approach uses the past average enrollment rate to estimate the expected

Figure 5. The Penalty Functions We Used in Our Markov Decision Process Formulation



Notes. (a) Reputation penalty: $f_1(W)$. (b) Gap penalty: $f_2(E_2)$.

Figure 6. (Color online) Two Validation Test Results Show Our Proposed Method Outperforms the Conventional Method in Predicting Student Enrollment Trends



Notes. (a) Estimating 2014 admission results from 2013 data. (b) Estimating 2015 admission results from 2013 and 2014 data.

number of enrollments. The proposed approach uses individual enrollment probabilities from the logistic regression model to estimate the expected number of enrollments. We conducted two validation tests for the following purposes: (1) to predict 2014 admissions by using 2013 data; and (2) to predict 2015 admissions by using 2013 and 2014 data. In these tests, we predicted the enrollment probability (p_i) of each applicant based on the past data.

Figure 6 shows the number of admissions offered (i.e., accepted applicants) in relation to the number of enrollments. The dashed line (Conventional) represents expected enrollments when a constant enrollment probability ($p_i = p$) is used. The dotted curve (Proposed) represents expected enrollments when individual enrollment probability (p_i) is used. When k applicants are accepted, the expected number of enrollment is the sum of p_i for $i = 1 \ldots k$. The solid black curve shows the number of actual enrollments obtained by counting the true enrollments up to rank k applicants. Note that the "real curve" and the "proposed curve" show a slightly convex downward shape, which indicates that higher-ranked applicants have lower enrollment probabilities in general. This coincides with the observation that applicants with higher scores tend to receive more admissions offers. Overall, the validation tests showed that the individual enrollment probability estimate matches real behavior.

We used this estimated probability to determine the number of admissions in the first and second admission stages. Before applying UAD to a real situation (2016 and 2017), we performed back-testing to 2014 and 2015 data. Table 1 shows the back-testing results of the UAD system decisions in comparison with human expert decisions.

Let us define the relative gap $= (|\boldsymbol{E_2}| - \text{Target})/\text{Target}$ as the measure of performance. For the 2014 test, the

Table 1. Back-Testing UAD Results Using Historical Data Shows that UAD Outperforms the Decisions of Human Experts

			UAD system					Human expert decision				
Training set	Test set	Target	A_1	W	$\|E_1\|$	A_2	$\|E_2\|$	A_1	W	$\|E_1\|$	A_2	$\|E_2\|$
2013	2014	670	1,206	64	690	62	657	1,162	61	659	61	637
2013 and 2014	2015	620	1,109	62	702	18	616	1,203	148	783	0	680

Table 2. Admission Decisions Based on Using the UAD Support System in 2016 and 2017

		Results					
Academic year	Targets	A_1	R	$\|E_1\|$	A_2	$\|E_2\|$	Relative gap
2016	570	996	75	627	35	581	1.93%
2017	550	952	91	584	62	579	5.27%

relative UAD gap was −1.94%, whereas that of human expert decisions was −0.93%. For the 2015 test, human experts showed a very large relative gap of +9.68%, which could have been improved to −0.65% had they used our UAD system. Based on these validation results, we were confident that the UAD support system can improve admissions decisions. Therefore, the KAIST admissions office used the UAD support system for its 2016 and 2017 decisions. We present the results in Table 2.

We determined that our proposed system works better than the existing system because it uses both probabilistic estimation and decision making. In comparison with existing empirical and qualitative methods, it predicts an applicant's enrollment probability better than the conventional method and also optimally solves the problem with a clear objective that minimizes the loss function as it approaches the target.

Conclusion

In this paper, we present two ideas to solve the UAD problem. First, we use logistic regression to obtain the enrollment probabilities of individual applicants. We then formulate the admission problem as a two-stage DP using this probability as the state-transition probability. We designed the loss functions by meeting with admissions experts. The UAD support system also includes manual intervention methods for adjusting enrollment probabilities to reflect changes in environmental factors that the probability model cannot know.

We performed validation tests first by using past data. Subsequently, KAIST has used the UAD support system for its admissions decisions since 2016. We believe that it is worthwhile to report and share our methodology and the results, and we hope that our work can provide useful insights for practitioners with similar problems.

Appendix A. Logistic Prediction Model for the Enrollment Probability of the i_{th} Applicant

As we mention above, adding higher-order terms did not meaningfully improve the prediction accuracy; therefore, we decided to use the following simple model with the coefficients shown in Table A.1 for the enrollment probability of the i_{th} applicant:

$$P(\mathbf{X}_i) = 1\Big/\left(1 + e^{-\beta^T \mathbf{X}_i}\right),$$

where $\mathbf{X}_i$ is the attribute vector of the i_{th} applicant, and β is the coefficient vector. Table A.1 shows the regressors (selected variables) and corresponding coefficients. Human experts in the admissions office confirmed the validity of the logistic regression model we obtained.

Table A.1. The Regressors and Coefficients of the Enrollment Probability Estimation Model

$\mathbf{X}_i$	β
$\mathbf{X}_{i0} = 1$: constant	0.0436
$\mathbf{X}_{i1}$: evaluation score (0~20)	−0.4977
$\mathbf{X}_{i2}$: scholarship (0 or 1)	3.5890
$\mathbf{X}_{i3}$: home location (0 or 1)—near Seoul or not	−0.3757
$\mathbf{X}_{i4}$: high school of type 1 (0 or 1)	2.1379
$\mathbf{X}_{i5}$: high school of type 2 (0 or 1)	1.2477
$\mathbf{X}_{i6}$: high school of type 3 (0 or 1)	−0.2705
$\mathbf{X}_{i7}$: high school of type 4 (0 or 1)	−1.2774

Appendix B. Markov Decision Process Formulation for the UAD Problem

Because we formulate the UAD problem as a two-stage dynamic-programming problem, it must contain states, actions, a reward (loss) function, and a transition function. The state at each stage t is defined as $\boldsymbol{E_t} = (e_{1,t}, \ldots, e_{N,t})$, where $e_{i,t} = 0$ or 1 for all i and $t = 0, 1, 2$. Here, N is the total number of applicants, and $e_{i,t} = 1$ means that the ith applicant has enrolled status at stage t; otherwise, it is 0. The initial state $\boldsymbol{E_0}$ is a zero vector (i.e., $e_{i,0} = 0$).

Please note that our enrollment probability model in Phase 1 (i.e., the training phase) predicts the final enrollment probability $p_i = P(e_{i,2} = 1)$ when an applicant has been admitted, because our enrollment data are the final enrollment data. However, in the following DP formulation, we must know the initial enrollment probability $q_i = P(e_{i,1} = 1|A_1)$ and the retention probability $r_i = P(e_{i,2} = 1|e_{i,1} = 1)$ for each i in the A_1 group. Although we would have preferred to be able to estimate individual q_i and r_i directly from the data, no appropriate data (i.e., initial enrollment data and retention data for individual applicants) were available during this study. However, we know the aggregated initial enrollment numbers; that is, among A_1 admissions offered, E_1^I enrolled initially and E_1^F were part of the final enrollment. For notational simplicity, let $e_{i,2}$ denote the event ($e_{i,2} = 1$) and $\bar{e}_{i,2}$ denote the complementary event ($e_{i,2} = 0$).

By applying Bayes' rule, we get

$$\begin{aligned} r_i = \mathrm{P}(e_{i,2}|e_{i,1}) &= \frac{\mathrm{P}(e_{i,1}|e_{i,2})\mathrm{P}(e_{i,2})}{\mathrm{P}(e_{i,1})} \\ &= \frac{\mathrm{P}(e_{i,1}|e_{i,2})\mathrm{P}(e_{i,2})}{\mathrm{P}(e_{i,1}|e_{i,2})\mathrm{P}(e_{i,2}) + \mathrm{P}(e_{i,1}|\bar{e}_{i,2})\mathrm{P}(\bar{e}_{i,2})} \\ &= \frac{\mathrm{P}(e_{i,2})}{\mathrm{P}(e_{i,2}) + \mathrm{P}(e_{i,1}|\bar{e}_{i,2})(1-\mathrm{P}(e_{i,2}))} \\ &= \frac{p_i}{p_i + \alpha_i(1-p_i)} \text{ for } i = 1, \ldots, A_1. \end{aligned}$$

Note that we use an obvious equation: $\mathrm{P}(e_{i,1}|e_{i,2}) = 1$. The unknown factor in the above equation is $\alpha_i = \mathrm{P}(e_{i,1}|\bar{e}_{i,2})$ for the A_1 group. Here, α_i is the probability that applicant i has initially enrolled, given that applicant i is not part of the final enrollment. If we assume that α_i is the same for all applicants, we can compute this with the aggregate admission, enrollment, and retention information, A_1, E_1^I, and E_1^F, as follows:

$$\alpha_i = \mathrm{P}(e_{i,1}|\bar{e}_{i,2}) = \frac{\mathrm{P}(e_{i,1} \cap \bar{e}_{i,2})}{\mathrm{P}(\bar{e}_{i,2})} \approx \frac{\#(e_{i,1} \cap \bar{e}_{i,2})}{\#(\bar{e}_{i,2})} = \frac{E_1^M - E_1^F}{A_1 - E_1^F} = \alpha.$$

With this, we can compute retention probability r_i and initial enrollment probability q_i:

$$\begin{aligned} r_i &= \frac{p_i}{p_i + \alpha(1-p_i)} \\ q_i &= \mathrm{P}(e_{i,1}|A_1) = \frac{\mathrm{P}(e_{i,2}|A_1)}{\mathrm{P}(e_{i,2}|e_{i,1})} = \frac{p_i}{r_i}. \end{aligned}$$

For convenience, we split $\boldsymbol{E_2}$ into the state vectors for the A_1 group and A_2 group—that is, $\boldsymbol{E_2} = (E_{21}, E_{22})$, where $\boldsymbol{E_{21}} = (e_{1,2}, \ldots, e_{A_1,2})$ and $\boldsymbol{E_{22}} = (e_{A_1+1,2}, \ldots, e_{A_1+A_2,2})$. By assuming the independence of the enrollment decision for each applicant, the joint probabilities, $\mathrm{P}(\boldsymbol{E_1}|A_1)$, $\mathrm{P}(\boldsymbol{E_{21}}|\boldsymbol{E_1})$, and $\mathrm{P}(\boldsymbol{E_{22}}|A_2)$ can be factored as follows:

$$\begin{aligned} \mathrm{P}(\boldsymbol{E_1}|A_1) &= \prod_{i \le A_1} q_i^{e_{i,1}}(1-q_i)^{1-e_{i,1}}, \\ \mathrm{P}(\boldsymbol{E_{21}}|\boldsymbol{E_1}) &= \prod_{i \le A_1} \mathrm{P}(e_{i,2}|e_{i,1}=1) = \prod_{i \le A_1} r_i^{e_{i,2}}(1-r_i)^{1-e_{i,2}}, \\ \mathrm{P}(\boldsymbol{E_{22}}|A_2) &= \prod_{i \in \le A_2-group} \mathrm{P}(e_{i,2}) = \prod_{i \in \le A_2-group} p_i^{e_{2,1}}(1-p_i)^{1-e_{2,1}}. \end{aligned}$$

Using the above state-transition probability, we can formulate the DP problem to minimize the total loss $L_{12}(A_1, W)$ in the first stage, and $L_2(\boldsymbol{E_1}, A_1, W, A_2)$ in the second stage. Here, L_2 is the expected gap penalty (f_2), and L_{12} is the sum of the reputation penalty (f_1) and the expected L_2. The final DP formulation is shown below. We can solve this optimization problem using backward induction.

Stage 1 (First Decision)

$$\begin{gathered} \left(A_1^*, W^*\right) = \operatorname*{argmin}_{A_1, W} L_{12}(A_1, W) \\ L_{12}(A_1, W) = f_1(W) + \sum_{\forall E_1} L_2\left(\boldsymbol{E_1}, A_1, W, A_2^*\right)\mathrm{P}(\boldsymbol{E_1}|A_1). \end{gathered}$$

Stage 2 (Second Decision)

$$\begin{gathered} A_2^* = \operatorname*{argmin}_{A_2} L_2(\boldsymbol{E_1}, A_1, W, A_2) \\ L_2(\boldsymbol{E_1}, A_1, W, A_2) = \sum_{\forall E_{21}} \sum_{\forall E_{22}} f_2(\boldsymbol{E_2})\mathrm{P}(\boldsymbol{E_{21}}|\boldsymbol{E_1})\mathrm{P}(\boldsymbol{E_{22}}|A_2). \end{gathered}$$

References

Bertsekas DP (1995) *Dynamic Programming and Optimal Control* (Athena Scientific, Belmont, MA).

Chade H, Lewis G, Smith L (2014) Student portfolios and the college admissions problem. *Rev. Econom. Stud.* 81(3):971–1002.

Cox DR (1958) The regression analysis of binary sequences (with discussion). *J. Royal Statist. Soc. B* 20(2):215–224.

Krstajic D, Buturovic LJ, Leahy DE, Thomas S (2014) Cross-validation pitfalls when selecting and assessing regression and classification models. *J. Cheminformatics* 6(10):1–15.

Maltz EN, Murphy KE, Hand ML (2007) Decision support for university enrollment management: Implementation and experience. *Decision Support Systems* 44(1):106–123.

Rebbapragada S, Basu A, Semple J (2010) Data mining and revenue management methodologies in college admissions. *Commun. ACM* 53(4):128–133.

Tibshirani R (1996) Regression shrinkage and selection via the lasso. *J. Royal Statist. Soc. B* 58(1):267–288.

Walczak S, Sincich T (1999) A comparative analysis of regression and neural networks for university admissions. *Inform. Sci.* 119(1): 1–20.

Verification Letter

Ji Hoon Kim, Manager of Undergraduate Admission, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea, writes:

"We were aware of submitting the manuscript entitled 'Optimizing the Multistage University Admission Decision Process' for publication in *Interfaces* for possible evaluation.

"We believe that this manuscript is appropriate for publication by *Interfaces* because this study suggests a solution of the University Admission Decision process, which is one of the important problems of OR/MS, and it has achieved good results by actually proceeding to implementation.

"This letter is to verify that we have used the method reported in this manuscript in the admission decisions of 2016 and 2017. It has had a very positive impact on our admission decision."

Donghyun Kim is a data scientist for NAVER in South Korea. He received a BS from Korea Advanced Institute of Science and Technology (KAIST) in mathematical science in 2012 and an MS and a PhD from KAIST in 2014 and 2018 in industrial and systems engineering. His recent research interests are in the machine learning and recommender systems for real services.

Namyong Kim is a PhD candidate in the Department of Industrial and System Engineering, Korea Advanced Institute of Science and Technology. He holds a BS in the division of industrial management engineering from Korea University. His research interests are in the application of machine learning to real problems such as university admission decision, scheduling of the manufacturing system, and traffic systems.

Junoh Cho is the admissions officer of the Korea Advanced Institute of Science and Technology. He is in charge of processing overall admission statistics. He received a PhD in education from Chungbuk National University in South Korea. He studied the curriculum and the statistics of education in detail.

Hayong Shin is a professor in the Department of Industrial and Systems Engineering at KAIST. He is currently serving as the dean of admissions office at KAIST. Before joining KAIST, he worked for Chrysler Corp., CubicTek Co., and LG Electronics. He received a BS from Seoul National University in 1985 and an MS and a PhD from KAIST in 1987 and 1991, all in industrial engineering. His recent research interests are in the area of machine learning, simulation, and modeling for manufacturing systems.