The College Application Problem

Max Kapur

Seoul National University
Department of Industrial Engineering
Management Science/Optimization Lab

May 3, 2022

Introduction

The optimal college application problem is a **novel combinatorial optimization problem**.

It involves **maximizing the expected maximum** of a portfolio of random variables subject to a budget constraint.

Methodological orientation:

- Investment with uncertain payoff, search for the efficient frontier recall **portfolio allocation** models.

- Generalizes the **knapsack problem**: Integral packing constraint, NP-completeness, approximation algorithms.

- Objective is a **submodular set function**.

Today's presentation: **define the problem** and briefly summarize our **solution algorithms**.

The admissions process

Consider a **single student's** college application decision.

Market contains m **schools**, indexed by $\mathcal{C} = \{1 \ldots m\}$. School j is named c_j.

Given a student's academic records, test scores, and demographic information, we can estimate her **admissions probability** f_j at each school.

Let the **random variable** $Z_j \sim \text{Bernoulli}(f_j) = 1$ if student is admitted, 0 otherwise. Assume independent.

Let $\mathcal{X} \subseteq \mathcal{C}$ denote the set of schools, or **application portfolio**, to which a student applies.

Application fees, time to write essays, and/or legal limits **constrain** applicant behavior. We consider a single knapsack constraint $\sum_{j \in \mathcal{X}} g_j \leq H$ where g_j is called c_j's **application cost**.

Utility model

Let $t_j \geq 0$ denote the **utility** the student receives if she attends c_j. Wlog, $t_j \leq t_{j+1}$.

Let t_0 denote her utility if she doesn't get into college. Wlog, $t_0 = 0$ (see paper).

The student's overall utility is the t_j-value associated with the **best** school she gets into:

$$\text{Utility} = \max\{t_0, \max\{t_j Z_j : j \in \mathcal{X}\}\}$$

When the student's application portfolio is $\mathcal{X}$, we refer to her expected utility as the portfolio's **valuation** $v(\mathcal{X})$.

Unpacking the portfolio valuation function

To get $v(\mathcal{X})$ into a tractable form, let $p_j(\mathcal{X})$ denote the probability that the student **attends** c_j.

This happens if and only if she **applies** to c_j, is **admitted** to c_j, and is **not admitted** to any school she prefers to c_j:

$$
p_j(\mathcal{X}) = \begin{cases} f_j \displaystyle\prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i), & j \in \{0\} \cup \mathcal{X} \\ 0, & \text{otherwise.} \end{cases}
$$

Therefore,

$$
v(\mathcal{X}) = \sum_{j=1}^{m} t_j p_j(\mathcal{X}) = \sum_{j \in \mathcal{X}} \Big(f_j t_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i) \Big).
$$

Problem statement

Problem 1 (The college application problem)

$$\text{maximize} \quad v(\mathcal{X}) = \sum_{j \in \mathcal{X}} \left(f_j t_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i) \right)$$

$$\text{subject to} \quad \mathcal{X} \subseteq \mathcal{C}, \quad \sum_{j \in \mathcal{X}} g_j \leq H$$

Problem 2 (The college application problem, INLP form)

$$\text{maximize} \quad v(x) = \sum_{j=1}^{m} \left(f_j t_j x_j \prod_{i > j} (1 - f_i x_i) \right)$$

$$\text{subject to} \quad x_j \in \{0, 1\}, j \in \mathcal{C}; \quad \sum_{j=1}^{m} g_j x_j \leq H$$

Do you trust the admissions consultant's advice?

Existing solutions

In practice, most use **distributive heuristics** such as distributing applications evenly among reach, target, and safety schools. Turns out to be a **risk-averse** approach.

Another intuitive idea is the **linearization heuristic**. Since the expected utility associated with applying to c_j (alone) is $f_j t_j$, solve the knapsack problem

$$\text{maximize} \quad \sum_{j \in \mathcal{X}} f_j t_j \qquad \text{subject to} \quad \sum_{j \in \mathcal{X}} g_j \leq H$$

as a surrogate. But this solution can be **arbitrarily bad**.

Fu (2014) solved a similar problem by **enumeration**, which is intractable for $m \geq 20$ or so.

Our algorithms provide both **time and accuracy guarantees**.

Our algorithms

In the **special case** where each $g_j = 1$, we provide an $O(m^2)$ algorithm.

The general problem is **NP-complete** (reduction from knapsack). We offer four algorithms:

- A linear relaxation and **branch-and-bound scheme**. Primarily of theoretical interest.

- A **dynamic program based on total expenditures**. Exact solution in $O(Hm + m \log m)$ time (pseudopolynomial). Very fast for "typical" instances in which g_j are small integers.

- A different DP based on truncated portfolio valuations. $(1 - \varepsilon)$-optimal solution in $O(m^3/\varepsilon)$ time: an **FPTAS**!

- A **simulated annealing heuristic**. Fast, typically within 2% of optimality.

Conclusion

"Maximax" form, integrality constraints make the college application problem **theoretically interesting**. Formally, it is a submodular maximization problem, but its approximability is more like knapsack.

Solutions to the college application problem have **practical value**: US admissions consultants charge an average of \$200/hr! $\Rightarrow$ open-sourcing our code is in the public interest.

Lots of extensions to consider: parametric risk aversion, distribution constraints, memory-usage improvements.

References

Fu, Chao. 2014. "Equilibrium Tuition, Applications, Admissions, and Enrollment in the College Market." *Journal of Political Economy* 122 (2): 225–81. https://doi.org/10.1086/675503.

Kapur, Max. 2022. "OptimalApplication." GitHub repository. https://github.com/maxkapur/OptimalApplication.

Kulik, Ariel, Hadas Shachnai, and Tami Tamir. 2013. "Approximations for Monotone and Nonmonotone Submodular Maximization with Knapsack Constraints." *Mathematics of Operations Research* 38 (4): 729–39. https://doi.org/10.1287/moor.2013.0592.

Sklarow, Mark. 2018. *State of the Profession 2018: The 10 Trends Reshaping Independent Educational Consulting.* Technical report, Independent Educational Consultants Association. https://www.iecaonline.com/wp-content/uploads/2020/02/IECA-Current-Trends-2018.pdf.