대학 지원 최적화 문제

Max Kapur & 홍성필

서울대학교
산업공학과
경영과학/최적화 연구실

May 2, 2022

서론

대학 지원 최적화 문제란 **새로운 조합 최적화 문제**이다.

예산 제약 조건 하에서, 다수 확률 변수로 이뤄진 포트폴리오의 **기대 최댓값을 최대화한다**.

발표 요약:

- 대학 지원 전략을 최적화 문제로 **모형화**.
- **현재 상황**: 입학 컨설팅 산업에서 흔히 사용하는 휴리스틱의 부족함.
- 본 연구가 제시하는 **이론적 결과, 해법, 알고리즘 구현**.

방법론적 지향

대학 지원 전략은 조합 최적화 중 여러 분야를 걸친다.

- 불확실한 성과, 효율적 투자선이 존재하는 일종의 **포트폴리오 배분** 모형.
- **배낭 문제**의 일반화: 정수 패킹 제약식, NP-completeness.
- 목적함수는 **submodular 집합 함수**이며, 근사 해법 결과에 따라 대학 지원 문제가 submodular 최적화의 비교적 쉬운 인스턴스로 해석할 수 있다.

모형

입학 과정

단 **한 학생**의 의사결정에 집중하자.

시장은 m개의 **대학교**를 포함하며, 그의 지표 집합은 $\mathcal{C} = \{1 \dots m\}$. j번째 학교의 이름은 c_j.

학생의 내신, 수능 점수, 기본 정보가 주어지면 각 학교의 **합격 확률** f_j를 추정할 수 있다.

확률 변수 $Z_j \sim \text{Bernoulli}(f_j)$는 학생이 합격하면 1, 아니면 0이다. 독립하다고 가정.

학생이 지원하는 학교의 집합 $\mathcal{X} \subseteq \mathcal{C}$를 **지원 포트폴리오**라고 부른다.

지원 수수료, 원서를 작성하는 시간, 또는 나라 정책에 따라 지원 행동이 **제한**된다. 본 논문은 단일 배낭 제약식 $\sum_{j \in \mathcal{X}} g_j \leq H$를 고려하며, 이때 g_j는 c_j의 **지원 비용**이라고 부른다.

효용 모형

c_j에 다니면 $t_j \geq 0$ 단위의 **효용**이 발생한다. Wlog, $t_j \leq t_{j+1}$.

어떤 대학에도 합격 안 하면 효용이 t_0이며, wlog $t_0 = 0$임을 가정할 수 있다 (논문 참고).

학생의 전체 효용은 그가 합격하는 **가장 좋은** 학교의 t_j-값:

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

지원 포트폴리오가 $\mathcal{X}$일 때, 기대 효용을 $\mathcal{X}$의 **가치**라고 부르며 $v(\mathcal{X})$처럼 표기한다.

포트폴리오 가치의 함수 형태

$v(\mathcal{X})$를 함수로 표현하기 위해, 학생이 c_j에 **진학하는** 확률을 $p_j(\mathcal{X})$라고 하자.

c_j에 진학하는 조건은 c_j에 **지원**하고, **합격**하고, c_j보다 선호하는 학교에 **합격하지 않는** 것이다:

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

따라서

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

문제 정의

문제 1 (대학 지원 최적화 문제)

$$\begin{aligned} \text{maximize} \quad & v(\mathcal{X}) = \sum_{j \in \mathcal{X}} \Bigl(f_j t_j \prod_{\substack{i \in \mathcal{X}: \\ i > j}} (1 - f_i) \Bigr) \\ \text{subject to} \quad & \mathcal{X} \subseteq \mathcal{C}, \quad \sum_{j \in \mathcal{X}} g_j \leq H \end{aligned}$$

문제 2 (대학 지원 최적화 문제, 정수 비선형 계획 모형)

$$\begin{aligned} \text{maximize} \quad & v(x) = \sum_{j=1}^{m} \Bigl(f_j t_j x_j \prod_{i > j} (1 - f_i x_i) \Bigr) \\ \text{subject to} \quad & x_j \in \{0, 1\}, j \in \mathcal{C}; \quad \sum_{j=1}^{m} g_j x_j \leq H \end{aligned}$$

기존의 해법

Timothy Peck — May 7, 2021 — 5 — 12th Grade, College Lists

Safety, Target, & Reach Schools: How to Find the Right Ones

What's Covered:

- What Are Reach, Target, and Safety Schools?
- Factors that Impact Your Chances
- Elements of a Balanced College List

Creating a school list is an important-yet-tricky step in the college application process. A strategically constructed school list weighs your desire to attend reach schools—the institutions you dream about going to—along with safety schools where you're very likely to secure admission. Consequently, the ideal school list is balanced between reach, target, and safety schools, allowing you to shoot for the stars while also ensuring admission into at least one school.

What Are Reach, Target, and Safety Schools?

"Reach," "safety," and "target" are common terms used in college applications to describe the odds a student has of getting accepted at a particular institution. Understanding these terms, and which categories colleges fall into, is a critical step in the application process.

What is a Reach School?

Reach schools are colleges where you have less than a 25% chance of admission (this is your

The JoongAng

[대입 수시 전략] 총 6번의 기회 …'상향·소신·안정' 분산 지원하라

중앙일보 | 업데이트 2015.08.26 10:15 지면보기

전민희 기자 구독

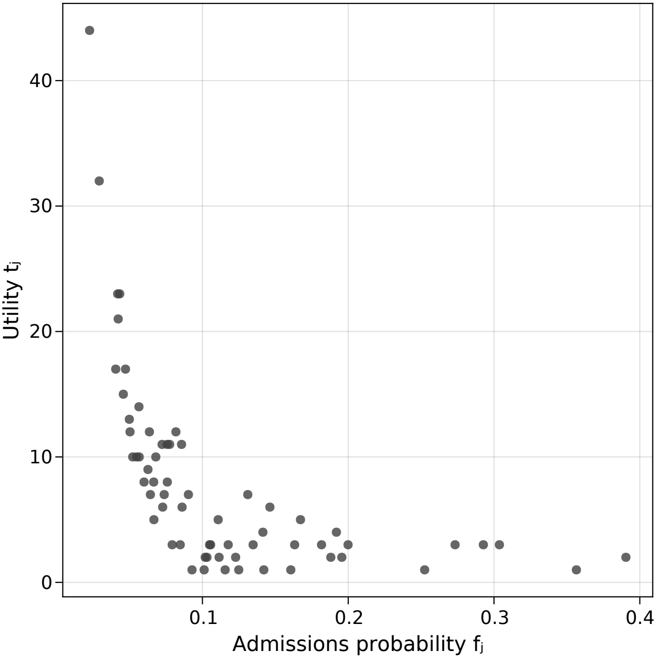
대학 최저학력기준 고려해 전략 지원
지난해 같은 전형 합격한 선배 내신 참고
수능 전 대학별고사 보는 곳은 최소화

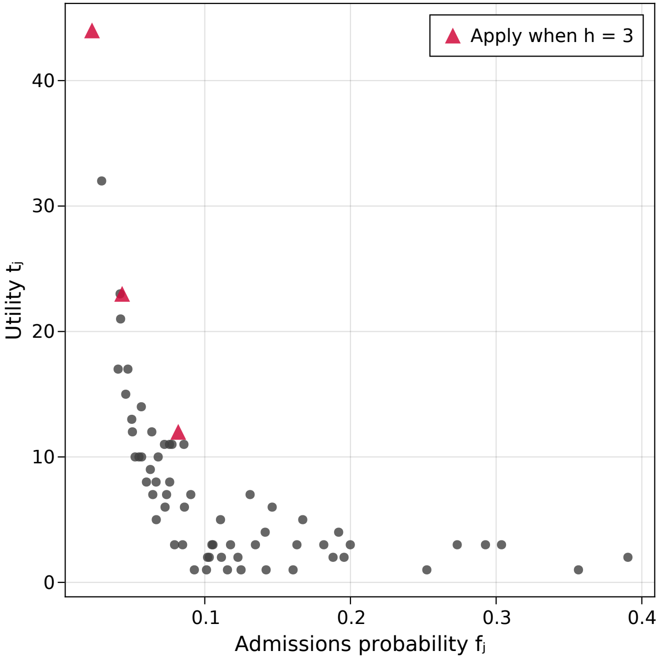
'지피지기 백전불태(知彼知己百戰不殆).' 적을 알고 나를 알면 백 번 싸워도 위태롭지가 않다는 뜻이다. 고대 중국의 병법서인 『손자』에 나온 말이지만 현대사회에서도 여러 가지 분야에서 회자된다. 그중 하나가 대학입시다. 특히 2주 앞으로 다가온 수시모집은 전형 종류가 다양해 '적'(모집전형) 을 알고, '나'(학생)에 대해 파악하는 게 무엇보다 중요하다.

자신의 학교생활기록부, 교과성적, 대학별고사 준비 상황, 예상 수능점수, 최저학력기준 통과 가능성에 대해 자세히 살핀 후 지원해야 합격률을 높일 수 있다. 수시모집 마무리 전략을 알아봤다.

논술전형도 학생부 성적 기준으로 지원

입학 컨설턴트의 조언, 믿을 만하는가?


0
10
20
30
40
Utility t_j
0.1
0.2
0.3
0.4
Admissions probability f_j



Apply when h = 3
Utility t_j
0
10
20
30
40
0.1
0.2
0.3
0.4
Admissions probability f_j

Existing solutions

In practice, most use **distributive heuristics** such as distributing applications evenly among attractive, selective "reach schools" and less-attractive "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**.

제시 알고리즘

다항 시간에 풀 수 있는 동일한 지원 비용의 경우

먼저 모든 $g_j = 1$이며 H는 단순한 지원 개수 제한 $h \leq m$이 되는 **특수한 경우**를 고려하자.

이것은 $h = 3, m = 202$인 한국 정시 입학 과정과 비슷한 상황이다.

지원 제한 h에 대응되는 최적 포트폴리오가 $\mathcal{X}_h$일 때, $\mathcal{X}_h \subset \mathcal{X}_{h+1}$을 증명할 수 있다. 이러한 **포함 사슬 관계 (nestedness)** 성질은 $v(\mathcal{X})$를 최대화하는 순서대로 학교를 하나씩 추가하는 **탐욕 해법**의 최적성을 의미한다.

변수 소거 기법은 개발하여 $v(\mathcal{X})$를 계산하는 시간을 환산 단위 시간을 줄임으로 $O(hm)$ 해법을 얻을 수 있다. 쥴리아 코드(Kapur 2022)를 사용하면 $m = 16384$인 인스턴스를 200 ms에 풀 수 있다.

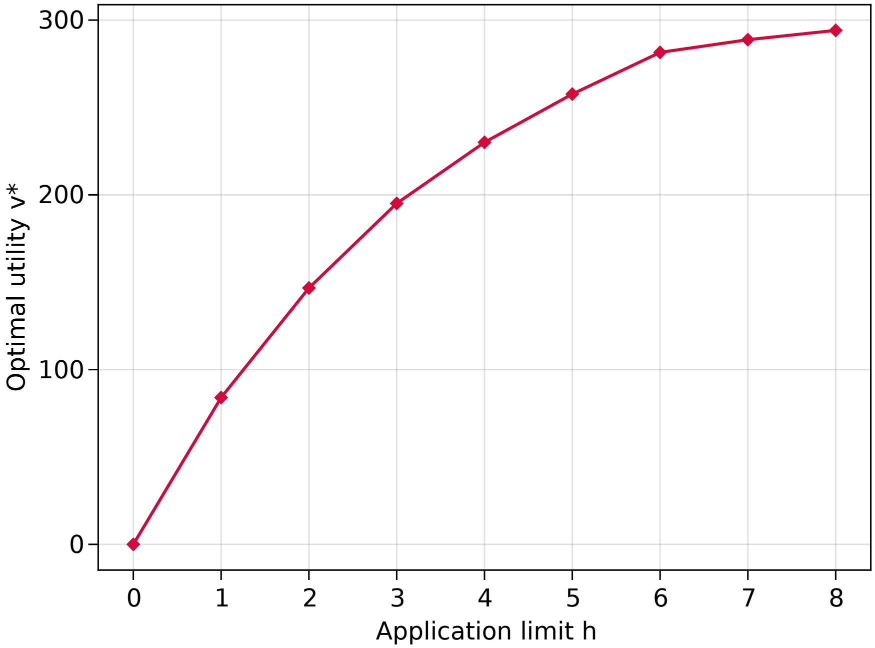
Fisher et al. (1978)에 결과를 강화한다.

작은 예제

$m = 8$개의 학교로 이루어진 가상 입학 시장의 대학교 자료와 최적 지원 포트폴리오.

지표 j	학교 c_j	합격 확률 f_j	효용 t_j	지원 순위	$v(\mathcal{X}_h)$
1	수성대	0.39	200	4	230.0
2	금성대	0.33	250	2	146.7
3	화성대	0.24	300	6	281.5
4	목성대	0.24	350	1	84.0
5	토성대	0.05	400	7	288.8
6	천왕성대	0.03	450	8	294.1
7	해왕성대	0.10	500	5	257.7
8	명왕성대	0.12	550	3	195.1

포함 사슬 관계 성질에 따라, 지원 제한이 h일 때 최적 포트폴리오는 지원 순위가 h 이하인 h개의 학교로 구성된다.


300
200
100
0
Optimal utility v*
0
1
2
3
4
5
6
7
8
Application limit h

Algorithms for the general problem

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.

Existence of FPTAS suggests college application is **a relatively easy instance of submodular maximization** (cf. Kulik et al. 2013).

결론

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.

The nestedness result for the $g_j = 1$ special case also resembles the knapsack problem, although the proof is more subtle.

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.

참고 문헌 I

Fisher, Marshall, George Nemhauser, and Laurence Wolsey. 1978. "An analysis of approximations for maximizing submodular set functions—I." *Mathematical Programming* 14: 265–94.

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.

Kapur, Max and Sung-Pil Hong. 2022. "The Optimal College Application Problem." Preprint. https://github.com/maxkapur/CollegeApplication.

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.

부록: 알고리즘 요약

알고리즘	문제	제한	정확도	계산 시간
나이브	동일한 지원 비용	없음	$(1/h)$-근사	$O(m)$
탐욕 해법	동일한 지원 비용	없음	정확	$O(hm)$
분지한계법	일반 문제	없음	정확	$O(2^m)$
지출액 동적 계획	일반 문제	g_j 정수	정확	$O(Hm + m \log m)$
FPTAS	일반 문제	없음	$(1-\varepsilon)$-근사	$O(m^3/\varepsilon)$
모의 담금질	일반 문제	없음	0-근사	$O(Nm)$