대학 지원 최적화 문제

Max Kapur

서울대학교
산업공학과
경영과학/최적화 연구실

May 3, 2022

서론

대학 지원 최적화 문제란 **새로운 조합 최적화 문제**이다.

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

방법론적 지향:

- 불확실한 성과, 효율적 투자선이 존재하는 일종의 **포트폴리오 배분** 모형.
- **배낭 문제**의 일반화: 정수 패킹 제약식, NP-completeness, 근사 해법.
- 목적함수는 **submodular 집합 함수**.

오늘 발표에서는 **문제를 정의하고 본 연구가 제시하는 해법**을 요약한다.

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 (논문에서 증명).

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

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

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)

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

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?

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?

기존의 해법

입학 컨설팅 산업에서는 주로 "상향 · 소신 · 안정 지원 학교" 각각 균일하게 지원하는 **배분적 휴리스틱**을 사용하며, **위험 회피적인** 전략임을 알 수 있다.

또한 **선형화 휴리스틱**이 있다. c_j에만 지원할 때 기대 효용이 $f_j t_j$이므로 다음 배낭 문제를 대리 문제로 푼다.

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

이의 **근사 계수가 0에 무한히 가까운** 예시를 만들 수 있다.

Fu (2014)는 비슷한 문제를 **열거법**으로 풀었으며, $m \geq 20$일 때 불현실적인 방법이다.

본 연구가 제시하는 알고리즘은 **계산 시간과 정확도를 보장한다**.

제시 알고리즘

모든 $g_j = 1$인 특수한 경우를 위한 $O(m^2)$ 알고리즘 제시.

일반적인 문제는 **NP-complete** (배낭 문제에서 변환). 4개의 알고리즘 제시:

- 선형 완화 문제와 해당 **branch-and-bound** 해법. 주로 이론적인 흥미가 있다.
- **지원 지출액 기반 동적 계획**. (의사 다항) $O(Hm + m \log m)$ 시간에 정확한 해를 구하며, g_j가 작은 정수가 되는 "전형적" 인스턴스에 대해 매우 효율적인 해법.
- 포트폴리오 가치의 라운딩을 이용한 다른 동적 계획. $O(m^3/\varepsilon)$ 시간에 $(1-\varepsilon)$-근사해를 출력하므로 **FPTAS**!
- **Simulated annealing** 휴리스틱. 빠르고 거의 항상 최적의 2% 이내에 해를 구한다.

Conclusion

“Maximax” 형태와 정수 조건 때문에 대학 지원 문제에는 **이론적으로 흥미**가 있다. Submodular 최대화 문제이지만, 근사 해법 성질은 배낭 문제에 더 가깝다.

좋은 대학 지원 전략에는 **금전적 가치**가 있다. 미국 입학 컨설컨트의 시간당 급료는 평균 200달러! $\Rightarrow$ 공공 이익을 위해 코드는 open source license로 공개.

문제를 확장할 수 있는 향후 연구: 위험 회피 모수, 배분적 (가나다군) 제약 조건, 메모리 소모량 절감.

참고 문헌

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.