

대학 입학 지원 최적화 문제

Max Kapur

서울대학교 산업공학과
경영과학/최적화 연구실
maxkapur@snu.ac.kr

지도 교수: 홍성필

July 1, 2022

대학 입학 지원 최적화 문제는 **새로운 조합 최적화 문제**이다.

예산 제약 조건 하에서, 다수 확률 변수로 이루어진 포트폴리오의 **기대 최댓값을 최대화**하는 문제이다.

방법론적 지향:

- 불확실한 성과, 효율적 투자선이 존재하므로 **일종의 포트폴리오 배분** 모형.
- **배낭 문제의 일반화**: 정수 채우기 (packing) 제약식, NP-completeness, 근사 해법의 필요.
- 목적함수는 **submodular** 집합 함수.

오늘 발표에서는 **문제를 정의하고 본 연구가 제시하는 해법을 요약**한다.

단 한 명의 학생의 의사결정에 집중하자.

시장은 m 개의 **대학교**를 포함하며, 지표 집합은 $\mathcal{C} = \{1 \dots m\}$. j 번째 학교의 이름은 c_j .

각 학교에 대해 학생의 **합격 확률** f_j 가 주어져 있다.

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

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

지원 전형료 예산, 원서를 작성하는 시간, 또는 나라의 정책에 따라 **지원 행동에 제약이 생긴다**. 본 논문은 단일 배낭 제약식 $\sum_{j \in \mathcal{X}} g_j \leq H$ 를 고려하며, 이때 g_j 는 c_j 의 **지원 비용**이라고 부른다.

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

어떤 대학에도 합격하지 않았을 때 효용은 t_0 이며, wlog $t_0 = 0$ 임을 가정할 수 있다 (논문에서 증명).

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

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

이의 기댓값은 \mathcal{X} 의 **가치**라고 부르며 $v(\mathcal{X})$ 로 표기한다.

포트폴리오 가치의 함수 형태

$v(\mathcal{X})$ 를 함수로 표현하기 위해, 학생이 c_j 에 **진학하는 확률**을 $p_j(\mathcal{X})$ 라고 하자.

c_j 에 진학하는 조건은 c_j 에 **지원**하고, **합격**하고, c_j 보다 선호하는 학교에는 **합격하지 않았을** 때이다.

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

따라서,

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

문제 정의

문제 1 (대학 지원 최적화 문제)

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

문제 2 (대학 지원 최적화 문제, 정수 비선형 계획 모형)

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

기존의 해법

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

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

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

그러나 이의 근사 계수는 0에 무한히 가까워질 수 있다.

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

본 연구는 **계산 시간과 정확도가 보장된** 알고리즘을 제시한다.

알고리즘 제시

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

일반적인 문제는 **NP-complete** (배낭 문제에서 변환).

4개의 알고리즘 제시:

- 선형 완화 문제와 해당 **branch-and-bound** 해법. 일반적인 INLP 문제에 대해 자주 쓰이는 방법이다.
- **총 지원 비용 기반 동적 계획**. (의사 다항) $O(Hm + m \log m)$ 시간에 정확한 해를 구하며, g_j 가 작은 정수가 되는 “전형적” 인스턴스에 대해 매우 효율적인 해법.
- 포트폴리오 가치의 라운딩을 이용한 동적 계획. $O(m^3/\epsilon)$ 시간에 $(1 - \epsilon)$ -근사해를 출력하므로 **FPTAS!**
- **Simulated annealing** 휴리스틱. 빠르고 대부분 98% 이상의 최적성을 얻었다.

“Maximax” 형태와 정수 조건 때문에 대학 지원 문제는 **이론적으로 흥미로운 문제이다**. Submodular 최대화 문제지만, 근사 해법의 성질은 배낭 문제에 더 가깝다 (cf. Fisher et al. 1978; Kulik et al. 2013; Kellerer et al. 2004).

좋은 대학 지원 전략에는 **금전적 가치가 있다**. 미국 입학 컨설턴트의 시간당 급료는 평균 200달러 (Sklarow 2018)!

⇒ 공공 이익을 위해 코드는 open-source license로 공개 (Kapur 2022).

향후 연구: 위험 회피 모수, 배분적 제약 조건 (가나다순), FPTAS의 메모리 소모량 절감.

참고 문헌

- 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.jl>.
- Kellerer, Hans, Ulrich Pferschy, and David Pisinger. 2004. *Knapsack Problems*. Berlin: Springer.
- 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>.