Knapsack Problem, Merit-Based Scholarship, Personalized Promotion

How to maximize resource allocation

Shuai Wang

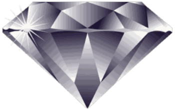
July 24, 2023

Presentation Outline

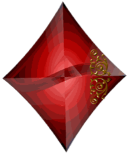
PACIFIC LIFE

Knapsack problem

Wt. = 5
Value = 10



Wt. = 3
Value = 20



Wt. = 8
Value = 25

Wt. = 4
Value = 8

Maximum wt. = 13

Knapsack problem

- You only bring one knapsack with a capacity limit to rob a bank.
- Different item has different amount of value and weight.
- Try to get as much value as possible.
- Which ones to choose with the capacity limit of the knapsack?

Knapsack problem math modeling

0-1 KP has the following Integer Programming (IP) formulation:

$$\text{maximize value} \quad \sum_{j \in N} p_j x_j \tag{1}$$

$$\text{subject to weight} \quad \sum_{j \in N} w_j x_j \leq c \tag{2}$$

$$x_j \in \{0, 1\}, \quad j \in N, \tag{3}$$

where each binary variable x_j, $j \in N$, is equal to 1 if and only if item j is selected.
p_j: price/value of each item;
w_j: weight of each item.
We cannot take all items because the total weight of the chosen items cannot exceed the knapsack capacity c.

PACIFIC LIFE

Reading

How the Mathematical Conundrum Called the ‘Knapsack Problem’ Is All Around Us

Smithsonian MAGAZINE

SCIENCE

How the Mathematical Conundrum Called the ‘Knapsack Problem’ Is All Around Us

A litany of issues in business, finance, container ship loading and aircraft loading derive from this one simple dilemma

Elizabeth Landau

Contributing Writer

March 9, 2020

Financial Aid

- In the United States, the 2012-2013 academic year, there were a total of 20.4 million students in degree-granting institutions.
- More than 80% of them received financial aid.
- Studies have shown that financial aid is one of the most important factors in attracting student and is vital to enrollment management.

PACIFIC LIFE

Background

Enrollment management consisted of approaches to help university to meet the established goal such as:

- Attract more high-caliber students.
- Diversify student body.
- Increase retention.
- Improve graduation rate.

Scholarship

- Scholarship is the focused type of financial aid of the study.
- Scholarship is the major marketing tool for targeting students.
- Scholarship helps the university:
  1. Giving more access to families who need help.
  2. Stimulating more students to major in area having labor shortage.
  3. Diversifying the student body.
- Merit-based scholarship.

PACIFIC LIFE

Motivation

1. **Non-optimal** usage of scholarship budget at university:
   - Over-spending scholarship budget would reduce revenue.
   - Under-spending scholarship would potentially undermine enrollment number and revenue.
2. The tuition income accounts for 48% of the yearly revenue of the university under study.
3. The optimal allocation of scholarship problem has **not** been widely studied in literatures.

PACIFIC LIFE

Research Questions and Contribution

- Research questions:
  1. How does scholarship affect student's decision?
  2. What is the optimal scholarship for each student under the overall budget?
  3. What is the ideal scholarship budget for school?

Enrollment and graduation prediction

- The outcome of enrollment and graduation problem is binary (yes or no).
- It is a two-class classification problem.
- Common methods for two-class classification problem:
  logistic regression, decision tree, svm, etc,.

PACIFIC LIFE

Variables

1. Academic: HS GPA, ACT/SAT, HS Percentile.
2. Financial: Pell Grant, EFC, Out-of-pocket, Scholarship, Unemployment rate.
3. Demographic: High School, Tier, Ethnicity.

Logistic Regression: Enrollment prediction results

GPA 2.9, ACT 19										
	Student	0	1000	2000	3000	4000	5000	6000	7000	8000
1	2.9-Tier1-19-White	59.55	64.63	69.39	73.77	77.73	81.24	84.31	86.96	89.22
2	2.9-Tier5-19-White	36.96	40.20	43.53	46.92	50.34	53.75	57.13	60.45	63.67
GPA 3.3, ACT 25										
3	3.3-Tier1-25-Hispanic	23.80	27.44	31.42	35.69	40.20	44.88	49.65	54.43	59.13
4	3.3-Tier1-25-White	55.60	59.32	62.94	66.42	69.72	72.84	75.75	78.43	80.90
GPA 3.8, ACT 28										
5	3.8-Tier1-28-White	42.29	46.05	49.85	53.65	57.41	61.08	64.63	68.03	71.25
6	3.8-Tier4-28-White	20.54	22.87	25.37	28.05	30.89	33.89	37.02	40.26	43.60

PACIFIC LIFE

Number of years prediction

- It is a regression problem.
- The following methods are compared to predict the years of stay in school:

	10-Fold Cross Validation		Test Data	
Model	RMSE	MAE	RMSE	MAE
GLM	1.40	1.2	1.53	1.26
SVM (Linear Kernel)	1.44	1.20	1.62	1.32
Decision Tree	1.43	1.24	1.43	1.23
Stochastic Gradient Boosting	1.40	1.19	**1.40**	**1.19**

PACIFIC LIFE

Why prediction model is not enough?

1. The prediction of enrollment and graduation provide some insights of how students response to the various scholarship.
2. They have not addressed the allocation of limited scholarship budget to students fundamentally.

Objective Model

Sets:

- I: set of applicants, indexed by i and j
- M: different levels of scholarship awards, indexed by m
  $m \in M = \{0, 1000, 2000, \ldots, 8000\}$

Variables:

- x_{im} : binary, whether a scholarship award m is allocated to applicant i or not

PACIFIC LIFE

Math notation

Parameters:

- p^e_{im}: probability of enrollment for applicant i, if given award m
- p^g_{im}: probability of graduation for applicant i, if given award m
- N_{im}: expected number of years student i stays at the institution, if given award m
- $d(i,j)$: 1 if applicant i dominates applicant j; 0 otherwise
- B: total budget for financial aid
- A_m: monetary value of award m
- T_i: tuition paid by applicant i
- SSI_i: government compensation for applicant i graduates

Objective

Maximize the total revenue: Tuition revenue + SSI income:

$$\max \quad \sum_{i\in I}\sum_{m\in M} x_{im}\cdot p^e_{im}\cdot (T_i - A_m)\cdot N_{im} + \sum_{i\in I}\sum_{m\in M} x_{im}\cdot p^e_{im}\cdot p^g_{im}\cdot SSI_i$$

- Revenue income: Prob E * (Tuition - Scholarship)* NumYears
- SSI income (The State Share of Instruction (SSI) is an allocation formula based on student outcomes): Prob E * Prob G * SSI

PACIFIC LIFE

Constraints

1. Each student only gets one scholarship:
   $\sum_{m \in M} x_{im} = 1 \quad \forall i \in I$

2. Total budget constraint:
   $\sum_{i \in I} \sum_{m \in M} x_{im} \cdot p^{e}_{im} \cdot A_m \leq B$

3. Dominance constraint:
   $\sum_{m \in M} x_{im} \cdot A_m \geq \sum_{m \in M} x_{jm} \cdot A_m \quad \forall (i,j) | d(i,j) = 1$

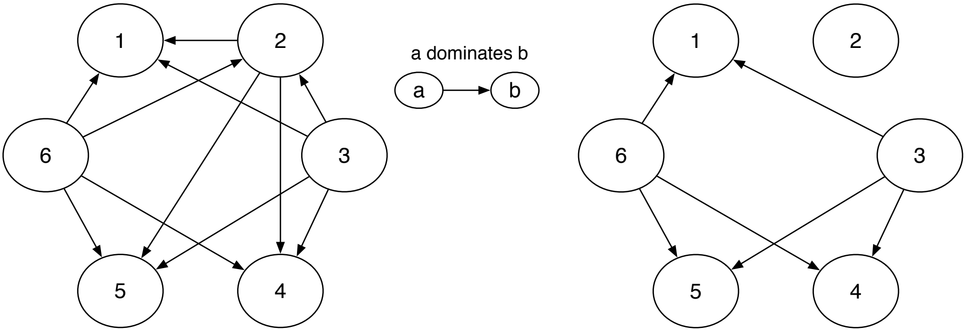
PACIFIC LIFE

Pair-wise Dominance Constraints

Around 5,500 applicants each year.
For the dominance constraints:
There are $(5,500 \times 5,500)/2$ or more than 15 million constraints.

Applicant	GPA	ACT
1	2.9	18
2	3.7	21
3	3.8	30
4	2.7	21
5	3.3	17
6	3.9	27

Full and redundant dominance

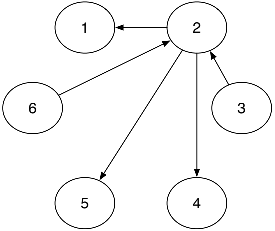


$\sum_{m \in M} x_{im} \cdot A_m \geq \sum_{m \in M} x_{jm} \cdot A_m \quad \forall (i,j) | d(i,j) = 1$

There are total 11 constraints in this case, 6 of them are redundant.

PACIFIC LIFE

Minimum dominance

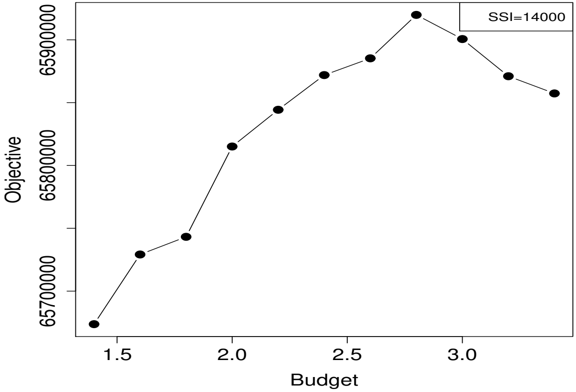


	1	2	3	4	5	6
1	0	0	0	0	0	0
2	1	0	0	1	1	0
3	0	1	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	1	0	0	0	0

Size of the optimization models

Model Components		Original Model	Reduced Model
Variables	Allocation (binary) x_i	57,860	57,860
Constraint	One Award per ID	5,260	5,260
	Dominance	**13,833,800**	**191,497**
	Total Budget	1	1
	Total Number of Constraints	13,839,061	196,758

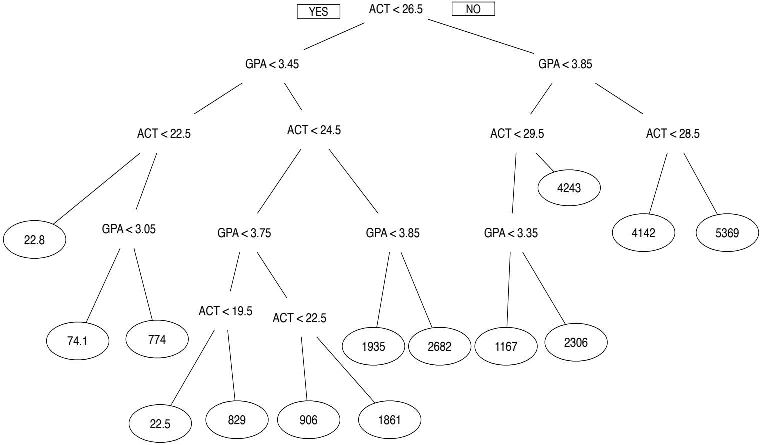
Optimization Results



Policy and Implementation

- First phase solves the prediction problems.
- Second phase solves optimal allocation problem.
- Enrollment administration needs a simple policy to implement.
- Decision tree and piecewise linear regression were used for this task.

Decision tree policy



Optimization Mean Scholarship vs GPA and ACT

GPA/ACT	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	Total
1																			0
1.1																			0
1.2		0																	0
1.3	0																		0
1.4		0																	0
1.5		0							0										0
2.5	0	0	0	0	0	0	0	0	0	0	0	0		0					0
2.6	0	0	0	0	0	0	0	0	0	0		1250	1500			1500			25.8
2.7	0	0	0	0	0	0	0	0	0	0	1500		2000						24.2
2.8	0	0	0	0	0	0	0	0	0	300	1500	2000				2500			39.4
2.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	818.2	1875.0	2000.0								63.4
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1166.7	1500.0	2000.0	2000.0	2200.0		3200.0			5200.0		216.3
3.1	0.0	0.0	0.0	0.0	0.0	0.0	523.8	1500.0	1727.3	2000.0	2000.0	2500.0	2500.0					7300.0	276.7
3.2	0.0	0.0	0.0	0.0	0.0	781.3	1000.0	1750.0	2000.0	2250.0	2500.0	2500.0	2780.0	3950.0	5200.0				573.7
3.3	0.0	0.0	0.0	0.0	647.1	1000.0	1613.6	2000.0	2315.8	2500.0	2920.0	3200.0	3866.7	4200.0		5200.0	5200.0		962.8
3.4	0.0	0.0	0.0	694.4	1000.0	1550.0	2000.0	2000.0	2500.0	2990.0	3200.0	3200.0	4200.0	4200.0					1185.1
3.5	0.0	0.0	695.7	1000.0	1724.1	2000.0	2000.0	2216.7	2500.0	3200.0	3200.0	3200.0	4200.0	4200.0		5200.0			1696.7
3.6	0.0	545.5	1000.0	1000.0	2000.0	2000.0	2357.1	2500.0	2864.0	3200.0	3200.0	3800.0	4200.0	4200.0	5200.0	5200.0	5200.0		2148.1
3.7	0.0	1000.0	1000.0	1500.0	2000.0	2285.7	2500.0	2945.5	3200.0	3200.0	3700.0	4200.0	4200.0	5200.0	5200.0		5200.0		2532.8
3.8	0.0	1000.0	1000.0	2000.0	2000.0	2500.0	3006.9	3200.0	3200.0	4014.8	4200.0	4200.0	4900.0	5200.0	5200.0	5200.0			3007.5
3.9	666.7	1000.0	1000.0	2000.0	2000.0	2500.0	3200.0	3200.0	3680.0	4200.0	4200.0	4644.4	5200.0	5200.0	5200.0	5200.0	5200.0		3360.4
4		1000.0	1000.0	2000.0	2000.0	2500.0	3200.0	3200.0	4200.0	4808.7	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	6542.9	7300.0	4186.1
4.1		1000.0	1000.0	2000.0	2000.0	2500.0	3200.0	3200.0	4200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	7300.0		3982.1
4.2	1000.0		1000.0	2000.0	2000.0	2500.0	3200.0	3200.0	4200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	7300.0		4175.0
4.3					2000.0	2500.0	3200.0	3200.0	4200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	7300.0		4607.9
4.4							3200.0	4200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	5200.0	7300.0		5125.0
4.5			1000.0		2000.0	2925.0		5200.0	5200.0		5200.0	5200.0		5200.0	5200.0	6233.3		7300.0	4605.3
4.6						4200.0			5200.0	5200.0	5200.0		5200.0		5900.0	7300.0	7300.0		5730.0
4.7				5200.0					5200.0	5200.0	6200.0			8400.0	8400.0	8400.0			6377.8
4.8										5200.0	6200.0		6200.0						5866.7
Grand Total	7.2	92.7	166.2	428.7	787.1	1206	1664.8	2112	2642.1	3407.5	3835.6	3981.3	4561.4	4784.9	5323.2	5258.8	6247.6	7300	1134.7

Piecewise linear regression based policy

Piecewise linear regression using composite score:

Composite Score	# of Applicants	Scholarship Amount
0-53.9	2,897	0
54-68.9	2,103	$309 \times CS - 16,380$
69-76.9	241	$101 \times CS - 2,024$
77-80	19	$711 \times CS - 48,910$

PACIFIC LIFE

Piecewise linear regression based policy

Simplified version of policy in piecewise regression form.

Composite Score	Scholarship Amount	# of Applicants
0-54.9	0	3,074
55-59.9	1,500	872
60-65.9	2,500	812
66-69.9	3,500	298
70-74.9	4,500	166
75+	6,000	38

Note: 41.6% applicants receive scholarship.

Business impact and Application

- The result of the study has been successfully implemented in the state university and has resulted in millions of financial benefits.
- The research would be applicable to many other institutions and offers a methodology, tools and insights into the solution of financial aid problems.

PACIFIC LIFE

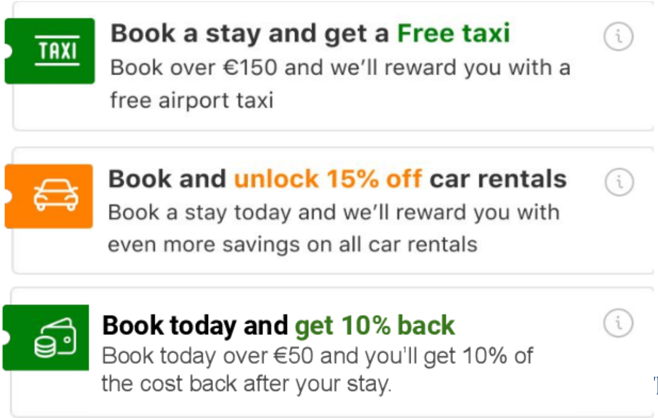
Business impact and Application

	2013	2014	# Increase	% Increase
Application	6,101	6,068	-43	-0.7%
Admitted	4,541	4,773	232	5.1%
Non-Scholarship	2,166	2,157	-9	-0.4%
Scholarship Award	2,375	2,616	241	**10.1%**
Matriculated	2,001	2,222	221	**11.0%**

Summary

1. A series of models are developed to predict:
   - Enrollment probability
   - Graduation probability
   - Number of years of stay
2. Developed minimum cardinality dominance table to reduce the model size.
3. An optimization model is developed with the objective to maximize the revenue.
4. A regression analysis is developed to translate the optimization results to managerial insights and derive a policy for implementation.

Personalized Coupons



Personalized Coupons

- Promotions are expected to generate a significant uplift in sales.
- Net revenue loss of a promotion: MSRP of a product \$100, discount \$25, break-even \$80, then the loss is $80 - (100 - 25) = \$5$.
- A dedicated budget usually limits this incremental net revenue loss.

PACIFIC LIFE

CATE

- The effect a promotion has on the probability of completing a purchase and on the expected net revenue loss varies from customer to customer.
- The Conditional Average Treatment Effect (CATE): the expected change in a metric of interest (conversion, revenue, click rate, churn, etc.) caused by a treatment, given the individual's characteristics, commonly known as Uplift Modeling.

ID	Gender	With Treatment	Without
1	M	40	30
2	M	20	20
3	F	10	15
4	F	30	30

- Average Treatment Effect: (10+0-5+0)/4 =5/4
- CATE(Male): (10+0)/2 = 5; CATE(Female): (-5+0)/2= -2/5

PACIFIC LIFE

Problem Formulation

- $Y_i(k)$: potential purchase if customer i is offered promotion k
- $R_i(k)$: potential net revenue if customer i is offered promotion k
  when k=0: potential outcomes if no promotion is offered to customer i.
- $CATE_Y(i, k)$: incremental effect on the expected purchase probability of customer i if presented with promotion k.
- $CATE_R(i, k)$: incremental effect on the expected net revenue
- $CATE_L(i, k)$: incremental effect on the expected net revenue loss, which equals to $-CATE_R(i, k)$

PACIFIC LIFE

Model: Multiple-Choice Knapsack Problem

- maximize revenue:

$$\sum_{i \in U} \sum_{k \in K_i} CATE_Y(i,k) * Z_{ik}$$

- subject to budget:

$$\sum_{i \in U} \sum_{k \in K_i} CATE_L(i,k) * Z_{ik} \leq C, \forall i \in U, k \in K_i$$

- Allow customer to pick at most one promotion from a finite set of eligible promotions:

$$\sum_{k \in K_i} Z_{ik} = 1, \forall i \in U$$

- Variable: $Z_{ik} \in \{0,1\} \quad \forall i \in U, k \in K_i$
- Reading:E-Commerce Promotions Personalization via Online Multiple-Choice Knapsack with Uplift Modeling @bookings.com

The End

Thanks!