Optimal Targeting of Customers for a Last-Minute Sale

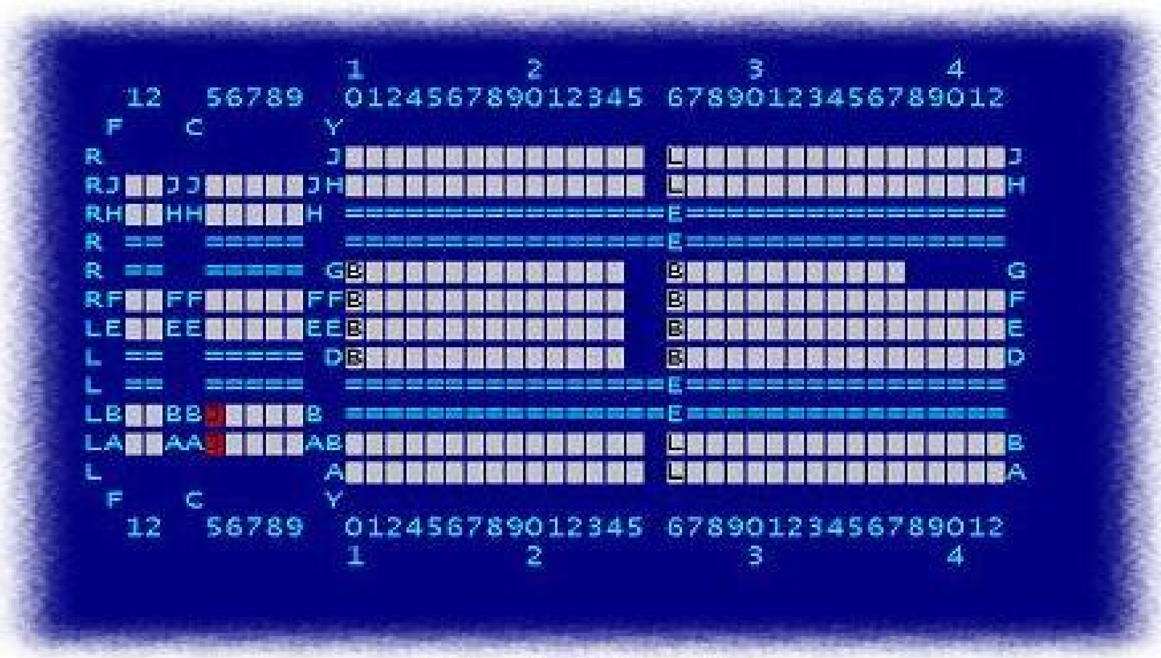
R. Cominetti, J. Correa, J. San Martín

Universidad de Chile

Journées Franco-Chiliennes d'Optimisation — Université de Toulon — Mai 2008

Want to sell business class upgrades

Who should get the offer?



Airbus A340-300 Seat Map

The Process — (1) Set of customers N

(2) Address the offer to $S \subseteq N$

(3) Customers accept/reject $A \subseteq S \subseteq N$

(4) Winner is chosen at random

The Single Item case

Setting	**Implication**
set of clients $i \in N$	select $S \subseteq N$ to offer
different revenue for each client v_i	want high revenue clients
different acceptance probability p_i	want high pbb clients
until sold out	first respondent wins
last minute	no time for regret

revenue $=$ discount price $-$ normal price $\times$ prob buys anyway

Goal

balance probabilities and revenues so that the selected $S \subseteq N$ maximizes expected revenue

The Problem — Discrete Version

- Expected revenue for $S \subseteq N$ is:

$$V_S = \sum_{A \subseteq S} \overbrace{\frac{v(A)}{|A|}}^{\text{REVENUE IF } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in A} p_i}^{\text{PROB THAT } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in S \setminus A} (1 - p_i)}^{\text{PROB THAT } S \setminus A \text{ REJECTS}}$$

Problem: find $V^* = \max_{S \subseteq N} V_S$

- V_S can be computed in $O(n^3)$ using convolutions

The Problem — Continuous Version

- Offer is made to each client with probability x_i

$$\mathbb{P}[i \text{ accepts}] = \mathbb{P}[Y_i = 1] = x_i \cdot p_i = y_i$$

$$V(y) = \sum_{A \subseteq N} \overbrace{\frac{v(A)}{|A|}}^{\text{REVENUE IF } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in A} y_i}^{\text{PROB THAT } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in N \setminus A} (1 - y_i)}^{\text{PROB THAT } N \setminus A \text{ REJECTS}}$$

Problem: find $V^* = \max_{0 \leq y_i \leq p_i} V(y)$

- Both problems are equivalent since $V(y)$ is linear in each variable y_i

Threshold Strategies

- If offer is made to a customer reporting v_i, shouldn't we also consider those customers with higher values?

- Find a threshold value V and offer to all clients such that $v_i \geq V$
  We assume $v_1 \geq v_2 \geq \cdots \geq v_n$

- Optimal threshold found in $O(n^4)$: $\max_{1 \leq i \leq n} V_{\{1,\ldots,i\}}$

- Typical in *revenue management*... but...

Threshold Strategies are not optimal!

$$V(1) = \tfrac{1}{2} \cdot 2 \qquad = 1$$

$$V(1,2) = \tfrac{1}{4} \cdot 2 + \tfrac{1}{4} \cdot 1 + \tfrac{1}{4} \cdot \tfrac{3}{2} + \tfrac{1}{4} \cdot 0 \qquad = 1.125$$

$$V(1,2,3) = \tfrac{1}{4} \cdot \tfrac{2.9}{2} + \tfrac{1}{4} \cdot \tfrac{1.9}{2} + \tfrac{1}{4} \cdot \tfrac{3.9}{3} + \tfrac{1}{4} \cdot 0.9 \qquad = 1.15$$

$$V(1,3) = \tfrac{1}{2} \cdot \tfrac{2.9}{2} + \tfrac{1}{2} \cdot 0.9 \qquad = 1.175$$

p_i	v_i
0.5	2
0.5	1
1	0.9

- Every subset can be optimal
- Sorting by probability or expected value is also sub-optimal

Heuristics: 10 customers, 200 instances

Algorithm	% opt	Min ratio	Avg ratio	Time
optimal	100.0	1.0000	1.0000	1611.2
THRESHOLD	93.0	0.9916	0.9998	23.5
LP-RELAX	47.5	0.8168	0.9771	0.4
LP2-RELAX	74.0	0.9775	0.9988	21.5
IN-OUT	99.0	0.9918	0.9999	15.4

Problem complexity is open

Threshold is $\frac{1}{2}$-optimal: LP relaxation

Rewrite objective function as

$$\boxed{V(y) = \sum_{i \in N} v_i \pi_i}$$

$$\pi_i = \mathbb{P}[i \text{ accepts and wins}] = y_i \, \mathbb{E}[\tfrac{1}{1+S_i}]$$

$$S_i = \sum_{j \neq i} Y_j \text{ (number of competitors)}$$

$$\Longrightarrow \quad 0 \leq \pi_i \leq p_i \quad \text{and} \quad \sum_{i \in N} \pi_i \leq 1$$

Threshold is $\frac{1}{2}$-optimal: LP relaxation

Rewrite objective function as

$$\boxed{V(y) = \sum_{i \in N} v_i \pi_i}$$

$$\pi_i = \mathbb{P}[i \text{ accepts and wins}] = y_i \, \mathbb{E}[\tfrac{1}{1+S_i}]$$

$$S_i = \sum_{j \neq i} Y_j \text{ (number of competitors)}$$

$$\Longrightarrow \quad 0 \leq \pi_i \leq p_i \quad \text{and} \quad \sum_{i \in N} \pi_i \leq 1$$

Consider the relaxation (upper bound)

$$V^* \leq V^{LP} = \max_{0 \leq y_i \leq p_i} \{\sum_{i \in N} v_i y_i : \sum_{i \in N} y_i \leq 1\}$$

and use it to get a $\frac{1}{2}$-optimal threshold strategy

Polynomial approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^{k} p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^{k} p_i & \text{if } i = k+1 \\ 0 & \text{otherwise} \end{cases}$$

Polynomial approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^{k} p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^{k} p_i & \text{if } i = k+1 \\ 0 & \text{otherwise} \end{cases}$$

- y^{LP} is a randomized strategy equivalent to

$$\begin{cases} \text{select } \{1, \ldots, k\} \text{ with probability } [\sum_{i=1}^{k+1} p_i - 1]/p_{k+1} \\ \text{select } \{1, \ldots, k+1\} \text{ with probability } [1 - \sum_{i=1}^{k} p_i]/p_{k+1} \end{cases}$$

Polynomial approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^{k} p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^{k} p_i & \text{if } i = k+1 \\ 0 & \text{otherwise} \end{cases}$$

- y^{LP} is a randomized strategy equivalent to

$$\begin{cases} \text{select } \{1, \ldots, k\} \text{ with probability } [\sum_{i=1}^{k+1} p_i - 1]/p_{k+1} \\ \text{select } \{1, \ldots, k+1\} \text{ with probability } [1 - \sum_{i=1}^{k} p_i]/p_{k+1} \end{cases}$$

- De-randomize in $O(n^3)$: max between $V_{\{1,\ldots,k\}}$ and $V_{\{1,\ldots,k+1\}}$

In the example:

p_i	v_i
0.5	2
0.5	1
1	0.9

$$\begin{aligned}
y_1^{LP} &= \tfrac{1}{2} \\
y_2^{LP} &= \tfrac{1}{2} \\
y_3^{LP} &= 0 \\
V^{LP} &= \tfrac{1}{2} \cdot 2 + \tfrac{1}{2} \cdot 1 && = 1.5 \\
V(y^{LP}) &= \tfrac{1}{4} \cdot 2 + \tfrac{1}{4} \cdot 1 + \tfrac{1}{4} \cdot \tfrac{3}{2} + \tfrac{1}{4} \cdot 0 && = 1.125
\end{aligned}$$

Theorem: ALG^{LP} **is a** $\frac{1}{2}$**−approximation algorithm**

Theorem: ALG^{LP} is a $\frac{1}{2}-$approximation algorithm

Proof: By Jensen's inequality

$$\mathbb{E}[\tfrac{1}{1+S_i}] \geq \frac{1}{1+\mathbb{E}(S_i)} = \frac{1}{1+\sum_{j\neq i} y_j^{LP}} \geq \frac{1}{1+1} = \frac{1}{2}$$

hence

$$V(y^{LP}) = \sum_{i\in N} v_i\, y_i^{LP}\, \mathbb{E}[\tfrac{1}{1+S_i}] \geq \tfrac{1}{2} \sum_{i\in N} v_i\, y_i^{LP}$$

so that

$$V^* \geq V(y^{LP}) \geq \tfrac{1}{2} V^{LP} \geq \tfrac{1}{2} V^*.$$

□

Alternative: Hyperbolic relaxation

$$\pi_i \geq \frac{y_i}{1+\sum_{j\neq i} y_j} \geq \frac{y_i}{1+\sum_{j\in N} y_j}$$

$$\boxed{V^* \geq \max_{0\leq y_i\leq p_i} \frac{\sum_{i\in N} v_i y_i}{1+\sum_{i\in N} y_i}}$$

- *Common-lines problem* in transit equilibrium (Chriqui&Robillard'75)
- Optimum is a threshold strategy
- Linear-time algorithm: $\max_k \, [v_1 p_1 + \cdots + v_k p_k]/[1 + p_1 + \cdots + p_k]$
- Also a $\frac{1}{2}$-approximation algorithm

Improved $\frac{2}{3}$-approximation

Let $x = \mathbb{P}[S = 0]$ so that $x + \sum_{i \in N} \pi_i = 1$

Moreover $\pi_i = y_i \, \mathbb{E}[\frac{1}{1+S_i}]$ with

$$\begin{aligned} \mathbb{E}[\tfrac{1}{1+S_i}] &\leq 1 \cdot \mathbb{P}[S_i = 0] + \tfrac{1}{2} \cdot \mathbb{P}[S_i > 0] \\ &= \tfrac{1}{2}(1 + \mathbb{P}[S_i = 0]) \\ &= \tfrac{1}{2}(1 + \tfrac{x}{1-y_i}) \end{aligned}$$

and then $y_i \leq p_i$ implies $\pi_i \leq \frac{p_i}{2}(1 + \frac{x}{1-p_i})$

Hence we get the alternative LP relaxation

$$
\begin{aligned}
V^* \leq V^{LP_2} \quad = \quad & \max \sum_{i \in N} v_i z_i \\
& z_i \leq \frac{p_i}{2}\left(1 + \frac{x}{1-p_i}\right) \\
& x + \sum_{i \in N} z_i = 1 \\
& x, z_i \geq 0
\end{aligned}
$$

Algorithm ALG^{LP_2}

- Find a *basic* optimal solution (z^*, x^*) for LP_2
- Set $y_i^{LP_2} = \frac{2z_i^*}{1+\frac{x^*}{1-p_i}}$ ...either 0 or p_i except for one value!
- De-randomize y^{LP_2} to get a set of the form $\{1, \ldots, k\}$

Theorem: ALG^{LP_2} is a $\frac{2}{3}$-approximation

Theorem: ALG^{LP_2} is a $\frac{2}{3}$-approximation

Proof: $V(y) = \sum_{i\in N} v_i y_i \mathbb{E}[\frac{1}{1+S_i}] \geq \sum_{i\in N} \frac{v_i y_i}{1+\sum_{j\neq i} y_j}$

Replacing $y_i^{LP_2}$ we get $V(y^{LP_2}) \geq \sum_{i\in N} v_i z_i^* \gamma_i$ with

$$\gamma_i = \frac{2(1+x^*)}{(3-x^*-2z_i^*)(1+\frac{x^*}{1-p_i})}.$$

Since $V^* \leq \sum_{i\in N} v_i z_i^*$ we need $\gamma_i \geq \frac{2}{3}$. This is obvious if $x^* = 0$. Else, since (z^*, x^*) is a basic solution exactly one of the two inequalities involving z_i^* is tight: if $z_i^* > 0$ then $z_i^* = \frac{p_i}{2}(1 + \frac{x^*}{1-p_i})$ so that

$$\gamma_i = \frac{2(1+x^*)}{(3-p_i-\frac{x^*}{1-p_i})(1+\frac{x^*}{1-p_i})} \geq \frac{2}{3}.$$

□

Multiple items

When there are m available items the value of a strategy y is

$$V(y) = \overbrace{\sum_{A\subseteq N} \min\left\{1, \frac{m}{|A|}\right\} v(A)}^{\text{REVENUE IF } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i\in A} y_i}^{\text{PROB THAT } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i\in S\setminus A} (1-y_i)}^{\text{PROB THAT } S\setminus A \text{ REJECTS}}$$

Multiple items

When there are m available items the value of a strategy y is

$$V(y) = \overbrace{\sum_{A\subseteq N} \min\left\{1, \frac{m}{|A|}\right\} v(A)}^{\text{REVENUE IF } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i\in A} y_i}^{\text{PROB THAT } A \text{ ACCEPTS}} \cdot \overbrace{\prod_{i\in S\setminus A} (1-y_i)}^{\text{PROB THAT } S\setminus A \text{ REJECTS}}$$

$$= \sum_{i\in N} v_i y_i \mathbb{E}[\min\{1, \tfrac{m}{1+S_i}\}]$$

> Problem: find $V^* = \max_{0\leq y_i\leq p_i} V(y)$

Complexity is open though $V(y)$ is computed in $O(n^3)$ by convolution

The trivial extension of LP gives an upper bound

$$V^{LP} = \max\left\{\sum_{i\in N} v_i y_i : \sum_{i\in N} y_i \le m \text{ and } 0 \le y_i \le p_i\right\}$$

The trivial extension of LP gives an upper bound

$$V^{LP} = \max \left\{ \sum_{i \in N} v_i y_i : \sum_{i \in N} y_i \le m \text{ and } 0 \le y_i \le p_i \right\}$$

which guarantees a constant factor of the optimum:

- $V(y) = \sum_{i \in N} v_i y_i \mathbb{E}[\min\{1, \frac{m}{1+S_i}\}] \ge \rho(y) \sum_{i \in N} v_i y_i$
- $\rho(y) = \mathbb{E}[\min\{1, \frac{m}{1+S}\}] \ge 1 - \frac{1}{m+1} \sqrt{1 + \mathrm{Var(S)}}\, \mathbb{P}(S \ge m)$

find lower bound for $\rho(y)$...

Asymptotically optimal bound

since $\mathrm{Var}(S) = \sum_i y_i(1-y_i) \leq m$ and $\mathbb{P}(S \geq m) \leq 1$, we get

$$\rho(y) \geq 1 - \frac{1}{\sqrt{m+1}}$$

so that $V(y^{LP})$ is within $\frac{1}{\sqrt{m+1}}$ from optimal ...but not better!

Sharper bound

$$\rho(y) \geq 1 - \frac{\sqrt{1+\sigma^2}}{m+1} \min\{1, \tfrac{1}{2} + \tfrac{M}{\sigma}\}$$

$$M = \max_{u \geq 0} \sqrt{2u}\, e^{-2u} \sum_{k=0}^{\infty} (\tfrac{u^k}{k!})^2 \sim 0.46882235549939533$$

"if the mean number of successes in n independent heterogeneous Bernoulli trials is an integer m then the median is also m"
(Jogdeo-Samuels'62, Siegel'01)

$$\mathbb{P}[S \geq m] \leq \tfrac{1}{2} + \mathbb{P}[S = m] \leq \tfrac{1}{2} + \tfrac{M}{\sigma}$$

(latter estimate from C+Vaisman'08)

Extensions — Buy by Bulk

Dean, Goemans, Vondrak 04

- The difference is that customer $i \in N$ demands $s_i \geq 1$ units
- Problem is now NP-hard: for $p_i = 1$ we get knapsack
- Computing $V(y)$ is actually $\#$P-complete
  (e.g. count the number of perfect matchings in a bipartite graph)

Extensions — Multistage

- Medium-term: address offers in k successive rounds.
- Find an adaptive online algorithm that computes an optimal subset of clients to address the offer in the next round.

Extensions — Pricing

- Personalized contact (email): price can be different for each client
- Pbb of accepting is a decreasing function of price $p_i : \mathbb{R}_+ \to [0, 1]$
- Select a price for each customer to maximize expected revenue