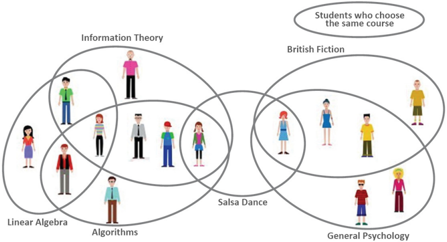
Stochastic Matching in Hypergraphs

Amit Chavan, Srijan Kumar and Pan Xu



May 13, 2014

ROADMAP

MATCHING

Definition

Given a (hyper)graph $G(V, E)$ a *matching* or *independent edge set* is a subset of E such that no two of them have a vertex in common.

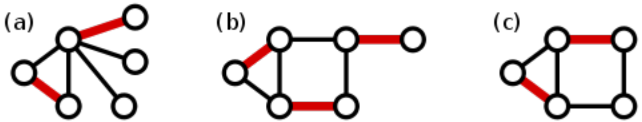


Figure: http://en.wikipedia.org/wiki/File:Maximum-matching-labels.svg

MATCHING

Resource allocation

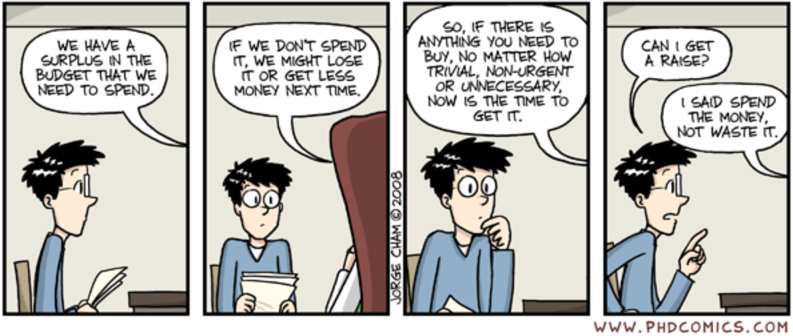


Figure: http://www.phdcomics.com/comics/archive/phd051908s.gif

MATCHING

Stable Marriage Problem

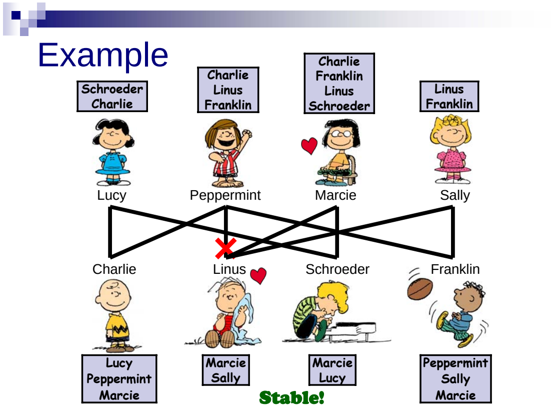


Figure: http://cramton.umd.edu/econ415/deferred-acceptance-algorithm.pdf

MATCHING

Latin Square Problem

5	3	4	6	7	8	9	1	2
6	7	2	1	9	5	3	4	8
1	9	8	3	4	2	5	6	7
8	5	9	7	6	1	4	2	3
4	2	6	8	5	3	7	9	1
7	1	3	9	2	4	8	5	6
9	6	1	5	3	7	2	8	4
2	8	7	4	1	9	6	3	5
3	4	5	2	8	6	1	7	9

Figure: http://upload.wikimedia.org/wikipedia/commons/thumb/3/31/Sudoku-by-L2G-20050714_solution.svg/250px-Sudoku-by-L2G-20050714_solution.svg.png

STOCHASTIC MATCHING

Setting:

- Each edge e is present independently with probability p_e.
- **Objective:** Maximum matching in graph given $p_e \quad \forall e \in E$.
- We don't know whether edge is present or not - just the probability.
- To find, query the edge, and if the edge is present, add it to matching – "probing" of edge.
- **Task:** Adaptively query the edge to maximize the expected matching weight.
- First introduced and studied by Chen, Immorlica, Karlin, Mahdian, and Rudra [2009].

WHY IS IT IMPORTANT?

Motivated by:

- Kidney exchange
- Online dating

Why is it important? - Kidney Exchange

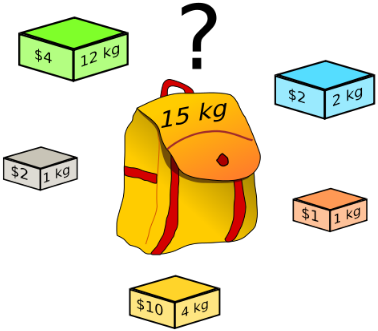
Figure: http://www.cartoonstock.com/lowres/animals-transplant-pig-kidney-transplantation-surgeon-dro0315l.jpg

WHY IS IT IMPORTANT? - ONLINE DATING

Figure: http://www.cartoonstock.com/newscartoons/cartoonists/bst/lowres/dating-wrong-conversations-arguments-issues-disagree-bstn86l.jpg

STOCHASTIC KNAPSACK PROBLEM

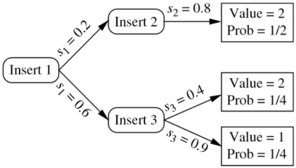
- Classical Knapsack
  - n items
  - Item i has size s_i and profit v_i
  - Knapsack capacity W
  - **Goal:** Compute the max profit feasible subset S
- Stochastic Knapsack
  - s_i are independent random variables with known distribution
  - **Goal:** Find a policy such that the expected weight of the inserted items is maximized
  - **Caveat:** Stop when knapsack overflows



ADAPTIVE AND NON-ADAPTIVE POLICIES

- Non-adaptive policy
  - An ordering $\mathcal{O} = \{i_1, i_2, \ldots, i_n\}$ of items
  - NON-ADAPT$(I) = \max_{\mathcal{O}} \mathbb{E}[val(\mathcal{O})]$
  - Optimal $\mathcal{O} = \{1, 2, 3\}, \mathbb{E}[val(\mathcal{O})] = 1.5$
- Adaptive policy
  - Function $\mathcal{P} : 2^{[n]} \times [0, 1] \to [n]$
  - Given a set of inserted items J and remaining capacity c, $\mathcal{P}(J, c)$ is the next item to insert
  - ADAPT$(I) = \max_{\mathcal{P}} \mathbb{E}[val(\mathcal{P}(\emptyset, 1))]$
  - Optimal expected value = 1.75

Item	Size distribution			
	s	p	s	p
1	0.2	1/2	0.6	1/2
2	0.8	1	-	-
3	0.4	1/2	0.9	1/2



ADAPTIVITY GAP

For an instance I,

$$\text{ADAPTIVITY-GAP}(I) = \sup \frac{\text{ADAPT}(I)}{\text{NON-ADAPT}(I)}$$

- Studied by Dean, Goemans, and Vondrák [2005].
- They show that for d-dimensional knapsack, the gap can be $\Omega(\sqrt{d})$.
- They also give a non-adaptive $O(d)$ approximation to the optimal adaptive.

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns
- Item i has random profit $v_i \in \mathbb{R}_+$, and a random d-dimensional size $s_i \in \{0,1\}^d$

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns
- Item i has random profit $v_i \in \mathbb{R}_+$, and a random d-dimensional size $s_i \in \{0,1\}^d$
- The probability distributions of different items are independent

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns
- Item i has random profit $v_i \in \mathbb{R}_+$, and a random d-dimensional size $s_i \in \{0,1\}^d$
- The probability distributions of different items are independent
- Each item takes non-zero size in at most k co-ordinates (out of d)

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns
- Item i has random profit $v_i \in \mathbb{R}_+$, and a random d-dimensional size $s_i \in \{0,1\}^d$
- The probability distributions of different items are independent
- Each item takes non-zero size in at most k co-ordinates (out of d)
- A capacity vector $b \in \mathbb{Z}_+$ into which the items must be packed

PRELIMINARIES – STOCHASTIC k-SET PACKING

An instance I consists of

- n items/columns
- Item i has random profit $v_i \in \mathbb{R}_+$, and a random d-dimensional size $s_i \in \{0, 1\}^d$
- The probability distributions of different items are independent
- Each item takes non-zero size in at most k co-ordinates (out of d)
- A capacity vector $b \in \mathbb{Z}_+$ into which the items must be packed
- **Goal:** Find an adaptive strategy of choosing items such that the expected profit is maximized

LP RELAXATION FOR OPTIMAL ADAPTIVE (BANSAL ET AL. [2010])

LP relaxation for optimal adaptive (Bansal et al. [2010])

Let $w_i = \mathbb{E}[v_i]$ be the mean profit, for each $i \in [n]$.

LP relaxation for optimal adaptive (Bansal et al. [2010])

Let $w_i = \mathbb{E}[v_i]$ be the mean profit, for each $i \in [n]$.
Let $\mu_i(j) = \mathbb{E}[s_i(j)]$ be the expected size of the i-th item in j-th co-ordinate.

LP relaxation for optimal adaptive (Bansal et al. [2010])

Let $w_i = \mathbb{E}[v_i]$ be the mean profit, for each $i \in [n]$.
Let $\mu_i(j) = \mathbb{E}[s_i(j)]$ be the expected size of the i-th item in j-th co-ordinate.

$$\text{maximize } \sum_{i=1}^{n} w_i y_i \tag{1}$$

$$\text{subject to } \sum_{i=1}^{n} \mu_i(j) y_i \leq b_j, \forall j \in [d] \tag{2}$$

$$y_i \in [0,1] \qquad , \forall i \in [n] \tag{3}$$

y_i is the probability that the adaptive algorithm probes item i.

A Solution Policy

- Let y^* denote an optimal solution to the linear program in 1.

A SOLUTION POLICY

- Let y^* denote an optimal solution to the linear program in 1.
- Fix a constant $\alpha \geq 1$ (to be specified later).

A SOLUTION POLICY

- Let y^* denote an optimal solution to the linear program in 1.
- Fix a constant $\alpha \geq 1$ (to be specified later).
- Pick a permutation $\pi : [n] \to [n]$ uniformly at random.

A SOLUTION POLICY

- Let y^* denote an optimal solution to the linear program in 1.
- Fix a constant $\alpha \geq 1$ (to be specified later).
- Pick a permutation $\pi : [n] \to [n]$ uniformly at random.
- Inspect items/columns in the order of π.

A SOLUTION POLICY

- Let y^* denote an optimal solution to the linear program in 1.
- Fix a constant $\alpha \geq 1$ (to be specified later).
- Pick a permutation $\pi : [n] \to [n]$ uniformly at random.
- Inspect items/columns in the order of π.
- Probe item c with probability y_c/α if and only if it is safe to do so.

APPROXIMATION RATIO

For any column $c \in [n]$, let $\{I_{c,l}\}_{l=1}^{k}$ denote the indicator random variable that the l-th constraint in the support of c is tight when c is considered in π.

APPROXIMATION RATIO

For any column $c \in [n]$, let $\{I_{c,l}\}_{l=1}^{k}$ denote the indicator random variable that the l-th constraint in the support of c is tight when c is considered in π.

$\Pr[c \text{ is safe when considered}] = \Pr[\wedge_{l=1}^{k} \neg I_{c,l}] \geq 1 - \sum_{l=1}^{k} \Pr[I_{c,l}]$

APPROXIMATION RATIO

For any column $c \in [n]$, let $\{I_{c,l}\}_{l=1}^{k}$ denote the indicator random variable that the l-th constraint in the support of c is tight when c is considered in π.
$\Pr[c \text{ is safe when considered}] = \Pr[\wedge_{l=1}^{k} \neg I_{c,l}] \geq 1 - \sum_{l=1}^{k} \Pr[I_{c,l}]$

Lemma

$$\Pr[I_{c,l}] \leq \frac{1}{2\alpha}$$

$2k$-APPROXIMATION

Let $j \in [d]$ be the l-th constraint in the support of c.

$2k$-APPROXIMATION

Let $j \in [d]$ be the l-th constraint in the support of c.
Let U_c^j denote the usage of constraint j, when c is considered.

$2k$-APPROXIMATION

Let $j \in [d]$ be the l-th constraint in the support of c.
Let U_c^j denote the usage of constraint j, when c is considered.

$$\begin{aligned}
\mathbb{E}[U_c^j] &= \sum_{a=1}^{n} \Pr[\text{column } a \text{ appears before } c \text{ AND } a \text{ is probed}]\mu_a(j) \\
&\leq \sum_{a=1}^{n} \Pr[\text{column } a \text{ appears before } c]\frac{y_a}{\alpha}\mu_a(j) \\
&= \sum_{a=1}^{n} \frac{y_a}{2\alpha}\mu_a(j) \\
&\leq \frac{b_j}{2\alpha}
\end{aligned}$$

$2k$-APPROXIMATION

Let $j \in [d]$ be the l-th constraint in the support of c.
Let U_c^j denote the usage of constraint j, when c is considered.

$$\begin{aligned}
\mathbb{E}[U_c^j] &= \sum_{a=1}^{n} \Pr[\text{column } a \text{ appears before } c \text{ AND } a \text{ is probed}]\mu_a(j) \\
&\leq \sum_{a=1}^{n} \Pr[\text{column } a \text{ appears before } c]\frac{y_a}{\alpha}\mu_a(j) \\
&= \sum_{a=1}^{n} \frac{y_a}{2\alpha}\mu_a(j) \\
&\leq \frac{b_j}{2\alpha}
\end{aligned}$$

Since $I_{c,l} = \{U_c^j \geq b_j\}$, by Markov's inequality, $\Pr[I_{c,l}] \leq \frac{\mathbb{E}[U_c^j]}{b_j} \leq \frac{1}{2\alpha}$.

$2k$-APPROXIMATION

- By union bound, the probability that a particular column c is safe when considered under π is at least $1 - \frac{k}{2\alpha}$.

$2k$-APPROXIMATION

- By union bound, the probability that a particular column c is safe when considered under π is at least $1 - \frac{k}{2\alpha}$.
- The probability of probing c is atleast $\frac{y_c}{\alpha}(1 - \frac{k}{2\alpha})$.

$2k$-APPROXIMATION

- By union bound, the probability that a particular column c is safe when considered under π is at least $1 - \frac{k}{2\alpha}$.
- The probability of probing c is atleast $\frac{y_c}{\alpha}(1 - \frac{k}{2\alpha})$.
- By linearity of expectation, the expected profit is at least $\frac{1}{\alpha}(1 - \frac{k}{2\alpha}) \sum_{c=1}^{n} w_c y_c$.

$2k$-APPROXIMATION

- By union bound, the probability that a particular column c is safe when considered under π is at least $1 - \frac{k}{2\alpha}$.
- The probability of probing c is atleast $\frac{y_c}{\alpha}(1 - \frac{k}{2\alpha})$.
- By linearity of expectation, the expected profit is at least $\frac{1}{\alpha}(1 - \frac{k}{2\alpha}) \sum_{c=1}^{n} w_c y_c$.
- Setting $\alpha = k$ implies an approximation ratio of $2k$.

RELATED WORK

- Stochastic Knapsack – Dean, Goemans, and Vondrák [2008]
  - A non-adaptive 4 approximation to the optimal adaptive
  - A adaptive $(3 + \varepsilon)$ approximation to the optimal adaptive

RELATED WORK

- Stochastic Knapsack – Dean, Goemans, and Vondrák [2008]
  - A non-adaptive 4 approximation to the optimal adaptive
  - A adaptive $(3 + \varepsilon)$ approximation to the optimal adaptive
- Stochastic k-Set Packing – Bansal et al. [2010]
  - $2k$ approximation in the general case
  - $k + 1$ approximation in the monotone outcome case

RELATED WORK

- Stochastic Knapsack – Dean, Goemans, and Vondrák [2008]
  - A non-adaptive 4 approximation to the optimal adaptive
  - A adaptive $(3 + \varepsilon)$ approximation to the optimal adaptive
- Stochastic k-Set Packing – Bansal et al. [2010]
  - $2k$ approximation in the general case
  - $k + 1$ approximation in the monotone outcome case
- Stochastic Matching in graphs (with patience constraints) – Bansal et al. [2010]
  - 3 approximation for bipartite graphs
  - 4 approximation for general graphs
- Matching in k-uniform hypergraph – Chan and Lau [2012]
  - $(k - 1 + 1/k)$ approximation in the deterministic case
  - The standard LP has an integrality gap of $(k - 1 + 1/k)$ – Füredi, Kahn, and Seymour [1993]

REFERENCES

Nikhil Bansal, Anupam Gupta, Jian Li, Julián Mestre, Viswanath Nagarajan, and Atri Rudra. When lp is the cure for your matching woes: Improved bounds for stochastic matchings. In *Algorithms–ESA 2010*, pages 218–229. Springer, 2010.

Yuk Hei Chan and Lap Chi Lau. On linear and semidefinite programming relaxations for hypergraph matching. *Mathematical programming*, 135(1-2): 123–148, 2012.

Ning Chen, Nicole Immorlica, Anna R Karlin, Mohammad Mahdian, and Atri Rudra. Approximating matches made in heaven. *Automata, Languages and Programming*, pages 266–278, 2009.

Brian C Dean, Michel X Goemans, and Jan Vondrák. Adaptivity and approximation for stochastic packing problems. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 395–404. Society for Industrial and Applied Mathematics, 2005.

Brian C Dean, Michel X Goemans, and Jan Vondrák. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research*, 33(4):945–964, 2008.

Zoltán Füredi, Jeff Kahn, and Paul D. Seymour. On the fractional matching polytope of a hypergraph. *Combinatorica*, 13(2):167–180, 1993.