Online Appendix to: Matching Models for Preference-sensitive Group Purchasing

TYLER LU, University of Toronto
CRAIG BOUTILIER, University of Toronto

A. MODEL OF RANDOM DISCOUNT MATCHING INSTANCES

This appendix describes the model of random discount matching instances generated for the experiments in Sec. 7. While this model may not be realistic in some ways, it represents what the authors believe to be plausible.

Product (Vendor) Model

The key idea here is that different vendors will offer goods of varying "quality", where quality can vary on several different attributes (including the special case of one quality measure). We use two attributes.

- For each vendor j, draw random quality $Q_a \sim \text{Beta}(4, 4)$ for each attribute a.
- Let overall quality Q_j for vendor j be $f(Q_{a_1}, Q_{a_2}, ...) = 20 \prod_a Q_a$.
- Set the initial price $p_j^0 = Q_j + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 0.05Q_j)$ is a small noise parameter

Discount Model

We keep the discount level to be fixed at 4 (including the default base threshold) for the experiments, although a straightforward extension of the description below can allow for arbitrary discount levels.

- Draw first discount percentage d_1 from the multinomial $[5\%\ 0.2; 10\%\ 0.6; 15\%\ 0.2]$ then set the discounted price $p_j^1 = (100 - d_1)\%$ of p_j^0.
- Draw discount percentage d_2 from multinomial $[5\%\ 0.25; 10\%\ 0.5; 15\%\ 0.25]$ then set the discount price $p_j^2 = (100 - d_1 - d_2)\%$ of p_j^0.
- Draw discount percentage d_3 from multinomial $[5\%\ 0.15; 10\%\ 0.5; 15\%\ 0.35]$ then set the discount price $p_j^3 = (100 - d_1 - d_2 - d_3)\%$ of p_j^0

To set the discount threshold, we note that each vendor would expect roughly $|M|/|N|$ buyers (ignoring variation in preferences, product quality, etc.) So we set thresholds according to the fraction of this expectation. First let $\varepsilon \sim \mathcal{N}(0, 0.02|M|/|N|)$.

- Set first threshold $\tau_j^1 = 0.75|M|/|N| + \varepsilon$.
- Set second threshold $\tau_j^2 = \tau_j^1 + 0.5|M|/|N| + \varepsilon$.
- Set third threshold $\tau_j^3 = \tau_j^2 + 0.75|M|/|N| + \varepsilon$.

User Preference Models

The basic intuition here is that different users care about different quality attributes, and that they may also care about the "brand" (vendor identity) independent of attributes, and that they may be more or less be price sensitive. So we might have some consumers that care mostly about price (care very little for quality); some who care mostly for specific attributes (possibly including brand), so are more price insensitive; and users who lie somewhere in between. Note: some consumers are clearly "discount" sensitive, and are inherently attracted to the discount level itself, but we do not model this. We will ignore price sensitivity, and instead focus on inherent preference.

EC'12, June 4–8, 2012, Valencia, Spain.

Suppose that there are k quality attributes. Each user i has an inherent utility for a vendor j on a $[0,1]$ scale generated by an additive utility function (leaving out i superscripts):

$$\left[\sum_{\ell=1}^{k} w_\ell a_{\ell j}\right] + w_b \cdot \text{BrandValue}_{ij}$$

where attribute/quality levels are normalized on a $[0,1]$ scale (best value set to one, worst value set to zero), weights $w_1, w_2, ...w_k$ plus brand weight w_b sum to one, and BrandValue_{ij} represents a component of a local value function over brands specific to buyer i.

We draw weights (including brand values) uniformly at random from $[0,1]$ and normalize their sum to one. Users who are more brand sensitive (with high w_b) are likely skewed toward specific brands, hence the local value function should reflect this. Generally speaking such users will have more brand loyalty to higher quality products. However, we will ignore this for now.

We now have a utility for each vendor j on a $[0,1]$ scale. We translate this to the willingness-to-pay as follows:

- Let j^* be most preferred product for buyer i: set willingness-to-pay for j^* to be $v_{ij^*} = p^0_{j^*} + \varepsilon_{j^*}$, where $\varepsilon_{j^*} \sim \mathcal{N}(0.5p^0_{j^*}/2, 0.1p^0_{j^*})$. The intuition here is that i is willing to pay roughly 50% more than j^*'s undiscounted base price, but with some (high) variance.
- Let j' be least preferred product for buyer i: set willingness-to-pay for $\tilde{j}$ to be $v_{ij'} = p^0_{j'} + \varepsilon_{j'}$, where $\varepsilon_{j'} \sim \mathcal{N}(0.15p^0_{j'}, 0.1p^0_{j'})$. Again the intuition is that i is willing to pay roughly 15% less than j''s undiscounted base price, but with some high variance.
- Then set the willingness-to-pay for all other products to be proportional to their utility in the range $[v_{ij'}, v_{ij^*}]$.

B. PROOFS

PROOF OF THEOREM 4.2. We can reduce the optimal allocation problem given threshold values to an instance of the minimum-cost maximum-flow problem. Construct a network consisting of a source node s and sink node t. Let nodes $a_1, \ldots, a_n$ represent buyers and add arcs (s, a_i) with capacity 1 and zero cost. Include nodes $r_1, \ldots, r_m$, and $r'_1, \ldots, r'_m$ and x. For all i, j, add arcs (a_i, r_j) and (a_i, r'_j), each with capacity 1 and cost $L - (v_{ij} - p_j)$, where L is the largest buyer valuation. Also add arcs (a_i, x) with capacity 1 and cost L. Flow from a_i to r_j represents an assignment of buyer i to vendor j s.t. i is one of the buyers needed to meet threshold τ_j; and flow from a_i to r'_j represents the assignment of i to j, but where i's demand is considered in excess of τ_j. Flow on arcs to x represents unassigned buyers.

For each node r_j, there must be at most τ_j buyers assigned to it. We enforce this by adding arcs (r_j, t) with capacity τ_j and zero cost. There must be, in total, $n - \tau_1 - \cdots - \tau_m$ buyers assigned to the "excess" nodes r'_j and x. We add a node e with arcs (r'_j, e) for $j \in M$, and arc (x, e), each with infinite capacity and zero cost, and arc (e, t) with capacity $n - \tau_1 - \cdots - \tau_m$ and zero cost.

A maximum flow of n can be achieved by routing the first $\tau_1 + \cdots + \tau_m$ buyers to the r_j's subject to capacity constraints, and remaining buyers to x. Recall the assignment of buyer i to vendor j corresponds to (integer) flows from a_i to either r_j or r'_j. In any max flow, there must be total flow of τ_j to r_j; otherwise total flow to the r_j's is less than $\tau_1 + \cdots + \tau_m$, hence total flow to t is less than n given the capacity of arc (e, t). This also means a max flow must reach full capacity at (e, t).

Because the cost of arcs (a_i, r_j), (a_i, r'_j), and (a_i, x) is proportional to the negative surplus of buyer i when assigned to vendor j at price $p_j(\tau_j)$, the SWM matching (fixing threshold requirements) corresponds to an integer optimal min-cost max-flow solution where flows on the aforementioned arcs correspond to matchings. Standard results for the min-cost max-flow problem state that there is always an optimal *integer* flow, which can be found in polynomial time using, e.g., a modification of the Ford-Fulkerson algorithm. □

PROOF OF THEOREM 4.3. The problem is in NP, since computing $SW(\mu)$ for a given matching μ is straightforward. For hardness, consider a reduction from the Knapsack problem where given integer capacity C, integer item weights $\{w_1, \ldots, w_T\}$ and values $\{v_1, \ldots, v_T\}$, and a number $y \geq 0$, we wish to determine whether there exists a subset of items $S \subseteq \{1, \ldots, T\}$ such that $\sum_{i \in S} v_i \geq y$ and $\sum_{i \in S} w_i \leq C$. In the reduction, construct:

- T vendors each with one threshold level, i.e. $D = 1$. They have thresholds $\tau_i^0 = 0$, $\tau_i^1 = w_i$ with corresponding prices $p_i^0 = v_i$, $p_i^1 = 0$.
- $n = C$ buyers $z_1, \ldots, z_n$ each of whom have zero valuation for all vendors, and
- T additional buyers $s_1, \ldots, s_T$ where $v_{s_i}(i) = v_i$ and $v_{s_i}(j) = 0$ for all $i, j \in \{1, \ldots, T\}$, $i \neq j$.
- Set $x = y$.

Suppose we have a "yes" instance of Knapsack, and that S is the set of feasible items such that its total value exceeds y. In the corresponding segmentation problem, we can allocate τ_i^2 of the z buyers, and an additional buyer s_i to each vendor $i \in S$. In such an allocation each vendor in S will set prices to zero. User s_i, $i \in S$, will obtain a payoff of v_i by being assigned to i. Thus the allocation has a payoff of $\sum_{i \in S} v_i$ which is at least x $(= y)$.

Suppose we have a "no" instance of Knapsack. Consider any allocation μ of buyers to vendors. For any vendor i in which the threshold τ_i^1 is not met, then the payoff of any buyer assigned to i cannot be positive. Consider the set S of vendors with at least τ_i^1 assigned buyers. Then the payoff at any vendor $i \in S$ is at most v_i because s_i has valuation v_i and all other buyers have zero valuation, while price of any vendor is zero. Thus the payoff of μ is at most $\sum_{i \in S} v_i$ which is strictly less than x, otherwise the corresponding item set S for Knapsack would be a feasible "yes" solution (since the number of buyers allocated to vendors in S satisfies $\sum_{i \in S} \tau_i^1 = \sum_{i \in S} w_i \leq n = C$ which is impossible by assumption). $\square$