MATHEMATICAL MODEL IN MARKETING: QUESTION 1

Jin Miao

December 11, 2017

1 Task 1

Denote the Story Set S = $\{S_1, S_2, S_3, ..., S_n\}$. Denote the candidate ranking $r_1 \underset{R}{\succ} r_2 \underset{R}{\succ} r_3 ... \underset{R}{\succ} r_n$ as R = $(r_1, r_2, r_3, ..., r_n)$ where $r_i \in S \quad \forall i = 1, 2, 3, .., n$ and $r_i \neq r_j \quad \forall i \neq j$. Denote the ranking by Algorithm t ($t = 1, 2, 3, .., m$) as $A_t = (A_{t1}, A_{t2}, A_{t3}, ..., A_{tn})$ where $A_{t1} \underset{A_t}{\succ} A_{t2} \underset{A_t}{\succ} A_{t3} ... \underset{A_t}{\succ} A_{tn}$ and $A_{ti} \in S \quad \forall i = 1, 2, 3, .., n$ and $A_{ti} \neq A_{tj} \quad \forall i \neq j$.

Define $O_i(x, y)$ as the ordering of the pair (x,y) based on Ranking i where

$$O_i(x, y) = \begin{cases} 1, & \text{if } x \underset{i}{\succ} y \\ 0, & \text{otherwise} \end{cases}$$

Using these notations, the Linear Ordering Problem can be formulated as

$$\max_{R=(r_1, r_2, r_3, ..., r_n)} \sum_{i=1}^{m} \sum_{(x,y) \subset R} O_R(x, y)$$

s.t.$\forall t \in [2, n] \quad 1 \leq O_i(x_1, x_2) + O_i(x_2, x_3) + ... + O_i(x_{t-1}, x_t) + O_i(x_t, x_1) \leq t - 1$

Given the notation for combinations ${}^nC_k = \frac{n!}{k!(n-k)!}$, for each algorithm with n unique stories, when $t = 2$,

$$\forall x \neq y, \quad O_i(x, y) + O_i(y, x) = 1$$

imposes ${}^nC_2$ constraints and when $t = 3$,

$$\forall x \neq y \neq z, \quad 1 \leq O_i(x, y) + O_i(y, z) + O(z, x) \leq 2$$

imposes ${}^nC_3$ constraints. Following this logic, the total number of constraints with respects to m and n is

$$\sum_{t=2}^{n} m({}^nC_t) = m({}^nC_2 + {}^nC_3 + ... + {}^nC_n) = m(2^n - n - 1)$$

2 Task 2

The term NP-hard refers to any problem that is at least as hard as any problem in NP. To date, no one has found a polynomial-time algorithm for any problem in NP. This implies that LOP cannot be solved with a polynomial-time algorithm.

3 Task 3

In order to obtain a reversal, we assume that the real preference ordering for the pair (S_1, S_2),is $S_1 \succ S_2$. Then the reversal is obtained when an algorithm ranks S_2 ahead of S_1.

$$\Pr(Reversal) = Pr(u_2 > u_1) = Pr(v_2 + \epsilon_2 > v_1 + \epsilon_1) = Pr(\epsilon_2 > v_1 + \epsilon_1 - v_2)$$

Given that each ϵ_k has an independent, extreme value distribution,

$$\Pr(Reversal) = Pr(\epsilon_2 > \epsilon_1 + v_1 - v_2) = \frac{e^{v_1 - v_2}}{e^{v_1 - v_2} + e^{v_2 - v_1}}$$

$$\Pr(Non - Reversal) = 1 - Pr(Reversal) = \frac{e^{v_2 - v_1}}{e^{v_1 - v_2} + e^{v_2 - v_1}}$$

Define Non-Reversal: $NR_{i,j}(x, y)$ as the ordering of the pair (x,y) based on Ranking i and Ranking j where

$$NR_{i,j}(x, y) = \begin{cases} 1, & \text{if } O_i(x, y) = O_j(x, y) = 0 \quad or \quad O_i(x, y) = O_j(x, y) = 1 \\ 0, & \text{otherwise} \end{cases}$$

We assume that each unique pair of stories from S is independent of each other and that rankings given by different algorithms are independent of each other, $\forall t$, $\forall i \neq j$, the probability of a non-reversal for the pair (S_i, S_j) is

$$\Pr\{NR_{R,A_t}(S_i, S_j)\}$$

$$= \frac{e^{v_j - v_i}}{e^{v_i - v_j} + e^{v_j - v_i}} \Pr\{O_R(S_i, S_j) = 1\} + \frac{e^{v_i - v_j}}{e^{v_i - v_j} + e^{v_j - v_i}} \Pr\{O_R(S_i, S_j) = 0\}$$

If we further assume that the candidate ranking is based on the deterministic part v_k such that $(r_1, r_2, ..., r_n)$ implies that $v_1 > v_2 > ... > v_n$, then $\forall t$, $\forall i \neq j$ and $v_i > v_j$, the probability of a non-reversal for the pair (S_i, S_j) is

$$\Pr\{NR_{R,A_t}(S_i, S_j)\} = \frac{e^{v_j - v_i}}{e^{v_i - v_j} + e^{v_j - v_i}}$$

On top of the probability of non-reversals, the expected number of non-reversals across all newsfeed algorithms is

$$\sum_{t=1}^{m} \sum_{(S_i, S_j) \subset S} Pr\{NR_{R,A_t}(S_i, S_j)\} = m \sum_{(S_i, S_j) \subset S} \frac{e^{v_j - v_i}}{e^{v_i - v_j} + e^{v_j - v_i}}$$

4 Task 4

Define Reversal: $R_{i,j}(x, y)$ as the ordering of the pair (x,y) based on Ranking i and Ranking j where

$$R_{i,j}(x, y) = \begin{cases} 1, & \text{if } O_i(x, y) \neq O_j(x, y) \\ 0, & \text{otherwise} \end{cases}$$

Given that reversal and non-reversal are mutually exclusive, $\forall(x, y) \subset S$,

$$R_{R,A_j}(x, y) + NR_{R,A_j}(x, y) = m$$

For the candidate ranking, R = $(r_1, r_2, ..., r_n)$, $\forall(x, y) \subset R$, whenever $O_i(x, y) = 1$, we have $NR_{R,A_j}(x, y) = 1$. Given that $x \underset{R}{\succ} y$, both $O_i(x, y) = 1$ and $NR_{R,A_i}(x, y) = 1$ imply that $x \underset{A_i}{\succ} y$. As a consequence, we have the same objective functions from maximizing the expected number of reversals and optimizing the discrete linear ordering problem.

The simplest example is $n = 2$. If among m algorithms, more than $\frac{m}{2}$ algorithms prefer S_1 to S_2, the optimal solution to the discrete linear ordering problem is S_1; and vice verse. From the perspective of non-reversals, when the candidate ranking agrees with the algorithm, we will have a non-reversal instead of a reversal. To maximize the number of non-reversals, we also need to find the candidate ranking that agrees with more than $\frac{m}{2}$ algorithms.

Admittedly, when $n > 2$, social preferences cannot be generated by pairwise majority viting. However, the intuition holds that $\forall i \forall(x, y) \in R$, the events leading to $O_i(x, y) = 1$ are exactly the same as those correpsonding to $NR_{R,A_i}(x, y) = 1$. The conclusion is stated as follows.

Assuming that $(r_1^*, r_2^*, r_3^*, ..., r_n^*)$ is the solution to the problem of maximizing the expected number of non-reversals, $(r_1^*, r_2^*, r_3^*, ..., r_n^*)$ will also maximize

$$\sum_{i=1}^{m} \sum_{(x,y) \subset R} O_i(x, y)$$

Assuming that $(r_1^*, r_2^*, r_3^*, ..., r_n^*)$ is the optimal solution to the discrete linear ordering problem, R = $(r_1^*, r_2^*, r_3^*, ..., r_n^*)$ will also maximize

$$\sum_{t=1}^{m} \sum_{(S_i,S_j) \subset S} Pr\{NR_{R,A_t}(S_i, S_j)\}$$

5 Task 5

Assuming that each unique pair of stories from S are independent of each other, given the fact that the probability of a non-reversal for the pair (S_i, S_j)

$$Pr\{NR_{R,A_t}(S_i, S_j)\}$$

the probability of non-reversal across all pairs between R and A_j is

$$\mathrm{P}r\{NP_{R,A_t}\} = \prod_{(S_i,S_j) \subset S} Pr\{NR_{R,A_t}(S_i, S_j)\}$$

Assuming that rankings given by different algorithms are independent of each other, the likelihood of non-reversal across all pairs between R and all algorithms is

$$\mathrm{L} = \prod_{t=1}^{m} Pr\{NP_{R,A_t}\} = \prod_{t=1}^{m} \prod_{(S_i,S_j) \subset S} Pr\{NR_{R,A_t}(S_i, S_j)\}$$

Maximizing lnL with respects to $(r_1, r_2, ..., r_n)$ is equivalent to Maximizing log-likelihood function with respects to $(r_1, r_2, ..., r_n)$.

$$\max_{R=(r_1,r_2,r_3,...,r_n)} lnL = \max_{R=(r_1,r_2,r_3,...,r_n)} \sum_{t=1}^{m} \ln\{Pr[NP_{R,A_t}]\}$$

$$= \max_{R=(r_1,r_2,r_3,...,r_n)} \sum_{t=1}^{m} \sum_{(S_i,S_j)\subset S} \ln\{Pr[NR_{R,A_t}(S_i, S_j)]\}$$

This formulation is equivalent to multinomial logit model with Maximum Likelihood Estimation method. McFadden(1974) shows that the log-likelihood function for MNL is globally concave for linear-in-parameter utility. Thus, maximizing the likelihood function is a convex optimization problem.

MATHEMATICAL MODEL IN MARKETING: QUESTION 2

Jin Miao

December 11, 2017

1 Task 1

1.1 Devise a polynomial-time algorithm that produces a stable assignment of the n houses to the n people.

Recursively implement the following two steps until all the n people is excluded.

Step 1: Denote $\boldsymbol{N}$ as the group of n people. Let every member declare their most desirable house. Consider the set $\boldsymbol{M} = \{$ j: Person j declares that House j is exactly his/her favourate house$\}$. Assume $m = |\boldsymbol{M}|$, then $m \leq 0$ and $m \geq n$. $\forall j \in \boldsymbol{M}$, the assignment of House j to Person j is coalition proof because Person j has already achieved the highest utility. Simultaneously, Person j is not willing to exchange House j with any other house. Update $\boldsymbol{N}$ to be $\boldsymbol{N} - \boldsymbol{M}$.

Step 2: Starting from any member N_1 of $\boldsymbol{N}$, N_1 declare the best house for him/her, which currently belongs to N_2. Create the set $\boldsymbol{Pool} = \{N_1\}$. Then N_2 declare the best house for him/her, which currently belongs to N_3. If N_3 is exactly N_1, then we can find a coalition $\boldsymbol{\{N_1, N_2\}}$ to exchange houses and then $\boldsymbol{N}$ should be updated as $\boldsymbol{N} - \boldsymbol{\{N_1, N_2\}}$. Every member in the new $\boldsymbol{N}$ will resay their favorite house. If $N_3 \neq N_1$, we update $\boldsymbol{Pool} = \{N_1, N_2\}$. Then N_3 declare the best house for him/her, which currently belongs to N_4. If $N_4 \in \boldsymbol{Pool}$, then we can find a coalition so that we can update $\boldsymbol{N}$ by excluding this coalition from $\boldsymbol{N}$. If $N_4 \notin \boldsymbol{Pool}$, we update $\boldsymbol{Pool} = \{N_1, N_2, N_3\}$. Proceed this procedure until we find a coalition.

1.2 Prove that the algorithm terminates and obtains a stable assignment.

This algorithm will always find a coalition regardless of the size of $\boldsymbol{N}$.

$\forall k \in \boldsymbol{N}$, when $N_k - 1$ declares N_k, N_k's favorite house is either in $\boldsymbol{Pool}$ or not. If N_k's favorite house is in $\boldsymbol{Pool}$, then this algorithm finds a coalition. If all of the first n-1 persons cannot find their favorite house in $\boldsymbol{Pool}$, N_n's favorite house must be in $\boldsymbol{Pool}$ given the fact that Person n is not perfectly satisfied with House n.

On top of this, the new group $\boldsymbol{N}$ updates their preferences every time we find a coalition and exclude every member in the coalition from $\boldsymbol{N}$. Thus, this algorithm terminates until no element remains in $\boldsymbol{N}$. The time complexity for this algorithm is polynomial because at most n steps are needed for each round.

This algorithm can obtain a stable assignment because every time we find a coalition and reassign houses, every one in this coalition will live in the house that he likes most. As a consequence, no one has any incentive to move after being reassigned to their favorite house (as declared).

2 Task 2

It is a dominant strategy for each person to truthfully reveal his or preferences.

Whenever we find a coalition, every one in this coalition will move to the house that he declares as his favorite one regardless of others' declarations. If he/she does not truthfully reveal his/her preferences, he/she will not be better off under the following two circumstances.

Case 1: he/she is included in the coalition in this round if he truthfully declares his preferences. In this case, he/she will get the house that is declared. Thus, the outcome from any untruthful declaration must be inferior to truthful declarations.

Case 2: he/she is not included in the coalition during this round. In this case, truthful declaration is still not inferior to untruthful declaration. For example, assume $\boldsymbol{N} = \{A, B, C, D\}$. Assume that Person A's preference ordering is $A \preceq D \preceq C \preceq B$. However, House B is not achievable for Person A because Person B likes House D most and Person D likes House B most. It is possible that Person A can achieve his second best by declaring that House C is his favorite. However, after excluding $\{B, D\}$ from $\boldsymbol{N}$, House C naturally becomes his favorite house when Person A updates his preferences in the next round.

To conclude, it is a dominant strategy to for each person to truthfully reveal his/her preferences.

MATHEMATICAL MODEL IN MARKETING: QUESTION 3

Jin Miao

December 11, 2017

1 Task 1

Define A as Adjacency Matrix where A_{ij} denotes the pair of Follower i and Leader j. $\mathbb{1}$ is One Vector so that each element of $\mathbb{1}$ is 1. Follower Vector represents the number of accounts that each user follows, which can be expressed as

$$F = A\mathbb{1}. \tag{1}$$

Similarly, Leader Vector L represents the number of accounts that each user leads , which can be expressed as

$$L = A^T\mathbb{1}. \tag{2}$$

The average number of followers is

$$F_1 = \frac{\mathbb{1}^T A\mathbb{1}}{\mathbb{1}^T\mathbb{1}} \tag{3}$$

The average number of leaders is

$$L_1 = \frac{\mathbb{1}^T A^T\mathbb{1}}{\mathbb{1}^T\mathbb{1}} \tag{4}$$

Given that both $\mathbb{1}^T A\mathbb{1}$ and $\mathbb{1}^T A^T\mathbb{1}$ are scalars, matrix transpose implies that

$$\{\mathbb{1}^T A\mathbb{1}\}^T = \mathbb{1}^T A^T\mathbb{1} \tag{5}$$

Thus,

$$F_1 = L_1 \tag{6}$$

Intuitively, there is one follower and one leader within each pair. Summing up all the pairs, the total number of followers equals to the total number of leaders. Dividing by the number of users, the average number of leaders equals to the average number of followers.

2 Task 2

Mathematical Explanations for the phenomenon that the average of followers is smaller than the average of followers of leaders.

The average number of followers of leaders is

$$F_2 = \frac{\mathbb{1}^T A^T A \mathbb{1}}{\mathbb{1}^T A \mathbb{1}} \tag{7}$$

From the Cauchy–Schwarz inequality,

$$\frac{\mathbb{1}^T A \mathbb{1}}{\mathbb{1}^T \mathbb{1}} \leq \frac{\mathbb{1}^T A^T A \mathbb{1}}{\mathbb{1}^T A \mathbb{1}} \tag{8}$$

Moreover, the two sides are equal if and only if $A\mathbb{1}$ and $\mathbb{1}$ are linearly dependent. This is consistent with the observation that if Adjacency Matrix A is fully connected (every user leads and follows each other without any exception), then $F_1 = F_2$. Otherwise, the average of followers is smaller than the average of followers of leaders.

Intuitively, we can recall the process of calculating the average of followers of leaders – first find the leaders for each user and then find the followers of each leader. On top of this, consider the fact that some celebrities, like Katy Perry, Justin Beiber and Barack Obama, have huge number of followers each on Twitter. Compared with ordinary users, these celebrities has two major differences: (1) From the perspective of being leaders, these celebrities have more followers; (2) From the perspective of followers, these celebrities are followed by more users. Thus, when calculating the average of followers of leaders, these two effects were added. This is consistent with the fact that, given $F = A\mathbb{1}$,

$$\mathbb{1}^T A^T A \mathbb{1} = \sum_{i=1} F_i^2 \tag{9}$$

$$F_2 = \frac{\sum_{i=1} F_i^2}{\sum_{i=1} F_i} \tag{10}$$

Take the example of Barack Obama, F_{Obama} is much larger than average, F_1, which makes the value of F_2 outnumber F_1.

3 Task 3

Mathematical Explanations for the phenomenon that the average number of leaders of followers is much higher than the average number of leaders.

The average number of leaders of followers is

$$L_2 = \frac{\mathbb{1}^T A A^T \mathbb{1}}{\mathbb{1}^T A^T \mathbb{1}} \tag{11}$$

From the Cauchy–Schwarz inequality,

$$\frac{\mathbb{1}^T A^T \mathbb{1}}{\mathbb{1}^T \mathbb{1}} \leq \frac{\mathbb{1}^T A A^T \mathbb{1}}{\mathbb{1}^T A^T \mathbb{1}} \tag{12}$$

Moreover, the two sides are equal if and only if $A^T\mathbb{1}$ and $\mathbb{1}$ are linearly dependent. This is consistent with the observation that if Adjacency Matrix A is fully connected (every user leads and follows each other without any exception), then $L_1 = L_2$. Otherwise, the average of followers is smaller than the average of followers of leaders.

Intuitively, we can think of the process of calculating the average of leaders of followers – first find the followers for each user and then find the leaders of each follower. On top of this, consider the fact that some people, like MarQuis Trill and Megamix Champion, follow millions of people on Twitter. Compared with ordinary users, these people has two major differences: (1) From the perspective of being followers, these people have more leaders; (2) From the perspective of leaders, these people are led by more users. Thus, when calculating the average of leaders of followers, these two effects were added. This is consistent with the fact that, given $L = A^T \mathbb{1}$,

$$\mathbb{1}^T A^T A \mathbb{1} = \sum_{i=1} L_i^2 \tag{13}$$

$$L_2 = \frac{\sum\limits_{i=1} L_i^2}{\sum\limits_{i=1} L_i} \tag{14}$$

Take the example of MarQuis Trill, L_{Trill} is much larger than average, L_1, which makes the value of F_2 outnumber F_1.

4 Task 4

Given the fact that $F_1 = L_1$ and $F_2 < L_2$, we have

$$\mathbb{1}^T A^T A \mathbb{1} > \mathbb{1} A A^T \mathbb{1} \tag{15}$$

The inequality above implied that

$$\sum_{i=1} L_i^2 > \sum_{i=1} F_i^2 \tag{16}$$

The inequality can be rewritten as

$$\sum_{i=1} \{L_i^2 - L_1\}^2 > \sum_{i=1} \{F_i - F_1\}^2 \tag{17}$$

The large discrepancy in the two averages ($F_2 << L_2$) arises from the fact that the variance of the number of leaders is larger than the variance of the number of followers. This is consistent with the observation that superstars and celebrities are at the center of viral marketing. It is common that people will follow superstars and celebrities while it is hard for ordinary people to get much attention, which accounts for the fact that Leader Vector (L) has larger variance compared with Follower Vector (F). Twitter created networks for its users and celebrities/superstars are widely connected within the Twitter network. This implies the potential of Tweeting as a Marketing Tool especially by influential tweets and retweets from celebrities and superstars.