Total Surplus Maximization - Online Peer to Peer Lending

Vinay Kumar Bysani
Rutgers University, New Brunswick
vinay.bysani@rutgers.edu

Nikita Balakrishnan
Rutgers University, New Brunswick
nikita.b@rutgers.edu

ABSTRACT

A fundamental component of e-commerce platform is ”Online Service Allocation (OSA)”, which matches service between producers and consumers. Existing systems focus on maximizing either of consumer utility or producer profit. We theorize, implement and test ”Total Surplus Maximization” [1] which embodies the idea of maximizing total surplus (sum of producer and consumer surplus) in Online Service Allocation. We intend to implement and test the hypothesis for online peer-to-peer lending data set and project results accordingly.

KEYWORDS

Online Service Allocation, Total Surplus Maximization, Law of diminishing marginal utility.

1 INTRODUCTION

Lot of businesses are going online in the past decade, from e-commerce, advertising, peer-to-peer lending, freelancing etc. One of the primary game changers in the online are sites like amazon who match buyers with sellers. In case of freelancing, sites like Upwork match workers with employers. Prosper which is a peer-to-peer matching website matches lenders with borrowers.

Since, consumers have a free will to choose the product/service they want, service allocation is implemented using recommender systems. Existing service allocation systems try to maximize profit of either producers or consumers, while the other is sacrificed. One of the widely used approach is Collaborative Filtering [2], which is favorable to consumers. This paper makes use of Total Surplus Maximization (TSM) framework which takes into consideration both producers and consumers surplus. Experimental Results obtained will prove that it will be beneficial to both parties. We run our experiments on Prosper loan data set [3], and the results obtained prove the proposed hypothesis. We programmed it in python programming language.

2 RELATED WORK

Social Surplus or Economic Surplus is a topic that has been existing from long back, but it is Economist Paul A. Baran who made the term popular [4]. With the advent of internet enabled services, the process of matchmaking between producers and consumers have been studied in detail [5, 6]. A lot of the research focuses on producer surplus or consumer surplus. One of the popular technique is Collaborative Filtering [6]. This paper addresses the issue that has been ignored, i.e., way to maximize both producer and consumer surplus.

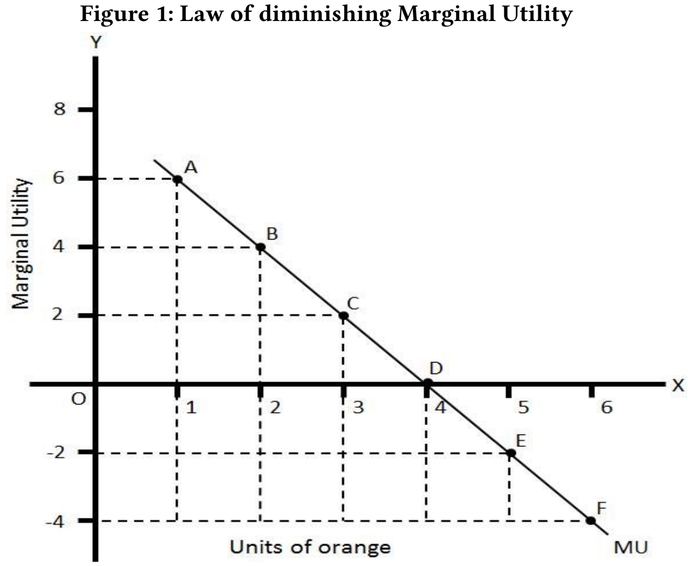
3 PRELIMINARIES

This section introduces couple of definitions which are very vital going forward with the theoretical aspect of framework.

3.1 Utility

Utility is a term used by economists to describe the measurement of satisfaction that a consumer gets from a good/service. It could be how much a person enjoys a food or how useful buying a product could be. It is a function of consumed quantity. Utility can be viewed as the maximum amount consumer is willing to pay for the good. Law of diminishing marginal utility which is proposed by economist Alfred Marshall, states that marginal utility decreases as consumption increases. In the following figure 1, a hungry person gets lot of utility when he is offered an orange, but when his marginal utility comes down when he is offered another orange and so on.

Figure 1: Law of diminishing Marginal Utility



3.2 Surplus

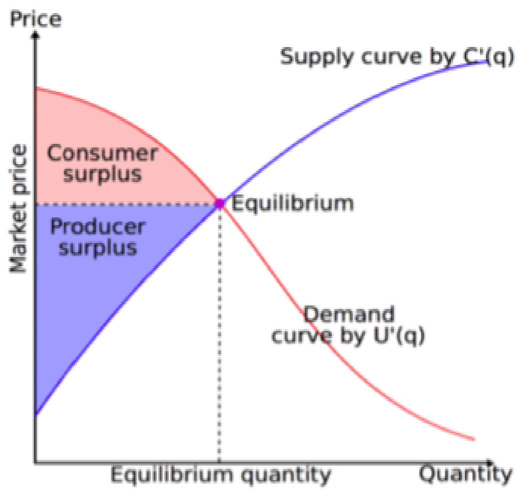
Surplus is the benefit obtained by obtaining a product/service. Producer Surplus (PS) is the profit a producer makes when he sells a good. Consumer Surplus (CS) is the utility a consumer obtains

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

through buying a good. In the following figure 2, the consumer surplus and producer surplus are shown. Demand curve is represented by marginal utility function, supply curve by marginal cost. Our aim is to find the equilibrium quantity and do the Online Service Allocation accordingly.

Figure 2: Marginal Utility and Marginal Cost curves



4 PROBLEM FORMALIZATION

We consider the problem of online service providing i.e distributing goods among given users so that the total surplus is maximized during this process.
Consider there are,
i) m consumers (lenders)

$$u_1, u_2, u_3, ..., u_m$$

ii) n goods (loans)

$$g_1, g_2, g_3, ..., g_n$$

iii) r producers (borrowers)

$$p_1, p_2, p_3, ..., p_r$$

Consider

$$1 \le i \le m$$
$$1 \le j \le n$$
$$1 \le k \le r$$

DEFINITION 1. Let $M = [M_1, M_2, M_3, ..., M_n]$ be the service quantity vector, where $M_j \ge 0$ is the total amount that good g_j can be provided by its producer in the system (the actual loan amount requested by the borrower) .

DEFINITION 2. The Online Service Allocation problem is to find an allocation matrix Q = $[Q_{ij}]_{mxn}$ where $Q_{ij} \ge 0$ is the quantity that consumer u_i is provided with good g_j i.e the lender u_i lends g_j.

5 TOTAL SURPLUS MAXIMIZATION

The best approach for Online Service Allocation is to try to find the best allocation matrix Q so as to maximize the Total Surplus under the constraints of service quantity.

$$\underset{Q}{\text{maximize}} \sum_i \sum_j (U_{ij}(Q_{ij}) - C_j(Q_{ij})) \tag{1}$$
$$s.t \;\; 1^T Q \le M, \;\; Q_{ij} \in S$$

where S is the set of possible legal values - for peer to peer lending, S = R_+, positive real numbers.
However, an important issue to consider is that the consumers may not always rationally choose the best quantities of goods (Hypothesis of Rational Man). Hence, we may lose the model generality and the learning power for data fitting if we restrict Q to real-valued matrices. Taking these issues into consideration, we relax the elements Q_{ij} in allocation matrix Q to random variables with a probability distribution. Thus, when $Q_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, consumer u_i chooses good g_j on quantity μ_{ij} with the highest probability but he/she may also make consumption on other quantities, but with a lesser probability.
Thus the previous equation becomes,

$$\underset{\theta(Q)}{\text{maximize}} \sum_i \sum_j \int (U_{ij}(Q_{ij}) - C_j(Q_{ij}))\, p(Q_{ij})\, dQ_{ij} \tag{2}$$
$$s.t \;\; 1^T \int Qp(Q)dQ \le M, \;\; Q_{ij} \in S$$

where $p(Q_{ij})$ is the probability density function of each quantity Q_{ij}, p(Q) = $[p(Q_{ij})]_{mxn}$ and the integral on Q is for each element. $\theta(Q)$ is the parameter set of all the Q_{ij}'s in Q.

6 THE PROPOSED MODEL

As mentioned in the previous section, in P2P lending services, the borrowers are loan request producers, where the loan requests can be viewed as financial products and the lenders are consumers of these loan requests that purchase these financial products by distributing their investments on these requests to make profits. Thus the OSA problem is that, how the lenders (i.e., consumers) should distribute their assets among the loan requests, so that both parties involved are benefited and the total surplus in the system is maximized.
The consumer surplus is calculated as

$$CS_{ij}(Q_{ij}) = (r_j - \hat{r})Q_{ij} \tag{3}$$

The producer surplus is calculated as

$$PS_{ij}(Q_{ij}) = (r_j^{max} - r_j)Q_{ij} \tag{4}$$

where r_j^{max} is the maximum rate of interest the borrower will accept for the loan amount, r_j is the highest interest rate among the winners and this is set as the final interest rate and $\hat{r}$ is the risk-free interest rate which is 0.01 in our case.
Now, the total surplus is calculated as the sum of the consumer and producer surplus

$$TS_{ij}(Q_{ij}) = CS_{ij}(Q_{ij}) + PS_{ij}(Q_{ij}) = (r_j^{max} - \hat{r})Q_{ij} \tag{5}$$

Now, we apply normal distribution on Q_{ij} as mentioned in the previous section to obtain the below equation,

$$\underset{U,\Sigma}{\text{maximize}} \sum_i \sum_j \int \frac{(r_j^{max} - \hat{r})Q_{ij}}{\sqrt{2\pi}\sigma_{ij}} exp\left(-\frac{(Q_{ij} - \mu_{ij})^2}{2\sigma_{ij}^2}\right) dQ_{ij} \tag{6}$$

$$s.t \quad 1^T \int \frac{Q}{\sqrt{2\pi}\Sigma} exp\left(-\frac{(Q - U)^2}{2\Sigma^2}\right) dQ \quad \leq M, \; Q_{ij} \in R_+$$

where $U = [\mu_{ij}]_{mxn}$ and $\Sigma = [\sigma_{ij}]_{mxn}$

Equation 6 is further simplified to the below equation:

$$\underset{U,\Sigma}{\text{maximize}} \sum_i \sum_j \mu_{ij}(r_j^{max} - \hat{r}) \tag{7}$$

$$s.t \quad 1^T U \leq M, \; \mu_{ij} \in R_+$$

This equation (7) can be solved to find the optimum values with **linear programming**. The linear system is created by taking into consideration, the borrower constraints i.e the amount of money a borrower needs (Loan amount) and the lender constraints i.e the maximum amount a lender i can lend totally (All the amount the lender has invested in different loans). Thus the total number of unknowns will be mxn i.e each element in the Q matrix. And we have to solve this linear system for these unknowns.

The expected value under Gaussian distribution is finally taken as the allocation matrix.

$$\overline{Q}_{ij} = \mu_{ij}$$

Interestingly, it's noticed that this method lets us allocate the investment/amount in a greedy manner based on the per capita surplus - $(r_j^{max} - \hat{r})$, which is indeed the intuitional rule for any investment.

7 EXPERIMENTS

We implement the hypothesis on Proper loans data set. Prosper is America's first peer-to-peer lending marketplace where individuals can either invest in personal loans or request to borrow money. Investors can consider borrower's credit scores, ratings, and histories and the category of the loan. Prosper handles the servicing of the loan and collects and distributes borrower payments and interest back to the loan investors. Here, borrower will specify two factors i.e., amount of money and max interest rate he is willing to accept for the loan. Lenders don't have to provide all the requested amount, they can only partly lend the amount requested by borrower.

7.1 Preprocessing

We filter the data based on the following three filters

(1) Date Filter (Nov 2005 to May 2009)
(2) Status Filter (Current, Late, Payoff in progress, Paid)
(3) Sort based on Time stamp

From 2009 third quarter, Prosper opened up an automatic bidding system for lenders. We intend to understand the behavior of consumers and producers, hence we filter the data as per the date. We consider listings of status **Current, Late, Payoff in Progress, or Paid**. We do not consider those listings whose status is Defaulted, Cancelled or Charge-of.

Here is the summary of data set before preprocessing and after preprocessing.

Figure 3: Data set before and after preprocessing

	Pre-Processing	Post-Processing
Transactions	\$3.3m	\$1.2m
Total Value	\$258m	\$108m
Borrowers	59k	47k
Lenders	35k	16k

7.2 Results

The following figure 4, shows TSM compared to TS for varying size of data set. As data set increases, constraints increase. Hence the amount can be distributed between borrowers maximizing total surplus. TSM is more improved as we have lot more data, since it will increase interactions, and hence optimization is possible given more relations.

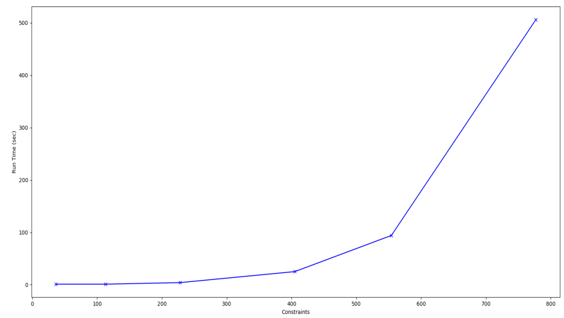
In figure 4, the last two columns compare the Total Surplus (TS) - i.e the surplus resulting from the actual allocation of amount in the given Prosper dataset and the Total Surplus Maximization (TSM) which is the allocation based on our method to maximize both producer and consumer surplus (Total Surplus). Clearly, TSM always has a better result when compared to TS in every case. This guarantees that our approach works far better than the existing p2p platforms' allocation approach.

Figure 4: Obtained Results

#Dataset	#Borrowers	#Lenders	#Constraints	#Total Value	#TS	#TSM
100	16	20	36	67,821	5772	6108
500	52	61	113	303,357	24,985	25958
1000	143	85	228	468,237	42,032	43227
2000	286	119	405	648,172	63,727	65,495
3000	402	152	544	799,142	80,443	82,490
5000	575	202	777	1,070,743	109,738	112,040

Figure 5, shows how run time is affecting given number of constraints. As constraints increase, the problem solving is computational heavy. We made use of Python Scipy module for solving linear programming equation. The below graph represents run time in python.

Figure 5: Constraints vs Run Time



8 CONCLUSIONS AND FUTURE WORK

Majority of research work in e-commerce is focused on improving recommender systems, increasing click through rate etc. However this paper stresses on increasing utility/profit for both consumers and producers by maximizing total surplus. This is a very interesting and important metric to consider to improve internet economy. This paper only focuses on peer-to-peer lending framework and data sets, however it can be extended for various different sites including e-commerce, freelancing etc.

ACKNOWLEDGEMENT

We would like thank professor Yongfeng Zhang for the course and on whose paper we implemented. We also would like to thank Teaching Assistant Yunqi Li who assisted us.

REFERENCES

[1] Economic recommendation with surplus maximization
http://yongfeng.me/attach/er-www16.pdf

[2] A Survey of Collaborative Filtering Techniques
https://www.hindawi.com/journals/aai/2009/421425/

[3] Prosper Loans Network Dataset
http://mlg.ucd.ie/datasets/prosper.html

[4] Economic Progress and Economic Surplus - Paul A Baran
https://www.jstor.org/stable/40400214?seq=1

[5] Introduction to Recommender Systems Handbook - F. Ricci, L. Rokach, and B. Shapira.
http://www.inf.unibz.it/ ricci/papers/intro-rec-sys-handbook.pdf

[6] Collaborative Filtering Recommender Systems - M. D. Ekstrand, J. T. Riedl, and J. A. Konstan.
https://www.nowpublishers.com/article/Details/HCI-009