Economic Recommendation with Surplus Maximization

Yongfeng Zhang[†‡],Qi Zhao[†*],Yi Zhang[†],Daniel Friedman[♯],Min Zhang[‡],Yiqun Liu[‡],Shaoping Ma[‡]
[†]School of Engineering, [♯]Economics Department, University of California, Santa Cruz, CA 95060, USA
[‡]Department of Computer Science & Technology, Tsinghua University, Beijing, 100084, China
zhangyf07@gmail.com,{qzhao2,yiz,dan}@ucsc.edu,{z-m,yiqunliu,msp}@tsinghua.edu.cn

ABSTRACT

The World Wide Web makes it easy for users to access various services online, such as online shopping in e-commerce, online financing in P2P lending networks, and even working online on freelancing platforms.

A fundamentally important function of web applications is *Online Service Allocation* (OSA), which means matching service *consumers* with 'appropriate' *goods* (i.e., products/loans/jobs) from service *producers*. Since consumers often have the freedom to choose goods, OSA function is normally provided in the form of personalized recommendation or search. Existing search and recommendation system metrics either measure the benefit of individual consumers or the profit of individual producers. However, a sustainable web platform needs to be beneficial for both consumers and producers, otherwise one party might stop using it.

In this paper, we aim to promote the *web intelligence for social good*. We show how to adapt economists' traditional idea of maximizing total surplus (the sum of consumer net benefit and producer profit) to the heterogeneous world of online service allocation (OSA). Modifications of traditional personalized recommendation algorithms enable us to apply total surplus maximization (TSM) to three very different types of real-world tasks (e-commerce, P2P lending, and freelancing). Our experimental results on three tasks suggest that TSM compares very favorably to currently popular approaches, to the benefit of both producers and consumers.

Categories and Subject Descriptors

M.5.4 [**Applied Computing**]: Law, Social and Behavioral Sciences- *Economics*

Keywords

Total Surplus Maximization; Online Service Allocation; Computational Economics; Recommendation Systems; Web-based Services

1. INTRODUCTION

Online applications and services have grown tremendously in recent years. Consumers find producers on E-commerce websites like Amazon or via social networks like Facebook, borrowers and lenders find each other via P2P lending services like Prosper, and freelancing websites like Amazon Mechanical Turk and Upwork match short term workers with employers. Such online service allocation seems destined to grow rapidly in the years ahead.

Because consumers are granted by law to choose freely among available services, enforced service allocation is usually not practical online. As a result, service allocation is usually implicitly realized through search [4] and recommendation [30] systems. Search engines such as Google or Amazon product search attempt to tackle cases where the system (at least partly) explicitly know the demands or intents of the consumers, while recommendation systems such as products or social network recommendations, try to infer consumer needs without requiring explicit user queries.

By its nature, service allocation is a two-sided matching activity, e.g., of consumers with producers. Economists since Adam Smith (1776) have taken a more balanced view of service allocation. The key insight is total surplus – the sum of producers' profit and consumers' net benefit. An efficient market – one that maximizes total surplus – is in the best interest of society, and potentially enables *both* sides to be better off than they would be in an inefficient market.

However, existing recommendation systems in online service matching platforms are designed to benefit only one side, while the benefits of the other side are seldom explicitly considered or even sacrificed due to the conflict of interests between consumers and producers governed by the economic theory of surplus [14]. For example, the widely adopted Collaborative Filtering (CF) [36] approach for recommendation is based on the preferences of consumers, while the benefits of producers play little role. Some online P2P lending systems focus on improving the revenue of lenders (i.e., consumers of the financial products), while ignoring the surplus of borrowers. These are problematic because the side who doesn't gain much benefit will stop using the system.

The purpose of this paper is to illustrate how to operationalize the economists' insight in online service allocation into personalized recommendation systems to solve the problem. We propose a Total Surplus Maximization (TSM) framework to integrate both consumer surplus and producer surplus into recommendation systems. By TSM, the system creates a bigger pie (total surplus) for consumers and producers to divide. There is a large gap between the traditional application of the economists' insight (a competitive market for a uniform commodity, with lots of small producers and consumers) and online allocation/recommendation of very personalized and heterogeneous services. To fill the gap, we develop surplus-oriented metrics for personalized recommendations in several online markets.

We will offer evidence that they can improve performance to the benefit of both sides. Indeed, our experimental results

*The first two authors contributed equally to this work.
*This work was conducted when the first author was visiting at University of California, Santa Cruz.

WWW 2016, April 11–15, 2016, Montréal, Québec, Canada.
ACM 978-1-4503-4143-1/16/04.
http://dx.doi.org/10.1145/2872427.2882973.

and analyses on three real-world datasets (e-commerce, P2P lending, online freelancing) conclude that total surplus maximization based recommendation approach performs competitively better than existing widely applied standard recommendation techniques. That is, society as a whole is better off while better satisfying the needs of both the consumers and the producers.

The rest of the paper is organized as follows: in Section 2 we introduce some basic definitions and concepts that form the basis of this work. In Section 3 we propose our Total Surplus Maximization (TSM) framework for Online Service Allocation (OSA), then tailored it to three typical applications in Section 4. We present experimental results on three tasks in Section 5, followed with some related work in Section 6 and conclusions in Section 7.

2. BASIC DEFINITIONS AND CONCEPTS

In this section, we introduce some of the key concepts and definitions in economics and recommendation systems. These will form the theoretical basis for our framework to be described later.

2.1 Utility

In economics, *utility* is a measure of one's preference over some set of goods or services. It is an important concept that serves as the underpinning of the rational choice theory [9]. A consumer's total utility for a given set of goods is the consumer's satisfaction experienced from consuming these goods.

Utility $U(q)$ is usually a function of the consumed *quantity* q. $U(q)$ is inherently governed by the *Law Of Diminishing Marginal Utility* [31], which states that as a person increases the consumption of a product, there is a decline in the marginal utility that the person derives from consuming each additional unit of the product, i.e., $U''(q) < 0$, while the marginal utility keeps positive, i.e., $U'(q) > 0$.

An example frequently used by economists is that, one who is extremely hungry may obtain a huge amount of satisfaction when consuming the first service of bread, but less satisfaction can be obtained when he/she continues to consume when feeling full.

Economists have introduced various functional forms for utility. Without of lose of generality, we introduce two simple and frequently adopted utility functions for easy understanding and exposition. They are the Exponential Utility:

$$U(q) = \frac{1 - \exp(-aq)}{a} \tag{1}$$

and the King-Plosser-Rebelo (KPR) utility:

$$U(q) = a \ln(1 + q) \tag{2}$$

where a in Exponential Utility represents a user's degree of risk aversion and a in KPR utility measures the risk aversion as well as the overall lift of the curve. Both of them are characterized by diminishing marginal utility and yield 0 utility when the quantity is zero, i.e., $U(0) = 0$.

Furthermore, utility can be viewed as a generalized price, and in the simplest sense, economists consider to reveal utility by eliciting consumer's Willingness To Pay (WTP) [14, 39], namely, the maximum amount of money to pay for acquiring another unit of good (i.e., $U'(q)$). This enables us to align the utility in the same scale as price so as to calculate the surplus of consumers and producers.

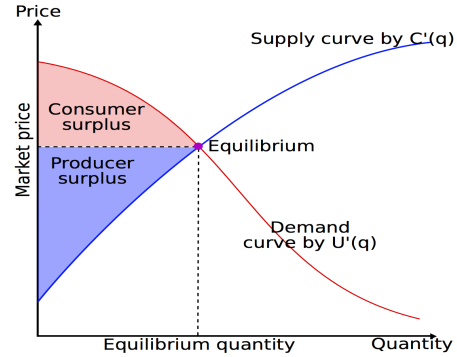

Figure 1: An intuitional explanation of surplus derived from marginal utility and marginal cost.

2.2 Surplus

Surplus is the net benefit (in dollar terms) associated with buying or selling a good or service[14, 1]. Intuitionally, Consumer Surplus (CS) is the amount of utility that one experiences beyond the amount that he/she pays (i.e., the price), and similarly, Producer Surplus (PS) is the amount that the producer earns beyond the cost.

The consumer and producer surpluses are indicated in Figure 1, where the demand curve features the marginal utility function $U'(q)$, which decreases according to the exposition in the previous section; meanwhile, the supply curve features the marginal cost $C'(q)$, which increases according to the Law of Diminishing Marginal Returns [31].

In competitive equilibrium, the price is determined by the intersection of the two curves. Given the quantity of consumption $q = q_c$, the Consumer Surplus CS is obtained by integrating the marginal utility that exceeds the price at each unit of consumption until q_c, which is denoted as follows:

$$CS = \int_0^{q_c} \big(U'(q) - P\big)\, \mathrm{d}q = U(q_c) - Pq_c \tag{3}$$

and the Producer Surplus PS is similarly determined by:

$$PS = \int_0^{q_c} \big(P - C'(q)\big)\, \mathrm{d}q = Pq_c - C(q_c) \tag{4}$$

Further more, the Total Surplus TS is defined as the sum surplus gained by the consumer and producer, which is:

$$TS = CS + PS = U(q_c) - C(q_c) \tag{5}$$

Finally, the total surplus of an economic system is the sum of the surpluses for all parties involved in all the transactions of the system.

We see in Eq.(5) that the price component offsets and does not affect the eventual total surplus in a single transaction. This is an important conclusion in Economics, which means that the surplus of a consumer (or producer) increases only at the expense of the other's, thus the only way for better social good is to increase the surplus of both parties together to obtain an increased total surplus.

2.3 Collaborative Filtering

Many online services involve the evaluation of producers by consumers (or vice versa) through ratings. For example, on E-commerce websites such as Amazon, users are allowed to rate the purchases with a numerical star rating of 1∼5; and on online freelancing websites such as Upwork or

Freelancer, such ratings are provided bidirectionally between employers and freelancers.

The numerical rating r_{ij} reveals the satisfaction that consumer u_i obtains from good g_j provided by producer p_k. As each consumer may only consume a small portion of all possible items, most of the r_{ij} ratings are missing in real-world systems, and it has been an inherently important task to predict these ratings, based on which to make personalized recommendations for consumers.

The most 'standard' and representative formalization of rating prediction could be the Collaborative Filtering (CF) [36] approach based on Latent Factor Models (LFM) [36, 17], which predicts the consumer-item ratings $\hat{r}_{ij}$ with consumer/item biases and latent factors:

$$\hat{r}_{ij} = \alpha + \beta_i + \gamma_j + \vec{x}_i^T \vec{y}_j \tag{6}$$

where α is the global offset, β_i and γ_j are the consumer and item biases, $\vec{x}_i$ and $\vec{y}_j$ are the K-dimensional latent factors of consumer u_i and item g_j, respectively, which are multiplied as an approximation to the rating. Based on a set of observed training records $\mathcal{R}$, the model is typically targeted with the goal of providing accurate rating predictions, where we determine the parameter set $\Theta = \{\alpha, \beta_i, \gamma_j, \vec{x}_i, \vec{y}_j\}$ with the following minimization problem,

$$\Theta = \underset{\Theta}{argmin} \sum_{r_{ij} \in \mathcal{R}} \left(r_{ij} - \hat{r}_{ij}\right)^2 + \lambda\Omega(\Theta) \tag{7}$$

and $\Omega(\Theta)$ is the ℓ_2-norm regularization term. The minimization of Eq.(7) can be easily accomplished with Stochastic Gradient Descent (SGD) or Alternating Least Squares (ALS) algorithms [17].

3. THE FRAMEWORK FOR OSA

In this section, we propose our Total Surplus Maximization (TSM) framework for the problem of Online Service Allocation (OSA). For clarity in organization and easy understanding, we introduce the key components in a logical order, and then unify these components to present the whole framework.

3.1 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.

Suppose there exist m consumers $\{u_1, u_2, \cdots, u_m\}$, n goods $\{g_1, g_2, \cdots, g_n\}$, which can be products in e-commerce, jobs in online freelancing networks, or loans in P2P lending websites; as well as r producers $\{p_1, p_2, \cdots, p_r\}$, where each good is and can only be provided by a single producer, but each producer may be able to provide multiple goods.

In the following, we use $1 \leq i \leq m, 1 \leq j \leq n$, and $1 \leq k \leq r$ to index the consumers, goods, and producers, correspondingly.

Definition 1. *Let $M = [M_1, M_2, \cdots, M_n]$ be the Service Quantity Vector, where $M_j \geq 0$ is the total amount that good g_j can be provided by its producer in the system.*

For example, $M_j = 1$ in online freelancing networks because each job can only be provided once to only one freelancer; in P2P lending networks like Prosper, we have $0 < M_j < \infty$, which is the amount of money requested by each loan; for e-commerce websites like Amazon, however, we can treat $M_j = \infty$ because for most of the normal goods, the producer can replenish the stock in case of a boosting market demand.

Definition 2. *The Online Service Allocation (OSA) problem thus aims to find an Allocation Matrix $Q = [Q_{ij}]_{m\times n}$, where $Q_{ij} \geq 0$ is the quantity that consumer u_i is provided with good g_j.*

To meet with the total possible service quantity restriction represented by vector M, we have $\sum_i Q_{ij} \leq M_j$ for each good g_j, i.e., $\mathbf{1}^T Q \leq M$, where $\mathbf{1}$ is a column vector of 1's. In different real-world application scenarios we may apply extra constraints on Q to meet specific task characteristics. For example, $Q_{ij} \in \mathbb{N}$ for e-commerce goods, or $Q_{ij} \in \{0, 1\}$ for online freelancing services.

The problem of OSA widely exists and finds its instantiation in a lot of online services or mobile applications wherever there is service consumption. Besides e-commerce, P2P lending, and freelancing services we exampled here, other applications include riding services such as Uber and Lyft, group purchase services such as Groupon, or even lodge renting services such as Airbnb, etc.

3.2 Personalized Utility

Different consumers may experience different utility even from the same quantity of the same good. In e-commerce websites, for example, one with an SLR camera in hand may obtain a high consumer surplus when supplied with an SLR lens, however, the surplus may be extremely low when a lens is provided to someone without a camera. For P2P lending, similarly, the same amount of money could mean a huge surplus to someone that is in an urgent need (thus willing to accept a higher interest rate), while the surplus may be lower for those who are not that thirsty for money (thus insists on lower interest rates).

This 'personalized' feature of utility makes up the inherent driving power for service allocation, which makes it reasonable for us to match the appropriate good with the appropriate consumer so as to maximize the potential total surplus in the whole system. And this process can come in the form of personalized recommendation or intelligent marketing assistance to decision makers in practical applications.

In this work, we adopt the personalized utility $U_{ij}(q)$ on a consumer-to-good level, namely:

$$U_{ij}(q) = \frac{1 - \exp(-a_{ij} q)}{a_{ij}}, \text{ or } U_{ij}(q) = a_{ij} \ln(1+q) \tag{8}$$

where $U_{ij}(q)$ indicates the utility when supplying a quantity q of good g_j to consumer u_i, which is parameterized by the personalized degree of risk aversion a_{ij}.

The robust estimation of a_{ij} can be derived differently for different scenarios, based on the availability of data and the applicable economic theories. For example, we adopt the *Law of Zero Surplus for the Last Unit* [13] for the inference of a_{ij} in e-commerce, while the property of percentage surplus is applied in freelancing. We will exposit in more details in the following section of Model Specification.

3.3 Total Surplus Maximization

Based on the personalized utility $U_{ij}(q)$ and the cost function $C_j(q)$ for each good g_j, the most direct approach for online service allocation tries to find an exact allocation matrix

Q so as to maximize the total social surplus under the constraints of service quantity:

$$\begin{aligned} &\underset{Q}{\text{maximize}} \sum_i \sum_j \big(U_{ij}(Q_{ij}) - C_j(Q_{ij})\big) \\ &\quad s.t.\ \mathbf{1}^T Q \le M,\ Q_{ij} \in \mathbb{S} \end{aligned} \tag{9}$$

where $\mathbb{S}$ is the set of possible legal values for a specific application, e.g., $\mathbb{S} = \mathbb{N}$ for e-commerce and $\mathbb{S} = \{0, 1\}$ for online freelancing.

However, an important issue to consider in practice is that the *Hypothesis of Rational Man* [9] may not always hold in real-world applications, and the consumers may not always rationally choose the best quantities of goods. As a result, we may lose the model generality and the learning power for data fitting if we restrict Q to real-valued matrices. Actually, consumers may indeed choose the best quantity with a greater probability, but they may also choose non-optimal quantities, although with lower probabilities.

To model this 'non-rationality' in consumer behavior, we relax the elements Q_{ij} in allocation matrix Q to random variables with a probability distribution. For example, when $Q_{ij} \sim \mathcal{N}(\mu_{ij}, \sigma_{ij})$, consumer u_i chooses good g_j on quantity μ_{ij} with the highest probability, but she may also make consumptions on other quantities, although the probability could be lower. In this sense, our final framework for service allocation maximizes the following expected total surplus:

$$\begin{aligned} &\underset{\Theta(Q)}{\text{maximize}} \sum_i \sum_j \int \big(U_{ij}(Q_{ij}) - C_j(Q_{ij})\big) p(Q_{ij}) \mathrm{d}Q_{ij} \\ &\quad s.t.\ \mathbf{1}^T \int Q p(Q) \mathrm{d}Q \le M,\ Q_{ij} \in \mathbb{S} \end{aligned} \tag{10}$$

where $p(Q_{ij})$ is the probability density function of each quantity Q_{ij}, $p(Q) = [p(Q_{ij})]_{m \times n}$, and the integral on Q is per element wise. Besides, $\Theta(Q)$ is the parameter set of all the Q_{ij}'s in Q. This probabilistic interpretation also simplifies the computation in some applications, which will be introduced in the following section of model specification.

The model produces the optimal density functions $p(Q)$ as the final output, and we take the expectation $\bar{Q} = \int Q p(Q) \mathrm{d}Q$ as the final allocation matrix to make system decisions.

4. MODEL SPECIFICATION

As noted before, our TSM framework for online service allocation can be specified into different online applications based on the availability of different data. To make expositions on this feature and also for practical application/evaluation, we specify our framework into three types of online services, i.e., e-commerce, P2P lending, and online freelancing, which correspond to different choices of the components in Eq.(10). For easy reference, Table 1 shows these choices in an integrated manner.

4.1 E-commerce

We first attempt to estimate the personalized utility $U_{ij}(q)$ based on the consumer purchasing records. Although $U_{ij}(q)$ is not directly observed in the data, it is subject to the *Law of Zero Surplus for the Last Unit* [13]. Let q_{ij} be the actual quantity of good g_j that consumer u_i purchased in the dataset, and let $CS_{ij}(q_{ij}) = U_{ij}(q_{ij}) - P_j q_{ij}$ be the consumer surplus obtained from such a purchasing behaviour, then the law of zero surplus gives us the following constraints:

$$\begin{aligned} \Delta CS_{ij}(q_{ij}) &= CS_{ij}(q_{ij}) - CS_{ij}(q_{ij} - 1) \ge 0 \\ \Delta CS_{ij}(q_{ij} + 1) &= CS_{ij}(q_{ij} + 1) - CS_{ij}(q_{ij}) < 0 \end{aligned} \tag{11}$$

The economic intuition here is that, a consumer decided to make a purchase of quantity q_{ij} because he/she can still obtain increasement on surplus with the last unit, but even a single more unit of purchase will decease the surplus.

In the spirit of collaborative filtering, we model the personalized parameter in Eq.(8) as $a_{ij} = \alpha + \beta_i + \gamma_j + \vec{x}_i^T \vec{y}_j$, where $\vec{x}_i$ is the K-dimensional consumer latent factor of consumer u_i, and $\vec{y}_j$ is similarly the latent factor corresponding to good g_j. Hence, the risk aversion parameters a_{ij} become intermediate parameters that can be derived from the actual parameters $\Theta = \{\alpha, \beta_i, \gamma_j, \vec{x}_i, \vec{y}_j\}$ in model optimization. Based on this, we maximize the following log-likelihood of observing the whole purchasing records dataset:

$$\begin{aligned} &\underset{\Theta}{\text{maximize}}\ \log p(D) \\ =& \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \log \big(Pr\big(\Delta CS_{ij}(q_{ij}) \ge 0\big) Pr\big(\Delta CS_{ij}(q_{ij} + 1) < 0\big)\big) \\ -& \lambda \big(\alpha^2 + \sum_{i=1}^{m} \beta_i^2 + \sum_{j=1}^{n} \gamma_j^2 + \sum_{i=1}^{m} \|\vec{x}_i\|_2^2 + \sum_{j=1}^{n} \|\vec{y}_j\|_2^2\big) \\ &\quad s.t.\ \vec{x}_i, \vec{y}_j \ge 0,\ \forall\ 1 \le i \le m,\ 1 \le j \le n \end{aligned} \tag{12}$$

where I_{ij} is an indicator whose value is 1 when consumer u_i purchased good g_j in the dataset, and 0 otherwise. The regularizer with coefficient $\lambda > 0$ is used to prevent model over-fitting. We apply the commonly used non-negative constraints [19, 18] on the latent factors $\{\vec{x}_i\}_{i=1}^m$ and $\{\vec{y}_j\}_{j=1}^n$, and adopt the sigmoid function to model the probabilities, which are:

$$Pr\big(\Delta CS_{ij}(q_{ij}) \ge 0\big) = \frac{1}{1 + \exp(-\Delta CS_{ij}(q_{ij}))} \tag{13}$$

and that,

$$Pr\big(\Delta CS_{ij}(q_{ij}+1) < 0\big) = 1 - Pr\big(\Delta CS_{ij}(q_{ij}+1) \ge 0\big) \tag{14}$$

An optimal solution of Eq.(12) can be obtained by gradient descent, which involves the computation of the gradients on the risk aversion parameter a_{ij}. To simplify the computation and make it possible for model estimation, we adopt the KPR utility function $U_{ij}(q) = a_{ij} \ln(1 + q)$. After the above model learning process, we can obtain the learning parameters in Θ, and thus the estimated risk aversion parameters $\hat{a}_{ij}$, which further give us the personalized utility functions $U_{ij}(q)$ as follows:

$$U_{ij}(q) = \hat{a}_{ij} \ln(1+q) = (\alpha + \beta_i + \gamma_j + \vec{x}_i^T \vec{y}_j) \ln(1+q) \tag{15}$$

For simplicity, we set the cost of selling e-commerce goods to be proportional to the quantity q with a ratio of c_j, i.e., $C(q) = c_j q$, where c_j is the cost of a unit service of good g_j.

At last, let the elements Q_{ij} in the allocation matrix Q of Eq.(10) follow a Poisson distribution because $Q_{ij} \in \mathbb{N}$, i.e., $p(Q_{ij} = q) = \lambda_{ij}^q e^{-\lambda_{ij}} / q!$, where λ_{ij} feature the distribution parameters. Finally, the framework for OSA based on total surplus maximization in Eq.(10) can be specified to the

Table 1: Parameter comparison of the three different specifications of the framework for OSA based on total surplus maximization, including the applications of e-commerce, P2P lending services, and online freelancing.

Application	$CS_{ij}(Q_{ij})$	$PS_{ij}(Q_{ij})$	$\mathbb{S}$	M	$p(Q_{ij})$	$\bar{Q}_{ij}$
E-commerce	$\hat{a}_{ij}\ln(1+Q_{ij})-P_jQ_{ij}$	$(P_j-c_j)Q_{ij}$	$\mathbb{N}$	$M_j=\infty$	$p(Q_{ij}=q)=\lambda_{ij}^q e^{-\lambda_{ij}}/q!$	λ_{ij}
P2P lending	$(r_j-\hat{r})Q_{ij}$	$(r_j^{max}-r_j)Q_{ij}$	$\mathbb{R}_+$	$0<M_j<\infty$	$Q_{ij}\sim\mathcal{N}(\mu_{ij},\sigma_{ij})$	μ_{ij}
Freelancing	$h(\hat{r}_{ij})s_jQ_{ij}$	$h(\hat{r}_{kj})s_jQ_{ij}$	{0,1}	$M_j=1$	$p(Q_{ij}=1)=\alpha_{ij}, P(Q_{ij}=0)=1-\alpha_{ij}$	$I_{\alpha_{ij}=\max\{\alpha_{i'j}\}_{i'=1}^m}$

following maximization problem:

$$\begin{aligned}\underset{\Lambda}{\text{maximize}} \sum_i\sum_j\sum_{q=0}^{\infty}\frac{\lambda_{ij}^q e^{-\lambda_{ij}}}{q!}\big(\hat{a}_{ij}\ln(1+q)-c_jq\big)\\ -\eta\sum_i\sum_j I_{ij}(\lambda_{ij}-q_{ij})^2\end{aligned}\tag{16}$$

where $\Lambda=[\lambda_{ij}]_{m\times n}$ is the parameter set, $\eta>0$ is regularizer coefficient, I_{ij} is still the indicator of whether u_i purchased g_j in the training set, and q_{ij} is the actual purchasing quantity. The quantity constraints are left out because $M_j=\infty$.

In practice, we do not have to sum over infinite q's to compute an expectation over Poisson distribution, but only need to consider sufficiently many choices. In this work, we choose to sum up from $q=0$ to 10 because $10!=3,628,800$ is sufficiently large to diminish the residuals according to the theory of Taylor series expansion.

The minimization of both Eq.(12) and Eq.(16) can be conducted based on gradient descent. Once the distribution parameters Λ are obtained from Eq.(16), we have the expected allocation matrix $\bar{Q}$ as:

$$\bar{Q}_{ij}=\sum_{q=0}^{\infty}q\cdot\frac{\lambda_{ij}^q e^{-\lambda_{ij}}}{q!}=\sum_{q=0}^{\infty}\frac{\lambda_{ij}^q e^{-\lambda_{ij}}}{(q-1)!}=\lambda_{ij}\tag{17}$$

which we take for service allocation and product recommendation. Note that in the regularizer of Eq.(16), λ_{ij} is actually the expectation of quantity Q_{ij} according to the nature of Poisson distribution (Eq.(17)). As a result, the regularization component applies a guidance to the learning process, so that the estimated allocation quantities for those observed transections in the training dataset would be close to their actual values.

4.2 Online Peer-to-Peer Lending

In P2P lending services like Prosper, the borrowers are loan request producers, where the loan requests can be viewed as financial products. 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 that we concern for P2P lending is that, how the lenders (i.e., consumers) should distribute their assets among the loan requests (i.e., determining the allocation matrix Q), so that the world becomes better, namely, the total surplus in the system is maximized.

In a standard online lending process, the borrower (request producer) p_k initiates a loan request g_j by specifying two basic factors, the first is the quantity of money to collect M_j, and the second is the maximal interest rate r_j^{max} that he/she would be willing to accept for this loan. Once a request is generated, the lenders (request consumers) u_i bid the request by providing the amount of money they would like to lend and the interest rates they ask for, which should be lower than or equal to r_j^{max}. When the total amount of money in bid exceeds the request in a given time period, the loan request then makes a deal, and the top bidders with the lowest interest rates whose money amounts to the request win the bid. Furthermore, the highest interest rate among the winners is set as the final interest rate r_j for the loan g_j.

As a result, the consumer surplus for the lenders is the interest they obtain from this loan r_jQ_{ij}, less the opportunity cost $\hat{r}Q_{ij}$ of investing the money with a risk-free investment, for example, to save the money in bank, where $\hat{r}$ is the risk-free interest rate. As a result, we have:

$$CS_{ij}(Q_{ij})=(r_j-\hat{r})Q_{ij}\tag{18}$$

Similarly, the producer surplus for the borrowers is the interest they would be willing to pay $r_j^{max}Q_{ij}$, less the actual interest they have to pay r_jQ_{ij}, namely,

$$PS_{ij}(Q_{ij})=(r_j^{max}-r_j)Q_{ij}\tag{19}$$

Thus the total surplus is:

$$TS_{ij}(Q_{ij})=CS_{ij}(Q_{ij})+PS_{ij}(Q_{ij})=(r_j^{max}-\hat{r})Q_{ij}\tag{20}$$

Because Q_{ij} represents the quantity of money that is a continuous variable, we apply a normal distribution to describe Q_{ij}, i.e., $Q_{ij}\sim\mathcal{N}(\mu_{ij},\sigma_{ij})$.

Finally, the problem of OSA for P2P lending boils down to the following specification of our framework in Eq.(10):

$$\begin{aligned}&\underset{U,\Sigma}{\text{maximize}}\sum_i\sum_j\int\frac{(r_j^{max}-\hat{r})Q_{ij}}{\sqrt{2\pi}\sigma_{ij}}\exp\left(-\frac{(Q_{ij}-\mu_{ij})^2}{2\sigma_{ij}^2}\right)\mathrm{d}Q_{ij}\\ &s.t.\ \mathbf{1}^T\int\frac{Q}{\sqrt{2\pi}\Sigma}\exp\left(-\frac{(Q-U)^2}{2\Sigma^2}\right)\mathrm{d}Q\le M, Q_{ij}\in\mathbb{R}_+\end{aligned}\tag{21}$$

where $U=[\mu_{ij}]_{m\times n}$ and $\Sigma=[\sigma_{ij}]_{m\times n}$ are the parameters. This can be further simplified to:

$$\begin{aligned}\underset{U,\Sigma}{\text{maximize}}\sum_i\sum_j\mu_{ij}(r_j^{max}-\hat{r})\\ s.t.\ \mathbf{1}^TU\le M,\ \mu_{ij}\in\mathbb{R}_+\end{aligned}\tag{22}$$

which can be solved to find the optima with linear programming. Finally, we take the expected quantity under Gaussian distribution as the allocation matrix, i.e.,

$$\bar{Q}_{ij}=\mu_{ij}\tag{23}$$

This result is interesting in that, it allows us to allocate the investments in a greedy manner according to the per capita surplus $(r_j^{max}-\hat{r})$ of each loan request, which is an intuitional rule for investment in practice and easily applicable in real-world systems.

4.3 Online Freelancing Platforms

In online freelancing networks like Mturk and Upwork, the employer (job producer) p_k post job g_j online, and the freelancers (job consumer) u_i apply for the jobs that they are willing to take. Because a job can only be assigned to a single freelancer and a freelancer can only decide to take a job or not rather than take part of a job, the elements Q_{ij} in allocation matrix Q can only be binary values in $\{0, 1\}$.

The employer and freelancer negotiate to decide the salary s_j for job g_j. After the job is accomplished, they make ratings on each other which indicates their satisfaction about the other side. We denote the rating given by freelancer u_i and employer p_k about the job g_j as r_{ij} and r_{kj}, respectively, which are integers in a specific rating scale.

To estimate the consumer and producer surplus experienced on a given job, we adopt the economic assumption that the percentage surplus against the price that the consumer pays or the producer obtains is proportional to the normalized ratings that they cast on each other [13, 39], i.e., a higher rating implies a higher percentage surplus.

To do so, we predict the freelancer-job ratings $\hat{r}_{ij}$ and employer-job ratings $\hat{r}_{kj}$, respectively, based on the Collaborative Filtering (CF) approach of Eq.(7) introduced in section 2.3. By the sigmoid function $h(x) = \frac{2}{1+\exp(-x)} - 1$, we further model the percentage surplus for freelancers as:

$$\frac{U_{ij}(Q_{ij}) - s_j}{s_j} = h(\hat{r}_{ij})Q_{ij} = \left(\frac{2}{1+e^{-\hat{r}_{ij}}} - 1\right) Q_{ij} \quad (24)$$

and the percentage producer surplus as:

$$\frac{s_j - C_j(Q_{ij})}{s_j} = h(\hat{r}_{kj})Q_{ij} = \left(\frac{2}{1+e^{-\hat{r}_{kj}}} - 1\right) Q_{ij} \quad (25)$$

where $Q_{ij} \in \{0, 1\}$ can be viewed as a binary indicator that whether or not a job is assigned, so that a surplus can be obtained for consumers and producers in Eq.(24) and (25).

As a result, the consumer, producer, and total surpluses implied in a specific job assignment u_i to g_j are:

$$\begin{aligned} CS_{ij}(Q_{ij}) &= U_{ij}(Q_{ij}) - s_j = h(\hat{r}_{ij})s_jQ_{ij} \\ PS_{ij}(Q_{ij}) &= s_j - C_j(Q_{ij}) = h(\hat{r}_{kj})s_jQ_{ij} \\ TS_{ij}(Q_{ij}) &= \left(h(\hat{r}_{ij}) + h(\hat{r}_{kj})\right) s_jQ_{ij} \end{aligned} \quad (26)$$

On considering that Q_{ij} is binary valued, we apply a Bernoulli distribution to model its probabilistic nature, i.e.:

$$p(Q_{ij} = 1) = \alpha_{ij}, P(Q_{ij} = 0) = 1 - \alpha_{ij} \quad (27)$$

where $0 \leq \alpha_{ij} \leq 1$. Let $A = [\alpha_{ij}]_{m\times n}$ be the parameter set, and let $M_j = 1$ because each individual job is by nature provided only once. The OSA problem for online freelancing services is thus specified as:

$$\begin{aligned} &\underset{A}{\text{maximize}} \sum_i \sum_j \left(h(\hat{r}_{ij}) + h(\hat{r}_{kj})\right) s_j\alpha_{ij} \\ &s.t.\ \mathbf{1}^T A \leq \mathbf{1},\ 0 \leq \alpha_{ij} \leq 1 \end{aligned} \quad (28)$$

Eq.(28) can be easily optimized using linear programming. Once the parameters in $A = [\alpha_{ij}]_{m\times n}$ are obtained, we assign the job g_j to the freelancer u_i of the maximum probability α_{ij} among $\alpha_{i'j}$ of all the freelancers on that job, namely:

$$\bar{Q}_{ij} = \begin{cases} 1, \text{ if } \alpha_{ij} = \max\{\alpha_{i'j}\}_{i'=1}^m \\ 0, \text{ otherwise} \end{cases} \quad (29)$$

This result is intuitional because it can also be achieved in a greedy manner by replacing α_{ij} with Q_{ij} in Eq.(28). In this way, we assign a given job g_j to the freelancer u_i who gains the highest value regarding $\left(h(\hat{r}_{ij}) + h(\hat{r}_{kj})\right) s_j$, which is actually a specification of the direct non-probabilistic framework in Eq.(9). Furthermore, this can be viewed as a surplus-augmented version of the traditional CF-based personalized recommendation algorithms, which will be discussed in the following together with the previous specifications.

4.4 Discussion

It is worthwhile to compare and contrast our framework with some traditional recommendation algorithms.

In the case of unlimited quantity where $M_j = \infty$, the quantity constraint $\mathbf{1}^T \int Qp(Q)\mathrm{d}Q \leq M$ in Eq.(10) can be removed and we obtain an unconstrained optimization function, just as shown in Eq.(16). In this case, the total surplus related to each consumer is independent from those of the others, and the optimal allocations for each consumer is independently isolated from each other. Take the e-commerce application for example, the allocation for a given consumer u_i can be obtained with the following equation:

$$\underset{\{\lambda_{ij}\}_{j=1}^n}{\text{maximize}} \sum_j \left(\sum_{q=0}^{\infty} \frac{\left(\hat{a}_{ij} \ln(1+q)\right)\lambda_{ij}^q e^{-\lambda_{ij}}}{q!} - \lambda_{ij}c_j\right) - \eta\Omega(\Theta) \quad (30)$$

This is similar to traditional Personalized Recommender System (PRS) [30] algorithms, where we consider the preferences of each targeted user and aim to provide the most relevant recommendations. The spirit of personalization has been inherently incorporated in the design of the personalized utility of Eq.(15), where $\hat{a}_{ij} = \alpha + \beta_i + \gamma_j + \vec{x}_i^T \vec{y}_j$ describes the consumer preference towards the goods in a collaborative manner based on the latent factors learned from the wisdom of the crowds, which is similar to the Collaborative Filtering approach in Section 2.3.

Similarly for online freelancing application denoted in Eq.(28), we see that for a given target job g_j, the employer-job rating $h(\hat{r}_{kj})$ (predicted by CF) and the hourly salary s_j would be known values. As a result, the greedy weight $\left(h(\hat{r}_{ij}) + h(\hat{r}_{kj})\right) s_j$ will only depend on the freelancer u_i. In this sense, we are actually assigning the job g_j to the freelancer u_i of the maximized $h(\hat{r}_{ij})s_j$. This is actually a generalization of CF-based algorithms that recommend job g_j to the freelancer u_i of the maximum predicted rating $\hat{r}_{ij}$, where the only difference is that we further take the hourly salary s_j into consideration for a maximized total surplus that is measured on a basis of money.

Another interesting yet intuitive conclusion from the existence of a non-infinity solution to Eq.(30) is that, larger quantity of products that the producers sell is not necessarily preferred by the system, although we assume the quantities that producers can supply are unlimited. This results from the diminishing marginal utility experienced by consumers, and this conclusion is verified by the disadvantages observed on dumping in practical tradings.

However, when the constraint on quantity exists, the consumer surpluses are correlated with each other, so that the allocation matrix that gains a globally maximized total surplus does not necessarily imply a maximized surplus for each consumer or producer.

5. EXPERIMENT

In this section, we present experimental results to verify the performance of our framework, on both the traditional tasks of purchasing prediction and personalized recommendation, as well as the completely new task of total surplus maximization.

For clarity in structure and also for an easier explanation to the insights of our framework, we first present a comprehensive exposition to the results on e-commerce application, which we think is one of the most representative and easy-to-understand application scenarios that match the economic theories. After that, we further introduce the empirical results on P2P lending and online freelancing applications, so as to provide a full-scope understanding of the performance of our framework.

5.1 E-commerce Dataset Description

We adopt the consumer purchasing records dataset from Shop.com[1] for experimental evaluation, because an important information source leveraged in our framework is the quantity of product that a consumer purchased in each transaction, which is absent in many of the public datasets. In the Shop.com dataset, however, we have both the product price information and the quantity that a consumer purchased in each record.

To avoid the problem of cold-start [21, 38], and to focus on our key research target of total surplus maximization, we select those consumers and products with at least five purchasing records, which is a frequently adopted pre-processing method in previous work [21, 20, 36]. Some statistics of our dataset are summarized in Table 2.

Table 2: Statistics of the Shop.com dataset

#Consumers	#Products	#Transactions	Density	Train/Test
34,099	42,691	400,215	0.03%	75%/25%

We see that the dataset is extremely sparse with a density of only 0.03%, which is similar to previously seen recommendation tasks. Furthermore, we randomly select 75% of the transactions from each consumer to construct the training set for model learning, and the rest 25% are used for testing. These amount to roughly $100k$ transactions by $34k$ consumers on $30k$ products in the testing dataset.

5.2 Parameter Selection and Experimental Setup

The personalized KPR utility function $U_{ij}(q)$ indicated in Eq.(15) is parameterized solely by parameter a_{ij}, and the estimation of a_{ij} boils down to the inference of consumer and product biases and latent factors by optimizing Eq.(12).

In the experiments, we set the hyper-parameter λ involved in Eq.(12) and η in Eq.(16) based on cross validation, and they are primarily set as $\lambda = 0.05$ and $\eta = 5$, unless we tune these parameters to investigate their influences on model performance. Throughout the experiments we set the number of latent factors (i.e., the dimensions of $\vec{x}_i$ and $\vec{y}_j$) $K = 20$ in Eq.(12), because we find that 20 factors are sufficiently enough to stable the model performance.

Once the estimated $\hat{a}_{ij}$'s are obtained with Eq.(12), we are able to evaluate the utility $U_{ij}(q)$ of an arbitrary consumer-product pair, which allows us to learn the average allocation quantities λ_{ij} in Eq.(16). Recognizing that λ_{ij} is consumer-product specific similar to a_{ij}, we once again parameterize it in a CF manner with $\lambda_{ij} = \alpha' + \beta'_i + \gamma'_j + \vec{x}'^T_i \vec{y}'_j$, and thus λ_{ij} can be estimated as a media parameter by gradient descending on $\Theta' = \{\alpha', \beta'_i, \gamma'_j, \vec{x}'_i, \vec{y}'_j\}$.

For simplicity, we set the cost $c_j = 0.5P_j$ for all the products in the dataset, where P_j is the price of a product g_j. The cost ratio 0.5 is an average estimation based on the surveys of 100 producers from different product categories. Besides, we find that the performance of our framework is not sensitive to different cost ratios in a reasonable range.

[1] `http://www.shop.com`

Please note that when the regularization term η in Eq.(16) is set sufficiently large, the effect of total surplus component will vanish and the equation turns into a mere CF problem to predict q_{ij}, which serves a baseline algorithm in our later comparative study.

The procedure ends up with the estimated values of λ_{ij} for any given consumer-product pair in our dataset. As suggested by Eq.(17), product recommendation list is thus provided to consumer u_i by ranking the products in descending order of λ_{ij}. For easy reference, the values of the involved hyper parameters are shown in Table 4.

Table 4: Summarization of the parameters. Note that the number of latent factors K and the CF regularization coefficient λ are identified by cross validation, and are fixed throughout the experiments; η varies in our experiment so as to examine its influence; c_j is the cost of product g_j.

#Latent factors K	λ in Eq.(12)	η in Eq.(16)	c_j in Eq.(16)
20	0.05	5	$0.5P_j$

5.3 Purchase Prediction and Recommendation

We investigated the performance of our TSM framework for the task of personalized purchase prediction and recommendation. For performance comparison, we adopt the widely used CF algorithm in Eq.(6) and (7) to predict the purchasing quantities directly, which are integer values ranging from 1 to 20. For fair comparison, the hyper-parameters K and λ are set the same as those in Table 4.

Similar to our TSM framework, once the predicted quantities are obtained, we construct the top-N recommendation list for a consumer from the testing set in descending order of the quantities. We adopt the measure Conversion-Rate@N (CR@N) for performance evaluation on top-N recommendation, which is a typical metric widely adopted in real-world e-commerce systems [8].

For a given number of testing consumers and the recommendation lists of length N for each of them, CR@N is the percentage of lists that 'hit' the purchase records in testing set for the target consumer. In our experiment, N runs from 1 up to 100. For each consumer in the testing set, there are as many as $30k$ candidate products for recommendation, and all the candidate products are present in the training

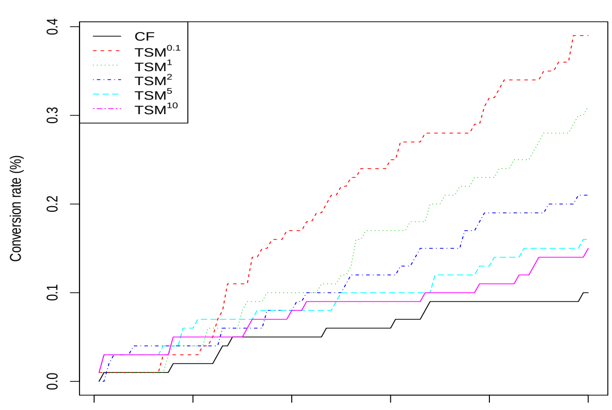

Figure 2: Comparison of the recommendation performance for CF and TSM$^\eta$. The y-axis is the conversion rate, and the x-axis is the length N of each recommendation list.

Table 3: Evaluation on Conversion Rate (CR@N) and Total Surplus (TS@N) for Top-N recommendation, where TSM^* stands for our TSM approach with regularization coefficient $\eta = *$ in Eq.(16).

N	5					10					20				
Method	CF	$TSM^{0.1}$	TSM^1	TSM^5	TSM^{10}	CF	$TSM^{0.1}$	TSM^1	TSM^5	TSM^{10}	CF	$TSM^{0.1}$	TSM^1	TSM^5	TSM^{10}
CR (%)	0.10	0.10	0.10	**0.30**	**0.30**	0.10	0.10	0.10	**0.30**	**0.30**	0.20	0.30	0.40	**0.60**	0.50
TS ($)	33.05	**1009.45**	**1009.45**	422.01	24.48	57.89	**2278.36**	2208.50	807.56	213.45	98.09	**2892.03**	3135.35	1137.89	676.65

dataset. For computational efficiency in evaluation, we randomly select 1000 users to evaluate average CR at each time, and the CR performance of 30 testing rounds are averaged to provide the final evaluation results.

The results for CF and our TSM framework with different choices on regularization coefficient $\eta = 0.1, 1, 5, 10$ with recommendation length $N = 5, 10, 20$ are presented in Table 3. And more complete results for N from 1 to 100 can be seen in Figure 2. The bolded improvements in Table 3 are significant at a 0.05 level.

The results show that our TSM framework outperforms CF for most choices of regularization coefficient η and recommendation length N. An interesting observation is that the performance of TSM$^\eta$ generally degrades with the increase of η on relatively long recommendation lists, all the way towards the performance given by the baseline algorithm of CF. This is actually reasonable as stated before, because when $\eta \to \infty$, TSM literally degenerates to CF and its performance will also converge to CF. This observation further emphasizes the importance of our surplus maximization component, and it suggests that maximizing with total surplus could be beneficial to the consumer experience on personalized recommendations.

Besides, the results also suggest that the choice of η should not be too small either, which would dismiss the quantity guidance of the observed purchases, especially for top precisions in shorter recommendation lists. One possible reason can be that without the quantity guidance, λ_{ij} would mostly depend on the personalized KPR utility U_{ij}. As KPR utility function is rather limited in terms of shape flexibility, it could fail to describe the actual consumer utility for some products. We actually confirmed this by predicting the purchase quantities using the constraints in Eq.(11) directly, and the predictions turned out rather inaccurate with larger *root mean squared error* (RMSE) than that by CF. In summary, η influences the performance by balancing the importance between total surplus and quantity guidance, and it should be properly selected in practical applications.

5.4 Evaluation on Total Surplus

In this section, we closely examine the performance of our framework under the *total surplus* metric, which is a core notion of this work. The evaluation is carried out based on the recommendation results from the above section. Similar to the Top-N conversion rate, we are interested in calculating the *accumulated tocial surplus* of a Top-N recommendation list for each user, which is defined as,

$$TS@N = \frac{1}{M}\sum_{i=1}^{M}\sum_{j\in\Pi_{i,N}}\left(\hat{a}_{ij}\ln(1+\lambda_{ij}) - c_j\lambda_{ij}\right) \tag{31}$$

where i and M are the index and the total number of testing consumers, respectively, N is the length of recommendation list, and $\Pi_{i,N}$ is the length-N personalized recommendation list for the i-th consumer.

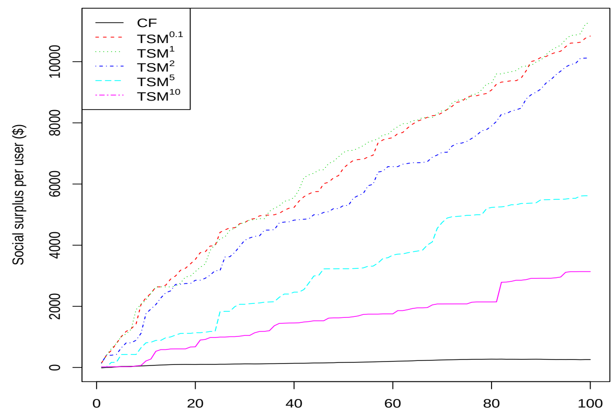

Figure 3: Comparison on Tocial Surplus (TS) for CF and TSM$^\eta$. Note that TS@N means the total social surplus of the top N recommendations when they are all accepted by the corresponding consumer.

Similarly, the results of $TS@N$ are reported in Table 3, and a full scope report under comprehensive choices of N can be found in Figure 3.

It can be seen from the results that our TSM approach constantly outperforms the CF method. This result is actually not surprising because our TSM framework is by nature able to maximize the total surplus by Eq.(16). Besides, we find that the smaller η is, the more total surplus our TSM approach gains. This observation on the influence of η further verifies the effects of the surplus maximization component and the quantity guidance in Eq.(16).

More interestingly, when combining this result with that on recommendation in the previous section, we find that our TSM framework can achieve decent results in terms of both total surplus and conversion rate when η is properly set. This is exciting because our framework is able to benefit the social good on total surplus, and at the same time improves the consumer experience in personalized recommendations.

5.5 P2P Lending Networks

To verify the performance on Peer-to-Peer lending networks, we use the datasets from a famous P2P lending website Prosper[2][6]. Beginning from the third quarter in 2009, Prosper introduced an automatic bidding mechanism that bids the listings (i.e., loan requests) on behalf of the lenders automatically once a listing is created. However, as we intend to investigate the behaviour of consumers and producers in an economic system, we prefer the decisions made directly by themselves, instead of those indirectly by the algorithms. As a result, we adopt those listing and bidding records before this mechanism was launched, which finally covers the period from November 9th 2005 to May 8th 2009.

As we do not consider risk control in our current model, we select those listings whose status are not *Defaulted*, *Cancelled* or *Charge-off* from the dataset, because these listings

[2] http://www.prosper.com

are meant to be ruled out from the system by the intelligent risk control mechanisms. Finally, our dataset involves those listings of the status *Current*, *Late*, *Payoff in Progress*, or *Paid*, which correspond to 46,680 listings, 1,814,503 biddings, and a total amount of $157,845,684 fundings. Some statistics of these records are summarized in Table 5.

Table 5: Statistics of the selected Prosper dataset, where 'rate' represents the interest rate of a loan.

#Listings	#Lenders	#Biddings	TotalAmount
46,680	49,631	1,814,503	$157,845,684
MinimumRate	MaximumRate	AverageRate	Amount/Listing
0.0001	0.4975	0.1662	$3,381.44

To calculate the total surplus reached by an arbitrary allocation $Q = [Q_{ij}]_{m\times n}$, we take the yearly average bank deposit interest rate $\hat{r} = 0.01$ as the risk-free interest rate, and the TS for P2P lending can be calculated as:

$$TS_{P2P} = \sum_i \sum_j Q_{ij}(r_j^{max} - \hat{r}) \tag{32}$$

Based on this, the results on total surplus for the actual allocations (Actual) and our Total Surplus Maximization (TSM) framework are shown as follows:

Table 6: Results on total surplus with and without our Total Surplus Maximization (TSM) framework.

	TS($)	TS/Listing($)	TS/capita($)
Actual	25,174,131	539.29	0.1595
TSM	33,838,364	724.90	0.2144

We see that our TSM framework achieves 34.42% higher total and per listing/capita surplus, from $0.16 per capita to $0.21 per capita, which is an exciting improvement in capital efficiency and social good on the online lending systems. Based on two-tailed t-test on the large amount of listings, the improvements are significant at a 0.01 level.

The improvement on total surplus is not surprising because our framework intends to achieve a maximized surplus among all the possible allocations. However, we should further verify that our allocations are acceptable to the lenders in practice. As a result, we calculate the Percentage of Paid (PoP) listings among all the funded listings in our dataset, which indicates the safety factor of a funding allocation.

Results show that the PoP among all the listings in our selected dataset is 69.37%, while the PoP among the funded listings of our TSM allocation is 73.32%, which is no lower than the actual PoP. This means that our TSM framework is able to gain a higher total surplus and benefit the social good without hurting the safety experience of the system.

5.6 Online Freelancing

We used the dataset from Zhubajie[3](ZBJ) for experimental verification of online freelancing applications. ZBJ is a famous Chinese online marketplace website that includes online jobs across various categories. Each employment record includes the employer, freelancer, and job IDs, the hourly salary, as well as the employer-job and freelancer-job ratings, which are integers ranging from 0 to 5. Some of the basic statistics of the dataset that we collected are summarized in Table 7.

Similar to experiments on e-commerce, we make job recommendations to freelancers based on the allocation matrix

[3]http://www.zbj.com

Table 7: Some key statistics of the ZBJ dataset.

#Employers	#Freelancers	#Jobs	AverageSalary
40,228	46,856	296,453	¥21.68/hr
#Employer Ratings	#Freelancer Ratings	Average Employer Rating	Average Freelancer Rating
276,103	241,638	2.336	2.405

produced by our framework, then verify the performance on this task. To do so, we take all the freelancer-job ratings, and conduct personalized recommendation based on Collaborative Filtering (CF). In CF, a job g_j is assigned to freelancer u_i who has the highest predicted rating $\hat{r}_{ij}$, while in our Total Surplus Maximization (TSM) framework, it is assigned to the freelancer where $Q_{ij} = 1$ according to Eq.(29).

We conduct five-fold cross-validation for both methods, and we still adopt the Conversion Rate (CR) for performance evaluation, which is the percentage of properly assigned jobs in the testing dataset. Results of TSM and CF methods are presented in Table 8, under different choices of the number of latent factors K used for rating prediction (see Eq.(6)).

Table 8: Conversion rate on job recommendation.

K	5	10	20	30	40	50
CF(%)	0.165	0.216	0.244	0.258	0.262	0.266
TSM(%)	0.384	0.421	0.453	0.486	0.507	0.512

Results show that our TSM framework gains consistently better performance on conversion rate for job recommendation. The improvements are significant at 0.01 level for all choices of latent factors K. According to the discussions in Section 4.4, the improvement comes from the inherent consideration of salary rate in our model, which implies that the salary could be an extremely important factor when freelancers seek for jobs. Besides, we see that the results tend to be stable when $K \geq 40$ for both methods, which means that a dimensionality of 40 could be sufficiently enough to describe the factors considered by freelancers.

We further calculate the total surplus for the allocations given by CF and TSM under different choices of K's. Once an arbitrary allocation $Q = [Q_{ij}]_{m\times n}$ is realized in practice, we obtain the total surplus as:

$$TS_{Fr} = \sum_i \sum_j \left(h(\hat{r}_{ij}) + h(\hat{r}_{kj})\right) s_j Q_{ij} \tag{33}$$

We calculate the total surplus for each of the five testing folds, where there are 59,291 job allocations on average in each fold. Finally, the averaged total surplus among the five folds are shown in Table 9, where the surplus is measured in CNY (¥) and 'm' is for 'million'.

Table 9: Total surplus of online freelancing job allocations under typical choices of latent factor K.

K	5	10	20	30	ActualAllocation
CF(¥)	1.562m	1.758m	1.824m	1.860m	2,593,618
TSM(¥)	3.235m	3.862m	4.270m	4.336m	

The improvements on total surplus are significant at 0.001 level for all choices of K. We see that our TSM framework consistently gains more surplus than CF. It even leads to more surplus than the actual surplus of the testing dataset. The TSM framework gains a total surplus of ¥73.13/job on a job-level when $K = 30$, while that for the CF approach and the actual allocation are ¥31.37/job and ¥43.74/job, respectively.

The fact that the total surplus of the actual allocation is less than that gained by our TSM approach implies the failure of market equilibrium, which is frequently observed by economists in the research of antitrust and market regulations. For online freelancing as an example, this comes from the problem of information asymmetry between freelancers and employers, because it could be impossible for the freelancers to browse millions of jobs to make a final decision. This further stresses the importance of personalized recommendation techniques in service allocation, which help to push the appropriate jobs to freelancers, so as to overcome the problem of information overload.

When putting the evaluation results on total surplus and recommendation together, we find it extremely exciting because our TSM framework leads to better market efficiency even than the practical market of the system, while at the same time it benefits the freelancers with more acceptable job recommendations. This means that our allocation solution may well be applied in practice for a better off in online markets compared with current actually adopted recommendation techniques.

6. RELATED WORK

In mainstream economics, economic surplus [14, 10, 1], also known as total welfare or Marshallian surplus [25] (named after Alfred Marshall), refers to three closely related quantities: consumer surplus, producer surplus, and social/total surplus, where social surplus is the sum of surpluses experienced by both consumers and producers. The research of surplus has had quite a long history in the progress of economical theories, dating back to as early as the 19th century with the initial understandings of Surplus Values [26, 27], when the gigantic increase in wealth and population brought by the First and Second Industrial Revolution drove economists to investigate the nature of economical increase [33].

A milestone comes later when Economist Paul A. Baran re-introduced the concept of Economic Surplus [1] in a modern supply-demand analysis framework, and he further elaborated the importance of this innovation and its consistency with the labor concept of value, as well as its supplementary relation to surplus values [2].

In modern economics, the concept of social surplus has been widely adopted by economists for economic system analysis and mechanism design, usually as a direct measure of *social good* to benefit the good of our human society [14, 28, 3]. However, although the Web has formed itself as a virtual society by continuously integrating the human activities from offline to online, the research community still has seldomly investigated the surplus nature of the Web as a social system.

Actually, a large number of Web-based services can be formalized as consumer-producer interaction systems, including the most commonly used E-commerce websites [23], online financing [22, 7], crowd-sourcing systems [11, 5], and even social networks [32, 15], where the consumers consume normal goods, financial products, freelancing jobs, or information from the corresponding producers therein.

These applications raise the practical problem of matching services from producers to consumers. Perhaps the most closely related tasks for such matching processes are Personalized Recommendation [12, 30, 16] and Search [4, 24], which feed the implicit or explicit needs of the users with recommendations and search results.

However, current approaches for such tasks mainly focus on the benefits of one side without explicitly modeling the benefits of the Web system as a whole. For example, the widely adopted Collaborative Filtering (CF) [12, 17, 37, 35, 29, 34] techniques for personalized recommendation inherently focus on the maximization of consumer satisfaction based on their preferences. Although the satisfaction of consumers intuitionally benefits the surplus of producers by improving the potential of user clicks, there is no direct guarantee that such a single-side oriented modeling can benefit both sides. In this work, however, we view the Web as a virtual society and propose to maximize the social surplus directly, based on well-developed and widely-accepted economic concepts and conclusions, which, to the best of our knowledge, is the first time to do so in the context of web-based applications.

7. CONCLUSIONS AND FUTURE WORK

Most existing literature on recommender systems focuses on developing new algorithms for standard evaluation metric such as RMSE, conversion rate or click through rate. There is little research on some fundamental questions, such as what metrics should be used to evaluate recommender systems and to what extent do the metrics reflect the goals of users, producers, platform providers, and the overall Web economy.

This paper is our first step towards finding principled answers to these questions based on established economic theory. Considering a recommender system as an information agent to support two-sided matching tasks, we introduce established economic surplus theory into recommender systems and meld it with recent data driven algorithmic approaches. Our proposed Total Surplus Maximization framework integrates the goals of users and suppliers, which can be a good metric to optimize for platform providers as it better reflects the overall economy of the online system. We have illustrated how to realize this framework for different recommendation systems. The experimental results and further analyze on multiple industry data demonstrated the effectiveness of the proposed framework.

This paper focuses on the broadest metric about efficiency, or maximization of total surplus, and this inherent principle is not restricted to recommendation tasks that we primarily investigated, but applicable to the whole research effort of web intelligence for social good. In the future, we will also examine performance metrics about its two major components: producer surplus and consumer surplus. We can also try the ideas on new datasets, compare different functional form and specifications of utilities and profits. We will implement it both in static (one-time) and dynamic (multi-period/session/page) recommendation or search settings, and evaluate with real users to see the short term and long term impact of the total surplus based framework.

Acknowledgement

This work was sponsored by the National Science Foundation under grant number CCF-1101741 and IIS-0953908 of the US, as well as the National Key Basic Research Program (2015CB358700) and Natural Science Foundation (61532011, 61472206) of China. Any opinions, findings, conclusions, or

recommendations expressed in this paper are the authors, and do not necessarily reflect those of the sponsors.

8. REFERENCES

[1] P. A. Baran. Economic Progress and Economic Surplus. *Science & Society*, pages 289–317, 1953.

[2] P. A. Baran. Monopoly Capital. *NYU Press*, 1966.

[3] T. Borgers, D. Krahmer, and R. Strausz. An Introduction to the Theory of Mechanism Design. *Oxford University Press*, 2015.

[4] S. Brin and L. Page. Reprint of: The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks*, 56(18):3825–3833, 2012.

[5] M. Buhrmester, T. Kwang, and S. D. Gosling. Amazon's Mechanical Turk: a New Source of Inexpensive, yet High-quality, Data? *Perspectives on psychological science*, 6(1):3–5, 2011.

[6] S. Ceyhan, X. Shi, and J. Leskovec. Dynamics of Bidding in a P2P Lending Service: Effects of Herding and Predicting Loan Success. *WWW*, pages 547–556, 2011.

[7] S. Ceyhan, X. Shi, and J. Leskovec. Dynamics of Bidding in a P2P Lending Service: Effects of Herding and Predicting Loan Success. *WWW*, pages 547–556, 2011.

[8] D. Chaffey, F. Ellis-Chadwick, R. Mayer, and K. Johnston. Internet Marketing: Strategy, Implementation and Practice. *Pearson Education*, 2009.

[9] J. S. Coleman and T. J. Fararo. Rational Choice Theory. *Nueva York: Sage*, 1992.

[10] J. M. Currie, J. A. Murphy, and A. Schmitz. The Concept of Economic Surplus and its Use in Economic Analysis. *The Economic Journal*, pages 741–799, 1971.

[11] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing Systems on the World-Wide Web. *Communications of the ACM*, 54:86–96, 2011.

[12] M. D. Ekstrand, J. T. Riedl, and J. A. Konstan. Collaborative Filtering Recommender Systems. *Foundations and Trends in Human-Computer Interaction*, 4(2):81–173, 2011.

[13] D. Friedman and S. József. Tractable Consumer Choice. *Theory and Decision*, pages 1–26, 2013.

[14] J. Gans, S. King, and N. G. Mankiw. Principles of Microeconomics. *Cengage Learning*, 2011.

[15] A. M. Kaplan and M. Haenlein. Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 53(1):59–68, 2010.

[16] J. A. Konstan and J. Riedl. Recommender Systems: from Algorithms to User Experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.

[17] Y. Koren, R. Bell, and C. Volinsky. Matrix Factorization Techniques for Recommender Systems. *Computer*, 2009.

[18] D. Lee and H. Seung. Learning the parts of objects with nonnegative matrix factorization. *Nature*, 401(6755):788–791, 1999.

[19] D. D. Lee and H. S. Seung. Algorithms for Non-negative Matrix Factorization. *NIPS*, pages 556–562, 2001.

[20] A. Levi, O. Mokryn, C. Diot, and N. Taft. Finding a Needle in a Haystack of Reviews: Cold Start Context-Based Recommender System. *RecSys*, pages 115–122, 2012.

[21] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades. Facing the Cold Start Problem in Recommender Systems. *Expert Systems with Applications*, 41(4):2065–2073, 2014.

[22] C. Luo, H. Xiong, W. Zhou, Y. Guo, and G. Deng. Enhancing Investment Decisions in P2P Lending :An Investor Composition Perspective. *KDD*, pages 292–300, 2011.

[23] B. Mahadevan. Business Models for Internet-based E-commerce. *California Management Review*, 42(4):55–69, 2000.

[24] C. D. Manning, P. Raghavan, and H. Schütze. Introduction to Information Retrieval. *Cambridge: Cambridge university press*, 1, 2008.

[25] A. Marshall and C. W. Guillebaud. Principles of Economics. *World Scientific*, 1961.

[26] K. Marx. *Wages, Prices and Value Added*. 1865.

[27] K. Marx, G. A. Bonner, and E. Burns. Theories of Surplus Value. *Capital*, Volume IV, 1863.

[28] E. S. Maskin. Mechanism Design: How to Implement Social Goals. *The American Economic Review*, pages 567–576, 2008.

[29] J. D. M. Rennie and N. Srebro. Fast maximum margin matrix factorization for collaborative prediction. *ICML*, pages 713–719, 2005.

[30] F. Ricci, L. Rokach, and B. Shapira. Introduction to Recommender Systems Handbook. *Springer US*, 2011.

[31] M. C. Rios, C. R. McConnell, and S. L. Brue. *Economics: Principles, problems, and policies.* McGraw-Hill, 2013.

[32] G. Ritzer and N. Jurgenson. Production, Consumption, Prosumption: The Nature of Capitalism in the Age of the Digital Prosumer. *Journal of Consumer Culture*, 10(1):13–36, 2010.

[33] M. Rothbard. An Austrian Perspective on the History of Economic Thought Volume II: Classical Economics. 1995.

[34] R. Salakhutdinov and A. Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. *ICML*, pages 880–887, 2008.

[35] B. Sarwar, G. Karypis, et al. Item-based collaborative filtering recommendation algorithms. *WWW*, pages 285–295, 2001.

[36] X. Su and T. M. Khoshgoftaar. A Survey of Collaborative Filtering Techniques. *Advances in Artificial Intelligence*, 4, 2009.

[37] G. Takacs, I. Pilaszy, B. Nemeth, and D. Tikk. Investigation of Various Matrix Factorization Methods for Large Recommender Systems. *Proc. ICDM*, 2008.

[38] Y. Zhang, M. Zhang, Y. Zhang, Y. Liu, and S. Ma. Understanding the Sparsity: Augmented Matrix Factorization with Sampled Constraints on Unobservables. *CIKM*, 2014.

[39] Q. Zhao, Y. Zhang, D. Friedman, and F. Tan. E-commerce Recommendation with Personalized Promotion. *RecSys*, pages 219–226, 2015.