Article

Matching Users' Preference under Target Revenue Constraints in Data Recommendation Systems

Shanyun Liu [1], Yunquan Dong [2],* , Pingyi Fan [1],* , Rui She [1] and Shuo Wan [1]

[1] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China;
liushany16@mails.tsinghua.edu.cn (S.L.); sher15@mails.tsinghua.edu.cn (R.S.);
wan-s17@mails.tsinghua.edu.cn (S.W.)

[2] School of Electronic and Information Engineering, Nanjing University of Information Science and
Technology, Nanjing 210044, China

* Correspondence: yunquandong@nuist.edu.cn (Y.D.); fpy@tsinghua.edu.cn (P.F.);
Tel.: +86-025-5873-1196 (Y.D.); +86-010-6279-6973 (P.F.)

check for
updates

Abstract: This paper focuses on the problem of finding a particular data recommendation strategy based on the user preference and a system expected revenue. To this end, we formulate this problem as an optimization by designing the recommendation mechanism as close to the user behavior as possible with a certain revenue constraint. In fact, the optimal recommendation distribution is the one that is the closest to the utility distribution in the sense of relative entropy and satisfies expected revenue. We show that the optimal recommendation distribution follows the same form as the message importance measure (MIM) if the target revenue is reasonable, i.e., neither too small nor too large. Therefore, the optimal recommendation distribution can be regarded as the normalized MIM, where the parameter, called importance coefficient, presents the concern of the system and switches the attention of the system over data sets with different occurring probability. By adjusting the importance coefficient, our MIM based framework of data recommendation can then be applied to systems with various system requirements and data distributions. Therefore, the obtained results illustrate the physical meaning of MIM from the data recommendation perspective and validate the rationality of MIM in one aspect.

Keywords: data recommendation; optimal recommendation distribution; utility distribution; message importance measure; importance coefficient

1. Introduction

Intelligent data recommendation has been one of the most fundamental techniques in wireless mobile Internet, and becomes more and more crucial in the era of big data. Note that, with the explosive growth of data, it is difficult to send all the data to the users within a tolerable time with traditional data processing technology [1,2]. By sending data in advance when the system is idle other than wasting time to wait for a clear request from users, the delay for users to acquire their desired data could be reduced significantly [3]. It is also arduous for users to find desired data among the mass of data available in the Internet [4]. In general, users can access their interested data faster and easier with the help of data recommendation systems, since they are usually well-designed based on the preferences of users [5,6]. Compared to search engines, recommendation systems are more convenient since they are less action/skill/knowledge demanding [4]. For example, many applications in mobile phones prefer to recommend some data based on the user's interests to improve user experience. In addition, the Internet enterprises can make a profit through mobile networks by using push-based advertisement technology [7].

Most previous works mainly discussed data push based on the content to solve the problem of data delivery [3–10]. Data push is usually regarded as a strategy of data delivery in distributed information systems [3]. The architecture for a mobile push system was proposed in [4,8]. In addition, Ref. [9] put forward an effective wireless push system for high-speed data broadcasting. The push and pull techniques for time-varying data network was discussed in [10]. Furthermore, on this basis, various content-based recommendation systems were provided as information-filtering systems which push data to users based on the knowledge about their preferences [11,12]. Ref. [11] discussed the joint content recommendation, and the privacy-enhancing technology in recommendation systems was investigated in [12]. In addition, Ref. [13] put forward a personalized social image recommendation method. Furthermore, the recommendation technology was also used to solve the problem in multimedia big data [14]. Recently, user customization becomes crucial. The personalized concern of users usually can be characterized by many data properties, such as data format and keywords [4,15,16]. Instead of discussing data delivery based on content, as an alternative, we shall discuss the distribution of recommendation sequence based on users' preference when the revenue of the recommendation system is required. In particular, we choose the frequency of data used to describe the preference of users since it has nothing to do with the concrete content.

Furthermore, we take the revenue of the recommendation system into account. That is, a recommendation process consumes resources while getting benefits. Different recommendations may bring different results. For example, the cost of omitting data that should be pushed may be much smaller than that of pushing invalid data to a user erroneously. On one hand, the users would rate the recommendation system higher if the desired data is correctly pushed. On the other hand, pushing the information that users do not ask, such as advertisements, may seriously degrade the user experience. Nevertheless, pushing advertisements does bring higher revenue. To balance the loss of different types of push errors, we shall weigh different types of push errors differently, which is similar to cost-sensitive learning [17–20].

For different application scenarios, the system focuses on different events according to the need. For example, the small-probability events captures our attention in the minority subsets detection [21,22], while someone prefers the event with high probability in support vector machine (SVM) [23]. For applications in the wireless mobile Internet, in the stage of user expansion, recommending the desired data accurately to attract more users is more important. However, in the mature stage of applications, more advertisements would be pushed to earn more, although it may degrade the user experience at some extent. This paper will mainly discuss both of these two cases.

We also present some new findings by employing the message importance [24–28], which characterizes the concern degree of events. Message importance measure (MIM) was proposed to characterize the message importance of events which can be described by discrete random variables where people pay most attention to the small-probability ones, and it highlights the importance of minority subsets [29]. In fact, it is an extension of Shannon entropy [30,31] and Rényi entropy [32] from the perspective of Fadeev's postulates [32,33]. That is, the first three postulates are satisfied for all of them, and MIM weakens the fourth postulates on the foundation of Rényi entropy. Moreover, the logarithmic form and polynomial form are adopted in Shannon entropy and Rényi entropy, respectively, while MIM uses the exponential form. Ref. [34] showed that MIM focuses on a specific event by choosing a corresponding importance coefficient. In fact, MIM has a wide range of applications in big data, such as compressed storage and communication [35] and mobile edge computing [36].

Note that a superior recommendation sequence should resemble that generated by the user himself, which means that the data recommendation mechanism agrees with user behavior. To this end, the probability distribution of the recommendation sequence and the utilization frequency of the user data should be as close as possible in the statistical sense. According to [37], this means that the relative entropy between the distribution of recommendation sequence and that of user data should be minimized. We assume that the recommendation model pursues best user performance with a certain revenue guarantee in this paper.

In this paper, we first find a particular recommendation distribution on the best effort in maximizing the probability of observing the recommendation sequence with the utilization frequency of the user data when the expected revenue is provided. Then, its main properties, such as monotonicity and geometrical characteristic, are fully discussed. This optimal recommendation system can be regarded as an information-filtering system, and the importance coefficient determines what events the system prefers to recommend. The results also show that excessively low expectation of revenue can not constrain recommendation distribution and exorbitant expectation of revenue makes the recommendation system impossible to design. The constraints on the recommendation distribution are true if the minimum average revenue is neither too small nor too large, and there is a trade-off between the recommendation accuracy and the expected revenue.

It is also noted that the form of this optimal recommendation distribution is the same as that of MIM when the minimum average revenue is neither too small nor too large. The optimal recommendation distribution is determined by the proportion of recommendation value of the corresponding event in total recommendation value, where the recommendation value is a special weight factor. The recommendation value can be seen as a measure of message importance, since it satisfies the postulates of the message importance. Due to the same form with MIM, the optimal recommendation probability can be given by the normalized message importance measure when MIM is used to characterize the concern of system. Furthermore, when importance coefficient is positive, the small-probability events will be given more attention, which is magnifying the importance index of small-probability events and lessening that of high-probability events. Therefore, we confirm the rationality of MIM from another perspective in this paper for characterizing the physical meaning of MIM by using data recommendation system rather than based on information theory. In addition, we expand MIM to the general case whatever the probability of systems interested events is. Since the importance coefficient determines what events set systems are interested in, we can switch to different application scenarios by means of it. That is, advertising systems are discussed if the importance coefficient is positive, while noncommercial systems are adopted if importance coefficient is negative. Compared with previous works about MIM [29,34,35], most properties of optimal recommendation distribution is the same, but a clear definition of the desired event set can be given in this paper. The relationship between utility distribution and MIM was preliminarily discussed in [38].

The main contributions of this paper can be summarized as follows. (1) We put forward an optimal recommendation distribution that makes the recommendation mechanism agree with user behavior with a certain revenue guarantee, which can improve the design of recommendation strategy. (2) We illuminate that this optimal recommendation distribution is normalized message importance, when we use MIM to characterize the concern of systems, which presents a new physical explanation of MIM from data recommendation perspective. (3) We expand MIM to the general case, and we also discuss the importance coefficient selection as well as its relationship with what events systems focus on.

The rest of this paper is organized as follows. The setup of optimal recommendation is introduced in Section 2, including the system model and the discussion of constraints. In Section 3, we solve the problem of optimal recommendation in our system model, and give complete solutions. Section 4 investigates the properties of this optimal recommendation distribution. The geometric interpretation is also discussed in this part. Then, we discuss the relationship between this optimal recommendation distribution and MIM in Section 5. It is noted that recommendation distribution can be seen as normalized message importance in this case. The numerical results is shown and discussed to certificate our theoretical results in Section 6. Section 7 concludes the paper. In addition, the main notations in this paper are listed in Table 1.

Table 1. Notations.

Notation	Description
S	The set of all the data
N	The number of data classes
S_i	The set of data belonging to the i-th class $S_i \cap S_j = \varnothing$ for $i \neq j$, and $S = S_1 \cup S_2 \cup ...S_{N-1} \cup S_N$
$Q = \{q_1, q_2, ..., q_N\}$	Raw distribution: the probability distribution of information source
$P = \{r_1, r_2, ..., r_N\}$	Recommendation distribution: the probability distribution of recommended data
$U = \{u_1, u_2, ..., u_N\}$	Utility distribution: the probability distribution of user's preferred data
$D(P \parallel U)$	The relative entropy or Kullback–Leibler (KL) distance between P and U
C_p	The cost of a single data push
R_p	The earning when the pushed data is liked by the user
C_n	The cost when the pushed data is not liked by the user
R_{ad}	The advertising revenue when the pushed data is not liked by the user
C_m	The cost missing to push a piece of user's desired data
β	The target revenue of a single data push
ϖ	The importance coefficient
α	$\alpha = \frac{\beta + C_p - R_p}{R_{ad} - R_p - C_n - C_m}$
γ_u	$\gamma_u = \sum_{i=1}^{N} u_i^2$
$g(\varpi, V)$	$g(\varpi, P) = \frac{\sum_{i=1}^{N} v_i^2 e^{\varpi(1-v_i)}}{\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}}$
$f(\varpi, x, U)$	$f(\varpi, x, U) = \frac{u_x e^{\varpi(1-u_x)}}{\sum_{i=1}^{N} u_i e^{\varpi(1-u_i)}}$
x	$x = 1, 2, \cdots, N$ is the index of classes

2. System Model

We consider a recommendation system with N classes of data, as shown in Figure 1. In fact, data is often stored based on its categories for the convenience of indexing. For example, the news website usually classifies news into the following categories: politics, entertainment, business, scientific, sports, and so on. At each time instant, the information source generates a piece of data, which belongs to a certain class with probability distribution Q. In general, the generated data sequence does not match the preference of the user. To optimize the information transmission process, therefore, a recommendation unit is used to determine whether the generated data should be pushed to the user with some deliberately designed probability distribution P. One the one hand, the recommendation unit can make predictions of the user's needs and push some data to the user before he actually starts the retrieval process. In doing so, the transmission delay can be largely reduced, especially when the data amount is large. On the other hand, the recommending unit enables non-expert users to search and to access their desired data much easier. Furthermore, we can profit more by pushing some advertisements to the user.
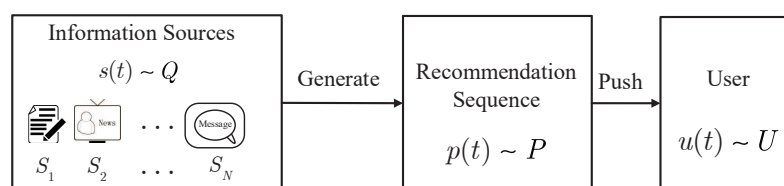

Figure 1. System model.

2.1. Data Model

We refer to the empirical probability mass function of the class indexes over the whole data set as the *raw distribution* and denote it as $Q = \{q_1, q_2, ..., q_N\}$. We refer to the probability mass function of users' preferences over the classes as the *utility distribution* and denote it as $U = \{u_1, u_2, ..., u_N\}$. That is, for each piece of data, it belongs to class i with probability q_i and would be preferred by the user with probability u_i. To fit the preference of the user under some target revenue constraint, the system will make random recommendations according to a *recommendation distribution* $P = \{p_1, p_2, ..., p_N\}$.

We assume that each piece of data belongs to one and only one of the N sets. That is, $S_i \cap S_j = \varnothing$ for $\forall i \neq j$, where S_i is the set of data belonging to the i-th class. Thus, the whole data set would be $S = S_1 \cup S_2 \cup ... S_{N-1} \cup S_N$. The raw distribution can be expressed as $q_i = \Pr\{d \in S_i\} = \mathrm{crad}(S_i)/\mathrm{crad}(S)$. In addition, the utility distribution U can be obtained by studying the data-using behavior of a specific group of users and thus, in this paper, is assumed to be known prior.

For traditional data push, we usually expect to make $|u(t) - s(t)|$ smaller than a given value [10]. Different from them, we do not consider this problem based on content. As an alternative, our goal is to find the optimal recommendation distribution P so that the recommended data would fit the preference of the user as much as possible. To be specific, each recommended sequence of data should resemble the desired data sequence of the user in the statistical sense. For a sequence of user's favorite data, let u^n be the corresponding class indexes. As n goes to infinity, it is clear that $u^n \in T(U)$ with probability one, where $T(U)$ is the typical set under distribution U. That is, $\Pr\{\frac{1}{n}\log\Pr(u^n) + H(U) = 0\} = 1$, where $\Pr(u^n)$ is the occurring probability of u^n and $H(U)$ is the joint entropy of U [37]. Since the class-index sequence r^n of recommended data is actually generated with distribution P, the probability that r^n falls in the typical set $T(U)$ of distribution U would be $\Pr\{r^n \in T(U)\} \doteq 2^{-nD(P\|U)}$, where $D(P \| U)$ is the relative entropy between P and U [37]. It is clear that the optimal P would maximizes probability $\Pr\{r^n \in T(U)\}$, which is equivalent to minimizing the relative entropy

$$D(P \| U). \tag{1}$$

In particular, our desired recommendation distribution P is not exactly the same as the utility distribution of the user because we would also like to intentionally push some advertisements to the users to increase our profit.

2.2. Revenue Model

We assume that the user divides the whole data set into two parts, i.e., the desired ones and the unwanted ones (e.g., advertisements). At the recommendation unit, the data can also be classified into two types according to whether it is recommended to the user. Different push types may strikingly lead to different results. For example, the cost of omitting data that should be pushed may be much smaller than that of erroneously pushing invalid data to a user. The user experience will be enhanced if the data needed by users is correctly pushed. Pushing some not needed content to users, such as some advertisements, may seriously degrade the user experience, but it still can bring in advertising revenue for the content delivery enterprise. Using a similar revenue model as that in cost-sensitive learning [17–20], we shall evaluate the revenue of the recommendation system as follows:

- The cost of making a recommendation is C_p;
- The revenue of a recommendation when the pushed data is liked by the user is R_p;
- The cost of a recommendation when the pushed data is not liked is C_n;
- The revenue of a recommendation when the pushed data is not liked (but can serve as an advertisement) is R_{ad};
- The cost of missing to recommend a piece of desired data of the user is C_m;

Therefore, the revenue of recommending a piece of data belonging to class i can be summarized in Table 2.

Table 2. The revenue matrix.

Preference \ Action	Recommend	Not Recommend
Desired	$R_p - C_p$	$-C_m$
Unwanted	$R_{ad} - C_p - C_n$	0

Moreover, the corresponding matrix of occurring probability is given by Table 3.

Table 3. The matrix of occurring probability.

Preference \ Action	Recommend	Not Recommend
Desired	$p_i u_i$	$(1 - p_i) u_i$
Unwanted	$p_i (1 - u_i)$	$(1 - p_i)(1 - u_i)$

In this paper, we assume $C_p, R_p, C_n, R_{ad}, C_m$ are constraints for a given recommendation system for simplicity. The expected system revenue can then be expressed as

$$\bar{R}(P) = \sum_{i=1}^{N} \left((R_p - C_p) p_i u_i + (R_{ad} - C_p - C_n) p_i (1 - u_i) - C_m (1 - p_i) u_i \right) \tag{2}$$

$$= -(R_p + C_n + C_m - R_{ad})\left(1 - \sum_{i=1}^{N} p_i u_i\right) + R_p - C_p \tag{3}$$

$$= -(R_p + C_n + C_m - R_{ad}) \sum_{i=1}^{N} p_i (1 - u_i) + R_p - C_p. \tag{4}$$

2.3. Problem Formulation

In this paper, we consider the following three kinds of recommendation systems: (1) the *adverting system* where recommending unwanted advertisements yields higher revenue; (2) the *noncommercial system* where recommending user's desired data brings higher revenue; (3) the *neutral system* where the system revenue is independent from recommendation probability P. For each given target revenue constraint $\bar{R}(P) \geq \beta$ and each kind of system, we shall optimize the recommendation distribution P by minimizing $D(P||U)$. In particular, the following auxiliary variable is used:

$$\alpha = \frac{\beta + C_p - R_p}{R_{ad} - R_p - C_n - C_m}. \tag{5}$$

2.3.1. Advertising Systems

In an *advertising system*, recommending a piece of user's unwanted data (an advertisement) yields higher revenue. Since the revenue of recommending an advertisement is the main source of income in this case, which is larger than other revenue and cost, the advertising system would satisfy the following condition:

$$\text{C1}: \quad R_p + C_n + C_m - R_{ad} < 0. \tag{6}$$

By combining Labels (4)–(6), it is clear that the constraint $\bar{R}(P) \geq \beta$ is equivalent to

$$\sum_{i=1}^{N} p_i (1 - u_i) \geq \alpha. \tag{7}$$

For advertising systems, therefore, the feasible set of recommendation distribution P can be expressed as

$$E_1 = \left\{ P : \sum_{i=1}^{N} p_i(1 - u_i) \geq \alpha \,\middle|\, R_{ad} > R_p + C_n + C_m \right\}. \tag{8}$$

For a given target revenue β, we shall solve the optimal recommendation distribution P of advertising systems from

$$\mathcal{P}_1 : \ \arg\min_{P \in E_1} \ D(P \| U) \tag{9}$$

$$\text{s.t.} \ \sum_{i=1}^{N} p_i(1 - u_i) \geq \alpha \tag{9a}$$

$$\sum_{i=1}^{N} p_i = 1. \tag{9b}$$

2.3.2. Noncommercial Systems

A *noncommercial system* is defined as a recommendation system where the revenue $R_p - C_p$ of recommending a piece of desired data is larger than the sum of revenue $R_{ad} - C_p - C_n$ of recommending an advertisement and the cost of not recommending a piece of desired data C_m. That is,

$$\text{C2} : \ R_p + C_n + C_m - R_{ad} > 0. \tag{10}$$

Accordingly, the constraint $\bar{R}(P) \geq \beta$ is equivalent to

$$\sum_{i=1}^{N} p_i(1 - u_i) \leq \alpha. \tag{11}$$

Therefore, the feasible set of recommendation distribution P for noncommercial systems can be expressed as

$$E_2 = \left\{ P : \sum_{i=1}^{N} p_i(1 - u_i) \leq \alpha \,\middle|\, R_{ad} < R_p + C_n + C_m \right\}. \tag{12}$$

Afterwards, we can solve the optimal recommendation distribution P through the following optimization problem:

$$\mathcal{P}_2 : \ \arg\min_{P \in E_2} \ D(P \| U) \tag{13}$$

$$\text{s.t.} \ \sum_{i=1}^{N} p_i(1 - u_i) \leq \alpha \tag{13a}$$

$$\sum_{i=1}^{N} p_i = 1. \tag{13b}$$

2.3.3. Neutral Systems

For the case $R_p + C_n + C_m - R_{ad} = 0$, the corresponding expected system revenue degrades to

$$\bar{R}(P) = R_p - C_p \tag{14}$$

and is independent from the recommendation distribution P. As long as the target revenue satisfies $\beta < R_p - C_p$, the constraint $\bar{R}(P) \geq \beta$ can be met by any recommendation distribution. Therefore, the recommendation distribution can be chosen as $P = U$.

2.3.4. Discussion of Systems

We note that in the scenario of the wireless mobile Internet, each new application needs to attract more users (i.e., grab larger market share) with excellent user experience in its early stage. After the market share and the user groups being stable, the application can earn money by pushing some advertisements at the cost of some degradations in user experience.

Noncommercial systems usually appear in the stage of user expanding. In order to enlarge market share, the main tasks of this stage are to attract more users and to convince them of the recommendation systems with excellent user experience. Therefore, the revenue by recommending desired data of users should be larger than that of recommending advertisements. To be specific, the desired data of users would be recommended with higher probability to increase the target revenue $\bar{R}(P)$. Since $\bar{R}(P)$ is decreasing with $\sum_{i=1}^{N} p_i(1 - u_i)$, those high-probability events are more important in this case.

Advertising systems are usually adopted in the mature stage of applications where users have been accustomed to the recommendation system and some advertisements are acceptable. Therefore, the applications would increase the revenue of pushing advertisements and the actual number of advertisement recommendations. To be specific, the desired data of users should be recommended with relatively small-probability while advertisements should be recommended with higher probability. In this sense, the small-probability events are more important here.

Remark 1. *Since* $\left\{ P : \bar{R}(P) \geq \beta \right\} = E_1 \cup E_2 \cup \left\{ P : \bar{R}(P) \geq \beta | R_{ad} = R_p + C_n + C_m \right\}$, *these three kinds of recommendation systems cover all the cases of this problem.*

Different from the other recommendation systems based on content [11,13,14], the recommendation systems in this paper discuss the distribution of recommendations based on the preference of users and the target revenue of the recommendation system. Thus, we have focused on the integrated planning of a sequence of recommendations other than the recommendation of a specific piece of data.

3. Optimal Recommendation Distribution

In this part, we shall present the optimal recommendation distribution for both advertising systems and noncommercial systems explicitly. We define an auxiliary variable and an auxiliary function as follows:

$$\gamma_u = \sum_{i=1}^{N} u_i^2, \tag{15}$$

$$g(\varpi, V) = \frac{\sum_{i=1}^{N} v_i^2 e^{\varpi(1-v_i)}}{\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}}, \tag{16}$$

where $\varpi \in (-\infty, +\infty)$ is a constant and $V = \{v_1, v_2, \cdots, v_N\}$ is a general probability mass function. Actually, we have $\gamma_u = e^{-H_2(U)}$ where $H_2(U)$ is the Rényi entropy $H_\alpha(\cdot)$ when $\alpha = 2$ [39].

In particular, we have the following lemma on $g(\varpi, V)$.

Lemma 1. *Function $g(\varpi, V)$ is monotonically decreasing with ϖ.*

Proof of Lemma 1. Refer to Appendix A. □

Lemma 2. $g(0, V) = \sum_{i=1}^{N} v_i^2$, $g(-\infty, V) = v_{\max}$, *and* $g(+\infty, V) = v_{\min}$, *where* $v_{\max} = \max\{v_1, v_2, \cdots, v_N\}$ *and* $v_{\min} = \min\{v_1, v_2, \cdots, v_N\}$.

Proof of Lemma 2. Refer to Appendix B. □

It is clear that we have $g(0, P) = \gamma_p$ and $g(0, U) = \gamma_u$.

3.1. Optimal Advertising System

Theorem 1. *For an advertising system with $R_p + C_n + C_m - R_{ad} < 0$, the optimal recommendation distribution is the solution of Problem $\mathcal{P}_1$ and is given by*

$$
p_x^* = \begin{cases}
u_x & \text{if } \alpha \le 1 - \gamma_u \\
\dfrac{u_x e^{\varpi^*(1-u_x)}}{\sum_{i=1}^N u_i e^{\varpi^*(1-u_i)}} & \text{if } 1 - \gamma_u < \alpha \le 1 \,, \\
NaN & \text{if } \alpha > 1,
\end{cases}
\tag{17}
$$

for $1 \le x \le N$, where α is defined in Label (5), γ_u is defined in Label (15), NaN means no solution exists, and $\varpi^* > 0$ is the solution to $g(\varpi, U) = 1 - \alpha$.

Proof of Theorem 1. First, if $\alpha > 1$, no solution exists since $\sum_{i=1}^N p_i(1 - u_i)$ is always smaller than or equal to one and the constraint (9a) can never be satisfied.

Second, if $\alpha \le 1 - \gamma_u$, we have $\sum_{i=1}^N p_i(1 - u_i) \ge \sum_{i=1}^N u_i(1 - u_i) = 1 - \gamma_u \ge \alpha$. That is, the constraint (9a) could by satisfied with any distribution P. Thus, the solution to Problem $\mathcal{P}_1$ would be $P = U$, which minimizes $D(P||U)$.

Third, if $1 - \gamma_u < \alpha \le 1$, we shall solve Problem $\mathcal{P}_1$ based on the following Karush–Kuhn–Tucher (KKT) conditions:

$$
\nabla_P L(P, \lambda, \mu) = \nabla_P \left(\sum_{i=1}^N p_i \ln \frac{p_i}{u_i} + \lambda \left(\alpha - \sum_{i=1}^N p_i(1 - u_i) \right) + \mu \left(\sum_{i=1}^N p_i - 1 \right) \right) = 0,
\tag{18}
$$

$$
\lambda \left(\alpha - \sum_{i=1}^N p_i(1 - u_i) \right) = 0,
\tag{18a}
$$

$$
\sum_{i=1}^N p_i - 1 = 0,
\tag{18b}
$$

$$
\alpha - \sum_{i=1}^N p_i(1 - u_i) \le 0,
\tag{18c}
$$

$$
\lambda \ge 0.
\tag{18d}
$$

Differentiating $\nabla_P L(P, \lambda, \mu)$ with respect to p_x^* and set it to zero, we have

$$
\ln p_x^* + 1 - \ln u_x - \lambda(1 - u_x) + \mu = 0,
\tag{19}
$$

and thus $p_x^* = u_x e^{\lambda(1-u_x)-\mu-1}$. Together with constraint (18b), we further have $e^{\mu+1} = \sum_{i=1}^N u_i e^{\lambda(1-u_i)}$ and

$$
p_x^* = \frac{u_x e^{\lambda(1-u_x)}}{\sum_{i=1}^N u_i e^{\lambda(1-u_i)}},
\tag{20}
$$

where $\lambda \ge 0$ is the solution to $\sum_{i=1}^N p_i^*(1 - u_i) = \alpha$, i.e., $g(\lambda, U) = 1 - \alpha$. By substituting λ with ϖ^*, the desired result in Label (17) would be obtained.

The condition $1 - \gamma_u < \alpha \le 1$ implies $g(\varpi^*, U) = 1 - \alpha \le \gamma_u = g(0, U)$. Since $g(\varpi, U)$ has been shown to be monotonically decreasing with ϖ in Lemma 1, we then have $\varpi^* > 0$. This completes the proof of Theorem 1. $\square$

Remark 2. *We denote*

$$
\beta_0 = (1 - \gamma_u)(R_{ad} - R_p - C_n - C_m) + R_p - C_p,
\tag{21}
$$

$$
\beta_{ad} = -(R_{ad} - R_p - C_n - C_m)u_{\min} + R_{ad} - C_p - C_n - C_m,
\tag{22}
$$

and have

- $\alpha \leq 1 - \gamma_u$ *is equivalent to* $\beta \leq \beta_0$, *which means that the target revenue is low and can be achieved by pushing data according to the preference of the user exactly.*
- $1 - \gamma_u < \alpha \leq 1$ *is equivalent to* $\beta_0 < \beta \leq \beta_{ad}$, *which means that the target revenue can only be achieved when some advertisements are pushed according to probability* p_x^*.
- $1 < \alpha$ *is equivalent to* $\beta > \beta_{ad}$, *which means that the target is too high to achieve, even when advertisements are pushed with probability one.*

3.2. Optimal Noncommercial System

Theorem 2. *For a noncommercial system with* $R_p + C_n + C_m - R_{ad} > 0$, *the optimal recommendation distribution is the solution of Problem* $\mathcal{P}_2$ *and is given by*

$$p_x^* = \begin{cases} NaN & \text{if } \alpha < 0, \\ \dfrac{u_x e^{\omega^*(1-u_x)}}{\sum_{i=1}^N u_i e^{\omega^*(1-u_i)}} & \text{if } 0 \leq \alpha < 1 - \gamma_u, \\ u_x & \text{if } 1 - \gamma_u \leq \alpha, \end{cases} \tag{23}$$

for $1 \leq x \leq N$, *where* α *is defined in Label* (5), γ_u *is defined in Label* (15), *NaN means no solution exists, and* $\omega^* < 0$ *is the solution to* $g(\omega, U) = 1 - \alpha$.

Proof of Theorem 2. First, if $\alpha < 0$, no solution exists since $\sum_{i=1}^N p_i(1 - u_i)$ is positive and cannot be smaller a negative number, and thus the constraint (13b) can never be satisfied.

Second, if $\alpha \geq 1 - \gamma_u$ and we set $p_x^* = u_x$, we would have $\sum_{i=1}^N p_i^*(1 - u_i) = 1 - \gamma_u \leq \alpha$, i.e., constraint (13a) is satisfied. Since setting $p_x^* = u_x$ minimizes $D(P||U)$, $P = U$ should be the solution of Problem 2.

Third, if $0 \leq \alpha < 1 - \gamma_u$, we shall solve Problem $\mathcal{P}_2$ using the following KKT conditions:

$$\nabla_P L(P, \lambda, \mu) = \nabla_P \left(\sum_{i=1}^N p_i \ln \frac{p_i}{u_i} + \lambda \left(\sum_{i=1}^N p_i(1 - u_i) - \alpha \right) + \mu \left(\sum_{i=1}^N p_i - 1 \right) \right) = 0, \tag{24}$$

$$\lambda \left(\sum_{i=1}^N p_i(1 - u_i) - \alpha \right) = 0, \tag{24a}$$

$$\sum_{i=1}^N p_i - 1 = 0, \tag{24b}$$

$$\sum_{i=1}^N p_i(1 - u_i) - \alpha \leq 0, \tag{24c}$$

$$\lambda \geq 0. \tag{24d}$$

Differentiating $\nabla_P L(P, \lambda, \mu)$ with respect to p_x^* and set it to zero, we have

$$\ln p_x^* + 1 - \ln u_x + \lambda(1 - u_x) + \mu = 0, \tag{25}$$

and $p_x^* = u_x e^{-\lambda(1-u_x)-\mu-1}$. Together with constraint (24b), we then have $e^{\mu+1} = \sum_{i=1}^N u_i e^{-\lambda(1-u_i)}$ and

$$p_x^* = \frac{u_x e^{-\lambda(1-u_x)}}{\sum_{i=1}^N u_i e^{-\lambda(1-u_i)}}, \tag{26}$$

where $\lambda > 0$ is the solution to $\sum_{i=1}^N p_i^*(1 - u_i) = \alpha$.

By denoting $\omega^* = -\lambda$, (26) turns to be $p_x^* = \frac{u_x e^{\omega^*(1-u_x)}}{\sum_{i=1}^N u_i e^{\omega^*(1-u_i)}}$ which is the desired result in Label (23).

Moreover, the condition $0 \le \alpha < 1 - \gamma_u$ implies $g(\varpi^*, U) = 1 - \alpha \ge \gamma_u = g(0, U)$. Since $g(\varpi, U)$ is monotonically decreasing with ϖ (cf. Lemma 1), we see that $\varpi^* < 0$. Thus, Theorem 2 is proved. □

Remark 3. *We denote*

$$\beta_{no} = -(R_{ad} - R_p - C_n - C_m)u_{\max} + R_{ad} - C_p - C_n - C_m, \tag{27}$$

and have the following observations.

- *$\alpha < 0$ is equivalent to $\beta \ge \beta_{no}$, which means that the target revenue is high and cannot be achieved by any recommendation distribution.*
- *$0 \le \alpha < 1 - \gamma_u$ is equivalent to $\beta_0 < \beta \le \beta_{no}$, which means that the target revenue is not too high and the information is pushed according to probability p_x^*. The user experience is limited by the target revenue in this case.*
- *$1 - \gamma_u \le \alpha$ is equivalent to $\beta \le \beta_0$, which can be achieved by pushing data according to the preference of the user exactly.*

3.3. Short Summary

We further denote

$$\beta_{ne} = R_p - C_p, \tag{28}$$

and the optimal recommendation distributions for various systems and various target revenues can then be summarized in Table 4.

Table 4. The optimal recommendation distribution.

	Case	β	α	p_x^*
Advertising system	①	$\beta \le \beta_0$	$\alpha \le 1 - \gamma$	u_x
	②	$\beta_0 < \beta \le \beta_{ad}$	$1 - \gamma < \alpha \le 1$	$\frac{u_x e^{\varpi(1-u_x)}}{\sum_{i=1}^{N} u_i e^{\varpi(1-u_i)}}$
	③	$\beta > \beta_{ad}$	$\alpha > 1$	NaN
Neutral system	④	$\beta \le \beta_{ne}$	NaN	u_x
	⑤	$\beta > \beta_{ne}$	NaN	NaN
Noncommercial system	⑥	$\beta_0 < \beta \le \beta_{no}$	$0 \le \alpha < 1 - \gamma$	$\frac{u_x e^{\varpi(1-u_x)}}{\sum_{i=1}^{N} u_i e^{\varpi(1-u_i)}}$
	⑦	$\beta \le \beta_0$	$\alpha \ge 1 - \gamma$	u_x
	⑧	$\beta > \beta_{no}$	$\alpha < 0$	NaN

Cases ③, ⑤, and ⑧ are extreme cases where the target revenue is beyond the reach of the system. For cases ①, ④, and ⑦, the target revenue is low ($\beta < \beta_{no}$ or $\beta < \beta_{ad}$), and thus is easy to achieve. In particular, the constraints (9a) and (13a) are actually inactive. Thus, the optimal recommendation distribution would be exactly the same with the utility distribution.

Cases ② and ⑥ are more practical and meaningful due to the appropriate target revenues used. To further study the property of the optimal recommendation distribution of these two cases, the following function is introduced:

$$f(\varpi, x, U) = \frac{u_x e^{\varpi(1-u_x)}}{\sum_{i=1}^{N} u_i e^{\varpi(1-u_i)}}, \tag{29}$$

where $\varpi \in (-\infty, +\infty)$.

In doing so, the optimal recommendation distribution of cases ② and ⑥ can be jointly expressed as

$$p_x^* = f(\varpi^*, x, U), \tag{30}$$

where ω^* is the solution to $g(\omega, U) = 1 - \alpha$.

In particular, $f(\omega^*, x, U)$ presents the optimal solution of case ② if $\omega^* > 0$ and presents the solution of case ⑥ if $\omega^* < 0$. Moreover, when $\omega = 0$, we have

$$f(0, x, U) = u_x, \tag{31}$$

which can be considered as the solution to cases ①, ④, and ⑦.

4. Property of Optimal Recommendation Distributions

In this section, we shall investigate how the optimal recommendation distribution diverges with respect to the utility distribution, in various systems and under various target revenue constraints. To do so, we first study the property of function $f(\omega, x, U)$, where ω_x and $\widetilde{\omega}_x \neq 0$ are, respectively, the solution to the following equations:

$$u_x = g(\omega, U), \tag{32}$$
$$u_x = f(\omega, x, U). \tag{33}$$

4.1. Monotonicity of $f(\omega, x, U)$

Theorem 3. *$f(\omega, x, U)$ has the following properties:*

(1) *$f(\omega, x, U)$ is monotonically increasing with ω in $(-\infty, \omega_x)$;*
(2) *$f(\omega, x, U)$ is monotonically decreasing with ω in $(\omega_x, +\infty)$;*
(3) *ω_x is decreasing with u_x, i.e, $\omega_y < \omega_x$ if $u_y > u_x$;*
(4) *$\omega_x < 0$ if $u_x > \gamma_u$;*
    $\omega_x = 0$ if $u_x = \gamma_u$;
    $\omega_x > 0$ if $u_x < \gamma_u$.

Proof of Theorem 3. (1) and (2) The derivative of (ω, x, U) with respect to ω can be expressed as

$$\frac{\partial\left(f(\omega, x, U)\right)}{\partial \omega} = \frac{u_x e^{\omega(1-u_x)}(1-u_x)\sum_{i=1}^{N} u_i e^{\omega(1-u_i)} - u_x e^{\omega(1-u_x)}\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}(1-u_i)}{\left(\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}\right)^2} \tag{34}$$

$$= \frac{u_x e^{\omega(1-u_x)} \sum_{i=1}^{N} u_i(u_i - u_x)e^{\omega(1-u_i)}}{\left(\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}\right)^2}. \tag{34a}$$

Since we have $u_x e^{\omega(1-u_x)} > 0$ and $\left(\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}\right)^2 > 0$, the sign of the derivative only depends on the term $\sum_{i=1}^{N} u_i(u_i - u_x)e^{\omega(1-u_i)}$.

Lemma 1 and Lemma 2 show that $g(\omega, U)$ is monotonically decreasing with ω and ω_x is the unique solution to equation $u_x = g(\omega, U)$. For $\omega < \omega_x$, therefore, we then have $g(\omega, U) > g(\omega_x, U) = u_x$, which is equivalent to $\sum_{i=1}^{N} u_i(u_i - u_x)e^{\omega(1-u_i)} > 0$. Thus, we have $\frac{\partial(f(\omega, x, U))}{\partial \omega} > 0$ and $f(\omega, x, U)$ is increasing with ω if $\omega < \omega_x$. Likewise, it can be readily proved that $f(\omega, x, U)$ is decreasing with ω if $\omega > \omega_x$.

(3) Suppose $g(\omega_x, U) = u_x$, $g(\omega_y, U) = u_y$ and $u_x < u_y$, we would have, $g(\omega_x, U) < g(\omega_y, U)$. Since $g(\omega)$ is decreasing with ω (cf. Lemma 1), we then have $\omega_x > \omega_y$, i.e., ω_x is decreasing with u_x.

(4) First, we have $g(0, U) = \gamma_u$ by the definition of $g(\omega, U)$ (cf. (16)). Using the monotonicity of $g(\omega, U)$ with respect to ω (cf. Lemma 1), we have $\omega_x < 0$ if $u_x > \gamma_u$ and $\omega_x > 0$ if $u_x < \gamma_u$.

Thus, the proof of Theorem 3 is completed. □

In particular, according to Lemma 2, if $u_x = u_{\min}$, ω_x will approach positive infinity, while ω_x will approach negative infinity if $u_x = u_{\max}$.

Remark 4.

(1) $f(\varpi, x, U)$ *is monotonously decreasing with ϖ if $u_x = u_{\max}$;*
(2) $f(\varpi, x, U)$ *is monotonously increasing with ϖ if $u_x = u_{\min}$.*

Remark 5. *We denote*

$$\beta_x = -(R_{ad} - R_p - C_n - C_m)u_x - C_p - C_n - C_m + R_{ad}, \tag{35}$$

and the relationships between $f(\varpi, x, U)$ and β in different systems are shown as follows.

- *For advertising systems,*

  (1) $f(\varpi, x, U)$ *is monotonically decreasing with β in (β_0, β_{ad}) if $u_x \geq \gamma_u$;*
  (2) $f(\varpi, x, U)$ *is monotonically increasing with β in (β_0, β_x) and monotonically decreasing in (β_x, β_{ad}) if $u_x < \gamma_u$.*

- *For noncommercial systems,*

  (1) $f(\varpi, x, U)$ *is monotonically decreasing with β in (β_0, β_{no}) if $u_x \leq \gamma_u$;*
  (2) $f(\varpi, x, U)$ *is monotonically increasing with β in (β_0, β_x) and monotonically decreasing in (β_x, β_{no}) if $u_x < \gamma_u$.*

4.2. Discussion of Parameters

Assume that there is unique minimum $p_{\min}$ and unique maximum $p_{\max}$ in utility distribution U. Without loss of generality, let $u_1 < u_2 \leq \ldots \leq u_t \leq \gamma_u < u_{t+1} \leq \ldots \leq u_{N-1} < u_N$, and $u_{\min} = u_1$ and $u_{\max} = u_N$. $P^* = (p_1^*, p_2^*, \ldots, p_N^*)$ is used to denote the optimal recommendation distribution. In addition, we only discuss the relationship between the optimal recommendation distribution and the parameters (β and ϖ) in Cases ② and ⑥ in this part.

In fact, we have the following proposition on $\widetilde{\varpi}_x$.

Proposition 1. $\widetilde{\varpi}_x$ *has the following properties:*

(1) $\widetilde{\varpi}_x$ *exists when $u_x \neq \gamma_u$, $u_x \neq u_{\max}$ and $u_x \neq u_{\min}$;*
(2) $\widetilde{\varpi}_x < 0$ *if $u_x > \gamma_u$;*
      $\widetilde{\varpi}_x > 0$ *if $u_x < \gamma_u$;*
(3) $\widetilde{\varpi}_x$ *is decreasing with u_x, i.e, $\widetilde{\varpi}_y < \widetilde{\varpi}_x$ if $u_y > u_x$.*

Proof of Proposition 1. Refer to Appendix C. □

For convenience, we denote $\widetilde{\varpi}_1 \overset{\Delta}{=} +\infty$ and $\widetilde{\varpi}_N \overset{\Delta}{=} -\infty$.

For advertising systems, the optimal recommendation distribution is given by (17) and $\varpi^* > 0$ (cf. Theorem 1). As $\varpi^* \to +\infty$, we have

$$p_1^* = f(+\infty, 1, U) = \lim_{\varpi^* \to +\infty} \frac{u_{\min} e^{\varpi^*(1-u_{\min})}}{\sum_{i=1}^{N} p_i e^{\varpi^*(1-u_i)}} \tag{36}$$

$$= \lim_{\varpi^* \to +\infty} \frac{u_{\min} e^{\varpi^*(1-u_{\min})}}{u_{\min} e^{\varpi^*(1-u_{\min})} + \sum_{u_i \neq u_{\min}} u_i e^{\varpi^*(1-u_i)}} \tag{36a}$$

$$= \lim_{\varpi^* \to +\infty} \frac{1}{1 + \sum_{u_i \neq u_{\min}} \frac{u_i}{u_{\min}} e^{\varpi^*(u_{\min}-u_i)}} \tag{36b}$$

$$= 1. \tag{36c}$$

Obviously, $p_k^* = f(+\infty, k, U) = 0$ when $k \geq 2$. Therefore, the utility distribution P^* is $(1, 0, 0, \ldots, 0)$ here.

Based on Proposition 1, if $0 < \widetilde{\omega}_{N_3+1} < \omega^* < \widetilde{\omega}_{N_3}$ ($1 \leq N_3 \leq N-1$), we have $p_x^* > u_x$ for $1 \leq x \leq N_3$ and $p_x^* < u_x$ for $N_3 + 1 \leq x \leq N$. The amount of optimal recommendation probability which is larger than corresponding utility probability is N_3. Let $N_4 = N - N_3$. N_3 decreases with increasing of ω^* (cf. Proposition 1). In particular, if $\omega^* > \widetilde{\omega}_2 > 0$, only the recommendation probability of the event with the smallest utility probability is enlarged and that of other events is reduced, compared to the corresponding utility probability. As the parameter ω^* approaches positive infinity, the recommendation distribution will be $(1, 0, 0, ..., 0)$. In conclusion, this recommendation system is a special information-filtering system which prefers to push the small-probability events.

For noncommercial systems, we can get the optimal recommendation distribution by (23) and $\omega^* < 0$ (cf. Theorem 2). As $\omega \to -\infty$, it is noted that

$$p_N^* = f(-\infty, N, U) = \lim_{\omega \to -\infty} \frac{u_{\max} e^{\omega(1-u_{\max})}}{\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}} \tag{37}$$

$$= \lim_{\omega \to -\infty} \frac{u_{\max} e^{\omega(1-u_{\max})}}{u_{\max} e^{\omega(1-u_{\max})} + \sum_{u_i \neq u_{\max}} u_i e^{\omega(1-u_i)}} \tag{37a}$$

$$= \lim_{\omega \to -\infty} \frac{1}{1 + \sum_{u_i \neq u_{\max}} \frac{u_i}{u_{\max}} e^{\omega(u_{\max}-u_i)}} \tag{37b}$$

$$= 1, \tag{37c}$$

and $p_k^* = f(-\infty, k, U) = 0$ when $k \leq N-1$. Hence, $P^* = (0, 0, .., 0, 1)$ in this case.

If $\widetilde{\omega}_{K+1} < \omega < \widetilde{\omega}_K < 0$ ($1 \leq K \leq N-1$), $p_x^* < u_x$ for $1 \leq x \leq K$ and $p_x^* > u_x$ for $K+1 \leq x \leq N$. The number of optimal recommendation probability which is larger than corresponding utility probability is $N - K$. Let $N_2 = N - K$ and $N_1 = K$. It is noted that N_2 decreases with decreasing of ω (cf. Proposition 1). In particular, if $\omega < \widetilde{\omega}_{N-1} < 0$, only the recommendation probability of the event with largest utility probability is enlarged and that of other events is reduced, compared to the corresponding utility probability. As the parameter ω approaches negative infinity, the push distribution will be $(0, 0, ..., 0, 1)$. In this case, the high-probability events are preferred to be pushed by this recommendation system.

Let the optimal recommendation distribution be equal to the utility distribution, i.e., $P^* = U$, and we have

$$u_j = \frac{u_j e^{\omega(1-u_j)}}{\sum_{i=1}^{N} u_i e^{\omega(1-u_i)}} = f(\omega, j, U) \quad , \quad 1 \leq j \leq N. \tag{38}$$

It is noted that $\omega = 0$ is the solution of (38) according to (31). Since $f(\omega, 1, U)$ is monotonously increasing with ω (cf. Remark 4), $f(\omega, 1, U) \neq f(0, 1, U) = u_1$ when $\omega \neq 0$. Thus, there exists one and only one root for (38), which is $\omega = 0$, and $P^* = U$ in this case. Here, all the data types are fairly treated.

For convenience, the relationship between parameters, the optimal recommendation distribution and the utility distribution is summarized in Table 5.

In fact, the recommendation system can be regarded as an information-filtering system which push data based on the preferences of users [12]. The input and output of this information-filtering system can be seen as utility distribution and optimal recommendation distribution, respectively. For advertising systems, compared to the input, on the output port, the recommendation probability of data belonging to the set $\{S_i | 1 \leq i \leq K\}$ is amplified, and that belonging to the set $\{S_i | K+1 \leq i \leq N\}$ is minified, where $1 \leq K \leq N-1$ and $0 < \widetilde{\omega}_{K+1} < \omega < \widetilde{\omega}_K$. Since $u_1 < u_2 \leq ... \leq ... \leq u_{N-1} < u_N$, the advertising system is a special information-filtering system which prefers to push the small-probability events. For noncommercial systems, the data with higher utility probability, i.e., $\{S_i | K+1 \leq i \leq N\}$, is more likely to be pushed, and the data with smaller utility probability, i.e., $\{S_i | 1 \leq i \leq K\}$, tends to be overlooked, where $\widetilde{\omega}_{K+1} < \omega < \widetilde{\omega}_K < 0$ and $1 \leq K \leq N-1$. Since

$u_1 < u_2 \leq ... \leq ... \leq u_{N-1} < u_N$, the high-probability events are preferred to be pushed by the noncommercial system.

Table 5. Optimal recommendation distribution with parameters. $\downarrow$ is used to denote $p_x^* < u_x$ and $\uparrow$ is used to denote $p_x^* > u_x$.

β	ω	$x = 1$	$2 \leq x \leq t$	$t+1 \leq x \leq N-1$	$x = N$	P^*
β_{no}	$-\infty$	0	$p_x^* < u_x$	$p_x^* < u_x$	1	$(0,0,...,0,1)$
(β_0, β_{no})	$(-\infty, 0)$	$p_1^* < u_1$	$p_x^* < u_x$	/	$p_N^* > u_N$	$(\underbrace{\downarrow,...,\downarrow}_{N_1}, \underbrace{\uparrow,...,\uparrow}_{N_2})$
$(\beta_0, \tilde{\beta}_x)$	$(-\infty, \widetilde{\omega}_x)$	/	/	$p_x^* < u_x$	$p_N^* > u_N$	
$(\tilde{\beta}_x, \beta_{no})$	$(\widetilde{\omega}_x, 0)$	/	/	$p_x^* > u_x$	$p_N^* > u_N$	
β_0	0	$p_1^* = u_1$	$p_x^* = u_x$	$p_x^* = u_x$	$p_N^* = u_N$	$(u_1, u_2, ..., u_N)$
$(\beta_0, \tilde{\beta}_x)$	$(0, \widetilde{\omega}_x)$	$p_1^* > u_1$	$p_x^* > u_x$	/	/	$(\underbrace{\uparrow,...,\uparrow}_{N_3}, \underbrace{\downarrow,...,\downarrow}_{N_4})$
$(\tilde{\beta}_x, \beta_{ad})$	$(\widetilde{\omega}_x, +\infty)$	$p_1^* > u_1$	$p_x^* < u_x$	/	/	
(β_0, β_{ad})	$(0, +\infty)$	$p_1^* > u_1$	/	$p_x^* < u_x$	$p_N^* < u_N$	
β_{ad}	$+\infty$	1	$p_x^* < u_x$	$p_x^* < u_x$	0	$(1,0,...,0,0)$

Remark 6. *The recommendation system can be regarded as an information-filtering system, and parameter ω is like an indicator which can reflect the system preference. If $\omega > 0$, the system is a advertising system, which prefers to recommend the small-probability events, while the system is a noncommercial system, which is more likely to recommend the high-probability events if $\omega < 0$. In particular, the system pushes data according to the preference of the user exactly if $\omega = 0$.*

4.3. Geometric Interpretation

In this part, we will give a geometric interpretation of optimal recommendation distribution by means of probability simplex. Let $\mathcal{P}$ denote all the probability distributions from an alphabet $\{S_1, S_2, ..., S_N\}$, and point U denote the utility distribution. In addition, P is the notation to denote recommendation distribution and P^* denotes the optimal recommendation distribution. With the help of the method of types [37], we have Figure 2 to characterize the generation of optimal recommendation distribution.
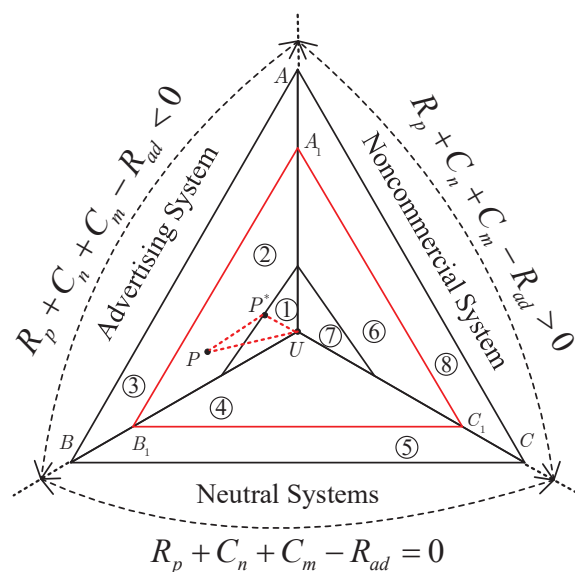
Figure 2. Probability simplex and optimal recommendation. Region ① denotes Case ① in Table 3 for $1 \leq i \leq 8$.

In Figure 2, all cases can be grouped into three major categories, i.e., $R_p + C_n + C_m - R_{ad} < 0$, $R_p + C_n + C_m - R_{ad} = 0$ and $R_p + C_n + C_m - R_{ad} > 0$. In fact, these three categories are advertising systems, neutral systems and noncommercial systems, which are denoted by triangles Δ_{ABU}, Δ_{BCU} and Δ_{ACU}, respectively. Triangles $\Delta_{A_1B_1C_1}$ denotes the region where the average revenue is equal to or larger than a given value. In this paper, our goal is to find the optimal recommendation distribution so that the recommended data would fit the preference of the user as much as possible when the average revenue satisfies the predefined requirement. Since the matching degree between the recommended data and the user's preference is characterized by $2^{-nD(P\|U)}$ (cf. Section 2.1), P^* should be a recommendation distribution on the border of or in the triangle $\Delta_{A_1B_1C_1}$, which is closest to utility distribution U in relative entropy. Therefore, there is no solution if recommendation distribution P falls within regions ③, ⑤ and ⑧.

For advertising systems, $\mathcal{P}$ can be divided into regions ①②③. Obviously, $P^* = U$ for region ① since U falls within region ①. P^* in ② is the recommendation distribution closest to U in relative entropy. It is noted that $2^{-nD(P^*\|U)} > 2^{-nD(P\|U)}$ if $P \neq P^*$ and $P \in$ ②. There is no solution in region ③ since $P \notin \Delta_{A_1B_1C_1}$. There is a similar situation for noncommercial systems, which are composed of regions ⑥, ⑦ and ⑧. There are only two regions for neutral systems. In this case, there is no solution when $P \in$ ⑤, and $P^* = U$ when $P \in$ ④.

Furthermore, triangle Δ_{ABU} characterizes the set E_1 and triangle Δ_{ACU} characterizes the set E_2.

5. Relationship between Optimal Recommendation Strategy and MIM

5.1. Normalized Recommendation Value

In this section, we shall focus on the relationship between optimal recommendation distribution and MIM in Cases ② and ⑥. In fact, the optimal recommendation distributions in other cases are invariable and it makes little sense to discuss the relationship between them.

Within a period of time, the parameters $C_p, R_p, C_n, R_{ad}, C_m$ can be seen as invariants of a given recommendation system, and they do not change with the users. The strategy recommendation is determined by the personalized user concern and the expected revenue in this paper. The former is characterized by the utility distribution, and the latter is denoted by β. Based on the discussion in Section 3.3, there is one-to-one mapping between the parameter ϖ and the minimum average revenue β. Thus, once the recommendation environment is determined, we only need to select the proper parameter ϖ^* to get the optimal recommendation distribution based on the utility distribution, and here ϖ^* is chosen to satisfy the expectation of the average revenue.

The form of optimal recommendation distribution in Label (29) suggests that we allocate recommendation proportion by weight factor $u_i e^{\varpi^*(1-u_i)}$, where $U = \{u_1, u_2, ..., u_N\}$ is the utility distribution. In a sense, we take the weight factor $u_i e^{\varpi^*(1-u_i)}$ as the recommendation value of data belonging to the i-th class, and we generate optimal recommendation distribution based on the recommendation value. It is noted that the optimal recommendation distribution is the normalized recommendation value. Furthermore, the total recommendation value of this user is $\sum_{i=1}^{N} u_i e^{\varpi(1-u_i)}$.

In fact, the recommendation values are the subjective view, and they are not objective quantity in nature. They show the relative level of recommendation tendency. If all the recommendation values multiply by the same constant, the suggested recommendation distribution will not change.

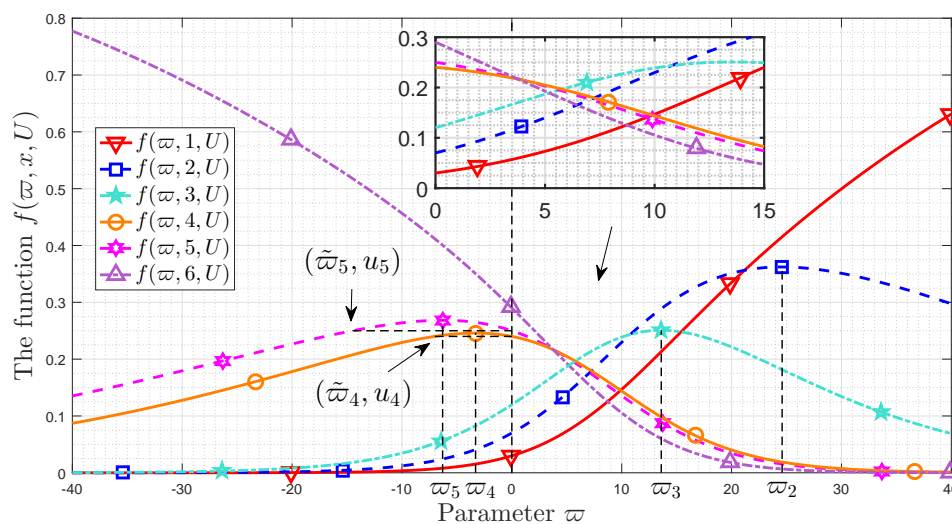### 5.2. Optimal Recommendation and MIM

Generally speaking, there must be a reason of the fact that the recommendation strategy prefers specific data. According to the discussion in Remark 6, advertising systems prefer the small-probability events (i.e., advertisement), and they push as many small-probability events as possible with the increasing of the minimum average revenue since pushing small-probability events is the main source of income for this system. Noncommercial systems prefer the high-probability events, and they push as many high-probability events as possible with the increasing of the minimum average revenue since

recommending a piece of desired data is the main source of income for this system. The preference of the recommendation system means that the system has its own evaluation of the degree of importance for different data. The system prefers to recommend the important data in order to achieve the recommendation target. Thus, the recommendation distribution is determined by the importance of data. That is, the recommendation probability of data belonging to the i-th class is the proportion of recommendation value $u_i e^{\omega^*(1-u_i)}$ in total recommendation value. In this sense, the recommendation value gives intuitive description of message importance. Furthermore, their relative magnitudes are more significant than their absolute magnitude. For a given parameter, the recommendation value is a quantity with respect to probability distribution. Hence, the recommendation value can be seen as a measure of message importance from this viewpoint, which agrees with the main general principles for definition of importance in previous literature [24–28]. Moreover, the total recommendation value characterizes the total message importance.

In fact, the form of this importance measure is extremely similar to that of MIM [29]. MIM is proposed to measure the message importance in the case where small-probability events contain most valuable information and the parameter ω in MIM is called *importance coefficient*. The importance coefficient is always larger than zero in MIM. Furthermore, the MIM highlights the importance of minority subsets and ignores the high-probability event by taking them for grant. As stated in Remark 6, the small-probability events are highlighted when $\omega > 0$. Therefore, MIM is consistent with the conclusion of this paper.

In addition, as the parameter increased to being sufficiently large, the message importance of the event with minimum probability dominates the MIM according to [29,35]. It is the same with recommendation value since $\lim\limits_{\omega \to +\infty} \frac{u_{\min} e^{\omega(1-u_{\min})}}{\sum_{i=1}^{N} p_i e^{\omega(1-u_i)}} = 1$ (cf. (36)). Furthermore, if ω is not a very large positive number, the form of $f(\omega, x, P)$ $(1 < x < N)$ is similar with Shannon entropy, which is discussed in [40]. Ref. [34] discussed the selection of importance coefficient, and it pointed out that the event with probability u_j becomes the principal component in MIM when $\omega = 1/u_j$. Due to the same form, the optimal recommendation distribution also has this conclusion. That is, $f(1/u_j, j, U)$ is larger than $f(1/u_j, i, U)$ if $i \neq j$, which is shown in Figure 3.

Figure 3. Function $f(\omega, x, U)$ vs. importance coefficient ω, when the utility distribution is $U = \{0.03, 0.07, 0.12, 0.24, 0.25, 0.29\}$.

Remark 7. *The optimal recommendation distribution is normalized message importance by means of MIM. In other words, the normalized message importance can also be seen as a distribution that is closest to utility distribution in relative entropy, when the average revenue of this recommendation system is larger than or equal to a predefined value.*

Although the MIM is proposed based on information theory, it can also be given from the viewpoint of data recommendation. In fact, Remark 7 characterizes the physical meaning of MIM from data recommendation perspective, which confirms the rationality of MIM in one aspect.

Remark 8. *We also expands MIM to general cases whatever the probability of systems interested event is. The importance coefficient ϖ plays a switch role on the event sets of systems attention, that is,*

(1) *If $\varpi > 0$, the importance index of small-probability events is magnified while that of high-probability events is lessened. In this case, the system prefers to recommend the small-probability events;*

(2) *If $\varpi < 0$, the importance index of high-probability events is magnified while that of small-probability events is lessened. In this case, the system prefers to recommend the high-probability events;*

(3) *If $\varpi = 0$, the importance of all events is the same.*

In addition, the value of this parameter ϖ can also give a clear definition of the set which recommendation systems are interested in. For example, for a given utility distribution U ($u_1 < u_2 \leq \ldots \leq u_{N-1} < u_N$), if $0 < \widetilde{\varpi}_{K+1} < \varpi < \widetilde{\varpi}_K$ ($1 \leq K \leq N-1$), the sparse events are focused on, and the set of these events is $\{S_i | 1 \leq i \leq K\}$ here. In fact, $\{S_i | 1 \leq i \leq K\}$ gives an unambiguous description of the definition of small-probability events in MIM. On the contrary, $\{S_i | K \leq i \leq N\}$ is the set of the events with high-probability which is highlighted if $\widetilde{\varpi}_{K+1} < \varpi < \widetilde{\varpi}_K < 0$ ($1 \leq K \leq N-1$). In particular, if $\varpi > \widetilde{\varpi}_2 > 0$, only the events with minimum probability will be focused on.

6. Numerical Results

In this section, some numerical results are presented to to validate the theoretical founds in this paper.

6.1. Property of $g(\varpi, V)$

Figure 4 depicts the function $g(\varpi, V)$ versus parameter ϖ. The probability distributions are $V_1 = \{0.1, 0.2, 0.3, 0.4\}$, $V_2 = \{0.05, 0.15, 0.3, 0.5\}$, $V_3 = \{0.13, 0.17, 0.34, 0.36\}$, $V_4 = \{0.01, 0.11, 0.12, 0.76\}$ and $V_5 = \{0.22, 0.25, 0.25, 0.28\}$. The scaling factor of ϖ is varying from -100 to 100.
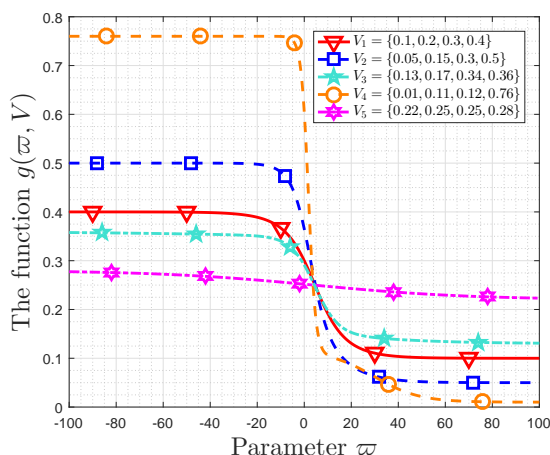

Figure 4. $g(\varpi, P)$ vs. ϖ.

Some observations are obtained in Figure 4. The functions in all the cases are monotonically decreasing with ϖ. When ϖ is small enough, i.e., $\varpi = -100$, $g(\varpi, V)$ is close to $\max(V_i)$ for $1 \leq i \leq 5$, while $g(\varpi, V)$ approaches $\min(V_i)$ for $1 \leq i \leq 5$ when $\varpi = 100$. In fact, we obtain $\min(V_i) \leq g(\varpi, V) \leq \max(V_i)$. These results validate Lemma 1 and Lemma 2. Furthermore, the velocity of change of $g(\varpi, V)$ in region $-20 < \varpi < 20$ is bigger than that in regions $\varpi < -20$ and $\varpi > 20$. The KL distances between these probability distributions and uniform distribution are

$0.1536, 0.3523, 0.1230, 0.9153, 0.0052$, respectively. Thus, the amplitude of variation of $g(\varpi, V)$ decreases with decreasing of KL distance with uniform distribution.

6.2. Optimal Recommendation Distribution

Then, we focus on proposed optimal recommendation distribution in this paper. The parameters of recommendation system are D1 = $\{C_p = 4.5, R_p = 2, C_n = 2, R_{ad} = 11, C_m = 2\}$ and D2 = $\{C_p = 1, R_p = 9, C_n = 2, R_{ad} = 3, C_m = 2\}$. The minimum average revenue β is varying from 0 to 3. The utility distributions are U1 = $\{0.1, 0.2, 0.3, 0.4\}$ and U2 = $\{0.05, 0.15, 0.3, 0.5\}$. Let $U_0 = \{0.25, 0.25, 0.25, 0.25\}$. Some results are listed in Table 6.

Table 6. The auxiliary variables in optimal recommendation distribution.

Case	$R_p + C_n + C_m - R_{ad}$	β_0	β_{no}	β_{ad}	γ_u
D1,U1	−5	1	/	2	0.3
D2,U1	10	1	2	/	0.3
D1,U2	−5	0.675	/	2.25	0.365
D2,U2	10	1.65	3	/	0.365

Figure 5 shows the relationship between optimal recommendation distribution and the minimum average revenue. Some observations can be obtained. In fact, the optimal recommendation distribution can be divided into three phases. In phase one, in which the minimum average revenue is small ($\beta < \beta_0$), the optimal recommendation distribution is exactly the same as utility distribution. In phase two, the minimum average revenue is neither too small nor too large ($\beta_0 < \beta < \beta_{no}$ or $\beta_0 < \beta < \beta_{ad}$). In this case, the optimal recommendation distribution changes with the increasing of minimum average revenue. In the phase three, in which the minimum average revenue is too large ($\beta > \beta_{no}$ or $\beta > \beta_{ad}$), there is no appropriate recommendation distribution.
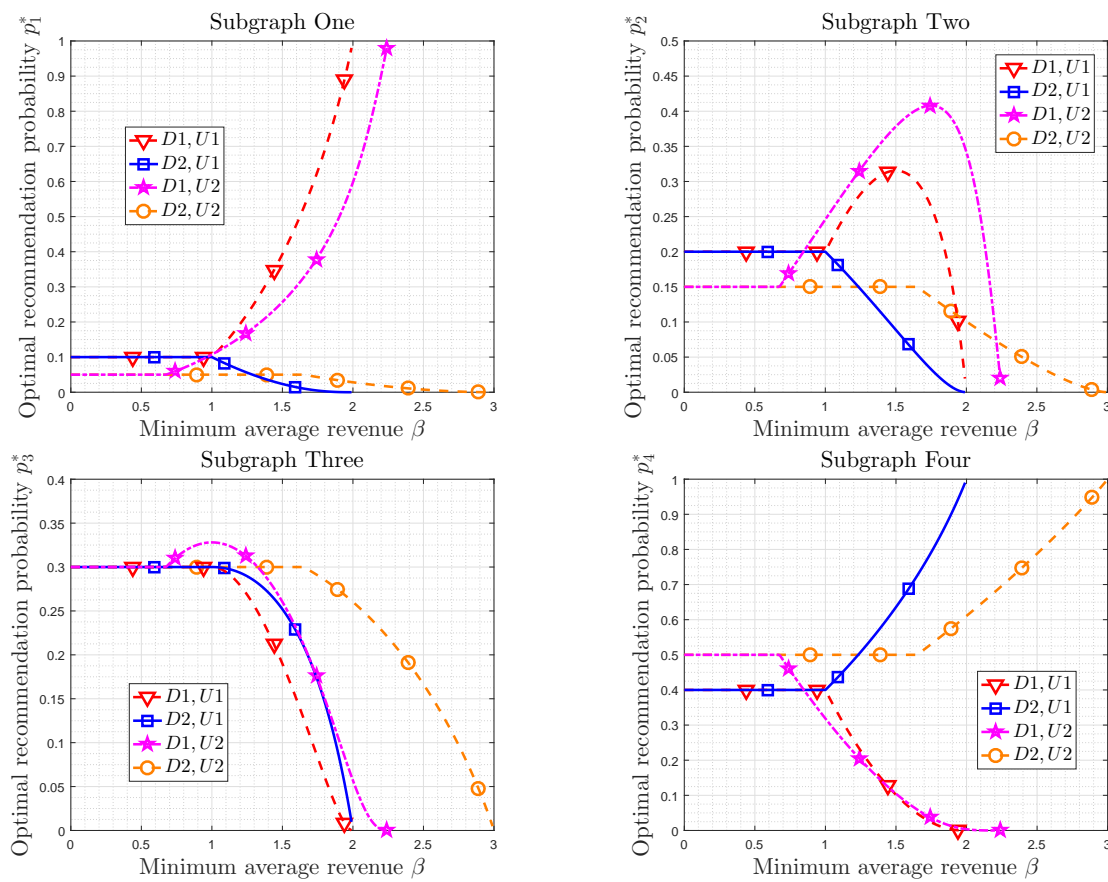
The optimal recommendation probability p_1^* versus the average revenue is depicted in subgraph one of Figure 5. If D1 is adopted, which means advertising systems, we obtain p_1^* increases with the increasing of minimum average revenue when $\beta_0 < \beta < \beta_{ad}$. p_1^* is larger than u_1 in this case. p_1^* approaches one when the average revenue is close to β_{ad}. If D2 is adopted, p_1^* decreases with the increasing of minimum average revenue in the region $\beta_0 < \beta < \beta_{no}$, and p_1^* is smaller than u_1. p_1^* approaches zero as the average revenue increasing to β_{no}. In addition, $p_1^* = u_1$ when $\beta < \beta_0$.

Subgraph two of Figure 5 depicts the optimal recommendation probability p_2^* versus the minimum average revenue. If $\beta < \beta_0$, the optimal recommendation probability is equal to the corresponding utility probability. If $\beta > \beta_{no}$ or $\beta > \beta_{ad}$, p_2^* in the four cases will decrease from u_2 to zero. In this case, p_2^* increases at the early stage and then decreases for D1, U1 and D1, U2, while p_2^* is monotonously increasing for D2, U1 and D2, U2. Since $u_2 < \gamma_u$, $R_p + C_n + C_m - R_{ad} < 0$ for D1 and $R_p + C_n + C_m - R_{ad} > 0$ for D2, these numerical results agree with the discussion in Section 4.1. Subgraph three is similar to the Subgraph two.

Subgraph four is contrary to the Subgraph one, which shows the optimal recommendation probability p_4^* versus the minimum average revenue. If $\beta < \beta_0$, p_4^* is equal to u_4. For D1, i.e., advertising systems, if the minimum average revenue is larger than β_0 and smaller than β_{ad}, p_4^* is smaller than u_4 and it decreases to zero as $\beta \rightarrow \beta_{ad}$. However, p_4^* is larger than u_4 and it increases to one as $\beta \rightarrow \beta_{no}$ for D2, i.e., noncommercial systems.

Figure 6 illustrates minimum KL distance between recommendation distribution and utility distribution versus the minimum average revenue when $\beta_0 < \beta < \beta_{no}$ or $\beta_0 < \beta < \beta_{ad}$. It is noted that the constraints on the minimum average revenue are true. The minimum KL distance increases with the increasing of the minimum average revenue. This figure also shows that the minimum average revenue can be obtained for a given minimum KL distance, when the utility distribution and recommendation system parameters are invariant. Since this KL distance represents the accuracy of

recommendation strategy, there is a trade-off between the recommendation accuracy and the minimum average revenue.

Figure 5. The optimal recommendation distribution vs. minimum average revenue. The parameters set $\{C_p, R_p, C_n, R_{ad}, C_m\}$ is denoted by D1 and D2, where D1 $= \{4.5, 2, 2, 11, 2\}$ and D2 $= \{1, 9, 2, 3, 2\}$. The utility distributions are U1 $= \{0.1, 0.2, 0.3, 0.4\}$ and U2 $= \{0.05, 0.15, 0.3, 0.5\}$.
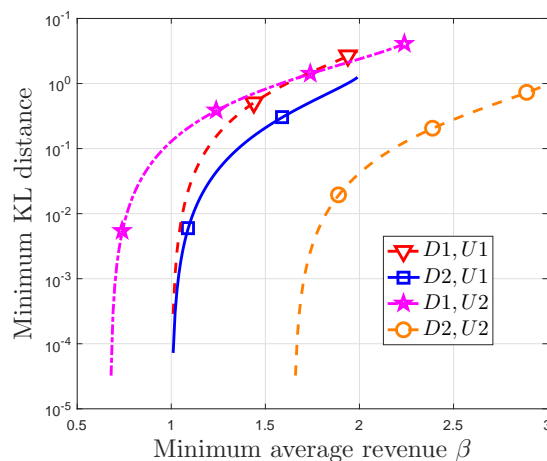
Figure 6. Minimum KL distance between recommendation distribution and utility distribution vs. minimum average revenue. The parameters set $\{C_p, R_p, C_n, R_{ad}, C_m\}$ is denoted by D1 and D2, where D1 $= \{4.5, 2, 2, 11, 2\}$ and D1 $= \{1, 9, 2, 3, 2\}$. The utility distribution is U1 $= \{0.1, 0.2, 0.3, 0.4\}$ and U2 $= \{0.05, 0.15, 0.3, 0.5\}$.

6.3. Property of $f(\varpi, x, U)$

Figure 3 shows that the importance coefficient ϖ versus the function $f(\varpi, x, U)$. The utility distribution is $\{0.03, 0.07, 0.12, 0.240.25, 0.29\}$ and the importance coefficient is varying from -40 to 40. It is easy to check that $u_1 < u_2 < u_3 < \sum_{i=1}^{6} u_i^2 < u_4 < u_5 < u_6$. We denote $\{\varpi_2, \varpi_3, \varpi_4, \varpi_5\} = \{24.6, 13.6, -3.3, -6.3\}$.

$f(\varpi, 1, U)$ is monotonically increasing with increasing of the importance coefficient. $f(\varpi, 1, U)$ is close to zero when $\varpi < -30$. $f(\varpi, i, U)$ ($i = 2, 3, 4, 5$) increases with the increasing of ϖ when $\varpi < \varpi_i$, and then it decreases with the increasing of ϖ when $\varpi > \varpi_i$. They achieve the maximum value when $\varpi = \varpi_i$. It is also noted that $\varpi_5 < \varpi_4 < 0 < \varpi_3 < \varpi_2$. If $\varpi < -30$, $f(\varpi, i, U)$ ($i = 2, 3$) is very close to zero, and $f(\varpi, i, U) > 0.05$ ($i = 2, 3$) if $\varpi > 30$, which means it changes faster when $\varpi < 0$. On the contrary, $f(\varpi, i, U)$ ($i = 4, 5$) changes faster with ϖ in $(0, +\infty)$ than that in $(-\infty, 0)$ since that it is still bigger than 0.5 when $\varpi = -40$ and it approaches zero when $\varpi > 30$. In addition, $f(\varpi, 6, U)$ decreases monotonically with the increasing of the importance coefficient, and $f(\varpi, 6, U)$ is close to zero when $\varpi > 30$.

Some other observations are also obtained. When $\varpi = 0$, we obtain $f(\varpi, i, U) = u_i$ ($i = 1, 2, 3, 4, 5, 6$). Without loss of generality, we take $f(\varpi, 5, U)$ as an example. There is a non-zero importance coefficient $\tilde{\varpi}_5$ which makes $f(\varpi, 5, U) = u_5$. If $0 > \tilde{\varpi}_4 > \varpi > \tilde{\varpi}_5$, we obtain $f(\varpi, i, U) > u_i$ for $i = 5, 6$ and $f(\varpi, i, U) < u_i$ for $i = 1, 2, 3, 4$. Compared with the utility distribution, the results of the elements in set $\{S_5, S_6\}$ are enlarged, and those in set $\{S_1, S_2, S_3, S_4\}$ are minified. The difference between these two sets is that the utility probability of S_5 or S_6 is larger than that in $\{S_1, S_2, S_3, S_4\}$. In addition. it is also noted that $f(\varpi, i, U) > u_i$ for $i = 6$ and $f(\varpi, i, U) < u_i$ for $i = 1, 2, 3, 4, 5$, when $\varpi < \tilde{\varpi}_5$. Here, only the function output of the element in high-probability set $\{S_6\}$ is larger than the corresponding utility probability.

Furthermore, when $\varpi = 1/p_i \in \{33.3333, 14.2857, 8.3333, 4.1667, 4, 3.4483\}$ for $1 \leq i \leq 6$, $f(\varpi, i, U) > f(\varpi, j, U)$ for $j \neq i$.

7. Conclusions

In this paper, we discussed the optimal data recommendation problem when the recommendation model pursues best user performance with a certain revenue guarantee. Firstly, we illuminated the system model and formulated this problem as an optimization. Then, we gave the explicit solution of this problem in different cases, such as advertising systems and noncommercial systems, which can improve the design of data recommendation strategy. In fact, the optimal recommendation distribution is the one that is the closest to the utility distribution in the relative entropy when it satisfies expected revenue. There is a trade-off between the recommendation accuracy and the expected revenue. In addition, the properties of this optimal recommendation distribution, such as monotonicity and geometric interpretation, were also discussed in this paper. Furthermore, the optimal recommendation system can be regarded as an information-filtering system, and the importance coefficient determines what events the system prefers to recommend.

We also obtained that the optimal recommendation probability is the proportion of corresponding recommendation value in total recommendation value if the minimum average revenue is neither too small nor too large. In fact, the recommendation value is a special weight factor in determining the optimal recommendation distribution, and it can be regarded as a measure of importance. Since its form and properties are the same as those of MIM, the optimal recommendation distribution is exactly the normalized MIM, where MIM is used to characterize the concern of the system. The parameter in MIM, i.e., importance coefficient, plays a switch role on the event sets of systems' attention. That is, the importance index of high-probability events is enlarged for the negative importance coefficient (i.e., noncommercial system), while the importance index of small-probability events is magnified by systems for the positive importance coefficient (i.e., advertising systems). In particular, only the maximum-probability event or the minimum-probability event are focused on as the importance coefficient approaches to negative infinity or positive infinity, respectively. These results give a new

physical explanation of MIM from the data recommendation perspective, which can validate the rationality of MIM in one aspect. MIM is also extended to the general case for whatever the probability of systems interested events is. One can adjust the importance coefficient to focus on desired data type. Compared with previous MIM, the set of desired events can be defined precisely. These results can help us formulate appropriate data recommendation strategy in different scenarios.

In the future, we may consider its applications in next generation cellular systems [41,42], wireless sensor networks [43] and very high speed railway communication systems [44] by taking the signal transmission mode into accounts.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MIM　　message importance measure
KL　　　Kullback–Leibler
SVM　　support vector machine

Appendix A. Proof of Lemma 1

The derivation of $g(\varpi, V)$ is given by

$$\frac{\partial g(\varpi, V)}{\partial \varpi} = \frac{\sum_{i=1}^{N} v_i^2(1-v_i)e^{\varpi(1-v_i)} \sum_{j=1}^{N} v_j e^{\varpi(1-v_j)} - \sum_{i=1}^{N} v_i^2 e^{\varpi(1-v_i)} \sum_{j=1}^{N} v_j(1-v_j)e^{\varpi(1-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1}$$

$$= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} v_i^2(1-v_i)v_j e^{\varpi(2-v_i-v_j)} - \sum_{i=1}^{N}\sum_{j=1}^{N} v_i^2 v_j(1-v_j)e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1a}$$

$$= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} v_i^2 v_j(v_j-v_i)e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1b}$$

$$= \frac{\sum\limits_{v_i<v_j} v_i^2 v_j(v_j-v_i)e^{\varpi(2-v_i-v_j)} + \sum\limits_{v_i>v_j} v_i^2 v_j(v_j-v_i)e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1c}$$

$$= \frac{\sum\limits_{v_j<v_i} v_j^2 v_i(v_i-v_j)e^{\varpi(2-v_j-v_i)} + \sum\limits_{v_i>v_j} v_i^2 v_j(v_j-v_i)e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1d}$$

$$= \frac{\sum\limits_{v_i>v_j} (v_j-v_i)v_i v_j(v_i-v_j)e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} \tag{A1e}$$

$$= \frac{\sum\limits_{v_i>v_j} -(v_j-v_i)^2 v_i v_j e^{\varpi(2-v_i-v_j)}}{\left(\sum_{i=1}^{N} v_i e^{\varpi(1-v_i)}\right)^2} < 0. \tag{A1f}$$

(A1d) is obtained by exchanging the notation of subscript. (A1f) follows from the fact that $v_i \geq 0$ for $1 \leq i \leq N$ and not all of v_i is zero.

Appendix B. Proof of Lemma 2

When $\varpi = 0$, we have

$$g(0, V) = \frac{\sum_{i=1}^{N} v_i^2}{\sum_{i=1}^{N} v_i} = \sum_{i=1}^{N} v_i^2. \tag{A2}$$

Let $g(-\infty, V) \overset{\Delta}{=} \lim\limits_{\varpi \to -\infty} g(\varpi, V)$, and we obtain

$$g(-\infty, V) = \lim_{\varpi \to -\infty} \frac{\sum_{i=1}^{n} v_i^2 e^{\varpi(1-v_i)}}{\sum_{i=1}^{n} v_i e^{\varpi(1-v_i)}} = \lim_{\varpi \to -\infty} \frac{v_{\max}^2 e^{\varpi(1-v_{\max})} + \sum\limits_{v_i \neq v_{\max}} v_i^2 e^{\varpi(1-v_i)}}{v_{\max} e^{\varpi(1-v_{\max})} + \sum\limits_{v_i \neq v_{\max}} v_i e^{\varpi(1-v_i)}}$$

$$= \lim_{\varpi \to -\infty} \frac{v_{\max} + \sum\limits_{v_i \neq v_{\max}} \frac{v_i^2}{v_{\max}} e^{\varpi(v_{\max}-v_i)}}{1 + \sum\limits_{v_i \neq v_{\max}} \frac{v_i}{v_{\max}} e^{\varpi(v_{\max}-v_i)}} = v_{\max}. \tag{A3}$$

Let $g(+\infty, V) \overset{\Delta}{=} \lim\limits_{\varpi \to +\infty} g(\varpi, V)$, and it is noted that

$$g(+\infty, V) = \lim_{\varpi \to +\infty} \frac{\sum_{i=1}^{n} v_i^2 e^{\varpi(1-v_i)}}{\sum_{i=1}^{n} v_i e^{\varpi(1-v_i)}} = \lim_{\varpi \to +\infty} \frac{v_{\min}^2 e^{\varpi(1-v_{\min})} + \sum\limits_{v_i \neq v_{\min}} v_i^2 e^{\varpi(1-v_i)}}{v_{\min} e^{\varpi(1-v_{\min})} + \sum\limits_{v_i \neq v_{\min}} v_i e^{\varpi(1-v_i)}}$$

$$= \lim_{\varpi \to +\infty} \frac{v_{\min} + \sum\limits_{v_i \neq v_{\min}} \frac{v_i^2}{v_{\min}} e^{\varpi(v_{\min}-v_i)}}{1 + \sum\limits_{v_i \neq v_{\min}} \frac{v_i}{v_{\min}} e^{\varpi(v_{\min}-v_i)}} = v_{\min}. \tag{A4}$$

Hence, $v_{\min} < g(\varpi, V) < v_{\max}$ when $\varpi \in (-\infty, +\infty)$ according to Lemma 1.

Appendix C. Proof of Proposition 1

(1) First, if $u_x = u_{\max}$ and $u_x = u_{\min}$, the non-zero solution of equation (33) does not exist since $f(\varpi, x, U)$ is monotonously changing (cf. Remark 4).

Second, if $u_x = \gamma_u$, $u_x \neq u_{\max}$ and $u_x \neq u_{\min}$, no solution exists since $f(0, x, U) > f(\varpi, x, U)$ for $\varpi \neq 0$ according to Theorem 3.

Third, if $u_x < \gamma_u$, $u_x \neq u_{\max}$ and $u_x \neq u_{\min}$, according to Theorem 3, we have $\varpi_x > 0$ and $f(\varpi, x, U)$ is increasing in $(0, \varpi_x)$ while decreasing in $(\varpi_x, +\infty)$, where ϖ_x is the solution to $g(\varpi, U) = 1 - \alpha$. It is easy to check that $f(+\infty, x, U) - f(0, x, U) = -u_x < 0$ (cf. (36)) and $f(\varpi_x, x, U) - f(0, x, U) > 0$. According to the zero point theorem, the non-zero solution to $f(\varpi, x, U) = u_x$, i.e., $\widetilde{\varpi}_x$ (cf. (33)), would be found in $(\varpi_x, +\infty)$.

Fourth, likewise, the non-zero solution $\widetilde{\varpi}_x$ can be found in $(-\infty, \varpi_x)$ if $u_x > \gamma_u$, $u_x \neq u_{\max}$ and $u_x \neq u_{\min}$.

(2) and (3) Consider $u_y < u_x < \gamma_u$ first. According to Theorem 3, we have $\varpi_x > 0$ and $f(\varpi, x, U)$ is increasing in $(0, \varpi_x)$ while decreasing in $(\varpi_x, +\infty)$, where ϖ_x is the solution to $g(\varpi, U) = 1 - \alpha$. Since we have $f(0, x, U) = u_x$, the non-zero solution to $f(\varpi, x, U) = u_x$, i.e., $\widetilde{\varpi}_x$ (cf. (33)), would only be found in $(\varpi_x, +\infty)$. Likewise, we obtain that $\widetilde{\varpi}_y > \varpi_y > 0$. Since $\varpi_y > \varpi_x$ when $u_y < u_x$ (cf. Theorem 3), we have $\widetilde{\varpi}_x > \varpi_x > 0$ and $\widetilde{\varpi}_y > \varpi_y > \varpi_x > 0$.

Furthermore, $f(\widetilde{\omega}_y, y, U) = u_y$ implies $\frac{e^{\widetilde{\omega}_y(1-u_y)}}{\sum_{i=1}^{N} u_i e^{\widetilde{\omega}_y(1-u_i)}} = 1$. Hence, we have

$$f(\widetilde{\omega}_y, x, U) = \frac{e^{\widetilde{\omega}_y(1-u_x)}}{\sum_{i=1}^{N} u_i e^{\widetilde{\omega}_y(1-u_i)}} \tag{A5}$$

$$= \frac{e^{\widetilde{\omega}_y(u_y-u_x)} e^{\widetilde{\omega}_y(1-u_y)}}{\sum_{i=1}^{N} u_i e^{\widetilde{\omega}_y(1-u_i)}} \tag{A5a}$$

$$< u_x = f(\widetilde{\omega}_x, x, U), \tag{A5b}$$

where (A5b) follows $\frac{e^{\widetilde{\omega}_y(1-u_y)}}{\sum_{i=1}^{N} u_i e^{\widetilde{\omega}_y(1-u_i)}} = 1$, $u_x > u_y$ and $\widetilde{\omega}_y > 0$.

Since $f(\omega, x, U)$ is decreasing with ω according to Theorem 3, we then have $\widetilde{\omega}_y > \widetilde{\omega}_x$.

Second, likewise, we can prove that $\widetilde{\omega}_x < \widetilde{\omega}_y < 0$ if $u_x > u_y > \gamma_u$.

Third, if $u_y < \gamma_u < u_x$, we shall have $\widetilde{\omega}_x < 0$ and $\widetilde{\omega}_y > 0$. Obviously, $\widetilde{\omega}_y > \widetilde{\omega}_x$ in this case.

Thus, the proof of Proposition 1 is completed.

References

1. Chen, M.; Mao, S.; Zhang, Y.; Leungm, V.C. Definition and features of big data. In *Big Data: Related Technologies, Challenges and Future Prospects*; Springer: New York, NY, USA, 2014; pp. 2–5.
2. Bi, S.; Zhang, R.; Ding, Z. Wireless communications in the era of big data. *IEEE Commun. Mag.* **2015**, *53*, 190–199. [CrossRef]
3. Franklin, M.; Zdonik, S. Data in your face: Push technology in perspective. In Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data, Seattle, WA, USA, 1–4 June 1998; pp. 516–519.
4. Hauswirth, M. Internet-Scale Push Systems for Information Distribution–Architecture, Components, and Communication. Available online: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=2C5856A9798C3085378770287B32D626?doi=10.1.1.7.4907&rep=rep1&type=pdf (accessed on 20 February 2019).
5. Tadrous, J.; Eryilmaz, A.; Gamal, H.E. Proactive content download and user demand shaping for data networks. *IEEE Trans. Netw.* **2015**, *23*, 1917–1930. [CrossRef]
6. Shoukry, O.; ElMohsen, M.A.; Tadrous, J. Proactive scheduling for content pre-fetching in mobile networks. In Proceedings of the 2014 IEEE International Conference on Communications, Sydney, Australia, 10–14 June 2014; pp. 2848–2854.
7. Kim, Y.; Lee, J.; Park, S.; Choi, B. Mobile advertisement system using data push scheduling based on user preference. In Proceedings of the IEEE Wireless Telecommunications Symposium (WTS), Prague, Czech Republic, 22–24 April 2009; pp. 1–5.
8. Podnar, I.; Hauswirth, M.; Jazayeri, M. Mobile push: Delivering content to mobile users. In Proceedings of the 22nd IEEE International Conference on Distributed Computing Systems Workshops, Vienna, Austria, 2–5 July 2002; pp. 563–568.
9. Nicopolitidis, P.; Papadimitriou, G.I.; Pomportsis, A.S. An adaptive wireless push system for high-speed data broadcasting. In Proceedings of the 14th IEEE Workshop on Local & Metropolitan Area Networks, Crete, Greece, 18 September 2005; pp. 1–5.
10. Bhide, M.; Deolasee, P.; Katkar, A.; Panchbudhe, A.; Ramamritham, K.; Shenoy, P. Adaptive push-pull: Disseminating dynamic web data. *IEEE Trans. Comput.* **2002**, *51*, 652–668. [CrossRef]
11. Li, Y.; Chen, L.; Shi, H.; Hong, X.; Shi, J. Joint content recommendation and delivery in mobile wireless Networks with Outage Management. *Entropy* **2018**, *20*, 64. [CrossRef]
12. Parra-Arnau, J.; Rebollo-Monedero, D.; Forné, J. Optimal Forgery and Suppression of Ratings for Privacy Enhancement in Recommendation Systems. *Entropy* **2014**, *16*, 1586–1631. [CrossRef]
13. Zhang, J.; Yang, Y.; Tian, Q.; Zhuo, L.; Liu, X. personalized social image recommendation method based on user-image-tag model. *IEEE Trans. Multimed.* **2017**, *19*, 2439–2449. [CrossRef]
14. Zhou, P.; Zhou, Y.; Wu, D.; Jin, H. Differentially private online learning for cloud-based Video recommendation with multimedia big data in social networks. *IEEE Trans. Multimed.* **2016**, *18*, 1217–1229. [CrossRef]

15. Verbert, K.; Manouselis, N.; Ochoa, X. Context-aware recommender systems for learning: A survey and future challenges. *IEEE Trans. Learn. Technol.* **2012**, *5*, 318–335. [CrossRef]

16. Cheng, Z.; Shen, J. On effective location-aware music recommendation. *ACM Trans. Inf. Syst.* **2016**, *34*, 13. [CrossRef]

17. Elkan, C. The foundations of cost-sensitive learning. In Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, Seattle, WA, USA, 4–10 August 2001; pp. 973–978.

18. Zhou, Z.; Liu, X. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 63–77. [CrossRef]

19. Lomax, S.; Vadera, S. A survey of cost-sensitive decision tree induction algorithms. *ACM Comput. Surv.* **2013**, *45*, 16:1–16:35. [CrossRef]

20. Du, J.; Ni, E.A.; Ling, C.X. Adapting cost-sensitive learning for reject option. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), Toronto, ON, Canada, 26–30 October 2010; pp. 1865–1868.

21. Zieba, A. Counterterrorism systems of spain and poland: Comparative studies. *Prz. Politol.* **2015**, *3*, 65–78. [CrossRef]

22. Ando, S.; Suzuki, E. An information theoretic approach to detection of minority subsets in database. In Proceedings of the Sixth International Conference on Data Mining (ICDM), Hong Kong, China, 18–22 December 2006; pp. 1–10.

23. Vapnik, V.; Kotz, S. *Estimation of Dependences Based on Empirical Data*; Jordan, M., Kleinberg, J., Schölkopf, B., Eds.; Springer: New York, NY, USA, 2006.

24. Ivanchev, J.; Aydt, H.; Knoll, A. Information maximizing optimal sensor placement robust against variations of traffic demand based on importance of nodes. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 714–725. [CrossRef]

25. Kawanaka, T.; Rokugawa, S.; Yamashita, H. Information security in communication network of memory channel considering information importance. In Proceedings of the IEEE International Conference on Industrial Engineering and Engineering Management (IEEM), Singapore, 10–13 December 2017; pp. 1169–1173.

26. Mönks, U.; Lohweg, V. Machine conditioning by importance controlled information fusion. In Proceedings of the IEEE 18th Conference on Emerging Technologies & Factory Automation (ETFA), Cagliari, Italy, 10–13 September 2013; pp. 1–8.

27. Li, Y.; Zhang, M.; Geng, X. Leveraging implicit relative labeling-importance information for effective multi-label learning. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Atlantic City, NJ, USA, 14–17 November 2015; pp. 251–260.

28. Masnick, B.; Wolf, J. On linear unequal error protection codes. *IEEE Trans. Inf. Theory* **1967**, *3*, 600–607. [CrossRef]

29. Fan, P.; Dong, Y.; Lu, J.; Liu, S. Message importance measure and its application to minority subset detection in big data. In Proceedings of the IEEE Globecom Workshops (GC Wkshps), Washington, DC, USA, 4–8 December 2016; pp. 1–6.

30. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [CrossRef]

31. Verdu, S. Fifty years of shannon theory. *IEEE Trans. Inf. Theory* **1998**, *44*, 2057–2078. [CrossRef]

32. Rényi, A. On measures of entropy and information. In Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1961; pp. 547–561.

33. Fadeev, D.K. Zum Begriff der Entropie ciner endlichen Wahrscheinlichkeitsschemas. In *Arbeiten zur Informationstheorie I*; Deutscher Verlag der Wissenschaften: Berlin, Germany, 1957; pp. 85–90.

34. She, R.; Liu, S.; Dong, Y.; Fan, P. Focusing on a probability element: parameter selection of message importance measure in big data. In Proceedings of the IEEE International Conference on Communications (ICC), Paris, France, 20–26 May 2017; pp. 1–6.

35. Liu, S.; She, R.; Fan, P.; Letaief, K.B. Non-parametric Message Important Measure: Storage Code Design and Transmission Planning for Big Data. *IEEE Trans. Commun.* **2018**, *66*, 5181–5196. [CrossRef]

36. She, R.; Liu, S.; Fan, P. Recognizing Information Feature Variation: Message Importance Transfer Measure and Its Applications in Big Data. *Entropy* **2018**, *20*, 401. [CrossRef]

37. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2006.

38. Liu, S.; She, R.; Wan, S.; Fan, P.; Dong, Y. A Switch to the Concern of User: Importance Coefficient in Utility Distribution and Message Importance Measure. In Proceedings of the IEEE International Wireless Communications & Mobile Computing Conference (IWCMC), Limassol, Cyprus, 25–29 June 2018; pp. 1–6.

39. Van Erven, T.; Harremoës, P. Rényi divergence and kullback-leibler divergence. *IEEE Trans. Inf. Theory* **2014**, *60*, 3797–3820. [CrossRef]

40. She, R.; Liu, S.; Fan, P. Information Measure Similarity Theory: Message Importance Measure via Shannon Entropy. *arXiv* **2019**, arXiv:1901.01137.

41. Xiong, K.; Fan, P.; Lu, Y.; Letaief, K.B. Energy efficiency with proportional rate fairness in multirelay OFDM networks. *IEEE J. Select. Areas Commun.* **2016**, *34*, 1431–1447. [CrossRef]

42. Xiong, K.; Chen, C.; Qu, G.; Fan, P.; Letaief, K.B. Group cooperation with optimal resource allocation in wireless powered communication networks. *IEEE Trans. Wirel. Commun.* **2017**, *16*, 3840–3853. [CrossRef]

43. Wang, Q.; Wu, D.O.; Fan, P. Delay-constrained optimal link scheduling in wireless sensor networks. *IEEE Trans. Veh. Technol.* **2010**, *59*, 4564–4577. [CrossRef]

44. Zhang, C.; Fan, P.; Xiong, K.; Fan, P. Optimal power allocation with delay constraint for signal transmission from a moving train to base stations in high-speed railway scenarios. *IEEE Trans. Veh. Technol.* **2015**, *64*, 5775–5788. [CrossRef]