Multi-Channel Sellers Traffic Allocation in Large-scale E-commerce Promotion

SHEN Xin
Nanyang Technological University
Singapore
xin.shen@ntu.edu.sg

Yizhou Ye
Alibaba Group
Hangzhou, China
yizhou.yyz@alibaba-inc.com

Martin Ester
Simon Fraser University
Burnaby, B.C., Canada
ester@sfu.ca

Cheng Long
Nanyang Technological University
Singapore
c.long@ntu.edu.sg

Jie Zhang
Nanyang Technological University
Singapore
zhangj@ntu.edu.sg

Zhao Li
Alibaba Group
Hangzhou, China
lizhao.lz@alibaba-inc.com

Kaiying Yuan
Zhejiang University
Hangzhou, China
yuanky@zju.edu.cn

Yanghua Li
Alibaba Group
Hangzhou, China
yichen.lyh@taobao.com

ABSTRACT

Large-scale online promotions, such as Double 11 and Black Friday, are of great value to e-commerce platforms nowadays. Traditional methods are not successful when we aim to maximize global Gross Merchandise Volume (GMV) in the promotion scenarios due to three limitations. The first is that the GMV of sellers varies significantly from daily scenarios to promotions. Second, these methods do not consider explosive demands in promotions, so that a consumer may fail to purchase some popular items due to sellers' limited capacities. Third, the traffic distribution over sellers presents divergence in different channels, thus rendering the performance of the traditional single-channel methods far from optimal in creating commercial values. To address these problems, we design a Multi-Channel Sellers Traffic Allocation (MCSTA) optimization model to obtain optimal page view (PV) distribution concerning global GMV. Then we propose a general constrained non-smooth convex optimization solution with a Multi-Objective Shortest Distance (MOSD) hyperparameter tuning method to solve MCSTA. This is the first work to systematically address this issue in the scenario of large-scale online promotions. The empirical results show that MCSTA achieves significant improvement of GMV by 1.1% based on A/B test during Alibaba's "Global Shopping Festival", one of the world's largest online sales events. Furthermore, we deploy MCSTA in other popular scenarios, including everyday promotion and video live stream service, to showcase that MCSTA can be widely applied in e-commerce and online entertainment services.

CCS CONCEPTS

• **Applied computing → Electronic commerce**; **Online shopping**; *Multi-criterion optimization and decision-making*; • **Information systems →** *Traffic analysis.*

ACM ISBN 978-1-4503-6859-9/20/10…$15.00
https://doi.org/10.1145/3340531.3412730

KEYWORDS

E-commerce, Multi-channel sellers traffic allocation, Online promotion, Online shopping

ACM Reference Format:
SHEN Xin, Yizhou Ye, Martin Ester, Cheng Long, Jie Zhang, Zhao Li, Kaiying Yuan, and Yanghua Li. 2020. Multi-Channel Sellers Traffic Allocation in Large-scale E-commerce Promotion. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3340531.3412730

1 INTRODUCTION

The e-commerce era is witnessing a rapid development of online platforms which have attracted an increasing number of people favoring online shopping. Customers explore items through multiple channels, such as search and recommendation. Business intelligence strategies applied in several channels increase Click Through Rate (CTR) and Conversion Rate (CVR). Meanwhile, large-scale online promotions have played a significant role in e-commerce platforms since Alibaba launched "Global Shopping Festival". In 2019, "Global Shopping Festival" has generated US$38.4 billion Gross Merchandise Volume (GMV).[1] Thus increasing GMV in a large-scale online promotion creates remarkable commercial values.

At first glance, it seems that lots of traditional models for daily scenarios, such as personalized search engines [8][22][12][9], recommendation algorithms [16][20], and seller ranking models [21], can handle the goal of maximizing global GMV in a large-scale promotion. However, the situation differs from existing daily scenarios in the following three important aspects:

First, the transaction features of sellers vary significantly from daily scenarios to promotions. We divide sellers on the e-commerce platform into 11 groups, S1 to S11, based on their business scales (S1 indicating the largest scale). As Figure 1 shows, within S5 to S9, more GMV is distributed in daily scenarios. While within S1 to S4, promotion scenarios have larger GMV distribution. Behaviors of the sellers and consumers change dramatically in promotion compared with the daily activities. Such discrepancies between daily and promotion activities pose a novel and critical challenge.

[1] https://www.alizila.com/by-the-numbers-2019-11-11-global-shopping-festival

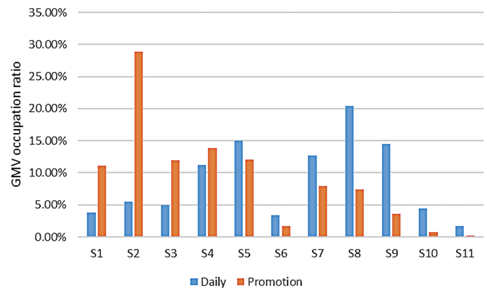

Figure 1: The distribution of sellers' GMV in daily scenarios and large-scale promotions

Second, "sold out" challenge widely exists in this problem. "Sold out" is caused by the false navigation to sellers with unsatisfied capacity. Some critical buyers have preferable sellers and will not buy items from other sellers. If their preferred sellers are sold out, we will lose these customers. To address this problem, we should balance traffic among different sellers during the promotion, to postpone and reduce cases of hot sellers' "sold out". Traditional models, including greedy and round-robin approaches, cannot handle such situations since they disregard signals of merchandise circulation. For example, sellers' real-time inventories and upper limits of logistics capacity are all sensitive in large-scale sales events.

Third, the skewness of the preference distribution over multiple sellers and multiple channels will be more severe when a promotion is launched. Such unbalanced behaviors will make traditional models for a single channel not so optimal in creating business values. In reality, a seller's PV can be routed to different channels and vice versa. When it comes to a promotion activity, balancing between multi-seller and multi-channel to achieve a global optimal GMV will be essential but greatly challenging.

Based on these characteristics, we aim to propose a framework to maximize the global GMV by optimizing traffic allocation over multi-seller and multi-channel in the scenario of a large-scale promotion. The traffic allocation means the Page View (PV) proportions. First of all, we use our previous model [17] to predict the GMV capacity of sellers. To maximize the global GMV, we design a Multi-Channel Sellers Traffic Allocation (MCSTA) optimization model, which combines three objectives: (1) maximize the platform global GMV, (2) ensure that each seller's GMV does not exceed the predicted capacity, (3) let the new allocation not deviate much from the original allocation to prevent the over-fitting phenomenon. Exactly solving this optimization problem is computationally infeasible. We provide a constrained non-smooth convex optimization algorithm, which is based on the augmented Lagrange and proximal gradient method, to obtain the optimal allocation results. The framework is deployed on "Tmall.com", Alibaba's real-world e-commerce platform, and evaluated through implementation of the control process in the A/B test system. The results show that the framework yields a significant additional gain of GMV, which is about 1.1% growth during Alibaba's "Global Shopping Festival".

Furthermore, to present potential commercial values of the proposed MCSTA model, we apply it in other scenarios, including everyday promotion and video live stream service. In everyday promotion, sellers' capacities are not limited. On an e-commerce platform, we try to ensure each seller participated in everyday promotion to achieve enough PV. In video live stream service, we attempt to use MCSTA to maintain a balance between popular hosts and new hosts, to solve the "cold start" problem and increase the attractiveness of new hosts.

We summarize major contributions of this paper as follows:

(1) We are the first to investigate a systematic method to increase the global GMV of a large-scale online promotion by assigning sellers traffic allocation over multiple channels, which differs from traditional personalized search or recommendation.
(2) We propose a Multi-Channel Sellers Traffic Allocation (MCSTA) optimization model to achieve optimal PV distribution concerning global GMV and sellers' capacities.
(3) We propose a general constrained non-smooth convex optimization solution with a Multi-Objective Shortest Distance (MOSD) hyperparameter tuning method to solve MCSTA.
(4) We extend the proposed MCSTA model, to the other two popular scenarios, everyday promotion in e-commerce and video live stream service in an online entertainment system, to show extensive commercial values of the model.

2 RELATED WORK

In recent years, with the development of e-commerce, the technology of search engine and recommendation system has been extensively studied. For example, personalized optimizing product ranking after user resolution [18] by deep learning [19] and reinforcement learning [8] models has achieved considerable success in improving traffic allocation. In recommendation channels, collaborative filtering [14], Naive Bayes [7], graph embedding [11], top-N subgroups [2], multiple objectives optimization [1], transfer learning [13], reinforcement learning [4][3], etc. have been proposed and applied widely. All these approaches can optimize traffic allocation and raise revenue by improving user experience. However, most of these approaches focus on daily shopping activities, where the models are generally updated daily. In the situation of large-scale online promotions in one day, they can neither adapt to fast-changing consumer preferences nor handle the explosive demands, resulting in sub-optimization in creating commercial values.

In addition, an e-commerce platform contains multiple channels. Most optimization algorithms are based on a single channel. Recently, multi-channel joint optimization [6] has been proposed, but the scope of its integration still focuses on a single scene, such as the main search engine and in-shop search information fusion. In contrast, our framework provides a multi-objective model in the large promotion scenario and tries to optimize global GMV by integrating various information, including PV, sellers' capacities, seller weights, etc., on multiple channels.

3 MULTI-CHANNEL SELLERS TRAFFIC ALLOCATION MODEL

For an e-commerce platform, global GMV is an important metric to evaluate an online promotion. To maximize global GMV, we carefully control the traffic over multiple channels and multiple sellers to avoid the "sold out" situation which wastes traffic of PV

Figure 2: A diagram of PV traffic allocation over multiple sellers and multiple channels

Table 1: Model notations and definitions

Notation	Definition
N_k	set of sellers belonging to the k-th category
$p_{i,j}$	number of PV of the i-th seller and the j-th channel
$v_{i,j}$	value per PV of the i-th seller and the j-th channel
t_k	upper boundary PV for the k-th category
c_i	GMV capacity of the i-th seller
w_i	importance of the i-th seller to the platform
$x_{i,j}$	PV proportion of the i-th seller in the j-th channel
$x^0_{i,j}$	original $x_{i,j}$
$u_{k,j}$	importance of the j-th channel to the k-th category
$\delta^0_{i,j}$	original PV of the i-th seller in the j-th channel
$\delta^1_{i,j}$	upper bound PV of the i-th seller in the j-th channel

and brings terrible user experience. We draw a diagram to show the PV traffic allocation over multiple sellers and multiple channels in Figure 2. Users can access items through various channels, for example, personalized recommendation channel 1 and search suggest drop-down list channel 2 in the diagram. Sellers have different influences on different channels. MCSTA allocates PV for multiple channels and multiple sellers and expose appropriate sellers to users, to optimize global GMV. Table 1 lists all the main symbols and definitions used throughout in the MCSTA model.

We formulate the seller traffic control model for multiple sellers and multiple channels as a joint optimization problem. To balance sellers' capacities and users' preferences to achieve optimization at the global scale, we define the loss function with three terms.

The first term focuses on maximizing the global GMV. In a real e-commerce platform, one seller's traffic usually concentrates in one category. In the "Tmall.com", each seller belongs to only one category. So we can model a seller's traffic in each category separately. The global GMV is defined as follows:

$$\sum_k \sum_{i \in N_k} \sum_j v_{i,j} p_{i,j} \tag{1}$$

where N_k is the set of sellers belonging to the k-th category, $v_{i,j}$ indicates the unit PV value of the i-th seller in the j-th channel, and $p_{i,j}$ indicates the number of PV of the i-th seller in the j-th channel.

We introduce the allocation factor $x_{i,j}$ representing the proportion of the i-th seller in the j-th channel, and the traffic capacity t_k representing the upper boundary of the number of PV for the k-th category. The first term of the objective function is written as:

$$L_{\text{max gmv}} = -\sum_k \sum_{i \in N_k} \sum_j v_{i,j} t_k x_{i,j} \tag{2}$$

The second term penalizes the oversell, i.e. it penalizes sellers whose GMV exceeds their own capacity to avoid "sold out":

$$L_{\text{min so}} = \sum_k \sum_{i \in N_k} w_i \left[t_k \sum_j v_{i,j} x_{i,j} - c_i \right]_+ \tag{3}$$

where c_i is the GMV capacity of the i-th seller, and w_i is the weight of i-th seller indicating its importance to the e-commerce platform. $t_k \sum_j v_{i,j} x_{i,j}$ represents the GMV of the i-th seller under the optimized allocation. Here we use the hinge loss $[.]_+$ to only penalize the sellers whose optimized GMV is in excess of their GMV capacity. The hinge loss function is defined as follows:

$$[z]_+ = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases} \tag{4}$$

Finally, we design a regularization term to reduce the gap between the new allocation and the original one. Its goal is to make sure that the optimal solution does not deviate too much from the original distribution, so as not to cause volatility or uncertainty. The third term of the loss function is defined as follows:

$$L_{\text{min var}} = \sum_k \sum_{i \in N_k} \sum_j t_k (x_{i,j} - x^0_{i,j})^2 \tag{5}$$

where $x^0_{i,j}$ denotes the original proportion of the i-th seller through the j-th channel without traffic control.

Combining the three terms above, we now present the objective function of the MCSTA model together with the constraints as follows:

$$\underset{x}{\arg\min} \quad L_{\text{max gmv}} + \lambda_1 L_{\text{min so}} + \lambda_2 L_{\text{min var}} \tag{6}$$

s.t.

$$\frac{\delta^0_{i,j}}{t_k} \leq x_{i,j} \leq \frac{\delta^0_{i,j} + u_{k,j}(\delta^1_{i,j} - \delta^0_{i,j})}{t_k} \tag{7}$$

$$\sum_{i \in N_k} \sum_j x_{i,j} = 1 \tag{8}$$

where

$$t_k = \sum_{i \in N_k} \sum_j \delta^0_{i,j} + u_{k,j}(\delta^1_{i,j} - \delta^0_{i,j}) \tag{9}$$

Equation 7 is the constraint to ensure that allocation $x_{i,j}$ does not exceed the upper bound of the proportion for the i-th seller through the j-th channel, and that allocation $x_{i,j}$ is not worse than the proportion in the original distribution. Here, we introduce a weight of channel, $u_{k,j}$, indicating the importance of the j-th channel to the k-th category. $u_{k,j}$ is obtained by historical conversion rate. Here, $\delta^0_{i,j}$ denotes the original number of PV of the i-th seller through the j-th channel, $\delta^1_{i,j}$ is the upper bound number of PV of the i-th seller through the j-th channel, and $u_{k,j}$ is the upper bound weight for the k-th category from the j-th channel. Equation 8 is the constraint that the traffic allocation of sellers and channels in the k-th category obeys the normalization condition. Equation 9 is to formulate that the limitation of the k-th category PV is determined by $u_{k,j}$, the importance of the j-th channel to the k-th category.

4 GENERAL SOLUTION TO MCSTA

The MCSTA model proposed above is a non-smooth convex optimization problem with constraints [10]. We propose a general solution based on the augmented Lagrangian method [5] and the proximal gradient method [15] to solve MCSTA.

We divide the MCSTA objective function into two parts, a differentiable term $f(x)$, and a non-differentiable term $g(x)$,

$$L(x) = \min_x f(x) + g(x) \tag{10}$$

$$f(x) = -\sum_k t_k \sum_{i\in N_k} \sum_j v_{i,j}x_{i,j} + \lambda_2 \sum_k t_k \sum_{i\in N_k} \sum_j (x_{i,j} - x^0_{i,j})^2 \tag{11}$$

$$g(x) = \lambda_1 \sum_k \sum_{i\in N_k} w_i \left[t_k \sum_j v_{i,j}x_{i,j} - c_i \right]_+ \tag{12}$$

Since different categories are independent of each other, we can split the problem into independent sub-problems.

For simplicity, we use $l_{i,j}$ and $h_{i,j}$ to represent lower bound and upper bound of each $x_{i,j}$ respectively. $|N_k|$ is the number of sellers in the k-th category. M is the number of channels. The vectorized representation is as follows:

$$\min_x x^T\lambda_2 x - (v^T + 2\lambda_2 {x^0}^T)x + w^T\,[Hx - c]_+ \tag{13}$$

$$s.t. \quad l \le x \le h \tag{14}$$

$$d^T x = 1 \tag{15}$$

where,

$$x = \left(x_{1,1}, \cdots, x_{1,M}, x_{2,1}, \cdots, x_{|N_k|,M}\right)^T \tag{16}$$

$$v = \left(v_{1,1}, \cdots, v_{1,M}, v_{2,1}, \cdots, v_{|N_k|,M}\right)^T \tag{17}$$

$$x^0 = \left({x^0}_{1,1}, \cdots, {x^0}_{1,M}, {x^0}_{2,1}, \cdots, {x^0}_{|N_k|,M}\right)^T \tag{18}$$

$$w = \lambda_1 \left(w_1, \cdots, w_{|N_k|}\right)^T \tag{19}$$

$$c = \left(\frac{c_1}{t_k}, \frac{c_2}{t_k}, \cdots, \frac{c_{|N_k|}}{t_k}\right)^T \tag{20}$$

$$l = \left(l_{1,1}, \cdots, l_{1,M}, l_{2,1}, \cdots, l_{|N_k|,M}\right)^T \tag{21}$$

$$h = \left(h_{1,1}, \cdots, h_{1,M}, h_{2,1}, \cdots, h_{|N_k|,M}\right)^T \tag{22}$$

$$H = \begin{pmatrix} \overbrace{v_{1,1}, v_{1,2}, \cdots, v_{1,M}}^{M}, 0,0,0,0,0,0,0,0,0, \cdots, 0 \\ \vdots \\ 0,0,0,0,0,0,0,0,0 \cdots, 0, \underbrace{v_{|N_k|,1} \cdots, v_{|N_k|,M}}_{M} \end{pmatrix} \tag{23}$$

$$d = (1, 1, \cdots, 1)^T \tag{24}$$

The objective function is convex, because both the differentiable part $f()$ and the non-differentiable part $g()$ are convex.[2]

Let

$$\begin{aligned} L_\beta(x,\alpha) = {} & x^T\lambda_2 x - (2\lambda_2 {x_0}^T + v^T)x + w^T\,[Hx - c]_+ \\ & - \alpha(d^T x - 1) + \frac{\beta}{2}(d^T x - 1)^2 \end{aligned} \tag{25}$$

$$s.t. \qquad l \le x \le h \tag{26}$$

The primal problem can be written as

$$\min_x \max_\alpha L_\beta(x,\alpha) \tag{27}$$

The dual problem is

$$\max_\alpha \min_x L_\beta(x,\alpha) \tag{28}$$

The iteration to solve the problem is

$$\begin{cases} x^t = \prod_{[l,h]} \arg\min_x L_\beta(x,\alpha^t) \\ \alpha^{t+1} = \alpha^t - \beta(d^T x^t - 1) \end{cases} \tag{29}$$

Given that the function is non-smooth and convex, we use the proximal gradient method to solve the formula above. Let

$$f(x) = x^T\left(\lambda_2 I + \frac{\beta}{2}dd^T\right)x - \left(2\lambda_2 {x^0}^T + v^T + \alpha^t d^T + \beta d^T\right)x \tag{30}$$

$$g(x) = w^T\,[Hx - c]_+ \tag{31}$$

$$\frac{\partial f}{\partial x} = (2\lambda_2 I + \beta dd^T)x - (2\lambda_2 p + v + \alpha^t d + \beta d) \tag{32}$$

The iteration is

$$\begin{cases} y^r = x^r - \theta^r \dfrac{\partial f}{\partial x} \\ x^{r+1} = \prod_{[l,h]} \arg\min_x \left(g(x) + \dfrac{1}{2\theta^r}(x - y^r)^T(x - y^r)\right) \end{cases} \tag{33}$$

Let

$$Q = diag(q_1, q_2, \cdots, q_{|N_k|}), q_i \in [0,1] \tag{34}$$

The problem is equivalent to

$$\min_x \max_Q w^T Q(Hx - c) + \frac{1}{2}x^T x - (y^r)^T x \tag{35}$$

The dual problem is

$$\max_Q \min_x w^T Q(Hx - c) + \frac{1}{2}x^T x - (y^r)^T x \tag{36}$$

The closed form solution is

$$\begin{cases} q_i = \prod_{[0,1]} (HH^T)^{-1}(Hy^r - c)_i / w_i \\ x = \prod_{[l,h]} y^r - \theta^r H^T Q w \end{cases} \tag{37}$$

[2] https://github.com/sxstar/MCSTA/blob/master/Proof.pdf

5 EVALUATION METRICS AND HYPERPARAMETER SELECTION

5.1 Evaluation Metrics

5.1.1 Offline Evaluation Metrics. The results of the MCSTA model depend on the choice of its two hyperparameters. For external evaluation of the quality of an MCSTA solution and to set the MCSTA hyperparameters, we define a set of offline evaluation metrics to measure three aspects of the allocation:

GMV Growth Rate: The improvement of GMV with regard to the same PV. The growth rate of GMV can be evaluated by comparing the MCSTA model with the original traffic distribution, which is defined as:

$$r_g = \frac{\sum_{i,j} x_{i,j} v_{i,j}}{\sum_{i,j} x^0_{i,j} v_{i,j}} - 1 \tag{38}$$

where $x_{i,j}$ is obtained from the MCSTA model.

Capacity Waste Rate: The mean of traffic waste for sellers. To evaluate the allocation from the perspective of the overall consumption rate of sellers to avoid the "sold out" situation, we define the overall waste rate of sellers as follows:

$$r_c = \underset{k}{\mathrm{mean}}(\max_i(\frac{t_k \sum_j x_{i,j} v_{i,j}}{c_i}, 1) - 1) \tag{39}$$

Allocation Variance: We define the volatility variance to characterize the optimized allocation compared with the original distribution, in order to prevent traffic allocation from collapsing to the small proportion of the modes, where the partial sellers or channels will get too much traffic resulting in global sub-optimization. The definition of allocation variance as follows:

$$v_a = \sum_{i,j} (\frac{x_{i,j}}{x^0_{i,j}} - 1)^2 / (|\text{sellers}|) \tag{40}$$

5.1.2 Online Evaluation Metrics. To test the effectiveness of the MCSTA model in the online platform, "Tmall.com", we perform an A/B test, in which 3% of the users are randomly selected as the base group, and 97% as the test group. We deploy the framework to optimize the traffic allocation. The metric of the GMV growth rate is defined as follows:

$$\frac{\text{GMV(test)} - \text{GMV(base)}}{\text{GMV(base)}} \tag{41}$$

This metric, the key indicator of online promotion, compares the growth rates of GMV between the test group and the base group. The performance is measured in terms of the GMV growth rate and the capacity waste rate on the day of the online promotion. In addition, we also use regular business metrics, including PV growth rate and sales volume growth rate as evaluation metrics.

5.2 Multi-Objective Shortest Distance (MOSD) Hyperparameter Tuning Algorithm

Traditionally, the hyperparameters are chosen manually and subjectively, while we try to select the hyperparameters automatically during the course of optimizing MCSTA. According to the design of the MCSTA function, the two hyperparameters (λ_1, λ_2) will influence the final results of traffic allocation significantly.

```
Input: μ1, μ2, n, low, up, ε(r_g), ε(r_c), ε(v_a)
Output: The hyperparameters (λ1, λ2) for each category
Initialize n^2 (λ1, λ2)'s by randomly sampling the regular
  points in the [low, up]^2 plane;
foreach category sellers data do
    foreach (λ1, λ2) do
        calculate the optimization allocation x through the
          MCSTA model with (λ1, λ2);
        calculate the three evaluations metrics r_g, r_c, v_a;
        filter hyperparameters groups that do not satisfy the
          three thresholds of evaluation metrics
          (ε(r_g), ε(r_c), ε(v_a));
    end
    calculate the three index scores (s(r_g), s(r_c), s(v_a));
    calculate the distance Dist(x) from each point x to the
      ideal point x**(1, 1, 1);
    select and save the best hyperparameters group (λ1, λ2)
      with the shortest distance as the optimal
      hyperparameters in this category;
end
return the hyperparameters (λ1, λ2) for each category;
```

Algorithm 1: Multi-Objective Shortest Distance (MOSD) hyperparameter tuning algorithm

We divide it into independent optimization tasks by category. The difference among the categories leads to a group of hyperparameter settings which cannot be applied to other categories. Therefore, a hyperparameter automatic adjustment algorithm is essential. Based on grid search, we design a Multi-Objective Shortest Distance (MOSD) hyperparameter tuning algorithm to select the best hyperparameters for each category.

We normalize the original three evaluation metrics into mapping to achieve a quantitative comparison among different evaluation indicators. We define three metrics as follows:

$$s(r_g) = \frac{r_g - \min(r_g)}{\max(r_g) - \min(r_g)} \tag{42}$$

$$s(r_c) = \frac{r_c - \min(r_c)}{\max(r_c) - \min(r_c)} \tag{43}$$

$$s(v_a) = 1 - \frac{v_a - \min(v_a)}{\max(v_a) - \min(v_a)} \tag{44}$$

We introduce three thresholds, $\epsilon(r_g), \epsilon(r_c)$ and $\epsilon(v_a)$, to ensure that each term meets its minimal goal. In addition, filtering unsatisfied points is also a pruning strategy in the search of hyperparameters. We show the pseudo-code of MOSD in Algorithm 1.

Input n is the number of samples. For each hyperparameter, low and up indicate the lower bound and upper bound of the search. To balance three evaluation indicators, we define the distance function $\text{Dist}(x)$ to be the distance between the point $x \equiv (s(r_g), s(r_c), s(v_a))$ and the ideal point $x^{**} \equiv (1, 1, 1)$ as follows:

$$\text{Dist}(x) = \sqrt{(1 - s(r_g))^2 + \mu_1(1 - s(r_c))^2 + \mu_2(1 - s(v_a))^2} \tag{45}$$

where (μ_1, μ_2) are the preference coefficients to control the bias of the final selection of hyperparameters. For example, when $\mu_1 =$

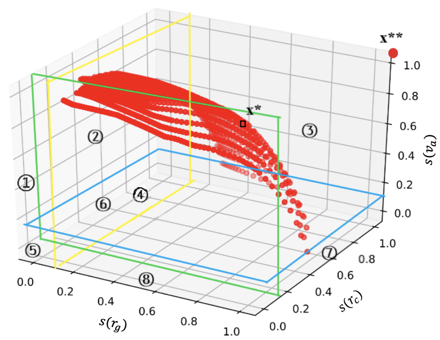

Figure 3: The multi-objective shortest distance hyperparameter tuning algorithm schematic

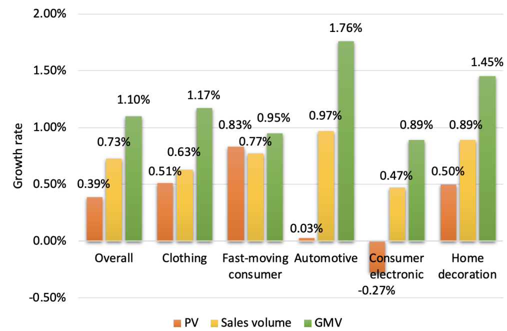

Figure 4: Online analysis of overall and category improvements on large-scale promotion

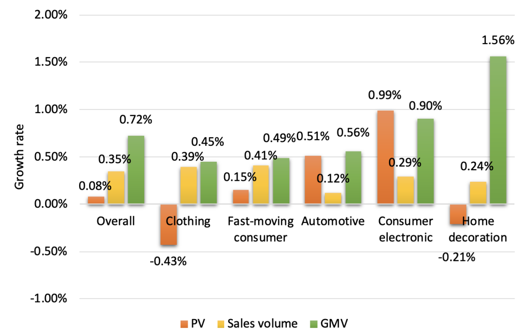

Figure 5: Online analysis of overall and category improvements on everyday promotion

$\mu_2 = 1$, overall effectiveness is attributed to three indicators equally. If μ_1 is bigger, allocation results are more biased to $s(r_c)$.

Figure 3 illustrates the MOSD algorithm. The three pre-set thresholds divide the space into eight quadrants. Point x^{**} is the ideal point in three-dimensional space. We define the Dist(x) function to measure the superiority of each hyperparameters group. We also add two coefficients μ_1 and μ_2 to the distance function, which can be flexibly determined according to business requirements. By flexibly adjusting the coefficients, the "optimal point" x^*, i.e. the best hyperparameters, can be found closest to the ideal point.

6 EXPERIMENTS AND RESULTS

To evaluate the effectiveness of the proposed framework, we deployed it on a large real-world e-commerce platform, "Tmall.com", to optimize the allocation over multi-channel and multi-seller. When reporting results, showing specific numbers may leak seller and user privacy. Therefore, we statistically show the results without involving personalized numbers.

6.1 Case Statement

This experiment is based on Alibaba's "Global Shopping Festival" on "Tmall.com". There are hundreds of millions of users and items in the platform. And over one billion transactions occurred during the promotion. We select 4 channels (search suggest drop-down list, search ranking, shopping cart recommendation and homepage recommendation) and more than 700 sellers from 5 categories in the platform. Meanwhile, we deploy a real-time data platform online to get real-time data such as GMV and PV of these sellers. And the value of PV is calculated by the ratio between the actual turnover of the seller and the actual PV. As we expected, the unit PV value varies greatly among different sellers and channels, and the initial PV proportion varies greatly among different sellers and channels, including some sellers with huge sales volumes.

First, we set the value of (μ_1, μ_2), preference coefficients of the Dist(x) function. Usually we set $\mu_1 = \mu_2 = 1$ with no special business settings, and then an optimal set of hyperparameters for each category is calculated by the MOSD algorithm. The optimal allocation results are obtained by inputting the optimal set of hyperparameters into the MCSTA algorithm to obtain the optimal allocation.

For the MCSTA model, timing interval is set to recalculate the optimal allocation by retrieving real-time data per hour. Finally, optimal allocation of the PV target is divided into hours granularity and fed to each channel to complete the control task.

6.2 Hyperparameters Setting

In the process of randomly sampling parameters, we first set up a set of λ_1 and λ_2 to randomly collect n. Here we set $n = 20$, then we obtain 400 groups of random hyperparameters. Through the MOSD function, we get a set of optimal hyperparameters C1 $=(\lambda_1', \lambda_2')$. Then, in a small range near C1, resample by $n = 20$, call the MOSD method again to obtain a set of optimal hyperparameters C2 $=(\lambda_1'', \lambda_2'')$. Because MCSTA performs threshold filtering on all hyperparameter groups in the second step, even if the three coefficients are not set reasonably, the final result would not exceed the acceptable range.

Assuming that here we have no special preference for the business bias, we set $\mu_1 = \mu_2 = 1$, and the number of samples is $20 * 20$, which is iterated in two rounds. The results of hyperparameter groups found for each category are shown in Table 2.

6.3 Results of Offline Experiments

We deploy MCSTA on "Tmall.com" to implement offline experiments. To the best of our knowledge, this is the first work to allocate sellers traffic among multiple channels to increase sales volume. We

Table 2: Results of hyperparameters experiments

Category ID	1	2	3	4	5
λ_1	40.74	0.27	5.57	2.86	0.72
λ_2	1.29	0.12	1.29	1.82	0.18

Table 3: Offline evaluation for large-scale promotion

Evaluation metrics	r_g	r_c	v_a
Overall	12.7%	23.9%	0.085

compare our model with current "Tmall.com" which has systematical business intelligence, including personal recommendation and personal search engine. The hyperparameters are obtained based on the MOSD hyperparameters tuning method. According to the three offline indicators, as Table 3 shows, the MCSTA model improves the GMV by 12.7%. The capacity waste rate decreases from 25.1% to 23.9%. The variance of the fluctuation is only 0.085.

The growth of GMV directly demonstrates that MCSTA facilitates more transactions and brings more commercial values. The decrease in capacity waste rate shows false navigation to sellers with unsatisfied capacity is reduced and customer experience is improved.

6.4 Results of Online Experiments

We analyze the online log data from the perspective of the whole platform and five categories, such as clothing, fast-moving consumer, automotive, consumer electronic, and home decoration.

In the process of group splitting evaluation, there is some fluctuation between base groups and test groups. Based on commercial experience, we consider that optimization effect is reliable when the GMV growth rate exceeds 0.5%.

The overall optimization effect is shown in Figure 4, in which the horizontal axis is the total and five categories subject to optimization testing, and the vertical axis is the GMV growth rate compared with the base group.

The platform PV and sales volume increase by 0.39% and 0.73%, respectively. The MCSTA also improves the overall GMV by 1.1%.

From the analysis of the category effect, the model has a positive effect on the volume of transactions in various categories, indicating that the allocation meets user's preference needs, and the increase in the transaction amount is more obvious.

In particular, GMV improvement ratios of the apparel, automobile, and home decoration categories exceed 1%.

We notice that the difference between offline and online experiments in terms of the GMV growth rate is large. The reason is as follows: In the offline experiment, the PV value is constant, and all users visit the items recommended by the model. In the online experiment, the PV value is being updated in real time, and some users may ignore the recommendation.

6.5 Deployments

MCSTA has been fully deployed on Alibaba's online scheduling platform since September 2019. It connects to two services, one is a real-time data engine, the other is a real-time inventory measurement system. MCSTA periodically obtains signals from these two services and allocates sellers traffic among multiple channels.

Table 4: Offline evaluation for everyday promotion

Evaluation metrics	r_g	r_t	v_a
Overall	7.9%	81.1%	0.098

7 FURTHER APPLICATIONS

7.1 Everyday Promotion

Even though large-scale online promotions play an important role in e-commerce, everyday promotion is more common. In everyday promotion, the "sold out" situation rarely happens. Instead of sellers' capacity limits, everyday promotion's participants often set a sales target. More sellers achieving their sales target indicates more successful everyday promotion.

In this case, we modify the second term of the MCSTA loss function as:

$$L_{\text{max target}} = \sum_k \sum_{i \in N_k} w_i \left[g_i - t_k \sum_j v_{i,j} x_{i,j} \right]_+ \tag{46}$$

where g_i is the sales goal of the i-th seller. Here we use the hinge loss $[.]_+$ to only penalize the sellers whose optimized GMV does not achieve their GMV target.

To evaluate the effectiveness of the MCSTA model, we introduce another evaluation metric, sales target completion rate:

$$r_t = \underset{k}{\text{mean}}(\min_i(\frac{t_k \sum_j x_{i,j} v_{i,j}}{g_i}, 1)) \tag{47}$$

We deploy the MCSTA model in "Tmall.com" for testing everyday promotion scenario. Experimental settings are similar to the previous experiments on the large-scale promotion. We select 4 channels and 285 sellers from 5 categories in the platform.

Offline experimental results are shown in Table 4. GMV increases by 7.9% after sellers traffic allocation. The sales target completion rate improves from 71.4% to 81.1%, while the variance of the fluctuation is only 0.098. Online experimental results are shown in Figure 5. The result of the A/B test shows that MCSTA brings a 0.72% improvement in terms of the overall GMV. Sales volume and GMV grow among all five categories. The reason for the difference between online and offline scenarios is similar to Section 6.4.

7.2 Video Live Stream Service in E-commerce

Video live stream service has been experiencing rapid development these years. Inspired by television advertisement, some e-commerce platform offers embedded video live stream service to extend their Internet commercial market. Customers can conveniently purchase an item by clicking the icon or link that appeared on the live system.

Typically, video live stream service providers benefit from advertisements and interactions between the audiences and video anchors. For e-commerce platforms with video live stream service, they also benefit from the increase of exposure of items and number of transactions.

We attempt to solve the "cold start" problem in the video live stream service scenario by the proposed MCSTA model. The first and second terms of the MCSTA loss function can be modified as:

$$L_{\max\text{tp}} = -\sum_{k}\sum_{i\in N_k}\sum_{j} p_{i,j} t_k x_{i,j} \tag{48}$$

$$L_{\max\text{target}} = \sum_{k}\sum_{i\in N_k} w_i \left[g_i - t_k \sum_{j} p_{i,j} x_{i,j} \right]_+ \tag{49}$$

where $p_{i,j}$ is the average time on page of the i-th host in the j-th channel, g_i is the visiting goal of the i-th host. Here we use the hinge loss $[.]_+$ to only penalize the hosts whose optimized time on page does not achieve their target.

To evaluate the effectiveness of the model, we introduce an evaluation metric, host's target completion rate. The definition of host's target completion rate is similar to the sales target completion rate in Equation 47. We assign a lower bound of total visiting time for each host to ensure that every video anchor has rational base traffic. New video anchors are able to be exposed to more audience.

We deploy the model in the embedded video live stream service on "Tmall.com". We design an online A/B test with tens of thousands of hosts in the real-world application. We allocate the PV through 3 different channels: search, recommendation and slide switch. The result of the A/B test shows that MCSTA brings an improvement of the host's goal completion rate from 48.3% to 70.4%.

8 CONCLUSIONS

In this paper, we investigate the novel problem of systematically improving global GMV in a large-scale e-commerce promotion by allocating PV distribution over multiple channels. We propose a Multi-Channel Sellers Traffic Allocation (MCSTA) optimization model to achieve optimal PV allocation for each seller and each channel. A general constrained non-smooth convex optimization solution with a Multi-Objective Shortest Distance (MOSD) hyperparameter tuning method is presented to solve MCSTA. The model demonstrates its superiority over original PV distribution through offline and online evaluations on one of the world's largest e-commerce platforms, and brings remarkable objective profit boosts to sellers. Specifically, empirical results show that MCSTA achieves a significant improvement of GMV by 1.1% based on the A/B test during Alibaba's "Global Shopping Festival". Moreover, we successfully deploy the multi-objective optimization model to other popular scenarios, including everyday promotion and video live stream service, demonstrating that MCSTA can be applied to other multi-objective optimization problems.

ACKNOWLEDGMENTS

This work was partly conducted within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc. and the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme.

We would like to thank Alibaba Group for valuable datasets and platforms and Alibaba-Zhejiang University Joint Research Institute of Frontier Technologies (AZFT) for technical support.

SHEN Xin would like to express the deepest appreciation to Jingyi Zhang for her continued assistance and encouragement.

REFERENCES

[1] Deepak Agarwal, Bee-Chung Chen, Pradheep Elango, and Xuanhui Wang. 2011. Click shaping to optimize multiple objectives. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*. 132–140.

[2] Jiajun Bu, Xin Shen, Bin Xu, Chun Chen, Xiaofei He, and Deng Cai. 2016. Improving collaborative recommendation via user-item subgroups. *IEEE Transactions on Knowledge and Data Engineering* 28, 9 (2016), 2363–2375.

[3] Qingpeng Cai, Aris Filos-Ratsikas, Pingzhong Tang, and Yiwei Zhang. 2018. Reinforcement Mechanism Design for e-commerce. In *Proceedings of the 2018 World Wide Web Conference*. 1339–1348.

[4] Shi-Yong Chen, Yang Yu, Qing Da, Jun Tan, Hai-Kuan Huang, and Hai-Hong Tang. 2018. Stabilizing reinforcement learning in dynamic environment with application to online recommendation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1187–1196.

[5] Andrew R Conn, Nicholas IM Gould, and Philippe Toint. 1991. A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM J. Numer. Anal.* 28, 2 (1991), 545–572.

[6] Jun Feng, Heng Li, Minlie Huang, Shichen Liu, Wenwu Ou, Zhirong Wang, and Xiaoyan Zhu. 2018. Learning to collaborate: Multi-scenario ranking via multi-agent reinforcement learning. In *Proceedings of the 2018 World Wide Web Conference*. 1939–1948.

[7] Rahul S Gaikwad, Sandeep S Udmale, and Vijay K Sambhe. 2018. E-commerce Recommendation System Using Improved Probabilistic Model. In *Information and Communication Technology for Sustainable Development*. Springer, 277–284.

[8] Yujing Hu, Qing Da, Anxiang Zeng, Yang Yu, and Yinghui Xu. 2018. Reinforcement learning to rank in e-commerce search engine: Formalization, analysis, and application. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 368–377.

[9] Shubhra Kanti Karmaker Santu, Parikshit Sondhi, and ChengXiang Zhai. 2017. On application of learning to rank for e-commerce search. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 475–484.

[10] Claude Lemaréchal. 2006. S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004 hardback, 65 US, ISBN 0 521 83378 7.

[11] Zhao Li, Xin Shen, Yuhang Jiao, Xuming Pan, Pengcheng Zou, Xianling Meng, Chengwei Yao, and Jiajun Bu. 2020. Hierarchical Bipartite Graph Neural Networks: Towards Large-Scale E-commerce Applications. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 1677–1688.

[12] Shichen Liu, Fei Xiao, Wenwu Ou, and Luo Si. 2017. Cascade ranking for operational e-commerce search. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1557–1565.

[13] Weike Pan. 2016. A survey of transfer learning for collaborative recommendation with auxiliary data. *Neurocomputing* 177 (2016), 447–453.

[14] Mirko Polato and Fabio Aiolli. 2017. Exploiting sparsity to build efficient kernel based collaborative filtering for top-N item recommendation. *Neurocomputing* 268 (2017), 17–26.

[15] Mark Schmidt, Nicolas L Roux, and Francis R Bach. 2011. Convergence rates of inexact proximal-gradient methods for convex optimization. In *Advances in neural information processing systems*. 1458–1466.

[16] Jizhe Wang, Pipei Huang, Huan Zhao, Zhibo Zhang, Binqiang Zhao, and Dik Lun Lee. 2018. Billion-scale commodity embedding for e-commerce recommendation in alibaba. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 839–848.

[17] Shen Xin, Martin Ester, Jiajun Bu, Chengwei Yao, Zhao Li, Xun Zhou, Yizhou Ye, and Can Wang. 2019. Multi-task based Sales Predictions for Online Promotions. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 2823–2831.

[18] SHEN Xin, Hongxia Yang, Weizhao Xian, Martin Ester, Jiajun Bu, Zhongyao Wang, and Can Wang. 2018. Mobile Access Record Resolution on Large-Scale Identifier-Linkage Graphs. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 886–894.

[19] Wayne Xin Zhao, Sui Li, Yulan He, Edward Y Chang, Ji-Rong Wen, and Xiaoming Li. 2015. Connecting social media to e-commerce: Cold-start product recommendation using microblogging information. *IEEE Transactions on Knowledge and Data Engineering* 28, 5 (2015), 1147–1159.

[20] Han Zhu, Xiang Li, Pengye Zhang, Guozheng Li, Jie He, Han Li, and Kun Gai. 2018. Learning tree-based deep model for recommender systems. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 1079–1088.

[21] Yu Zhu, Junxiong Zhu, Jie Hou, Yongliang Li, Beidou Wang, Ziyu Guan, and Deng Cai. 2018. A brand-level ranking system with the customized attention-GRU model. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3947–3953.

[22] Tao Zhuang, Wenwu Ou, and Zhirong Wang. 2018. Globally optimized mutual influence aware ranking in e-commerce search. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*. 3725–3731.