



A prediction-oriented optimal design for visualisation recommender systems

Yingyan Zeng, Xinwei Deng, Xiaoyu Chen & Ran Jin

To cite this article: Yingyan Zeng, Xinwei Deng, Xiaoyu Chen & Ran Jin (2021) A prediction-oriented optimal design for visualisation recommender systems, *Statistical Theory and Related Fields*, 5:2, 134-148, DOI: [10.1080/24754269.2021.1905376](https://doi.org/10.1080/24754269.2021.1905376)

To link to this article: <https://doi.org/10.1080/24754269.2021.1905376>



Published online: 30 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 660



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



A prediction-oriented optimal design for visualisation recommender systems

Yingyan Zeng^a, Xinwei Deng ^b, Xiaoyu Chen^a and Ran Jin^a

^aGrado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA; ^bDepartment of Statistics, Virginia Tech, Blacksburg, VA, USA

ABSTRACT

A good visualisation method can greatly enhance human-machine collaboration in target contexts. To aid the optimal selection of visualisations for users, visualisation recommender systems have been developed to provide the right visualisation method to the right person given specific contexts. A visualisation recommender system often relies on a user study to collect data and conduct analysis to provide personalised recommendations. However, a user study without employing an effective experimental design is typically expensive in terms of time and cost. In this work, we propose a prediction-oriented optimal design to determine the user-task allocation in the user study for the recommendation of visualisation methods. The proposed optimal design will not only encourage the learning of the similarity embedded in the recommendation responses (i.e., users' preference), but also improve the modelling accuracy of the similarities captured by the covariates of contexts (i.e., task attributes). A simulation study and a real-data case study are used to evaluate the proposed optimal design.

ARTICLE HISTORY

Received 8 November 2020
Revised 28 February 2021
Accepted 16 March 2021

KEYWORDS

Design of experiments;
space-filling design;
user-task allocation;
V-optimality; visualisation
recommender system

1. Introduction

Visualisation methods have been widely investigated to graphically represent complex data sets for effective comprehension and generation of insights (Spence, 2001). With multiple visualisation methods available (e.g., use of colour, layout and labelling, user-system interaction, visual encodings), the effectiveness and efficiency of a visualisation is largely determined by the contexts (e.g., tasks, environments, devices, modalities) and individual differences (e.g., personal preference, perceptual and cognitive capabilities) (Bednarik et al., 2012; Chen & Jin, 2017; Voigt et al., 2012). Beyond visualisation, recommender systems have been widely adopted to provide personalised suggestions based on attributes of the contexts and historical information of users' interests. Therefore, in the literature, recommender systems for personalised visualisation have been developed to provide the right visualisation methods to the right people given specific contexts. Among the various approaches of visualisation recommender systems, there are three major categories: (i) content-based filtering that predicts the recommendation scores (e.g., user's rating, preference score, completion time, cognitive workload, etc.) based on the similarity among users and contexts captured by the corresponding covariates (Bogdanov et al., 2013); (ii) collaborative filtering that makes the recommendation based on the similarity hidden in the historical responses (Kunkel et al., 2017); and (iii) a hybrid approach as a combination of (i) and (ii), which considers both similarities

from the users and contexts of a visualisation, as well as the historical rating of a visualisation method (Mutlu et al., 2016).

This research focuses on the hybrid approach for the personalised visualisation recommendation by formulating the problem as a recommender system considering both responses and covariates. Specifically, the recommendation score in a visualisation recommender system is modelled as a compound response of (i) collaborative information that quantifies the implicit similarities among users' ratings over tasks and visualisation methods, and (ii) content-based information that quantifies the explicit similarities represented by covariates (i.e., user input logs and task attributes). It is commonly assumed that the implicit similarities can be represented by a latent low-rank matrix, and the relation between covariates and the content-based response can be modelled by a linear regression term. For example, Chen et al. (2020) developed a Personalised Recommender system for Information visualisation Methods via Extended matrix completion (PRIME) for the recommendation of optimal visualisation methods catering to users' characteristics and contexts.

The development of such a recommender system relies on the collection of training data including users' and tasks' information as covariates, and the users' ratings on tasks as the response. Here we assume there is no learning effect for users conducting tasks in order. However, the collection of training data by the user study is often an expensive process in terms of time and

cost. A good offline experimental design is therefore desired to optimise the user-task allocation to collect data efficiently and effectively for the user study. With a limited budget on the number of experiments, it is crucial to design an optimal user-task allocation such that the collected training data is highly informative for constructing a recommender system.

The objective of this work is to create an optimal experimental design, i.e., an optimal user-task allocation, to enhance the recommendation accuracy of a visualisation recommender system. Here we focus on constructing experimental designs to collect the initial training data set offline from user studies. To this end, the offline Design of Experiments (DoE) are commonly used for similar design problems. For example, optimal designs, based on some optimality criteria, have been applied in collaborative filtering methods to evaluate a new item on existing users (Anava et al., 2015). Another design criterion, space-filling design criterion, is commonly adopted in computer simulations with design points evenly spread over the experiment region (Joseph, 2016). For the matrix completion as techniques for recommender systems, space-filling designs were applied as an initial design strategy to sample the initial observed entries for accurate recovery of the low-rank matrix (Mak & Xie, 2017). However, these design methods either focus on the implicit similarities in the low-rank response matrix or the explicit similarities in the covariates. Hence, they may not work well in the visualisation recommender system due to the lack of joint consideration for both types of similarities.

To overcome the aforementioned challenges, we propose a prediction-oriented optimal design to collect an informative training data set for the visualisation recommender system through jointly considering the collaborative information (implicit similarity) and content-based information (explicit similarity). As both types of similarities are important for the visualisation recommendation, the proposed optimal design can efficiently enhance the performance of the recommender system. The key idea of the proposed design is to first explore the overall design space to find the candidate design region to recover the low-rank matrix by space-filling design criterion, and then exploit this candidate design region by the optimality criterion such that the selected design achieves superior prediction accuracy over all candidate designs. The contributions of the proposed method are in two folds. First, a constrained optimality criterion is proposed to balance the optimal design for both the low-rank matrix response and the linear response decomposed from the recommendation score. This lays a foundation of initial training data sampling for the recommender system with a limited number of experiments. Second, an efficient design construction algorithm is proposed to find the optimal design to achieve the high recommendation accuracy over user and context space in the practical application.

The remainder of this paper is organised as follows. Section 2 provides a literature review. Section 3 gives a brief introduction to the PRIME model and then details the proposed prediction-oriented optimal design. Section 4 evaluates the performance of the proposed method by a numerical study. Section 5 validates the proposed design via a real case study of the visualisation recommender system. We conclude this work with some discussions in Section 6.

2. Literature review

In this section, we review existing recommender systems with a focus on visualisation recommendation, and experimental design methods for recommender systems.

2.1. Recommender systems and applications in visualisations

Recommender systems have become ubiquitous in e-commerce to generate customised recommendations for users (Ricci et al., 2015). In the context of visualisation recommendation, a visualisation recommender system aims to provide accurate suggestions of visualisation methods for users given specific contexts. Existing methods for visualisation recommender systems broadly fall into three categories: (i) content-based filtering; (ii) collaborative filtering; and (iii) hybrid filtering (Jones et al., 2015; Vartak et al., 2017).

The content-based filtering matches the attributes of users' profiles with those of items. Hence, visualisation recommender systems built by content-based filtering predict the recommendation score based on the covariates of users and contexts (Mutlu et al., 2016). However, the content-based filtering cannot provide satisfactory performance when limited information is available from the covariates. In contrast, collaborative filtering-based recommender systems suggest items based on the implicit similarity of a group of users on a set of items among the historical users' ratings. With a small sample size, the implicit similarity is usually modelled by a low-rank matrix with random Gaussian noise, and thus formulated as a matrix completion problem (Candès & Recht, 2009). But the collaborative filtering approach suffers from the so-called the cold-start problem (Schein et al., 2002). The cold-start problem in the context of visualisation recommendation refers to providing recommendations for new users or visualisation tasks without historical responses available, where the pure collaborative filtering cannot provide good results (Melville & Sindhvani, 2010; Schein et al., 2002). Therefore, a hybrid filtering method considering both explicit and implicit similarities is desired to leverage the advantages of both methods. For example, Mutlu et al. (2016) developed VizRec as a hybrid visualisation recommender system considering the information

of both the recommendation context and score via a weighted combination. However, the ambiguity of weights limits its practical application. In PRIME (Chen et al., 2020), the recommendation score is formulated as the compound response of two kinds of similarities, i.e., a low-rank matrix and a covariate-based linear regression term. In this work, we adopt PRIME, a more generic hybrid approach, as the underlying model for the visualisation recommender system.

2.2. Experimental designs for recommender systems

In the literature, few works focussed on offline experiment designs for recommender systems to collect informative training data under limited resources (e.g., limited runs, limited budget for user study, limited time). Most of the existing methods fall into the offline experimental design, including optimal designs and space-filling designs. The optimal designs belong to a class of experimental designs with respect to some statistical criteria (Pukelsheim, 2006). Several prediction-oriented criteria have been developed in the optimal design literature, including I-optimality criterion for minimising the average variance in prediction over the entire design space (Box & Draper, 1959), G-optimality criterion for minimising the maximum variance of the estimation across the design space (Kiefer & Wolfowitz, 1959), and V-optimality criterion for minimising the integrated variance over the region of interest within the design space (Laake, 1975). In terms of their related application to recommender systems, Anava et al. (2015) formulated the cold-start problem in collaborative filtering as a model-based optimal design to minimise the prediction error based on A-optimality criterion. However, their method only considers the optimality of the linear interaction between users and items captured by the latent factor model. Bhat et al. (2020) adopted optimal design for the offline user-treatment allocation problem in A/B testing under the linear relationship between the observed covariates of treatments and their impact on users. Similarly, this design only considers the explicit similarities represented by covariates. Thus, these methods may not work well for the visualisation recommendation due to their lack of joint consideration of both important similarities.

Another DoE strategy that has been adopted in related experimental design problems in recommender systems is the space-filling design (Joseph, 2016; Santner et al., 2003). Intuitively, such a design is suitable for enhancing the prediction accuracy over the design region with deterministic response since the prediction error can be considered as a function of its location relative to the design point. There is no direct application of this design criterion to recommender systems. As a related work, the idea of space-filling design criterion

has been adopted by Mak and Xie (2017) for an active matrix completion problem, which actively selects the entries to be observed to recover the low-rank matrix (Chakraborty et al., 2013; Ruchansky et al., 2015). Their study validated the effectiveness of the balanced property in the initial design (see the detail in Section 3), and as a design with balanced property, the space-filling design was thus adopted as the initial sampling strategy in the proposed active sampling scheme. Although the use of space-filling designs can enhance the estimation of low-rank matrix, it may not work well for the recommender system when the content-based information (i.e., linear response) is dominant in users' ratings and the number of runs of experiments is small. Moreover, without an accurate predictive model obtained by an initial training data set, it is difficult for the active sampling strategy to quantify the information of the entry to be observed on the matrix. Thus, the online schema does not fit the offline design problem. Except active matrix sampling, there is few work focussing on the offline design of optimal observation allocation for matrix completion in the context of recommendation problems.

In summary, there is limited research on the experimental design for offline initial training of recommender systems. The proposed method could be a pioneer work on the optimal experimental design, i.e., user-task allocation, for the user study of visualisation recommender systems catering to both implicit and explicit similarities.

3. The proposed prediction-oriented optimal design

As discussed in Section 2, an offline prediction-oriented optimal design criterion is desired to consider both low-rank matrix estimation and the linear estimation in order to enhance the recommendation performance over the user and context space. In this section, we first formulate the user-task allocation in the user study as an optimal design problem. Then we detail the prediction-oriented optimal design and the design construction algorithm.

3.1. Problem setup

Suppose that $Y = (y_{i,j})_{U \times T}$ is the response matrix, where U is the number of users and T is the number of tasks. That is, $y_{i,j}$ is the response to be obtained when the i th user is assigned to complete the j th task. The value of $y_{i,j}$ represents the recommendation score (e.g., user's rating, preference score, cognitive workload). Denote the whole indices of the response matrix Y as $\Omega = \{(i,j), i = 1, \dots, U; j = 1, \dots, T\} = [U] \times [T]$ which contains UT design points. The user-task allocation is to select an appropriate size- S subset $\Omega_S \subset \Omega$ such that experimenters can assign users to complete

the selected tasks to obtain the responses. Specifically, we define the design indicator variable $z_{i,j}$ as

$$z_{i,j} = \begin{cases} 1, & \text{the } i\text{th user is assigned to complete the} \\ & \text{}j\text{th task;} \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Then the response entry $y_{i,j}$ will be collected by conducting the corresponding experiment with $z_{i,j} = 1$. The user-task allocation is to select $\Omega_S = \{(i,j) : z_{i,j} = 1, i = 1, \dots, U; j = 1, \dots, T\}$.

Denote $\mathbf{x}_{i,j}$ to be a p -dimensional covariate vector (e.g., users' input logs, task's information) associated with the i th user and the j th task, which is observable before conducting the experiment. Thus, the whole set of covariate vectors can be written as $\mathbf{X} \in \mathbb{R}^{UT \times p}$ with each row being $\mathbf{x}_{i,j} \in \mathbb{R}^p$, $i = 1, \dots, U; j = 1, \dots, T$. Note that $\mathbf{x}_{i,j}$ is a column vector. Here we assume \mathbf{X} to be available before designing the experiment since they are gathered from the candidate users and candidate tasks. To prevent one user from conducting one task for multiple times, we do not consider the repetition in the design. Table 1 illustrates an example of one possible design.

As an experimental design problem, instead of optimising the level of factors (i.e., the value of $\mathbf{x}_{i,j}$), here our goal is to select the optimal design runs (i.e., $z_{i,j} = 1$) according to the covariates such that the obtained experimental outcomes can be used to accurately predict the full response matrix \mathbf{Y} . Clearly, such an optimal design will depend on the model in use for quantifying the relationship between the covariates and the response. Here we consider the PRIME model based on matrix completion with the consideration of covariates information:

$$\mathbf{Y} = \mathbf{R} + \mathcal{A}(\mathbf{X}\beta) + \mathbf{E}, \quad (2)$$

where $\mathbf{E} = (\epsilon_{i,j})_{UT \times p}$ is an error matrix with $\epsilon_{i,j} \stackrel{i.i.d}{\sim} N(0, \sigma^2)$, $\mathbf{R} \in \mathbb{R}^{U \times T}$ is the low-rank matrix modelling the implicit similarity, $\beta \in \mathbb{R}^p$ are the linear regression coefficients quantifying the explicit similarity among covariates \mathbf{X} . Here $\mathcal{A}(\cdot)$ is a linear mapping operator from \mathbb{R}^{UT} to $\mathbb{R}^{U \times T}$.

Table 1. Illustration of one possible design: the coloured rows represent the selected runs.

Design Indicator ($z_{i,j}$)	Possible Run(k)	User(i)	Task(j)	$\mathbf{x}_{i,j}$
1	1	1	1	$\mathbf{x}_{1,1}$
0	2	1	2	$\mathbf{x}_{1,2}$
1	3	1	3	$\mathbf{x}_{1,3}$
...
0	$T+1$	1	T	$\mathbf{x}_{1,T}$
1	$T+2$	2	1	$\mathbf{x}_{2,1}$
...
1	$2T$	2	T	$\mathbf{x}_{2,T}$
0
1	$(U-1)T+1$	U	1	$\mathbf{x}_{U,1}$
...
0	UT	U	T	$\mathbf{x}_{U,T}$

3.2. The proposed optimal design criterion

As a user-task allocation problem, the optimal design is to choose S optimal design runs from the candidate of UT design points, with the aim to enable optimal prediction accuracy for the PRIME model. Denote \mathbf{z} to be a matrix of indicators with $z_{i,j}$ as elements. In the setting of visualisation recommendation, we propose the prediction-oriented optimal design formulation as:

$$\min_{\mathbf{z}} \quad \text{tr} \left\{ \left[\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T f(\mathbf{x}_{i,j}) \right] \times \left(\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T \right)^{-1} \right\} \quad (3a)$$

$$\text{s.t.} \quad \sum_i \sum_j z_{i,j} = S, \quad z_{i,j} \in \{0, 1\}, \\ i = 1, \dots, U, j = 1, \dots, T; \quad (3b)$$

$$\sum_i z_{i,j} \geq \left\lfloor \frac{S}{T} \right\rfloor, \quad \text{for } j = 1, \dots, T; \quad (3c)$$

$$\sum_j z_{i,j} \geq \left\lfloor \frac{S}{U} \right\rfloor, \quad \text{for } i = 1, \dots, U. \quad (3d)$$

The solution of \mathbf{z} corresponds to the indices (i,j) of the selected design points in the optimal design space Ω_{S^*} . In the proposed design criterion, the objective function in (3) is a V-optimality criterion over the selected design points (Crary et al., 1992). The constraints in (3) are to pursue the space-filling characteristics on the selected design of size S . The proposed formulation takes both the low-rank matrix estimation and the linear estimation into consideration through the optimality objective and space-filling constraints.

Specifically, the PRIME model at selected design points \mathbf{X}_{Ω_S} can be expressed as

$$y_{i,j} = \mathbf{x}_{i,j}^T \beta + r_{i,j} + \epsilon_{i,j}, \quad (i,j) \in \Omega_S, \quad (4)$$

where $r_{i,j}$ is the element of the unknown low-rank matrix \mathbf{R} . When \mathbf{R} is given, it is easy to show that the model in (4) is a linear model, of which the V-optimal criterion (Crary et al., 1992) can be expressed as:

$$\phi_V(\mathbf{X}_{\Omega_S}) = \int_{(i,j) \in \Omega_V} \mathbf{x}_{i,j}^T \left(\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T \right)^{-1} \\ \times \mathbf{x}_{i,j} df(\mathbf{x}_{i,j}) \\ = \text{tr} \left[\mathbf{B} \left(\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T \right)^{-1} \right], \quad (5)$$

where $\mathbf{B} = \int_{(i,j) \in \Omega_V} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T df(\mathbf{x}_{i,j})$ and Ω_V represents the region of interest. Here $f(\mathbf{x}_{i,j})$ is a probability measure as a weighting for different regions of inputs, which

can be specified based on the selection of the region of interest and the available information about covariates. Note that when one chooses $\Omega_V = \Omega_S$, it is improper to select the uniform distribution for the probability measure since it will lead to a constant objective function, resulting in a randomly selected space-filling design. It should be noticed that the objective of the original V-optimality is to minimise the average prediction variance over the region of interest Ω_V . The PRIME model is distinguished from the linear regression due to the unknown term \mathbf{R} . The alternating direction method of multiplier (ADMM) was utilised to solve the PRIME model (Chen et al., 2020), in which the low-rank \mathbf{R} and regression parameters β are iteratively estimated. Here we adopt the use of V-optimality criterion for constructing designs for the recommender system, which is new to the literature.

Regarding the region of interest Ω_V , we adopted the choice to make $\Omega_V = \Omega_S$, under the consideration of the optimisation objective function and constraints. The constraints (3c) and (3d) pose space-filling characteristics to one feasible design Ω_S , which implies that the selected design points in Ω_S could enhance the estimation of low-rank matrix. Therefore, we apply V-optimality to Ω_S to further improve the estimation of the linear term. Moreover, since the region of interest with a larger size will be more informative when we want to optimise the estimation variance with a limited number of runs available, we will select all points in Ω_S as Ω_V instead of a subset of Ω_S .

It should be noticed that the region of interest Ω_V can be selected among the whole experiment space Ω , which does not necessarily need to be the same as the selected design space Ω_S . Hence, with $\Omega_V = \Omega_S$, the V-optimal criterion in (5) becomes (3a) in the proposed design criterion.

With the V-optimality as the objective function in (3), we further impose some constraints to enhance the prediction accuracy with respect to the low-rank matrix \mathbf{R} . Specifically, we consider the constraints in (3c) and (3d) for a space-filling and balanced design. That is, we constrain the number of entries in each row (i.e., observed responses for each user) to be similar and the number of entries in each column (i.e., observed responses for each task) to be similar. Thus, the constraints in (3d) balance the number of selected experiments among users and the constraints in (3c) balance the number of selected experiments among tasks. It was proven that sampling an entry in every row and column (i.e., the balance property) improves low-rank matrix completion (Mak & Xie, 2017).

We would like to remark that the proposed design criterion in (3) is a constrained V-optimality with the balanced constraints on the number of entries in the rows and columns of the observed response matrix. It was discussed by Mak and Xie (2017) that when all row and column spaces are equally likely, all the

observed response matrices with balanced characteristics yield the same performance on average in recovering \mathbf{R} from the perspective of the expected information gain. Meanwhile, the objective of V-optimality criterion is to enhance the prediction from the perspective of minimising the average prediction variance over Ω_S in terms of β .

3.3. Design construction algorithm

Note that the V-optimality criterion in (3) itself is NP-hard. Here we consider an efficient practical construction algorithm to obtain the optimal design. Specifically, a multi-step optimisation procedure is proposed as the construction algorithm. The key idea is to effectively seek and reduce the feasible region by the constraints so that the objective function can be optimised efficiently. In Step 1, a candidate pool is generated with designs feasible to the size constraint (3b). The candidate pool should be sufficiently large to support a wide searching space for space-filling designs. Denote the searching size of the candidate pool as K (i.e., to generate K candidate designs in the pool). In Step 2, we apply (3c) and (3d) as a screening step to keep the designs satisfying the balance constraints in the candidate pool. Denote the space-filling filter as $\mathcal{P}_{\Omega_{SF}}(\cdot)$, where $\mathcal{P}_{\Omega_{SF}}(\mathbf{X}_{\Omega_S}) = 1$ if a design \mathbf{X}_{Ω_S} satisfies the space-filling constraints (3c) and (3d). In Step 3, the objective of V-optimality criterion is optimised within the reduced candidate design pool efficiently through calculating and sorting the V-optimality criterion value. The design with the smallest V-optimality criterion value will be selected as the optimal design $\mathbf{X}_{\Omega_S^*}$. A pseudo-code of the algorithm is reported in Algorithm 1.

For Algorithm 1, the searching size K typically should be set large enough to find enough space-filling designs. In order to construct a larger candidate design set \mathcal{F} , an alternative approach is to find the feasible solutions to constraints (3b), (3c) and (3d) by the interior point method or directly by off-the-shelf solvers (e.g., CPLEX, Gurobi).

4. Numerical simulation

4.1. Simulation setting

In this section, we evaluate the performance of the proposed optimal design through several simulated data sets. For each run of experiment, suppose that the underlying PRIME model $y_{ij} = \mathbf{x}_{i,j}^T \beta + r_{ij} + e_{ij}$ is used to generate the response, where $\mathbf{x}_{i,j} \in \mathbb{R}^{1 \times P}$; y_{ij} , r_{ij} , e_{ij} are the (i, j) th elements of \mathbf{Y} , \mathbf{R} and \mathbf{E} , respectively. Here $e_{ij} \sim N(0, 0.1)$. The covariates $\mathbf{x}_{i,j}$, $i = 1, \dots, U, j = 1, \dots, T$, are independently and identically generated from a multivariate normal distribution $N(\mathbf{0}, \Sigma)$, where the covariance matrix

Algorithm 1: The Construction Algorithm of Prediction-oriented Optimal Design

- 1 Initialize: candidate design pool $\mathcal{F} = \emptyset$, candidate objective value set $\mathcal{V} = \emptyset$
 - 2 Step 1: **for** $k=1,2,\dots,K$ **do**
 - 3 generate random candidate design indicator matrix \mathbf{z}^k with run size S , and the corresponding design $\mathbf{X}_{\Omega_S^k}$
 - 4 $\mathcal{F} := \mathcal{F} \cup \mathbf{z}^k$
 - 5 Step 2: **for** $k=1,2,\dots,K$ **do**
 - 6 **if** $\mathcal{P}_{\Omega}(\mathbf{X}_{\Omega_S^k}) \neq 1$ **then**
 - 7 $\mathcal{F} := \mathcal{F} \setminus \{\mathbf{z}^k\}$
 - 8 Step 3: **for** $\mathbf{z}^m \in \mathcal{F}$ **do**
 - 9 $V_m = \text{tr}\{[\sum_{i=1}^U \sum_{j=1}^T z_{i,j}^m \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T f(\mathbf{x}_{i,j})](\sum_{i=1}^U \sum_{j=1}^T z_{i,j}^m \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T)^{-1}\}$
 - 10 $\mathcal{V} := \mathcal{V} \cup V_m$
 - 11 $m^* = \text{argmin} \mathcal{V}$
 - 12 Output: $\mathbf{X}_{\Omega_{S^*}} = \mathbf{X}_{\Omega_{S^*}^{m^*}}, \mathbf{z}^* = \mathbf{z}^{m^*}$
-

$\Sigma = (\sigma_{i,j})_{p \times p}$ with $\sigma_{i,j} = \rho^{|i-j|}$ follows the simulation setup in Deng and Jin (2015). In this simulation study, we set the dimension of covariates $p = 12$, the number of users and tasks $U = T = 10$, and $\rho = 0.5$, where ρ controls the correlation among covariates. Given the covariates \mathbf{X} , we choose a multivariate Gaussian distribution for the weighting function $f(\cdot)$. That is, $f(\mathbf{x}_{i,j}) = (2\pi)^{-\frac{p}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(\mathbf{x}_{i,j}-\boldsymbol{\mu})^T \Sigma (\mathbf{x}_{i,j}-\boldsymbol{\mu})}$, where $\boldsymbol{\mu} \in \mathbb{R}^p$ and $\Sigma \in \mathbb{R}^{p \times p}$ are the mean vector and covariance matrix to be estimated from \mathbf{X} . Specifically, we use the sample mean and sample covariance matrix as the estimates. Other covariance matrix estimation methods (Deng & Tsui, 2013; Kang & Deng, 2020) can also be adopted here. With all $U \times T = 100$ design points, the searching size K in Algorithm 1 is set to be 10,000, which works well empirically to balance the optimality and computation speed. Figure 1 shows a graphical representation of the proposed optimal design with PRIME as the underlying model.

Based on the setting of the covariates, we consider different scenarios to generate the simulation data by varying the number of runs of experiment S , the sparsity level of β , the low-rank structure of \mathbf{R} , and the relative significance of implicit and explicit similarities (i.e., magnitude ratio) of \mathbf{R} over $\mathbf{X}\beta$. Since the maximal number of runs of experiments is $S = U \times T = 100$, we vary S in two levels: $S \in \{40, 60\}$, such that the cost for 60% and 40% experiment runs can be saved, respectively. For the explicit similarity modelled by a linear term $\mathbf{X}\beta$ in PRIME model, the sparsity is defined as the percentage of zero elements in β . We consider the sparsity with two levels $\{0\%, 66\%\}$ with the zero elements in β independently and identically from a normal distribution $N(\xi, 1)$. Here $\xi \in \{-1, 1\}$ is randomly generated from a Bernoulli distribution $Ber(0.5)$ for each scenario.

To generate \mathbf{R} with a low-rank structure, we set the rank $q = 2$ in all scenarios and consider the following

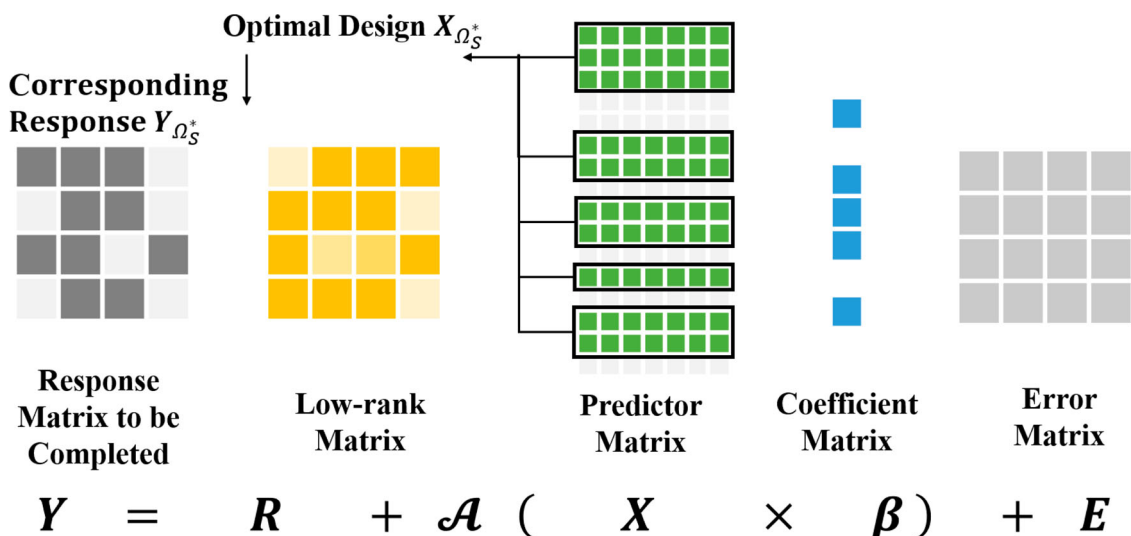


Figure 1. A graphical representation of the proposed optimal design with PRIME as the underlying model.

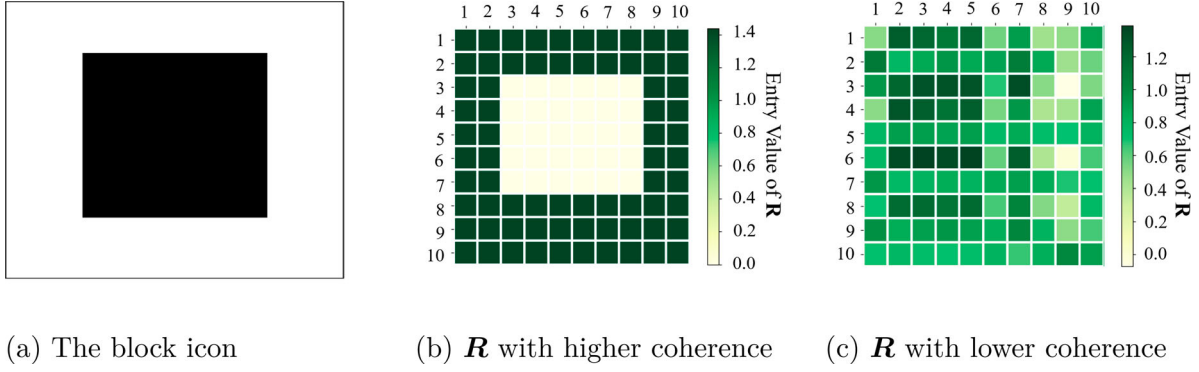


Figure 2. (a) The block icon used to generate \mathbf{R} with higher coherence; (b) The value of entries in \mathbf{R} generated by the block icon; (c) The value of entries in \mathbf{R} generated by \mathbf{QV} .

two approaches based on coherence. Here, coherence measures the extent to which the singular vectors of a matrix are correlated with the standard basis, which can characterise the ability to extract information from a small subset of columns (Mohri & Talwalkar, 2011). The first method generates \mathbf{R} from a block icon (see Figure 2(a)) to mimic the low-rank structure used in tensor regressions (Zhou et al., 2013). The second method generates $\mathbf{R} \in \mathbb{R}^{U \times T}$ with rank q by the product of two base matrices: $\mathbf{R} = \mathbf{QV}$, where $\mathbf{Q} \in \mathbb{R}^{U \times q}$ and $\mathbf{V} \in \mathbb{R}^{q \times T}$ are two randomly generated matrices. Note that the first method provides a higher coherence (Donoho et al., 2005) for \mathbf{R} , which is easier to recover given the same number of observations. Examples of the generated \mathbf{R} from the two methods are presented in Figure 2(b,c), respectively.

The magnitude ratio indicates the dominant type of similarity in the response \mathbf{Y} , which is denoted as mr . We select mr with two levels $\{0.1, 0.5\}$ following the simulation setup in Chen et al. (2020). The ratio is controlled through rescaling the elements in \mathbf{R} so that the value of the (i, j) th element in the rescaled \mathbf{R}' is: $r'_{ij} = r_{ij} \cdot mr \cdot \frac{\max(\mathbf{X}\beta) - \min(\mathbf{X}\beta)}{\max(\mathbf{R}) - \min(\mathbf{R})}$. Note that all the low rank matrices \mathbf{R} mentioned later have been rescaled in their generation process.

To examine the performance of the proposed method, we evaluate the recommendation accuracy for the PRIME model with the training data specified by the optimal design $\mathbf{X}_{\Omega_S^*}$ and the corresponding observed response matrix $\mathbf{Y}_{\Omega_S^*}$. The Root-Mean-Square Error (RMSE) over the non-allocated entries (i.e., empty entries) in \mathbf{Y} predicted by PRIME model is adopted as the performance measure. The performance of the proposed optimal design is systematically evaluated for (EV1) evaluating the performance of V-optimality criterion as the objective Function (3a) over other optimal design criteria with the same set of space-filling constraints, (EV2) evaluating the performance of the space-filling Constraints (3c), (3d) over the same set of unconstrained optimality criteria, and (EV3) comparing the performance of the proposed optimal design (3) with randomly generated designs.

Here I-optimality (Box & Draper, 1959) and G-optimality criteria (Kiefer & Wolfowitz, 1959), both of which are related to the model prediction accuracy, are adopted as benchmarks to compare with the selected V-optimality criterion for the objective function (3). The I-optimality is to minimise the average variance of prediction over the entire experiment region Ω . Given a design \mathbf{X}_{Ω_S} with its corresponding indicator variable \mathbf{z} , the I-optimality value can be computed as:

$$\phi_I(\mathbf{X}_{\Omega_S}) = \int_{(i,j) \in \Omega} \mathbf{x}_{i,j}^T \left(\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T \right)^{-1} \times \mathbf{x}_{i,j} df(\mathbf{x}_{i,j}). \quad (6)$$

The G-optimality aims at minimising the maximum variance of the predicted values, which gives the G-optimality value for design \mathbf{X}_{Ω_S} as:

$$\phi_G(\mathbf{X}_{\Omega_S}) = \max \text{diag} \left[\mathbf{X} \left(\sum_{i=1}^U \sum_{j=1}^T z_{i,j} \mathbf{x}_{i,j} \mathbf{x}_{i,j}^T \right)^{-1} \mathbf{X}^T \right]. \quad (7)$$

Based on the benchmark criteria, four variants of the proposed optimal design method will be evaluated from the aforementioned three perspectives, which results in 12 constructions of optimal designs for each simulation scenario. Specifically, taking V-optimality criterion as the objective function (3), for each simulation scenario, the four variants include:

- V-Opt-SF: the design optimal to the proposed V-optimality criterion with space-filling constraints (denoted as Opt-SF in Table 2);
- V-Opt: the design optimal to V-optimality criterion without considering space-filling constraints (denoted as Opt);
- V-SF: one design randomly selected from the candidate pool of K designs which satisfies the space-filling constraints (denoted as SF), where K is the searching size in Algorithm 1;

Table 2. The average RMSEs for recommendation accuracy in simulation study.

Setting				Run size $S = 40$				Run size $S = 60$			
Generation of \mathbf{R}	Magnitude ratio	Sparsity	Criteria	Opt-SF	Opt	SF ^a	Random ^b	Opt-SF	Opt	SF ^a	Random ^b
QV	0.1	0%	V	0.446	0.458	0.455	0.462	0.313	0.316	0.337	0.376
			I	0.450	0.477	0.425	0.453	0.359	0.335	0.362	0.368
			G	0.448	0.380	0.447	0.449	0.380	0.333	0.341	0.342
		66%	V	0.261	0.283	0.306	0.403	0.253	0.282	0.271	0.307
			I	0.317	0.287	0.294	0.293	0.282	0.302	0.290	0.284
			G	0.316	0.294	0.290	0.324	0.263	0.196	0.273	0.257
	0.5	0%	V	2.557	2.705	2.689	2.901	1.241	1.324	1.393	1.340
			I	2.633	2.417	2.558	3.003	1.403	1.524	1.417	1.397
			G	2.761	2.450	2.471	2.763	1.319	1.243	1.451	1.311
		66%	V	1.491	1.408	1.580	1.610	1.080	1.171	1.182	1.332
			I	1.417	1.266	1.540	1.551	1.245	1.484	1.208	1.244
			G	1.579	1.446	1.482	1.619	1.235	1.089	1.271	1.296
Block	0.1	0%	V	0.677	0.825	0.777	0.829	0.520	0.814	0.687	0.752
			I	0.826	0.720	0.812	0.979	0.736	0.557	0.743	0.903
			G	0.706	0.815	0.804	0.748	0.709	0.622	0.672	0.707
		66%	V	0.454	0.534	0.476	0.513	0.306	0.452	0.352	0.415
			I	0.470	0.441	0.488	0.540	0.331	0.314	0.362	0.349
			G	0.448	0.491	0.490	0.548	0.329	0.340	0.375	0.402
	0.5	0%	V	3.309	3.800	3.691	4.110	2.685	3.927	3.124	3.511
			I	3.814	3.444	3.827	3.963	3.769	2.841	3.138	2.898
			G	3.283	4.024	3.813	3.552	3.356	2.979	3.296	3.024
		66%	V	2.135	2.230	2.324	2.483	1.879	2.493	2.086	2.176
			I	2.104	1.868	2.143	2.876	2.457	1.990	2.467	1.955
			G	1.829	2.124	2.302	2.303	2.450	2.049	2.578	2.308

Note: Best results for all 16 scenarios are highlighted in **bold**.

^aThree replicates for the (-SF) designs in each scenario.

^bThree replicates for the (-Random) designs in each scenario

V-Random: one design randomly selected from the candidate pool of K designs (denoted as Random).

Similarly, we can define I-Opt-SF to I-Random with respect to I-optimality criterion and G-Opt-SF to G-Random with respect to G-optimality criterion for each simulation scenario. Note that the V-Opt-SF method is the proposed optimal design. It should also be mentioned that the construction methods for (-SF) designs in each scenario are the same which do not need to consider the optimality criterion, thus can be viewed as three replicates. Same for the (-Random) designs.

4.2. Simulation results

Table 2 summarises the RMSEs as the recommendation performance metric in 16 simulation scenarios. Overall, it can be observed that the proposed method (i.e., Opt-SF with V-optimality criterion) gives consistently better performance over benchmark criteria and variants by comparing the RMSEs when run size $S = 60$. However, for the scenarios when run size $S = 40$, the proposed method does not consistently outperform the benchmark methods, which could be due to the limited sample size. Xu (2018)'s study on noisy matrix completion showed that $4nq - 4q^2$ is the tight lower bound for the number of entries to be observed to obtain a unique solution for the matrix completion for $\mathcal{C}^{n \times n}$ with its rank $q \leq n/2$. With rank $q = 2$ to generate $\mathbf{R}^{10 \times 10}$ in the simulation, $S = 64$ is the theoretical least number of entries to be observed to successfully recover the low-rank structure. It implies that, when

$S = 40$, it is theoretically difficult to complete the low-rank matrix \mathbf{R} in PRIME, especially when the linear regression coefficients β should also be estimated.

Regarding to the performance evaluation $EV1$, it can be observed that the proposed V-optimality criterion with space-filling constraints (V-Opt-SF) outperforms I-Opt-SF and G-Opt-SF in most (i.e., 12 out of 16) scenarios by comparing RMSEs within Opt-SF columns in Table 2. For $EV2$, it is seen that the use of space-filling constraints significantly enhances the recommendation accuracy by comparing RMSEs between V-Opt-SF and V-Opt for 15 out of 16 scenarios. In terms of performance evaluation $EV3$, the proposed V-Opt-SF presents its superior performance in all scenarios in comparison with V-Random. The advantage of the proposed prediction-oriented optimal design method V-Opt-SF demonstrated from all three evaluation perspectives can be attributed to the integration of V-optimality criterion and space-filling constraints. This integration effectively balances the enhancement in estimating linear regression coefficients from V-optimality criterion and the enhancement in estimating low-rank matrix from space-filling constraints.

This simulation study also demonstrates the inferior performance of adopting I- and G-optimality criteria as the objective function in (3). For example, I-optimality criterion does not perform well compared to another two criteria by investigating Opt-SF columns in Table 2. Another disadvantage of I-optimality criterion is the increasing computation burden caused by the integral over the entire design space (6). In terms of G-optimality criterion, with a small run size (i.e., $S = 40$), it leads to the design with better prediction

accuracy when \mathbf{R} is generated by the block icon and the significance of implicit similarity is higher (i.e., $mr = 0.5$). Since G-optimality criterion aims at optimising the worst estimation variance, it could provide a robust design against overfitting and other possible worse cases with a small observation and an underlying \mathbf{R} with lower coherence. Correspondingly, when \mathbf{R} is generated by the base matrices, the designs with the optimal I-optimality criterion value perform better without considering the space-filling constraints. A possible explanation can be that, due to the small run size and high coherence on \mathbf{R} , directly optimising the overall estimation variance will lead to better overall estimation. As mentioned above, the three criteria are all related to prediction accuracy with similar expression. Therefore, it is not realistic for one optimality criterion to dominate all scenarios.

The advantage of V- over I- and G-optimality criteria can also be demonstrated by investigating the prediction errors for each entry of \mathbf{Y} in PRIME model. Figures 3 and 4 present the absolute error of each entry $|y_{i,j} - \hat{y}_{i,j}|$ in V-Opt-SF, I-Opt-SF, and G-Opt-SF designs with run size $S \in \{40, 60\}$, where the optimal design points are marked with 'X'. It can be observed that the prediction errors in the proposed V-Opt-SF

design are more evenly distributed compared to the other two optimality criteria, which indicates its robustness in terms of the recommendation accuracy for all potential users and tasks.

Furthermore, the advantage of space-filling constraints can be clearly demonstrated by comparing the performance of the designs with (i.e., SF) and without (i.e., Opt and Random) space-filling constraints. Figure 5 shows the box plots of RMSEs by adopting SF designs, Opt designs and Random designs. Note that the SF designs represented by the box plots are generated from the candidate K designs in Algorithm 1 which satisfy the space-filling constraints. Correspondingly, we picked 500 designs with top-ranked criterion value for Opt designs. Similarly, 500 designs are randomly picked from the candidate pool of K designs for Random designs. In Figure 5, the box plots for SF designs have a narrow range with much fewer outliers than those for Opt and Random designs. It implies the robustness provided by the space-filling constraints. The plausible explanation is that these space-filling constraints are effective in restricting the feasible region for design generation to prevent PRIME from making extreme predictions that directly lead to low recommendation accuracy.

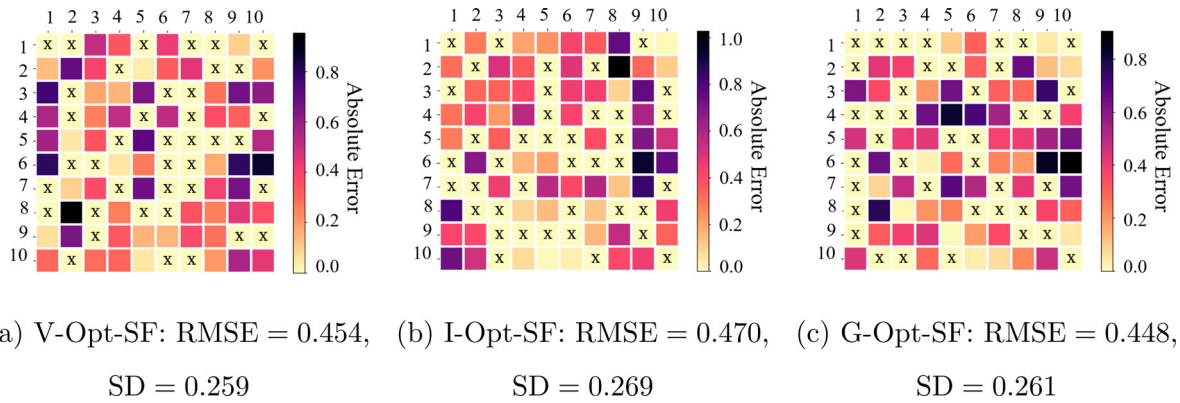


Figure 3. Absolute value of prediction error of Opt-SF designs with $S = 40$, \mathbf{R} generated by block, Sparsity = 66%, $mr = 0.1$, annotated with RMSE and its Standard Deviation (SD). (a) V-Opt-SF: RMSE = 0.454, SD = 0.259. (b) I-Opt-SF: RMSE = 0.470, SD = 0.269 and (c) G-Opt-SF: RMSE = 0.448, SD = 0.261.

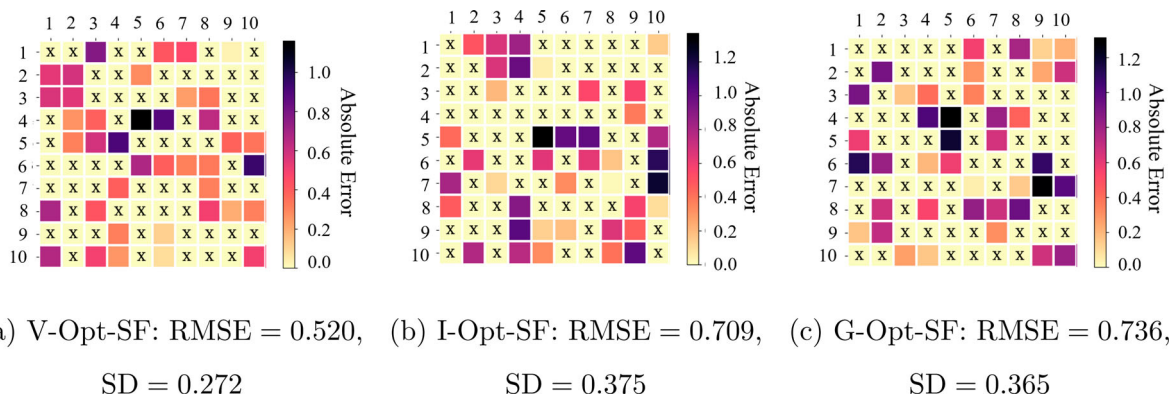


Figure 4. Absolute value of prediction error of Opt-SF designs with $S = 60$, \mathbf{R} generated by block, Sparsity = 0%, $mr = 0.1$, annotated with RMSE and its Standard Deviation (SD). (a) V-Opt-SF: RMSE = 0.520, SD = 0.272. (b) I-Opt-SF: RMSE = 0.709, SD = 0.375 and (c) G-Opt-SF: RMSE = 0.736, SD = 0.365.

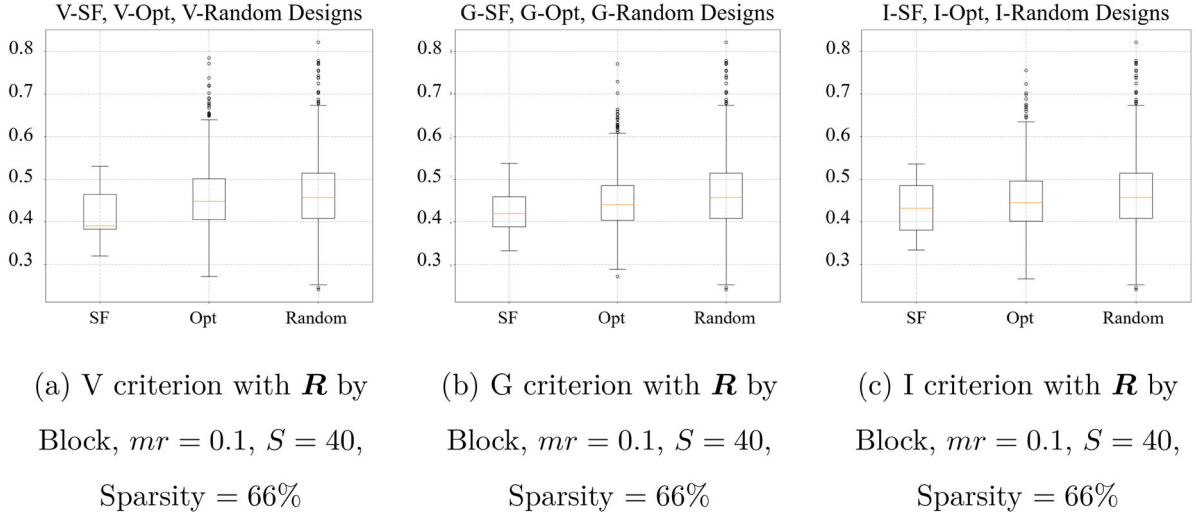


Figure 5. Box plots of the RMSE of SF designs, Opt designs and Random designs. (a) V criterion with \mathbf{R} by Block, $mr = 0.1$, $S = 40$, Sparsity = 66%. (b) G criterion with \mathbf{R} by Block, $mr = 0.1$, $S = 40$, Sparsity = 66% and (c) I criterion with \mathbf{R} by Block, $mr = 0.1$, $S = 40$, Sparsity = 66%.

4.3. An approximate sampling algorithm for scalability

The proposed design construction algorithm (Algorithm 1) is suitable for the small-scale user study in visualisation recommendation. However, in applications with large size of users and tasks, the searching size K should be set large enough to ensure the optimality, which can cause high computation workload. To address this challenge of scalability, we develop an approximate sampling algorithm as an alternative for an efficient implementation.

Firstly, we consider an approximation solution for the proposed optimal design criterion (3) which relaxes the decision variable \mathbf{z} to be continuous (i.e., $z_{ij} \in [0, 1]$) and relaxes the size constraints accordingly (Boyd et al., 2004). Denote $\lambda_{ij} = \frac{z_{ij}}{S}$ as the relative frequency of the (i, j) th design. The relaxed problem can be expressed as follows:

$$\min_{\mathbf{z}} \operatorname{tr} \left\{ \left[\sum_{i=1}^U \sum_{j=1}^T \lambda_{ij} \mathbf{x}_{ij} \mathbf{x}_{ij}^T f(\mathbf{x}_{ij}) \right] \times \left(\sum_{i=1}^U \sum_{j=1}^T \lambda_{ij} \mathbf{x}_{ij} \mathbf{x}_{ij}^T \right)^{-1} \right\} \quad (8a)$$

$$\text{s.t. } \sum_i \sum_j \lambda_{ij} \leq 1, \quad \lambda_{ij} \in [0, 1], \quad i = 1, \dots, U, \\ j = 1, \dots, T; \quad (8b)$$

$$\sum_i \lambda_{ij} \geq \left\lfloor \frac{1}{T} \right\rfloor, \quad \text{for } j = 1, \dots, T; \quad (8c)$$

$$\sum_j \lambda_{ij} \geq \left\lfloor \frac{1}{U} \right\rfloor, \quad \text{for } i = 1, \dots, U. \quad (8d)$$

In this way, the solution λ^* can be viewed as the sampling probability where λ_{ij} is the probability of sampling the (i, j) th design point. Also, we have $z_{ij}^* = \lambda_{ij}^* \cdot S$. Secondly, we sample from the obtained distribution for $M = 1000$ times to get 1000 candidate designs. Finally, the design with the smallest V-optimality criterion value will be selected as the selected design $\mathbf{X}_{\Omega_{S^*}}$.

We applied this approximate sampling algorithm to all simulation scenarios and summarised the corresponding results in Table 3. We evaluated the recommendation accuracy for the PRIME model as well as the computational efficiency of the proposed algorithm and the approximate sampling algorithm. The continuous constrained optimisation problem (8) is solved numerically by Nelder-Mead method (Nelder & Mead, 1965). From the results in the table, although the designs obtained by the approximate sampling (V-Approximate Sampling) are not as good as those obtained by the proposed algorithm (V-Opt-SF) in 12 out of 16 scenarios, it is seen that the performance of the V-Approximate Sampling is generally better than those of Opt, SF and Random in Table 2. Moreover, the approximate sampling algorithm achieves higher computational efficiency when comparing the computational time of these two methods in Table 3.

We further investigated the performance of the approximate sampling method using the relaxed formulation in (8). Figure 6 demonstrates the optimal design obtained by the proposed algorithm (V-Opt-SF) with run size $S \in \{40, 60\}$, where the optimal design points are marked with 'X' and the colour shows the value of each entry z_{ij}^* obtained from the relaxed problem (8). The covariates \mathbf{X} for these two scenarios are the same, which causes similar patterns in the heatmaps. It can be shown that in general the V-Opt-SF design points are allocated to entities z_{ij}^* with higher value, which indicates the proposed Algorithm 1 and the

Table 3. The average RMSEs for recommendation accuracy and the corresponding computational time in simulation.

Setting				RMSE		Computational time (sec) ^a	
Run size S	Generation of R	Magnitude ratio	Sparsity	V-Opt-SF	V-approximate sampling	V-Opt-SF	V-approximate sampling
40	QV	0.1	0%	0.446	0.442	22.172	12.673
			66%	0.261	0.277	35.417	17.665
		0.5	0%	2.557	2.375	83.159	17.757
	Block	0.1	0%	1.491	1.478	51.563	16.888
			66%	0.677	0.722	45.450	24.898
		0.5	0%	0.454	0.462	25.654	15.623
60	QV	0.1	0%	3.309	3.644	26.425	15.541
			66%	2.135	2.119	46.912	17.010
		0.5	0%	0.313	0.327	69.846	24.519
	Block	0.1	0%	0.253	0.278	32.910	24.272
			66%	1.241	1.494	98.129	32.505
		0.5	0%	1.080	1.351	56.572	26.284
Block	0.1	0%	0.520	0.532	47.665	23.603	
		66%	0.306	0.334	51.871	23.192	
	0.5	0%	2.685	2.870	80.968	26.133	
			66%	1.879	2.038	84.045	26.489

^aThe simulation is implemented in Python 3.7.6 and R 3.6.1 on a laptop with 1.80 GHz Intel Core i7-8565U Processor, 16.0 GB RAM and Windows 10.

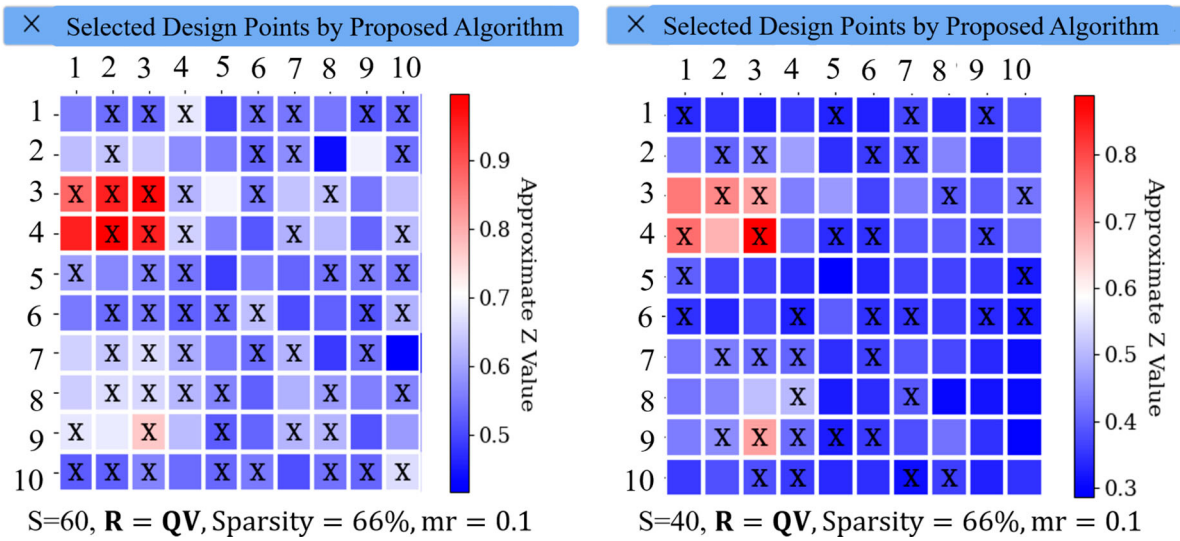


Figure 6. The optimal designs obtained by the proposed algorithm (V-Opt-SF) with different run size S and the corresponding z_{ij}^* value obtained from the relaxed problem (8).

approximate sampling algorithm optimise the design in a similar way. The slightly inferior performance of the V-Approximate Sampling designs might be caused by the existence of multiple local optimums in the optimisation process of problem (8), which indicates that the final solution can be easily affected by the choice of the starting point.

In a short summary, the approximate sampling algorithm can be an efficient solution with less computation workload, and it can be easily extended to large-scale scenarios. To further improve the computational efficiency, one can consider a tractable convex programme formulation such as the second-order conic program (SOCP) (Sagnol & Harman, 2015) for the V-optimal criterion in (8).

5. Case study

In this section, we demonstrate the merits of the proposed optimal design to support data collection in a user study for the visualisation recommender system

PRIME. We refer Chen and Jin (2017) for the details of the visualisation evaluation study. In this study, three interactive visualisation methods (i.e., static node-link tree, collapsible node-link tree, and zoomable layout) were evaluated for 14 participants performing 11 predefined visual searching tasks (Chen et al., 2020). Here we only focus on the visualisation method of static node-link tree as shown in Figure 7, and we target at making accurate predictions on recommendation scores for user i performing task j with this visualisation method.

For the 11 tasks, their contents are about searching for the required information from a hierarchical data set with simple calculations. The 11 tasks can be classified into three types: (1) finding the nodes with target names; (2) finding the parent nodes of those with target names; (3) counting the number of children nodes of those with target names. The difficulty of the tasks can be divided into three levels, which directly result in different responses (i.e., perceived task difficulty rated by participants). These information will

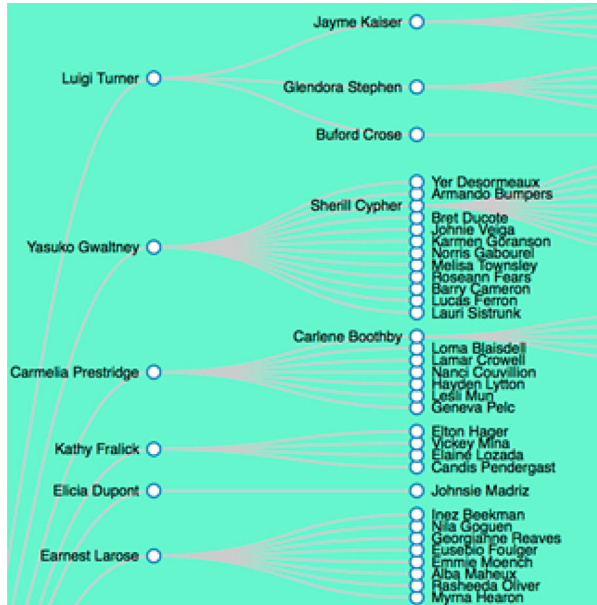


Figure 7. The visualisation method of static node-link tree, where hierarchical relationships between names were reflected by the node links.

be utilised to generate covariates characterising each task. After performing the assigned tasks, the participants would rate the complexity of the performed tasks, which would form the response matrix representing the users' evaluation on the task complexity.

To develop the recommender system for visualisation tasks, covariates X characterising the tasks and users are collected for the estimation of explicit similarity. Covariates X are mainly generated by the descriptive information related to users and tasks (i.e., gender, age, major information of the users, and the type, difficulty level of the tasks) so that they can be observed before conducting the experiments. The multivariate Gaussian distribution is adopted for the weight function $f(x_{i,j})$ for each design point, which is similar as the setting in the simulation study. Same to the setting in the simulation section, the PRIME is adopted as the underlying model for the visualisation recommender system. In Chen et al. (2020), the data of all 14 users performing all 11 tasks have been collected, which is helpful for us to evaluate the performance of the proposed design. When an optimal design is used to form the training data for the PRIME model, we still consider the Root-Mean-Square Error (RMSE) over the non-allocated entries (i.e., empty entries) in Y predicted by PRIME model as the performance measure. We evaluated the proposed optimal design criterion on different numbers of runs of experiments. With total $U \cdot T = 154$ ($U = 14$, $T = 11$) design points, the number of runs of experiments S is set as five levels $S \in \{60, 70, 80, 100, 120\}$ with a focus on smaller run size. For each size of runs of experiments, four designs will be evaluated on their prediction RMSE: (1) the proposed optimal design (Opt-SF); (2) the design optimal to the criterion without considering space-filling constraints (Opt); (3) one randomly

Table 4. The RMSE recommendation accuracy in the case study.

Run size S	Evaluation				
	Criterion	Opt-SF	Opt	SF	Random
60	V	1.515	1.563	1.536	2.117
	I	1.707	1.543		
	G	1.662	1.730		
70	V	1.518	1.896	1.539	1.815
	I	1.547	1.660		
	G	1.529	1.801		
80	V	1.461	1.745	1.648	1.739
	I	1.640	1.523		
	G	1.693	1.872		
100	V	1.512	1.645	1.554	1.616
	I	1.390	1.482		
	G	1.614	1.418		
120	V	1.322	1.614	1.538	1.726
	I	1.326	1.628		
	G	1.534	1.658		

selected space-filling design (SF); (4) one randomly generated design (Random). Here, K is also selected as $K = 10,000$.

Table 4 summarises the performance of all designs with different levels of run size, where the results in bold represent the best design with the highest recommendation accuracy in each run size. In general, the proposed optimal design achieves the best performance in four out of five scenarios, especially when the run size is small. With its high computational efficiency of V-optimality criterion, the proposed method establishes the superiority by its efficiency and robustness in practical applications. It is also seen that the designs achieving the optimality criterion with space-filling constraints (Opt-SF), perform better than the designs only considering the optimality (Opt) or the designs only considering the space-filling constraints (SF). Besides, the advantage of random space-filling designs over random designs demonstrates the effectiveness of imposing space-filling constraints. It can also be shown that the uncertainty of the recommendation accuracy rapidly increases when the run size is quite limited (i.e., 60/154). In such scenarios, the data collected by a randomly generated design will lead to very low recommendation accuracy (e.g., $RMSE = 2.117$). A recommender system developed by such less informative data may not be able to provide trustworthy recommendations for new users, which could imply the necessity and importance of a prediction-oriented optimal design. In a short summary, the case study for visualisation method recommendation shows that the proposed optimal design can be efficiently and effectively adopted in practice to enable informative data collection and the development of an accurate recommender system with limited experiments.

6. Discussion

This work studies the optimal design problem for the visualisation recommender system. A V-optimality

design criterion with space-filling constraints are integrated to form the proposed prediction-oriented optimal design criterion. The proposed optimal design jointly considers the low-rank property of the deterministic collaborative information and the content-based information with linear response. By constructing the proposed optimal design to collect informative data with limited runs of experiments, the resultant recommender system can achieve high recommendation accuracy over the possible users and contexts.

It is worth remarking that the proposed design strategy can be applied to other recommender systems with limited resources to collect the training data. For example, it can be adopted to the visualisation system based on augmented reality (AR) for worker's training and guidance in manufacturing processes (Chen et al., 2016). The proposed design can also be extended to support recommender systems in other fields with abundant covariates information of the tasks and users to be matched, such as personalised knowledge discovery (Vozniuk et al., 2016).

Note that in this work, the users' rating scores are used as the response in the proposed design for the recommender system. However, the ranking information can be more effective in suggesting the top-ranked visualisation methods to users. In this case, the pairwise loss can be adopted instead of the prediction error in the underlying model to provide pairwise comparisons of visualisation methods (Chen & Jin, 2018, 2020). Then the objective of the proposed optimal design criterion (3a) can be modified accordingly to reduce the average variance in the ranking prediction, thus enhancing the accuracy of ranking recommendation. Besides, although we selected the selected design space Ω_S as the region of interest Ω_V based on the space-filling constraints and the objective function, further investigation is expected to find the optimal region of interest with theoretical support. Moreover, the proposed method is suitable for a small-size user study in visualisation recommendation. However, in practical scenarios with large size of users and tasks, a screening step is useful to select the users from a large user pool to conduct the user study. Various sampling strategies such as cluster sampling and stratified sampling can be used for the screening step before the proposed optimal design (Sharma, 2017). The ethical consideration such as the fairness for the screening process is also of great importance in practice.

There are several future research directions. First, the proposed method can be extended to optimal A/B testing to support the recommendation of the optimal visualisation method for a user given a task. Second, visualisation methods will be parameterised based on the characteristics (i.e., layout parameters, colour coding, etc.) to generate a continuous design space. Thus, more effective and efficient optimal design methods can be investigated based on the continuous design

space to improve the recommendation performance for visualisation methods. Third, in practical applications, besides the observable covariates such as survey results and users' input logs, unobservable covariates are usually available, which can only be collected during the experiments (e.g., wearable sensor signals). Including unobservable covariates to better quantify the explicit similarity will enhance the performance of the recommender system, which can be considered by online optimal design methods. Besides, after conducting offline experiments, an online active sampling method can be investigated to adaptively enhance the performance of the recommender system. In the online sampling, the ensemble modelling method (Jin & Deng, 2015) can be utilised to ensemble the online schema with the initial offline design criterion.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Yingyan Zeng is a Ph.D. candidate in the Grado Department of Industrial and Systems Engineering at Virginia Tech. She received a B.E. degree from Shanghai Jiao Tong University, China in 2019. She is working as graduate research assistant and teaching assistant at Virginia Tech. Her research focused on data analytics and statistical modeling in smart manufacturing systems. Ms. Zeng is a member of INFORMS.

Dr. Xinwei Deng is an associate professor in the department of statistics at Virginia Tech. His research interests include machine learning, design of experiment, and interface between experimental design and machine learning.

Xiaoyu Chen is a Ph.D. candidate in the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received a B.E. degree from School of Optoelectronics at Beijing Institute of Technology, China in 2015. He previously worked as Research assistant and Grado teaching assistant at Virginia Tech. His research focused on human-AI collaboration and Fog-Cloud computing in smart manufacturing systems. Mr. Chen is a member of INFORMS, IISE, and IEEE. His awards and honors include Doctoral Student of the Year in the Grado Department of Industrial and Systems Engineering at Virginia Tech, 2019.

Dr. Ran Jin is an Associate Professor and the Director of Laboratory of Data Science and Visualization at the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his Ph.D. degree in Industrial Engineering from Georgia Tech, Atlanta, his Master's degrees in Industrial Engineering, and in Statistics, both from the University of Michigan, Ann Arbor, and his bachelor's degree in Electronic Engineering from Tsinghua University, Beijing. His research focuses on supervised learning and computation services in smart manufacturing and cognitive-based interactive visualization. His research includes lightweight middleware design in Industrial Internet, machine learning pipelines and deep neural networks in manufacturing, data fusion modeling for variation analysis, monitoring, prognosis, and virtual inspection, and Augmented Reality and

web-based manufacturing visualization systems design. His research has been used by many manufacturing companies in additive manufacturing, thermal spray coating, broaching, semiconductor, printed electronics, optical and continuous fiber manufacturing industries. Dr. Jin is currently serving as an Associate Editor for IISE Transactions (Design and Manufacturing), Associate Editor for ASME Journal of Manufacturing Science and Engineering, and Associate Editor for INFORMS Journal on Data Science. For more information about Dr. Jin, please visit his faculty website at Virginia Tech: <http://ise.vt.edu/ran-jin>.

ORCID

Xinwei Deng  <http://orcid.org/0000-0002-1560-2405>

References

- Anava, O., Golan, S., Golbandi, N., Karnin, Z., Lempel, R., Rokhlenko, O., & Somekh, O. (2015). Budget-constrained item cold-start handling in collaborative filtering recommenders via optimal design. In *Proceedings of the 24th international conference on world wide web* (pp. 45–54). Florence, Italy: International World Wide Web Conferences Steering Committee.
- Bednarik, R., Vrzakova, H., & Hradis, M. (2012). What do you want to do next: A novel approach for intent prediction in gaze-based interaction. In *Proceedings of the symposium on eye tracking research and applications* (pp. 83–90). Association for Computing Machinery.
- Bhat, N., Farias, V. F., Moallem, C. C., & Sinha, D. (2020). Near-optimal AB testing. *Management Science*, 66(10), 4477–4495. <https://doi.org/10.1287/mnsc.2019.3424>
- Bogdanov, D., Haro, M., Fuhrmann, F., Xambó, A., Gómez, E., & Herrera, P. (2013). Semantic audio content-based music recommendation and visualization based on user preference examples. *Information Processing & Management*, 49(1), 13–33. <https://doi.org/10.1016/j.ipm.2012.06.004>
- Box, G. E., & Draper, N. R. (1959). A basis for the selection of a response surface design. *Journal of the American Statistical Association*, 54(287), 622–654. <https://doi.org/10.1080/01621459.1959.10501525>
- Boyd, S., Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.
- Candès, E. J., & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6), 717. <https://doi.org/10.1007/s10208-009-9045-5>
- Chakraborty, S., Zhou, J., Balasubramanian, V., Panchanathan, S., Davidson, I., & Ye, J. (2013). Active matrix completion. In *2013 IEEE 13th international conference on data mining* (pp. 81–90). IEEE.
- Chen, X., & Jin, R. (2017). Statistical modeling for visualization evaluation through data fusion. *Applied Ergonomics*, 65, 551–561. <https://doi.org/10.1016/j.apergo.2016.12.016>
- Chen, X., & Jin, R. (2018). Data fusion pipelines for autonomous smart manufacturing. In *2018 IEEE 14th international conference on automation science and engineering (CASE)* (pp. 1203–1208). IEEE.
- Chen, X., & Jin, R. (2020). AdaPipe: A recommender system for adaptive computation pipelines in cyber-manufacturing computation services. *IEEE Transactions on Industrial Informatics*, 1–1. <https://doi.org/10.1109/TII.2020.3035524>
- Chen, X., Lau, N., & Jin, R. (2020). PRIME: A personalized recommendation for information visualization methods via extended matrix completion. *ACM Transactions on Interactive Intelligent Systems*, 11(1), 1–30. <https://doi.org/10.1145/3366484>
- Chen, X., Sun, H., & Jin, R. (2016). Variation analysis and visualization of manufacturing processes via augmented reality. In *Presented at the IISE annual conference. Proceedings*. IISE.
- Crary, S. B., Hoo, L., & Tennenhouse, M. (1992). I-optimality algorithm and implementation. In *Computational statistics* (pp. 209–214). Springer.
- Deng, X., & Jin, R. (2015). QQ models: Joint modeling for quantitative and qualitative quality responses in manufacturing systems. *Technometrics*, 57(3), 320–331. <https://doi.org/10.1080/00401706.2015.1029079>
- Deng, X., & Tsui, K. W. (2013). Penalized covariance matrix estimation using a matrix-logarithm transformation. *Journal of Computational and Graphical Statistics*, 22(2), 494–512. <https://doi.org/10.1080/10618600.2012.715556>
- Donoho, D. L., Elad, M., & Temlyakov, V. N. (2005). Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1), 6–18. <https://doi.org/10.1109/TIT.2005.860430>
- Jin, R., & Deng, X. (2015). Ensemble modeling for data fusion in manufacturing process scale-up. *IIE Transactions*, 47(3), 203–214. <https://doi.org/10.1080/0740817X.2014.916580>
- Jones, B., Silvestrini, R. T., Montgomery, D. C., & Steinberg, D. M. (2015). Bridge designs for modeling systems with low noise. *Technometrics*, 57(2), 155–163. <https://doi.org/10.1080/00401706.2014.923788>
- Joseph, V. R. (2016). Space-filling designs for computer experiments: A review. *Quality Engineering*, 28(1), 28–35. <https://doi.org/10.1080/08982112.2015.1100447>
- Kang, X., & Deng, X. (2020). On variable ordination of Cholesky-based estimation for a sparse covariance matrix. *Canadian Journal of Statistics*. <https://doi.org/10.1002/cjs.11564>
- Kiefer, J., & Wolfowitz, J. (1959). Optimum designs in regression problems. *The Annals of Mathematical Statistics*, 30(2), 271–294. <https://doi.org/10.1214/aoms/1177706252>
- Kunkel, J., Loepf, B., & Ziegler, J. (2017). A 3D item space visualization for presenting and manipulating user preferences in collaborative filtering. In *Proceedings of the 22nd international conference on intelligent user interfaces* (pp. 3–15). Association for Computing Machinery.
- Laake, P. (1975). On the optimal allocation of observations in experiments with mixtures. *Scandinavian Journal of Statistics*, 2, 153–157. <https://www.jstor.org/stable/4615596>
- Mak, S., & Xie, Y. (2017). *Active matrix completion with uncertainty quantification*. Preprint. arXiv:1706.08037.
- Melville, P., & Sindhvani, V. (2010). Recommender systems. *Encyclopedia of Machine Learning*, 1, 829–838.
- Mohri, M., & Talwalkar, A. (2011). Can matrix coherence be efficiently and accurately estimated? In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 534–542). JMLR Workshop and Conference Proceedings.
- Mutlu, B., Veas, E., & Trattner, C. (2016). Vizrec: Recommending personalized visualizations. *ACM Transactions on Interactive Intelligent Systems*, 6(4), 1–39. <https://doi.org/10.1145/2983923>
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The Computer Journal*, 7(4), 308–313. <https://doi.org/10.1093/comjnl/7.4.308>
- Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM.

- Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In *Recommender systems handbook* (pp. 1–34). Springer.
- Ruchansky, N., Crovella, M., & Terzi, E. (2015). Matrix completion with queries. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1025–1034). Association for Computing Machinery.
- Sagnol, G., & Harman, R. (2015). Computing exact d -optimal designs by mixed integer second-order cone programming. *Annals of Statistics*, 43(5), 2198–2224. <https://doi.org/10.1214/15-AOS1339>
- Santner, T. J., Williams, B. J., Notz, W. I., & Williams, B. J. (2003). *The design and analysis of computer experiments* (Vol. Springer).
- Schein, A. I., Popescul, A., Ungar, L. H., & Pennock, D. M. (2002). Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 253–260). Association for Computing Machinery.
- Sharma, G. (2017). Pros and cons of different sampling techniques. *International Journal of Applied Research*, 3(7), 749–752.
- Spence, R. (2001). *Information visualization* (Vol. Springer).
- Vartak, M., Huang, S., Siddiqui, T., Madden, S., & Parameswaran, A. (2017). Towards visualization recommendation systems. *ACM SIGMOD Record*, 45(4), 34–39. <https://doi.org/10.1145/3092931.3092937>
- Voigt, M., Pietschmann, S., Grammel, L., & Meißner, K. (2012). Context-aware recommendation of visualization components. In *The fourth international conference on information, process, and knowledge management (EKNOW)* (pp. 101–109). IARIA XPS Press.
- Vozniuk, A., Rodriguez Triana, M. J., Holzer, A. C., & Gillet, D. (2016). Combining content analytics and activity tracking to identify user interests and enable knowledge discovery. In *6th international workshop on personalization approaches in learning environments (PALE)*. ACM.
- Xu, Z. (2018). The minimal measurement number for low-rank matrix recovery. *Applied and Computational Harmonic Analysis*, 44(2), 497–508. <https://doi.org/10.1016/j.acha.2017.01.005>
- Zhou, H., Li, L., & Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association*, 108(502), 540–552. <https://doi.org/10.1080/01621459.2013.776499>