Online Clustering of Bandits With High-Dimensional Sparse Relevant User Features

Chien Ming Chi
National Taiwan University

Hsuan Tien Lin
National Taiwan University

Ching Kang Ing
National Tsing Hua University

Abstract

We propose HSB, a novel learning method that takes into account user heterogeneity by clustering users into groups according to their item preferences. HSB incorporates the idea that user heterogeneity in item preferences can be observed by investigating their *sparse relevant user features* (SRF) w.r.t. each item. Regarding each item, SRF is a subset of all user features; users with the same SRF (i.e., the values of their SRF are the same) have the same level of interest in this item. HSB clustering puts users with the same SRF into the same group w.r.t. each item and make recommendations accordingly, so heterogeneity in preferences is addressed. Moreover, HSB uses statistical feature selection tools to ensure predictive performance when exposed to high-dimensional features. Theoretical analysis for such novel combination of bandit learning methods and statistical tools as well as regrets analysis are given. Real data as well as synthetic data studies are conducted to demonstrate the competitive prediction performance of HSB against that of a pool of state-of-the-art heterogeneity-sensitive learning methods.

1 INTRODUCTION

In comparison to traditional commerce, ecommerce such as online web services tends to have an advantage in that every person with Internet access is a potential customer. This advantage is also a difficulty, however, because the brick-and-mortar strategy of hiring salespeople to recommend and sell business products to each online visitor is almost impossible. Recommendation systems (RSs) are proposed as a solution to this situation. A main task of an RS is to – based on historical user activities – suggest products that are likely to be favored by the user and at the same time identify potentially popular products (or items) in the inventory. Among other approaches, the multi-armed bandit has proven to be an efficient way to deal with this problem, in the literature known as the 'exploration-exploitation (EE) trade-off problem'. Due to the promising results of applications of bandits, a substantial amount of quality work has contributed to the study of multi-armed and contextual bandits [Auer et al., 2001, Abbasi-Yadkori et al., 2011, Auer, 2002, Chu et al., 2011, Krause and Ong, 2011, Lai and Robbins, 1985, Langford and Zang, 2007, Li et al., 2010] (and references therein). The efficiency of these benchmark learning methods more or less assumes that all users have a uniform item preference, which largely simplifies the computation with learning performance maintained when applications are not very involved.

In reality, however, consumers with distinct features, e.g., genders, job classes, or other individual status, tend to have different preferences when facing a given set of choices. In light of this, our work aims at improving the quality of recommendation by considering the effect of inherent user heterogeneity in their choice-making process when information on user features is available. Specifically, we propose HSB (Heterogeneity-Sensitive Bandit), a learning method that allows two users with distinct user features to have possibly different item preferences. On the other hand, distinct users can have the same features, and hence the same preference; we call these users 'same-type users'.

Defining user type w.r.t. user features is common for many learning methods. Our novel improvement is based on the idea that only a small subset of user features are needed to decide the user type. In most learning situations, how many and which user features are included in the system does not depend on learners.

Proceedings of the 22$^{\text{nd}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2019, Naha, Okinawa, Japan. PMLR: Volume 89. Copyright 2019 by the author(s).

Without any prior information, there is a tendency to record more features than are needed for designing a reliable learning machine. As a result, only a subset of user features are informative in the sense that they can be used for identifying the type of a user. Such a subset of informative user features are referred to as *sparse relevant user features*(SRF).

The concept of sparse relevant user features is crucial in our work. It has two traits: *i)* often the number of relevant features is very small compared with that of all features (sparse); *ii)* since each item may attract different types of users, the set of relevant user features for this item may be different from those for other items (item-dependent). Sparsity of relevant features prevents users from being clustered into impractically many groups. For example, given a typical learning situation with a set of 30 binary user features, identifying user types by the values of their features results in many distinct yet meaningless user types (2^{30} in this case). On the other hand, the second trait can be seen from this example: the 'genders' factor may be relevant for a purchase decision for the novel series *Twilight*, whereas it can be less decisive w.r.t. the *Harry Potter* series. The precise relevance of such features shall be decided numerically by the learner, on the fly; we hence do not presume all items share the same set of relevant features.

If users are the same type regarding an item, these users share the same level of interest in this item; hence they are clustered into the same group (w.r.t. this item). The heterogeneity of two users can be reflected in terms of interest levels in items when they are in different groups w.r.t. these items. With users clustered, the novel *clustering of bandits with user features*, which is rooted in the multi-armed bandit framework, is then used to address the EE problem. HSB therefore inherits the competence of LinUCB [Chu et al., 2011] and UCB1 [Auer et al., 2001] to balance the EE problem and improves learning by the idea of sparse relevant user features.

Specifically, HSB addresses a practical learning situation in which aggregate observations are many but the received samples of each individual user are few. In other words, HSB clustering inference does not hinge on how many observations the learner has in a single user's historical data. More precisely, what HSB needs to select relevant features (and hence make the clustering inference) of a given item is data consisting of dependent variables (user responses or payoffs when items are assigned) and covariates (user features) w.r.t. the item; individual information (e.g., user ID) utilized for storing each user's historical activities plays no role in the selection. HSB therefore takes advantage of learning user heterogeneity by identifying relevant user features, yielding improved learning performance with benchmark bandits in practical situations such as in the Yahoo! R6B dataset. Roughly 50k distinct users visit the Yahoo! website in the first 70k visits in a Yahoo! dataset (hence the averaged visiting times for each user is no more than 2; see simulation section). Such learning situations are common in applications.

Moreover, the statistical selection tools we use enable our learning method to select relevant features from high-dimensional features in practice. Most learning methods demand preprocessed user features to function normally when the number of features is so many as to be proportional to total learning rounds. With binary user features and finitely many user features and items, we show that with high probability, the regret upper bound for HSB is the sum of multiple classic multi-armed bandits' regret upper bounds; the summation depends on how many relevant features (hence user groups) each item corresponds to. A key result for regret analysis which may be of independent interest is the *almost surely* feature selection of the statistical tool; *consistent* feature selection is not enough for sequential data(see section 4). To the authors' knowledge, the idea of combining statistical feature selection given sequential data has never been formally studied.

1.1 RELATED WORK

In terms of how users with different preferences are identified, HSB forms a contrast to state-of-the-art learning methods that also consider heterogeneity in item preferences. In [Cesa-Bianchi et al., 2013], the user network (clusters) are predefined and static over time. DynUCB, COFIBA, CLUB and CAB[Nguyen and Lauw, 2014, Li et al., 2016, Gentile et al., 2014, Gentile et al., 2017][1] cluster users based on *individual* (hence user IDs are involved) item preferences estimated on the fly. For CLUB, each user is associated with estimated item preferences; for this user, the estimation is based on her own historical activity data before the current round. Clustering is then performed by separating two users when two estimated preferences differ from each other by a non-trivial data-driven amount (an estimated confidence bound). Clustering in CAB shares most spirit with that of CLUB but CAB forms user groups w.r.t. a given item as well as the coming user (seen by the learner at the current round): the estimated users group includes users that tend to give a similar payoff to the given item as the coming user. Therefore,

[1] To see how their estimated (items') context vectors can be explained as user item preferences, let context vectors' there be one-hot encoding vectors; also note that in this work we use the term 'users' contextual information' (see section 2) as an alternative to 'user features'.

the clustering inference made by CAB is similar to that by HSB in the sense that the inference is item-dependent. COFIBA also makes item-dependent clustering in a way similar to CAB but is somewhat involved. Among these methods, DynUCB is the only one has fixing number of (predefined constant) user clusters; DynUCB re-assigns the coming user to the closest group (the measurement is the distances between the item-independent estimated preferences, which are updated over time, of this user and each group) at each round.

In terms of how the clustering inference evolves as the observations accumulate, CAB also resembles HSB. At each round, both HSB and CAB 'forget' the clustering results made in previous rounds, and make new clustering inferences based on the data in hand. This flexibility allows the methods to react quickly to a sudden missing set of data such as when an influential user drops off the account or a popular product is currently not for sale; they simply perform another clustering at the next round using the updated data. On the other hand, CLUB, COFIBA and DynUCB's clustering inferences depend on the entire sequential data collected prior to the current time.

Overall HSB is unique among state-of-the-art methods for its clustering inference is SRF-based (instead of individual-based). Besides, the flexible algorithm designs (both item-dependent and "forgetting" clustering inference) equip HSB as well as some other methods, especially CAB, the ability to handle dynamic learning situations. In section 5 we shall see the competitive performance of HSB against other benchmark and state-of-the-art methods.

2 Learning Model, Clustering of Bandits With User Features, and Statistical Tools

2.1 Learning Model

We consider a bandit learning environment where an algorithm receives one user i_t and available items set $C_t = \{\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,c_t}\}$ at discrete time t; all users are assumed to be in a user set $\mathcal{U} = \{\boldsymbol{x}_1^{\mathcal{U}}, \ldots, \boldsymbol{x}_n^{\mathcal{U}}\} \subset \{0,1\}^{d_{\mathcal{U}}}$ and items in item set $C_t \subset \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M\} \equiv \mathcal{I} \subset \mathbb{R}^{d_{\mathcal{I}}}$. Once a user i_t is assigned an item $\boldsymbol{x} \in C_t$ by the algorithm, she delivers a payoff, $y_t \in \{0,1\}$, indicating she likes the item ($y_t = 1$, or click) or not ($y_t = 0$).

For each item m, we model the relation between the user's features (or, equivalently, the user's contextual information) and her interest level for the assigned items by a logistic model. We define $f_m(\boldsymbol{x}^{\mathcal{U}})$ as the probability for a user $\boldsymbol{x}^{\mathcal{U}}$ to like item m,

$$f_m(\boldsymbol{x}^{\mathcal{U}}) = \frac{\exp\left(\boldsymbol{\beta}^{m'}\boldsymbol{x}^{\mathcal{U}}\right)}{1+\exp\left(\boldsymbol{\beta}^{m'}\boldsymbol{x}^{\mathcal{U}}\right)}, \tag{1}$$

where $\boldsymbol{x}^{\mathcal{U}} \in \mathcal{U}$ and $\boldsymbol{\beta}^{m'} = (\beta_1^m, \ldots, \beta_{d_{\mathcal{U}}}^m)'$ is the constant (but unknown) coefficient vector. Conditional on i_t, $\{y_t = 1\}$ is independent of other varialbles and with probobility $f_m(i_t)$. We refer to $f_m(\boldsymbol{x}^{\mathcal{U}})$ as $\boldsymbol{x}^{\mathcal{U}}$'s interest level for item m. In addition, the model assumes there are subsets $J^m \subset \{1, \ldots, d_{\mathcal{U}}\}$ such that $\beta_i^m = 0, i \in J^m$, $m = 1, \ldots, M$.

(1) provides a general framework for modeling item preferences while taking into account user group information (e.g., user $\boldsymbol{x}^{\mathcal{U}}$'s item preference is $(f_1(\boldsymbol{x}^{\mathcal{U}}), \ldots, f_M(\boldsymbol{x}^{\mathcal{U}}))'$). In practice, the use of the logistic model gives our algorithm access to modern statistical feature selection methods as well as a straightforward model interpretation when selecting relevant features for each item. Under (1) and given an upcoming user $i_t \in \mathcal{U}$, the machine's goal is to recommend item m such that $m = \arg\max_{m' \leq M} f_{m'}(i_t)$. The regret incurred by recommending item $\bar{m}$ is defined to be (at time t)

$$r_t = \max_{m' \leq M} f_{m'}(i_t) - f_{\bar{m}}(i_t) \tag{2}$$

When the performance is evaluated by (2) (in terms of minimizing the regret), a competitive learning algorithm must be able to balance the samples used for estimating user preference means and those used for deciding the best $\bar{m}$ at each round t: this is known as the exploration-exploitation (EE) problem. To address the EE problem under framework (1) given $J^j, j = 1, \ldots, M$, we introduce our novel approach *clustering of bandits with user features* in Section 2.2. Note that J^j is essentially unknown to the machine ahead of time in most practical cases. In section 2.3, we introduce statistical sparse relevant feature selection methods that can be used to estimate J^j under modeling framework (1).

2.2 Clustering of Bandits With User Features

The relevant feature index set J^j is used to cluster users w.r.t item j. For any two distinct users with same values of relevant features ($x_i^{\mathcal{U}} = x_i^{\mathcal{U}'}$ for $i \in J^j$, where $x_i^{\mathcal{U}}, x_i^{\mathcal{U}'}$ are the i-th component in $\boldsymbol{x}^{\mathcal{U}}, \boldsymbol{x}^{\mathcal{U}'}$, respectively), their interest levels for the item are the same; otherwise they can have different interest levels for this item. Therefore, for each item j, there are at most $2^{\#J^j}$ (our feature space is assumed to be $\{0,1\}^{d_{\mathcal{U}}}$) different groups of users in the sense that users in the same group exhibit the same interest in the item and may or may not do so with users from different groups. We cluster users into these $2^{\#J^j}$ groups w.r.t item j. After the clustering, each item faces groups of users such that *i)* the user groups are mutually exclusive and collectively exhaustive (collectively equivalent to $\mathcal{U}$); *ii)* these groups change across items if and only if the sets of relevant features do so.

When J^j is given and users are clustered in such a fashion, a version of the multi-armed bandit method can be used to balance the EE problem. Define $g_j(\boldsymbol{x}^{\mathcal{U}})$ to be the user group that $\boldsymbol{x}^{\mathcal{U}}$ belongs to w.r.t item j. Let $A_{j,t}(D), C_{j,t}(D)$, $D \subset \mathcal{U}$ denote the number of assignments and clicks, respectively, w.r.t. item j before round t based on historical activity data of D. A naive estimator for $f_j(\boldsymbol{x}^{\mathcal{U}})$ is then $\frac{C_{j,t}(g_j(\boldsymbol{x}^{\mathcal{U}}))}{A_{j,t}(g_j(\boldsymbol{x}^{\mathcal{U}}))}$. The upper confidence bound [Lai and Robbins, 1985, Auer et al., 2001], given upcoming user i_t and item j, is calculated by

$$UCB_{j,t} = \frac{C_{j,t}(g_j(i_t))}{A_{j,t}(g_j(i_t))} + \alpha\sqrt{\frac{\log(1+t)}{1+A_{j,t}(g_j(i_t))}},$$

where $\alpha > 0$ is the exploration tuning parameter. The machine then assigns the item maximizing user i_t's UCBs. Notice that if the values of J^j are identical, the bandit (a traditional UCB1 ([Auer et al., 2001]) for example) is essentially clustered into $2^{\#J^1}$ independent bandits run at the same time, each of which corresponds to a unique group of users.

2.3 Statistical Sparse Relevant Feature Selection

Logistic Trimming

Logistic regression is typically applied to differentiate relevant features from less informative features [Fan and Li, 2006]. Given the sample $x_1, \ldots, x_n \in \mathbb{R}^p$ for some constant p and the corresponding binary responses $y_1, \ldots, y_n$, we define the logistic loss function (log-likelihood) as

$$l_n(\beta) = \frac{1}{n}\sum_{t=1}^{n}\left(y_t x_t^{'}\beta - \log\left(1+\exp\left(x_t^{'}\beta\right)\right)\right),$$

where $\beta = (\beta_1, \ldots, \beta_p)^{'} \in \mathbb{R}^p$; and the mean vector estimator is $\hat{\beta}_{J,n} = \arg\max_{\beta \in \Theta; \beta_i = 0, i \in J^c} l_n(\beta)$, where $J \subset \{1, \ldots, p\} \equiv J_p$, $J^c = \{1, \ldots, p\} \backslash J$ and $\Theta \subset \mathbb{R}^p$ is the parameter space. To retrieve the information for the relevant features, we define the information criterion (IC, the sum of the estimated goodness of fit and a penalty proportional to the complexity of the fitted model) as $\mathrm{IC}_n(J) = l_n(\hat{\beta}_{J,n}) + \#JR_n$, where $R_n = O\left(\frac{(\log n)^{1+\varepsilon}}{n}\right)$ is some deterministic sequence with arbitrary small $\varepsilon > 0$. The selection procedure is described in Algorithm 1.

Algorithm 1 Logistic Trimming

```
    Input: $\{x_i, y_i\}_{i=1}^n$, $\mathcal{RF} = \emptyset$; Output: $\mathcal{RF}$;
for $j = 1, \ldots, p$ do
    if $n \geq p$ and $\mathrm{IC}_n(J_p) \leq \mathrm{IC}_n(J_p \backslash \{j\})$ then
        $\mathcal{RF} = \mathcal{RF} \cup \{j\}$
    end if
end for
```

Logistic Orthogonal Matching Pursuit

Logistic Orthogonal Matching Pursuit(LOMP) outputs a set of selected relevant features and meanwhile addresses selection from high-dimensional features(i.e. a large number of features). The algorithm LOMP in [Chen et al., 2018] adopts orthogonal matching pursuit and attempts to select the variable minimizing the objective function while fixing all other variables (gradient descent) at each step. The algorithm eventually stops the selection process and holds only a minor portion of the whole variables considered as those most likely to "explain" the dependent variable. The last step of the algorithm trims off unnecessary variables retained from previous selection in a way similar to Logistic Trimming, and reports the resulted set of variables; we refer the reader to [Chen et al., 2018] for a specific description of the algorithm. Processing the algorithm in this way sidesteps the dimensional issue, as only a handful of variables is under consideration at any one time, while still making efficient use of the information for all variables.

3 Heterogeneity-Sensitive Bandit

Heterogeneity-sensitive bandit (HSB) consists of two parts: A protection mechanism in Algorithm 2 that protects the information quality of the collected data from deteriorating, and Algorithm 3, which neatly integrates the clustering of bandits with user features and statistical feature selection methods for better prediction performance.

Data Collection and Protection Mechanism

Protection is mainly about storing sets of data while ensuring for the smooth application of logistic regression. Algorithm 2 keeps tracks of and updates two kinds of data collections – $X(\boldsymbol{x})$, $y(\boldsymbol{x})$, $\bar{X}(\boldsymbol{x})$, $\bar{y}(\boldsymbol{x})$, $\boldsymbol{x} \in \mathcal{I}$ – for usage in UCB calculations and relevant feature selection. At round t (line 13), the algorithm stores information on the coming user $i_t \in \mathcal{U}$, assignment $\boldsymbol{x}_{a_t'}$, and delivered payoff y_t by appending them accordingly to the end of $X(\boldsymbol{x}_{a_t'}), y(\boldsymbol{x}_{a_t'})$, all of which start out as empty sets. As for the sophisticated collections of $\bar{X}(\boldsymbol{x}), \bar{y}(\boldsymbol{x})$, for all $\boldsymbol{x} \in \mathcal{I}$, HSB assesses the quality of information associated with $\bar{X}(\boldsymbol{x}_j), a_t^{'} = j$, and appends i_t, y_t to $\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j)$, respectively, if they are qualified (the minimum eigenvalue such that $\lambda_{\min}(\lambda_j) \geq c_0$, i.e., if the average minimum eigenvalue of sample covariate-variance matrix is non-degenerate). If not, the algorithm stops appending data and, at rounds $t \in Q$ (lines 4, 5), the algorithm assigns item j to the coming users $d_{\mathcal{U}}$ times (if $\boldsymbol{x}_j \in C_t$; lines 3, 5) without optimizing UCBs. If too many items are disqualified, the algorithm randomly selects one from among them.

```
Algorithm 2 Heterogeneity-Sensitive Bandit
    Input
   - Set of users $\mathcal{U}$, set of items $\mathcal{I}$;
   - Recorded data $\bar{X}(\boldsymbol{e}), \bar{y}(\boldsymbol{e}), X(\boldsymbol{e}), y(\boldsymbol{e}) = \emptyset$, $\boldsymbol{e} \in \mathcal{I}$;
   - $a_0 = 0$, $f_j = 0$, $\lambda_j = \beta \boldsymbol{I}_{d_{\mathcal{U}}}$, $t_j = 1$ for all $j \leq M$,
     $Q = \{i^b | i \in \mathbb{N}\}$ for any $b > 2$;
 1: for $t = 1, \ldots, T$ do
 2:     if $f_j > 0$ for any $j \in \{i : \boldsymbol{x}_i \in C_t\}$ then
 3:         $a_t^{'} = j$; $a_t$ such that $\boldsymbol{x}_{t,a_t} = \boldsymbol{x}_j$;
 4:     else if $t \in Q$ and $\lambda_{\min}(\lambda_j) < c_0$ for some
    $j \in \{i : \boldsymbol{x}_i \in C_t\}$ then
 5:         $a_t^{'} = j$, $f_{a_t^{'}} = d_{\mathcal{U}}$, $a_t$ such that $\boldsymbol{x}_{t,a_t} = \boldsymbol{x}_j$;
 6:     else
 7:         Set $a_t, a_t^{'}$ = Algorithm 3;
 8:         Set flag to TRUE;
 9:     end if
10:     Observe payoff $y_t$;
11:     Append $i_t$, $y_t$ to $X(\boldsymbol{x}_{t,a_t}), y(\boldsymbol{x}_{t,a_t})$, respectively;
12:     if not flag or (flag and $\lambda_{\min}(\lambda_{a_t^{'}}) \geq c_0$) then
13:         Append $i_t$, $y_t$ to $\bar{X}(\boldsymbol{x}_{t,a_t}), \bar{y}(\boldsymbol{x}_{t,a_t})$, respectively;
14:         Set $\left(\lambda_{a_t^{'}}, t_{a_t^{'}}\right) = \left(\frac{\lambda_{a_t^{'}} t_{a_t^{'}} + i_t i_t^{'}}{t_{a_t^{'}} + 1}, t_{a_t^{'}} + 1\right)$;
15:         Set flag to FALSE;
16:     end if
17:     $f_{a_t^{'}} = \max\{f_{a_t^{'}} - 1, 0\}$;
18: end for
$\lambda_{\min}(A)$ denotes the minimum eigenvalue of $A$.
```

Learning User Preferences and Making Predictions

Having $\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j)$ in hand at round t (line 1), Algorithm 3 applies statistical feature selection methods('Logistic(X, y)' denotes the application of either Logistic Trimming or Logistic Orthogonal Matching Pursuit to datasets X, y) to each $j = 1, \ldots, M$ of $\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j)$ to obtain and record the corresponding relevant users' features, $\hat{\mathcal{RF}}_j \subset \{1, \ldots, d_{\mathcal{U}}\}, j = 1, \ldots, M$. The user subset w.r.t. items s given at time t, $\hat{P}_s$, can then be constructed (line 2). The UCBs are calculated accordingly. Notice that the users clustering inference made in Algorithm 3 is updated at each round.

For notation, we define $\boldsymbol{x}^{\mathcal{U}}(D)$, non-empty $D \subset \{1, \ldots, d_{\mathcal{U}}\}$, to be a real number index such that $\boldsymbol{x}^{\mathcal{U}}(D) = \boldsymbol{x}^{\mathcal{U}'}(D)$ if and only if the D-th features in $\boldsymbol{x}^{\mathcal{U}}$ and $\boldsymbol{x}^{\mathcal{U}'}$ are identical; if $D = \emptyset$, define $\boldsymbol{x}^{\mathcal{U}}(D) = \boldsymbol{x}^{\mathcal{U}'}(D)$ for any $\boldsymbol{x}^{\mathcal{U}}, \boldsymbol{x}^{\mathcal{U}'}$. Let $y_j(D), D \subset \mathcal{U}$, denote the subset of $y(\boldsymbol{x}_j)$ associated with users D only. Moreover, we define a function, $l(.)$, for measuring the length of $y_j(\boldsymbol{x}^{\mathcal{U}})$ and the ones vector, $\boldsymbol{1} = (1, \ldots, 1)^{'}$, with a length equivalent to that of whichever vector it is multiplied by.

analysis section).

```
Algorithm 3
Input
   - Set of users $\mathcal{U}$, set of items $\mathcal{I}$;
   - Exploration parameter $\alpha$;
   - Users' historical activity data;
   - User $i_t$;
Output Decision at time $t$: $a_t, a_t^{'}$;
 1: Set $\hat{\mathcal{RF}}_j = \text{Logistic}(\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j)), j = 1, \ldots, M$;
 2: Set $\hat{P}_s = \left\{\boldsymbol{x}^{\mathcal{U}} | \boldsymbol{x}^{\mathcal{U}}(\hat{\mathcal{RF}}_{k_s}) = i_t(\hat{\mathcal{RF}}_{k_s}), \boldsymbol{x}^{\mathcal{U}} \in \mathcal{U}\right\}$,
    where $k_s$'s are such that $\boldsymbol{x}_{k_s} = \boldsymbol{x}_{t,s}, s = 1, \ldots, c_t$;
 3: Recommend $a_t, a_t^{'}$ such that $\boldsymbol{x}_{a_t^{'}} = \boldsymbol{x}_{t,a_t}$ and
        $a_t = \arg \max_{j=1,\ldots,c_t} \frac{y_j^{'}(\hat{P}_j)\boldsymbol{1}}{l(y_j(\hat{P}_j))} + \hat{CB}_j$,
        $\hat{CB}_j = \alpha \sqrt{\frac{\log\left(1 + \sum_{s \leq c_t} l(y_s(\hat{P}_s))\right)}{1 + l(y_j(\hat{P}_j))}}$;
```

4 Regrets Analysis

We show with high probability that HSB almost surely selects relevant features. The regrets of deploying HSB can hence be bounded by a sum of multiple multi-armed regrets bounds plus strategic assignments and a deterministic number.

Condition 1. A sequence of $\{z_{t,1}, z_{t,2}\}$ such that

$$P(z_{t,2} = q | z_{t,1}) = \left(\frac{\exp\left(z_{t,1}^{'} \beta^*\right)}{1 + \exp\left(z_{t,1}^{'} \beta^*\right)}\right)^q \left(\frac{1}{1 + \exp\left(z_{t,1}^{'} \beta^*\right)}\right)^{1-q},$$

$q = 1, 0$, where $z_{t,1} \in \mathbb{R}^{d_{\mathcal{U}}}$ and $\beta^* = (\beta_1^*, \ldots, \beta_{d_{\mathcal{U}}}^*)^{'} \in \Theta \subset \mathbb{R}^{d_{\mathcal{U}}}, |\Theta| < \infty$; ii) $z_{t,2}$ conditional on $z_{t,1}$ is independent of all other variables.

Condition 2. A sequence $\{x_i\}$ such that x_i's are independently and identically distributed with $0 < \|x_1\| < \infty$ and $\lambda_{\min}\left(E(x_1 x_1^{'})\right) > 0$.

Contrary to standard consistent feature selection [Zhao and Yu, 2007, Fan and Li, 2001, Zou, 2006, Chen et al., 2018], in a sequential context, selection is an ongoing activity as sample size increases while preserving all previous data. Theorem 1 adapts statistical feature selection for sequential data by a result of almost sure feature selection. The relation between users' contextual information and responses to assigned item j is assumed to satisfy Condition 1 with $\beta^* = \boldsymbol{\beta}^j$.

Theorem 1. *Fix an item j. Assume the i_t's drawn from $\mathcal{U}$ satisfy Condition 2. Let $\hat{J}_t = Logistic(\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j))$ (Logistic Trimming is applied). Then*

$$P(\hat{J}_t = J^j \text{ eventually}) = 1.$$

Note that at times t when $f_j \neq 0$, Logistic Trimming is not applied to data. Regrets analysis of HSB requires the result in Theorem 1 and the standard results of regrets bound for multi-armed bandit problems [Abbasi-Yadkori et al., 2011]. Theorem 2 gives a regrets upper bound for HSB.

Theorem 2. *Consider an environment with a sequence of $\{i_t, C_t\}$ such that $C_t = \{\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,c_t}\}$, $1 \leq c_t$ for all t, is arbitrarily drawn from $\mathcal{I}$; i_t's drawn from $\mathcal{U}$ satisfy Condition 2. Then with Logistic Trimming and*

$$C\hat{B}_j = \sqrt{\frac{(l(y_j(\hat{P}_j)) + 1)}{l(y_j(\hat{P}_j))^2} \left(1 + 2\log\left(\frac{2^{K_{\mathcal{U}}} M (1 + l(y_j(\hat{P}_j)))^{1/2}}{\delta}\right)\right)},$$

where $K_{\mathcal{U}} = \# \cup_k J^k$, HSB has regrets, of probability $1 - \delta - M\delta_2$, such that for all large T,

$$\sum_{t=1}^{T} r_t \leq 2^{K_{\mathcal{U}}} M \left(3\Delta_M + \frac{16}{\Delta_m} \log\left(\frac{2^{K_{\mathcal{U}}+1} M}{\Delta_m \delta}\right)\right) + T_{\delta_2, d_{\mathcal{U}}, b} + d_{\mathcal{U}} T^{\frac{1}{b}}, \tag{3}$$

where $\Delta_M = \max E$, $\Delta_m = \min E \backslash \{0\}$, $E = \{|f_j(\boldsymbol{x}^{\mathcal{U}}) - f_k(\boldsymbol{x}^{\mathcal{U}'})| : j, k \leq M; \boldsymbol{x}^{\mathcal{U}}, \boldsymbol{x}^{\mathcal{U}'} \in \mathcal{U}\}$, and $T_{\delta_2, d_{\mathcal{U}}, b}$ is a constant depending on subscript parameters and δ_2 is due to the application of Theorem 1.

By letting δ be $\frac{1}{T}$, the first term in (3) yields a classical $\log T$ upper bound with high probability; the third term has a trade-off relation to the constant term $T_{\delta_2, d_{\mathcal{U}}, b}$. Built upon our novel perspective toward the usage of users' contextual information, HSB maintains a regrets upper bound of standard multi-armed bandit problems plus a constant resulting from the application of Theorem 1 and strategic assignments. Theoretically, b can be arbitrarily large; we can set it to a constant greater than 2 so as to guarantee a better bound than $\sqrt{T}$, an upper bound for standard contextual bandits.

5 Experiments

5.1 Datasets

Artificial datasets

In the artificial datasets there are 500 items, and $\#\mathcal{U} = I \in \{50, 5000\}$ users, each associated with 100 user features represented by a vector of 100 binary numbers ($\boldsymbol{x}^{\mathcal{U}} \in \{0, 1\}^{100}$). All user features are generated uniformly at random from $\{0, 1\}$. A user's item preference is described by a vector of 500 probabilities, each of which indicates the interest level of this user for the corresponding item, i.e. the chance this user clicks on the corresponding item (if the item is assigned). The interest levels in a preference vector are decided by a simple linear probability model. Specifically, item j is associated with a coefficient vector of length 100, $\boldsymbol{\beta}^j = (\beta_1^j, \ldots, \beta_{100}^j)'$, and a set of indexes of relevant user features, $J^j \subset \{1, \ldots, 100\}$. The coefficient vector is sparse in the sense that $\beta_i^j = 0$ if $i \notin J^j$ and $\#J^j$ is a relatively small number. In the datasets, one of two index subsets, J_1^* and J_2^*, is randomly assigned to J^j with equal chance. On the other hand, without replacement we draw four numbers from $\{1, \ldots, 100\}$ and let J_1^*, J_2^* be sets of first two and last two numbers, respectively. We set the non-zero coefficients in $\boldsymbol{\beta}^j$ to $(p, -p)$ with $p \in \{0.05, 0.1\}$. The interest level of user $\boldsymbol{x}^{\mathcal{U}}$ for item j is $\boldsymbol{\beta}^{j'} \boldsymbol{x}^{\mathcal{U}} + \varepsilon_j$, where ε_j is drawn from a uniform $(0.1, 0.15)$ distribution that is independent of all other variables. There are in total $2^{\#(J_1^* \cap J_2^*)} = 16$ types of users.

The total learning rounds $T = 10,000$; at each round t, a i_t is randomly drawn from $\mathcal{U}$ and served to the machine. The machine assigns one item $\bar{m}$ at time t and the user clicks through ($y_t = 1$) with probability $\boldsymbol{\beta}^{\bar{m}'} i_t + \varepsilon_{\bar{m}}$. At time t, we evaluate the performance of a learning algorithm run on the artificial datasets by the averaged total click-throughs (click-through rate, CTR) given by $i_1, \ldots, i_t$, i.e., $\frac{\sum_{i \leq t} y_i}{t}$.

Yahoo dataset

Two real datasets, 18k and 550k, were extracted from the Yahoo! R6B dataset [Li et al., 2011]. Yahoo! R6B records the website visitor click logs for news articles displayed on Yahoo!'s front page. The dataset is a text file consisting of millions of lines, each of which contains the information about a single visit. The information includes a timestamp, the visitor's user features, whether the visitor clicked through, the new articles available to be displayed, and the displayed article. The visiting activities were recorded chronologically from October 2 to 16, 2011. At each line t, the available article set C_t contains roughly 30 to 50 articles. At the same time, there is an article drawn from C_t uniformly at random and displayed to user i_t (on the Yahoo! front page) by the website server. Moreover, i_t's user features were recorded as a binary vector of dimension 136. For more details, see [Gentile et al., 2014, Li et al., 2016, Li et al., 2011].

We prepared our 18k dataset using the following steps. We used only data from the first 7 days and removed visits for which the user features were all zeros (except the first feature, which is always 1). As these users are guests to the website, they have no features. Since the original dataset provided no information about user IDs, we viewed visitors with the same user features (all 136 user features) as the same user when preparing the datasets. Filtering out users that visited the website less than 50 times resulted in our first dataset, 18k. Thus this dataset recorded visiting activities for roughly 18k users.

In addition to the 18k dataset, we also sought to evaluate the algorithms on a less conditional environment, in the sense that the visit times were not considered when preparing the dataset. More precisely, we again began with the first 7 days and took out guest activities. For the 550k dataset, we retained only the i-th line such that $\frac{i}{3}$ is an integer. The resulting new dataset contained 2,787,635 lines and 552,588 distinct users (hence the name 550k). Only one-third of the lines were retained for 550k, so 18k and 550k contained roughly the same number of text lines.

We followed [Li et al., 2011] for the offline evaluation of learning algorithms. As the item (article) assigned by the machine at time t sometimes did not coincide with the assigned item in the recorded activities at line t, we deleted such lines and moved on to the next line whenever mismatches occurred. Roughly 70,000 rounds remained after running these algorithms on the datasets. The averaged click-throughs until time $t = 1, \ldots, T$, $\frac{\sum_{i \le t} y_i}{t}$, were used for evaluating the performance of the learning algorithms.

LastFM dataset

LastFM contains users, music artists, tags made by users for artists, and information about music listening activities. In the dataset, there are 1892 users and 17632 artists along with a document recording which user has listened to which artists and another file with information about tagging activities.

We generated a dataset of $T =$ 15,000 rounds with the following steps. 100 artists were randomly drawn from 12133 artists who were tagged by at least one user; these 100 artists made up the available items set, C_t, for all t. The visitor at each time, i_t, was drawn uniformly at random (with replacement) from 1892 users. The payoff of assigning artist (item) $j \in C_t$ to i_t was 1 if the artist had been listened to by i_t and 0 otherwise. At each round, one artist was assigned to i_t and payoff y_t was observed by the machine; $\frac{\sum_{i \le t} y_t}{t}$, $t = 1, \ldots, T$ was used to evaluate the learning methods.

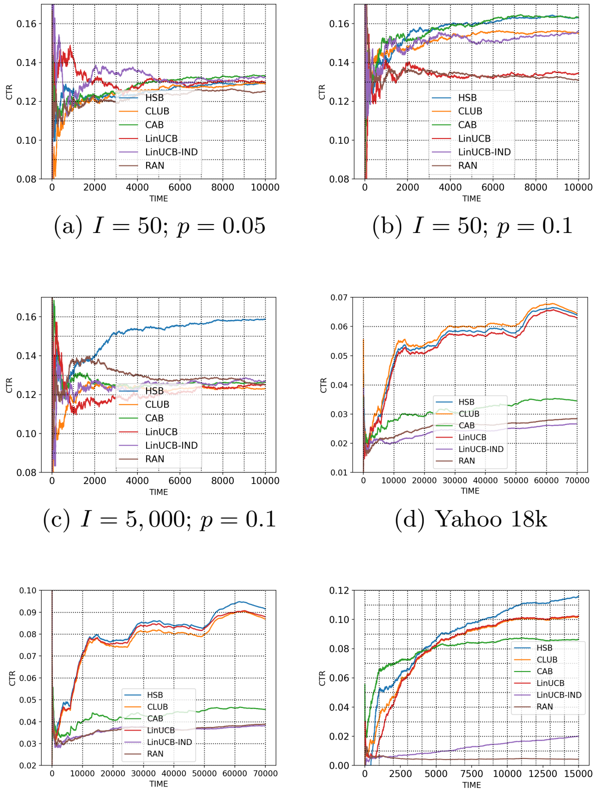
In addition, we used user tags to form user features. All tagged words were split by spaces, dashes, hyphens, and carets, yielding 9264 distinct 'finer tags'. A version of TFIDF was used to generate user features: the i-th feature was 1 for user $\boldsymbol{x}^{\mathcal{U}}$ if she tagged the i-th finer tag for any artist and 0 otherwise. Each user was associated with a 9264-dimensional binary vector representing the user features.

5.2 Algorithms

The tests results were averaged over 6 and 3 runs for CAB on lastFM and the rest experiments, respectively. The exploration-exploitation parameter α for every algorithm (except for RAN) and edge deletion parameter α_2 for CLUB were tuned by grid searching for the best setting on $\{0, 0.01, \ldots, 0.2\}$ and $\{0, 0.1, \ldots, 0.5\}$, respectively. We set $\gamma = 0.2$ for CAB. For tuning we used the first 60,000 text lines in Yahoo! datasets and 1,000 rounds for the artificial and lastFM datasets. With the tuned parameters, the reported tests results were based on the rest of the Yahoo datasets and another independently generated $t - t_0 =$ 10,000, 15,000 rounds for the artificial and lastFM datasets, respectively. In all of our datasets, item contexts were given by one-hot encoding binary vectors, i.e. where the i-th item's context is a all-zeros vector of length M with the i-th component set to 1.

- HSB: We evaluated a variant of HSB, where the protection mechanism was suppressed by setting the conditions of the if/else statements in line 2 and 4 to FALSE[2] and LOMP was used. Intercept was added into user features if it was not included.
- CLUB [Gentile et al., 2014]: An algorithm that specializes in learning user clusters on the fly. We implemented a variant of CLUB with an Erdos-Renyi graph as the initial user graph.
- CAB [Gentile et al., 2017]: At each round, CAB estimates the preference similarity between the coming individual and the other users regarding each item; it makes the prediction based on this similarity information.
- LinUCB & LinUCB-IND: Variants of benchmark algorithms. One single instance of LinUCB

[2] The protection mechanism can play a crucial role in particular practical situations. However, it is mainly of theoretical use in our experiments.



(a) $I = 50$; $p = 0.05$

(b) $I = 50$; $p = 0.1$

(c) $I = 5,000$; $p = 0.1$

(d) Yahoo 18k

(e) Yahoo 550k

(f) lastFM

[Chu et al., 2011] was used in LinUCB and each distinct individual used her own instance of LinUCB ing LinUCB-IND.

- RAN: Recommends an item uniformly drawn from C_t at each round.

5.3 Results

In the artificial datasets there was a uniform item preference if p was set to 0. When I is relativly small, LinUCB-IND effectively addresses heterogeneity in preference. Compared with (b) and (c), the results in (a) suggests seeing case $p = 0.05$ as a case with almost uniform user preference, due to the closeness in performance of LinUCB-IND and LinUCB. In light of this, the cases $p = 0.1$ exhibit appropriate levels of heterogeneity, as the performance of LinUCB is essentially equivalent to that of RAN when $p = 0.1$.

As expected, HSB's performance is independent of I, whereas all others are essentially RAN when there are $5,000$ distinct users given a heterogeneity level of $p = 0.1$. This result demonstrates the difference between the clustering methods adopted in CLUB, CAB, and HSB. CLUB and CAB clusterings tend to work very well when every user has relatively rich historical activity data (compared to T) such as the cases in (a) and (b); on the other hand, the HSB clustering relies on the sparsity of relevant user features for competitive performance.

With relatively few distinct users, only CAB is sensitive enough to compete with LinUCB-IND in an environment with mild heterogeneity in user preference ($p = 0.05$). LinUCB-IND is not at its most efficient state in the sense that there are only 16 user types but the information of 50 distinct users are used; adaptive learning algorithms (CAB, HSB, CLUB) in this case ($p = 0.1; I = 50$) use the user types information more efficiently and outperform LinUCB-IND.

Sparse relevant features were assumed in our simulation setting. We used Yahoo! and lastFM datasets to further evaluate the practical aspect of this assumption. Yahoo! is a popular, large e-company: its website portal is visited by all types of users everyday. All news articles on this website must be well-written to attract the most types of users, with as many click-throughs as possible. This means it is very difficult to compete with or outperform LinUCB on the Yahoo! dataset. For lastFM, as an online music streaming service, it is expected to have users with heterogeneous music tastes. However, songs from a small number of popular artists account for a large portion of online listening activity; we do not presume a great level of heterogeneity. ID information is available in lastFM but not in Yahoo!; thus in Yahoo!, we assign a unique ID to users with a distinct set of user features.

On the 18k dataset, CLUB dominates thoroughly along with HSB and LinUCB: CLUB and HSB (especially CLUB) prove their ability to learn and use the information of such mild heterogeneity in user preferences. For the 550k dataset, the average visit times for each user are too few; CLUB no longer retains its learning advantage. HSB, on the other hand, still maintains its learning power especially during the last 20 thousand rounds.

In the lastFM case, the average visit time for each user is $\frac{15,000}{1892} = 7.93$ by $t = T = 15,000$, which is far too little for LinUCB-IND, CAB, and CLUB to learn the intended heterogeneous item preferences. Clustering based on sparse relevant user features mitigates the problem of insufficient samples for each distinct user; overall, HSB outperforms LinUCB and CLUB. The results of CLUB essentially coincide with that of LinUCB, which is the winner between the results of LinUCB and LinUCB-IND. On the other hand, CAB addresses the cold-start issue at the first 4000 rounds.

The results in the lastFM case also attest the competence of HSB in selecting relevant features from 9264 user features. On the whole, these results show that the idea of sparse relevant user features is practical.

References

[Abbasi-Yadkori et al., 2011] Abbasi-Yadkori, Y., Pal, D., and Szepesvari, C. (2011). Improved algorithms for linear stochastic bandits. In *Proc. NIPS*.

[Auer, 2002] Auer, P. (2002). Using confidence bounds for exploration-exploitation trade-offs. *Journal of Machine Learning Research*, 3:397–422.

[Auer et al., 2001] Auer, P., Cesa-Bianchi, N., and Fischer, P. (2001). Finite-time analysis of the multiarmed bandit problem. *Machine Learning*.

[Cesa-Bianchi et al., 2013] Cesa-Bianchi, N., Gentile, C., and Zappella, G. (2013). A gang of bandits. In *Proc. NIPS*.

[Chen et al., 2018] Chen, Y. L., Dai, C. S., and Ing, C. K. (2018). Model selection for high-dimensional sparse nonlinear models using chebyshev greedy algorithm. *Technical Report*.

[Chu et al., 2011] Chu, W., Li, L., Reyzin, L., and Schapire, R. E. (2011). Contextual bandits with linear payoff functions. In *Proc. AISTATS*.

[Fan and Li, 2001] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*.

[Fan and Li, 2006] Fan, J. and Li, R. (2006). Statistical challenges with high dimensionality: Feature selection in knowledge discovery. In *International Congress of Mathematicians*.

[Gentile et al., 2017] Gentile, C., Li, S., Kar, P., Karatzoglou, A., Zappella, G., and Etrue, E. (2017). On context-dependent clustering of bandits. In *Proc. 34th ICML*.

[Gentile et al., 2014] Gentile, C., Li, S., Kar, P., and Zappella, G. (2014). Online clustering of bandits. In *Proc. 31th ICML*.

[Krause and Ong, 2011] Krause, A. and Ong, C. S. (2011). Contextual gaussian process bandit optimization. In *Proc. 25th NIPS*.

[Lai and Robbins, 1985] Lai, T. and Robbins, H. (1985). Asymptotically efficient adaptive allocation rules. *Advanced in Applied Mathematics*, 6:4–22.

[Langford and Zang, 2007] Langford, J. and Zang, T. (2007). The epoch-greedy algorithm for contextual multi-armed bandits. In *Proc. NIPS*.

[Li et al., 2010] Li, L., Chu, W., Langford, J., and Schapire, R. E. (2010). A contextual-bandit approach to personalized news article recommendation. In *Proc. WWW*, pages 661–670.

[Li et al., 2011] Li, L., Chu, W., Langford, J., and Wang, X. (2011). Unbiased offline evaluation of contextual-bandit-based news article recommendation algorithms. In *Proc. WSDM*.

[Li et al., 2016] Li, S., Karatzoglou, A., and Gentile, C. (2016). Collaborative filtering bandits. In *Proc. 39th SIGIR*.

[Nguyen and Lauw, 2014] Nguyen, T. T. and Lauw, H. W. (2014). Dynamic clustering of contextual multi-armed bandits. In *23rd CIKM*, pages 1959–1962.

[Zhao and Yu, 2007] Zhao, P. and Yu, B. (2007). On model selection consistency of lasso. *Journal of Machine Learning Research*, 7:2541–2563.

[Zou, 2006] Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*.

A Appendix

A.1 Proof of Theorem 1 and 2

Condition 1. A sequence of $\{z_{t,1}, z_{t,2}\}$ such that *i)*

$$P(z_{t,2} = q|z_{t,1}) = \left(\frac{\exp\left(z_{t,1}^{'}\beta^*\right)}{1+\exp\left(z_{t,1}^{'}\beta^*\right)} \right)^q \left(\frac{1}{1+\exp\left(z_{t,1}^{'}\beta^*\right)} \right)^{1-q},$$

$q = 1, 0$, where $z_{t,1} \in \mathbb{R}^{d_{\mathcal{U}}}$ and $\beta^* = (\beta_1^*, \ldots, \beta_{d_{\mathcal{U}}}^*)^{'} \in \Theta \subset \mathbb{R}^{d_{\mathcal{U}}}, |\Theta| < \infty$; *ii)* $z_{t,2}$ conditional on $z_{t,1}$ is independent of all other variables.

Condition 2. A sequence $\{x_i\}$ such that x_i's are independently and identically distributed with $0 < \|x_1\| < \infty$, a.s. and $\lambda_{\min}\left(E(x_1 x_1^{'})\right) > c_0 > 0$.

Condition 3. This condition defines a process $\{z_j\}$. Let c_0 be defined as in Condition 2 and z_1 be arbitrary but bounded (by some $C > 0$) vector of length $d_{\mathcal{U}}$. Define $\eta_0 = \lambda_{\min}(\beta \boldsymbol{I}) > c_0$ and

$$\eta_1 = \frac{1}{t}\lambda_{\min}\left(\sum_{s\leq 1} z_s z_s^{'} + \beta \boldsymbol{I}\right).$$

For $t > 1$, we recursively define z_{t+1} and η_{t+1}. Set *counter* to 0 at $t = 2$,

$$z_{t+1} = \begin{cases} \text{if } (\eta_t \geq c_0) \text{ and } (\textit{counter} = 0 \text{ or } \textit{counter} = d_{\mathcal{U}}): \\ \quad \text{arbitrarily } z_{t+1} \in \mathbb{R}^{d_{\mathcal{U}}} \text{ with } \|z_{t+1}\| < C; \\ \quad \text{set } \textit{counter} \text{ to } 0. \\ \text{else if } (\textit{counter} = d_{\mathcal{U}}) \text{ and } (\eta_t < c_0): \\ \quad \text{an independent } x_1 \text{ satisfying Condition 2;} \\ \quad \text{set } \textit{counter} = 1. \\ \text{else : an independent } x_1 \text{ satisfying Condition 2;} \\ \quad \text{set } \textit{counter} = \textit{counter} + 1. \end{cases}$$

and

$$\eta_{t+1} = \frac{1}{t}\lambda_{\min}\left(\sum_{s\leq t+1} z_s z_s^{'} + \beta \boldsymbol{I}\right).$$

Let $T_1 = \inf\{t : \eta_t < c_0\}$, $T_i = \inf\{t : \eta_t < c_0, t \geq T_{i-1} + d_{\mathcal{U}}\}, i > 1$, and $T_0 = 1 - d_{\mathcal{U}}$. For a process satisfying Condition 3, we define a corresponding random sequence $\{s_i\}$ by assigning the k-th (chronologically) element in $\cup_{j>0}\{i : T_{j-1} + d_{\mathcal{U}} \leq i \leq T_j\}$ to s_k. Notice that these rounds exclude the strategical assignment rounds (those rounds at which items are randomly assigned) but include $T_{j-1} + d_{\mathcal{U}}$'s.

To prove Theorem 1, we first introduce Theorem 3, which says under the given conditions, Logistic Trimming (Algorithm 1) can almost surely select relevant features (along with a subsequence). We defer the proof of Theorem 3 to the last paragraph after we introduce Lemma 1 to 3 and notation.

Theorem 3. *Given a sequence $\{x_i, y_i\}$ satisfying Condition 1 and $\{x_i\}$ satisfying Condition 3 (hence Condition 2). Let $J_p^* = \{i : \beta_i^* \neq 0, i \leq d_{\mathcal{U}}\}$ and $\hat{J}_t$ = Logistic(X_t, Y_t), where $X_t = (x_1, \ldots, x_t)^{'}$, $Y_t = (y_1, \ldots y_t)^{'}$ (Logistic Trimming is applied). Then*

$$P(\hat{J}_{s_n} = J_p^* \textit{ eventually}) = 1.$$

Proof of Theorem 1. With probability 1, $l(\bar{y}(\boldsymbol{x}_j))$ increases to infinity as t in Algorithm 2 passes to infinity. Consider a subsequence of $\bar{X}(\boldsymbol{x}_j), \bar{y}(\boldsymbol{x}_j)$ where Algorithm 2 applies Logistic Trimming to them (only at the rounds when $f_j = 0$). With repeats taken off from this subsequence we see it has the same properties as that of $\{x_{s_i}, y_{s_i}\}$. Therefore, the proof is finished by an application of Theorem 3. □

Proof of Theorem 2. We first assume $C_t = \mathcal{I}$ for all t. As T grows, each item cluster eventually receives infinitely many samples of visting users' features and their responses, satisfying the requirements for Theorem 1. By Theorem 1, there is an event of probability $1 - M\delta_2$ where the information of relevant features w.r.t. each item is retrieved (by Logistic Trimming) at all rounds (this is not saying when Logistic Trimming is applied at $f_j \neq 0$ for any j, Theorem 1 still holds; nevertheless, at those strategic assignment rounds, the machine makes recommendations without using users' contextual information) after round $T_{\delta_2, d_{\mathcal{U}}, b}$ for some $T_{\delta_2, d_{\mathcal{U}}, b} > 0$. On this event, after round $T_{\delta_2, d_{\mathcal{U}}, b}$, HSB is essentially the reference algorithm. Given i_t and define $\hat{P}(\boldsymbol{x}_j, i_t) = \left\{\boldsymbol{x}^{\mathcal{U}} | \boldsymbol{x}^{\mathcal{U}}(\hat{\mathcal{RF}}_j) = i_t(\hat{\mathcal{RF}}_j), \boldsymbol{x}^{\mathcal{U}} \in \mathcal{U}\right\}$ similar to that in Algorithm 3 ($\hat{\mathcal{RF}}_j$ indicates the set of relevant user features w.r.t. items j). By an application of Lemma 6 in [Abbasi-Yadkori et al., 2011] we have that conditional on the event all relevant features are retrieved, with probability $1 - \delta^{'}$, given any $\boldsymbol{x}_j \in \mathcal{I}$ and for all $i_t \in \mathcal{U}$, $t > 0$,

$$\left| \frac{y_j^{'}(\hat{P}(\boldsymbol{x}_j, i_t))\mathbf{1}}{l(y_j(\hat{P}(\boldsymbol{x}_j, i_t)))} - f_j(i_t) \right| \leq \sqrt{\frac{(l(y_j(\hat{P}(\boldsymbol{x}_j, i_t))) + 1)}{l(y_j(\hat{P}(\boldsymbol{x}_j, i_t)))^2}\left(1 + 2\log\left(\frac{(1 + l(y_j(\hat{P}(\boldsymbol{x}_j, i_t))))^{1/2}}{\delta'}\right)\right)}. \tag{4}$$

Note that the assignments of item j to users $g_j(\boldsymbol{x}^{\mathcal{U}})$ are counted by and accumulated in $l(y_j(\hat{P}(\boldsymbol{x}_j, \boldsymbol{x}^{\mathcal{U}})))$. As argued in Theorem 7 in [Abbasi-Yadkori et al., 2011] using (4), the total number of assignments of item j to users $g_j(\boldsymbol{x}^{\mathcal{U}})$, $l(y_j(\hat{P}(\boldsymbol{x}_j, \boldsymbol{x}^{\mathcal{U}})))$, provided item j is not the best choice to this group, is bounded by

$$3 + \frac{16}{\Delta_m}\log\left(\frac{2}{\Delta_m \delta^{'}}\right)$$

with probability $1 - \delta^{'}$ (Δ_m can be sharpened if $C_t = \mathcal{I}$ for all t; see [Abbasi-Yadkori et al., 2011]).

We can then finish the proof by arguments similar to those in Lemma 6 and Theorem 7 (with a little bit more algebraic calculation: since $C_t \subset \mathcal{U}$ may not be the same through each round, the best item w.r.t. i_t can differ over time; Δ_M, Δ_m are needed for the argument) in the same paper. We sum over at most $2^{\#\cup_k J^k} M$ (the number of distinct user groups w.r.t item j is smaller than $2^{\#\cup_k J^k}$ for all j; $j \leq M$) regrets upper-bounds as well as the strategic assignments and let $\delta^{'} = \frac{\delta}{2^{\#\cup_k J^k} M}$ to finish the proof. □

Remark 1. *The first and third term in (3) can be improved. The $2^{K_{\mathcal{U}}}$ in the first term can be further sharpened by exploiting the assumption of uniform drawn i_t and the specification of relevant features in each item cluster as done by [Gentile et al., 2014]. On the other hand, the $T^{\frac{1}{b}}$ can be viewed as an increasing function of T, $h(T)$ with $h(x) = x^{\frac{1}{b}}$. We note that theoretically any $h(x)$ such that $\lim_{x\to\infty} h(x) = \infty$ suffices. Nevertheless, such improvements introduce unnecessary notation burden so we leave (3) as it is. For the reader who is interested in or needs to pursue a better bound in the first term, we refer to the discussions both in the article and the supplementary file of [Gentile et al., 2014].*

Lemma 1. *Let $\{z_i\}$ be a process satisfying Condition 3. Then there exists $c_1 > 0$ such that*

$$P\left(\lambda_{\min}\left(\frac{\sum_{t\leq s_n} z_t z_t^{'}}{s_n}\right) \geq c_1 \text{ eventually}\right) = 1. \tag{5}$$

Proof. Recall that

$$\eta_t = \frac{1}{t}\lambda_{\min}\left(\sum_{s\leq t} z_s z_s^{'} + \beta \boldsymbol{I}\right);$$

and $\{s_i\}$: For a process satisfying Condition 3, we define a corresponding random sequence $\{s_i\}$ by assigning the k-th (chronologically) element in $\cup_{j>0}\{i : T_{j-1} + d_{\mathcal{U}} \leq i \leq T_j\}$, where $T_0 \equiv 1 - d_{\mathcal{U}}$, to s_k.

By the definition of $\{z_i\}$, on $\{\eta_{s_n} < c_0\}$, $z_{s_n+1}, \ldots, z_{s_{n+1}}$ will be such that

$$E\left(\lambda_{\min}\left(\sum_{i=s_n+1}^{s_{n+1}} z_i z_i^{'}\right)\middle| z_j, j \leq s_n\right) = E\left(\lambda_{\min}\left(\sum_{i=1}^{d_{\mathcal{U}}} x_i x_i^{'}\right)\right) > 0,$$

where x_i's satisfy Condition 2. By this, $d_{\mathcal{U}} < \infty$, and the fact that if $\eta_{s_{n-1}} < c_0$, η_{s_n} is the minimum eigenvalue of the arithmetic mean of all the terms in the numerator of $\eta_{s_{n-1}}$ and $\sum_{i=s_n+1}^{s_{n+1}} z_i z_i^{'}$. By this observation and tedious calculation, there exist $\beta, c_0, k_0, \Delta > 0$ such that for all $n \geq k_0$,

$$E(\eta_{s_n} - \eta_{s_{n-1}} | \eta_{s_{n-1}} < c_0) \geq \frac{\Delta}{n}. \tag{6}$$

A proper choice of β and the boundedness of z_j imply that for all large n (say, $n > k_0$ for some $k_0 > 0$), if $\eta_{s_n} \geq c_0$, then $\eta_{s_{n+1}} \geq c_3$ for some $c_3 < c_0$; we pick an arbitrary c_2 such that $c_2 < c_3$. Moreover, we introduce a process $\{\eta^*_{i,t}\}_{i\geq 1, t\geq 1}$ whose existence is guaranteed by (6): $\eta^*_{i,t}$ such that on $\{\eta_{s_{i+t-1}} < c_0\}$,

$$\eta_{s_{i+t-1}} + \eta^*_{i,t} \leq \eta_{s_{i+t}} \text{ and } E\left(\eta^*_{i,t} | \eta_{s_{i+t-k}}, k \geq 1\right) = \frac{\Delta}{i+t};$$

or otherwise $\eta^*_{i,t}$ is independent of all other random variables and

$$E\left(\eta^*_{i,t}\right) = \frac{\Delta}{i+t}, \left\|\eta^*_{i,t}\right\| \leq \frac{C}{i+t}.$$

For any $C_{\bar{\Delta}} \geq k_0$ and a large β depending on $C_{\bar{\Delta}}, c_0, c_2$, we have, by the boundedness of z_i's and the definitions of $\eta^*_{i,t}$, c_3,

$$\begin{aligned}\sum_{i\geq C_{\bar{\Delta}}} \mathbf{1}_{\eta_{s_i} < c_2} - C_{\bar{\Delta}} &\leq \sum_{i\geq C_{\bar{\Delta}}} \sum_{j\geq 1} \mathbf{1}_{\eta_{s_{i-1}} \geq c_0, \eta_{s_i} < c_0} \mathbf{1}_{\eta_{s_{i+j}} < c_2} \\ &\leq \sum_{i\geq C_{\bar{\Delta}}} \sum_{j\geq 1} \mathbf{1}_{\eta_{s_{i-1}} \geq c_0, \eta_{s_i} < c_0} \mathbf{1}_{\eta_{s_i} + \sum_{t=1}^{j} \eta^*_{i,t} < c_2} \\ &\leq \sum_{i\geq C_{\bar{\Delta}}} \sum_{j\geq 1} \mathbf{1}_{c_3 + \sum_{t=1}^{j} \eta^*_{i,t} < c_2}, \text{ a.s.}\end{aligned} \tag{7}$$

$C_{\bar{\Delta}}$, whose value will be specified, is used to overcome initial exceptions when bounding the last term probabilistically. Hoeffding's inequality is employed to bound the sum of probability of events in (7); typical requirements for this inequality to work include conditional boundedness and conditional mean being equivalent to unconditional mean, properties that have been satisfied by $\eta^*_{i,t}$ since it is bounded by $\frac{C}{i+t}$ and $E\left(\eta^*_{i,t} | \eta^*_{i,s}, s < t\right) = \frac{\Delta}{i+t}$. By a version of Hoeffding's inequality, for all $i \geq 1$, $j \geq 1$,

$$\begin{aligned}&P\left(\left|c_3 + \sum_{t=1}^{j} \eta^*_{i,t} - \left(c_3 + E\left(\sum_{t=1}^{j} \eta^*_{i,t}\right)\right)\right| \geq \bar{\Delta}\left(\log(i+j) - \log(i+1)\right)\right) \\ &\leq \exp\left(\frac{-2\bar{\Delta}^2 \left(\log(i+j) - \log(i+1)\right)^2}{\sum_{t=1}^{j} (i+t)^{-2} C^2}\right).\end{aligned} \tag{8}$$

To bound the summation of (8) over i and j, we categorize it into three parts with custom argument for each one. Without loss of generality, $C = 1$. Given $C_{\bar{\Delta}}, \bar{\Delta}, d, \alpha, p_1$ such that $0 < d < 1$; $2\bar{\Delta}^2\alpha \geq 1$; $C_{\bar{\Delta}} \geq \max\{3, k_0, (\exp(\alpha) - 1)^{1/d} - 1, \exp(\frac{(1+d)\alpha}{d^2}) - 1\}$; and for $i \geq C_{\bar{\Delta}}$, $\sum_{k=i}^{p_1 i+1} k^{-1} \leq c_3 - c_2$. One more condition for $C_{\bar{\Delta}}$ is that given α, d satisfying previously stated conditions; for $i \geq C_{\bar{\Delta}}$, $j \geq (1+i)^{1+d}$, we have

$$\left(\log(i+j) - \log(i+1)\right)^2 \geq \alpha \log j. \tag{9}$$

The proof of (9) is omitted as it involves only simple but tedious calculation.

- $1 \leq j \leq p_1 i$, $C_{\bar{\Delta}} \leq i$: Ignored. As $\eta^*_{i,j}$'s are bounded by $\frac{\Delta}{i+j}$, a wise pick of p_1 such that

$\log p_1 \leq \frac{c_3 - c_2}{2}$ prevents the series $(c_3 + \sum_{t=1}^{j} \eta_{i,t}^*)$ from 'touching' the lower bound c_2. Note that given p_1, for $j \geq p_1 i$, $\log(i+j) - \log(i+1) \geq c_4$ for some $c_4 > 0$.

- $p_1 i \leq j \leq (1+i)^{1+d}$, $C_{\bar{\Delta}} \leq i$: By the definition of c_4,

$$\begin{aligned}\sum_{i \geq C_{\bar{\Delta}}} \sum_{(1+i)^{1+d} \geq j \geq p_1 i} & \exp\left(\frac{-2\bar{\Delta}^2 (\log(i+j) - \log(i+1))^2}{\sum_{t=1}^{j} (i+t)^{-2}}\right) \\ & \leq \sum_{i \geq C_{\bar{\Delta}}} (1+i)^{1+d} \exp(-(1+i)^{1-d} 2\bar{\Delta}^2 c_4^2) \\ & < \infty.\end{aligned}$$

- $(1+i)^{1+d} \leq j$, $C_{\bar{\Delta}} \leq i$: Using (9), the definition of α, and some basic algebraic calculation,

$$\begin{aligned}\sum_{i \geq C_{\bar{\Delta}}} \sum_{j \geq (1+i)^{1+d}} & \exp\left(\frac{-2\bar{\Delta}^2 (\log(i+j) - \log(i+1))^2}{\sum_{t=1}^{j} (i+t)^{-2}}\right) \\ & \leq \sum_{i \geq C_{\bar{\Delta}}} \sum_{j \geq (1+i)^{1+d}} \exp\left(\frac{-2\bar{\Delta}^2 \alpha \log j}{\sum_{t=1}^{j} (i+t)^{-2}}\right) \\ & \leq \sum_{i \geq C_{\bar{\Delta}}} \sum_{j \geq 2} j^{-i} < \infty.\end{aligned}$$

Note that for a proper choice of $\bar{\Delta}$, we have, on the complement event of that in (8), $\sum_{t=1}^{j} \eta_{i,t}^* \geq E\left(\sum_{t=1}^{j} \eta_{i,t}^*\right) - \bar{\Delta}(\log(i+j) - \log(i+1)) > 0$. Combining this, a proper choice of c_4, we have bounded the summation of (8) over all i, j. By the boundedness of the summation of (8) (over i, j), (7), and the Borel-Cantelli Lemma, we have

$$\lim_{m \to \infty} P(\cap_{n \geq m} \{\eta_{s_n} \geq c_2\}) = 1. \tag{10}$$

By (10), the convexity of $\lambda_{\min}$, and $\beta < \infty$, there exists $c_1 > 0$ such that (5). □

Notation

For Lemmas 1 to 3 and the proof of Theorem 3, we define the following notation. Let $\sigma(x) = \exp(x)(1 + \exp(x))^{-1}, x \in \mathbb{R}^1$; β^* stands for the true parameters; $\delta(n) = s_0 n^{-1/2} (\log n)^{1/2+\varepsilon}$ for some $s_0 > 0$ and arbitrary $\varepsilon > 0$; let $R_n = O\left(\frac{(\log n)^{1+2\varepsilon}}{n}\right)$. $\hat{\beta}_{J,n} = \arg\max_{\beta \in \Theta; \beta_i = 0, i \in J^c} l_n(\beta)$, where $J \subset \{1, \ldots, p\}, J^c = \{1, \ldots, p\} \backslash J$. Define the ball $B_y(x) = \{z : \|z - x\| \leq y\}$. For c_i's, we reuse these notations.

The following lemmas concern processes satisfying (11). We state and prove the lemmas in context without special process indexes for generality; replacing $\sum_n$ with $\sum_{s_n}$ makes the lemma applicable (the dimension $p = d_{\mathcal{U}} < \infty$) to processes satisfying Condition 3 with special round indexes, $\{x_{s_i}\}$.

Lemma 2. *Let the process $\{x_i\}$ satisfy the condition for $n \geq n_0$,*

$$\lambda_{\min}\left(\frac{\sum_{t=1}^{n} x_t x_t^{'}}{n}\right) > c_1, \tag{11}$$

for some $c_1, n_0 > 0$. Define J to be any subset of $\subset J_p^* = \{i | \beta_i^* \neq 0, i = 1, \ldots, p\}$ and $\mathcal{B}_n = \left\{\left\|\hat{\beta}_{J,n} - \beta^*\right\| \leq \delta(n)\right\}$. Then

$$\lim_{m \to \infty} P(\cap_{n \geq m} \mathcal{B}_n) = 1, \tag{12}$$

for a proper choice of s_0 in $\delta(n)$.

Proof. Define $\Theta_n = \Theta \cap B_{\delta(n)}^c(\beta)$; given any $v_i \in \mathbb{R}^p, i = 1, \ldots, p$ with $v_i \perp v_j$ if $i \neq j$; $\|v_i\| = 1$. Moreover,

$$\begin{aligned}E_{1,n} &= \left\{\lambda_{\min}\left(\frac{\sum_{t=1}^{n} x_t x_t^{'}}{n}\right) > c_1\right\}, \\ E_{2,n} &= \left\{\inf_{\beta \in \Theta_n} \max_{i=1}^{p} \frac{1}{n} \left|\sum_{t=1}^{n} \left(\sigma(x_t^{'}\beta^*) - \sigma(x_t^{'}\beta)\right) x_t^{'} v_i\right| \right. \\ & \qquad \left. \geq c_2 n^{-1/2} (\log n)^{1/2+\varepsilon}\right\}, \\ E_{3,n} &= \left\{\max_{i=1}^{p} \frac{1}{n} \left|\sum_{t=1}^{n} \left(y_t - \sigma(x_t^{'}\beta^*)\right) x_t^{'} v_i\right| < c_3 n^{-1/2} (\log n)^{1/2+\varepsilon}\right\}.\end{aligned}$$

By assumption,

$$\sum_n P(E_{1,n}^c) < \infty. \tag{13}$$

By the definition of y_t's; the boundedness of x_i's and Θ; $p < \infty$; an arbitrarily small $c_3 > 0$; and a version of Hoeffding's inequality, we have

$$\sum_n P(E_{3,n}^c | \mathcal{F}_x) < \infty, \tag{14}$$

where $\mathcal{F}_x$ stands for the sigma field generated by the process $\{x_i\}$.

For any $v_i \in \mathbb{R}^p, i = 1, \ldots, p$ with $v_i \perp v_j$ if $i \neq j$; $\|v_i\| = 1$; and $z \in \mathbb{R}^p$; we have $\max_{i=1}^{p} \left\|z^{'} v_i\right\| \geq \frac{\|z\|}{p}$. Therefore there exists $c_4 > 0$ (depending on c_1) such that on $E_{1,n}$,

$$\inf_{\beta \in \Theta_n} \max_{i=1}^{p} \frac{1}{n} \left|(\beta - \beta^*)' \sum_{t=1}^{n} x_t x_t^{'} v_i\right| \geq c_4 n^{-1/2} (\log n)^{1/2+\varepsilon}. \tag{15}$$

By (15); the boundedness of x_i and Θ;

$$\sigma(x_t^{'}\beta^*) - \sigma(x_t^{'}\beta) = \frac{\sigma(x_t^{'}\beta_c)}{1 + \sigma(x_t^{'}\beta_c)} x_t^{'}(\beta - \beta^*)$$

for some $\beta_c \in B_{\|\beta^* - \beta\|}(\beta^*)$ by Taylor's expansion; and a proper choice of c_2, s_0 (start by fixing a $s_0 > 0$;

then pick a small enough c_2. In a later usage we need $c_2 > c_3$; it is possible since $c_3 > 0$ is arbitrarily small in (14)), and for all large n,

$$E_{1,n} \subset E_{2,n}. \tag{16}$$

On $E_{2,n} \cap E_{3,n}$,

$$\begin{aligned}
&\inf_{\beta\in\Theta_n} \max_{i=1}^{p} \frac{1}{n}\left|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta\right)\right)x_t^{'}v_i\right| \\
&\geq \inf_{\beta\in\Theta_n} \max_{i=1}^{p} \frac{1}{n}\left|\sum_{t=1}^{n}\left(\sigma\left(x_t^{'}\beta^*\right) - \sigma\left(x_t^{'}\beta\right)\right)x_t^{'}v_i\right| \\
&\qquad - \max_{i=1}^{p} \frac{1}{n}\left|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta^*\right)\right)x_t^{'}v_i\right| \\
&\geq (c_2 - c_3)n^{-1/2}(\log n)^{1/2+\varepsilon} > 0.
\end{aligned} \tag{17}$$

Note that $|\nabla l_n(\beta) v_i| = n^{-1}\left|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta\right)\right)x_t^{'}v_i\right|$; if it is not zero then it is not the solution of the optimization. Hence, by (13), (14), (16), (17), and the Borel-Cantelli Lemma we have (12). □

Lemma 3. *Let $\{x_i\}$ be the process satisfying (11), $J_p^* \subset J$, and $E_n = \left\{\left\|\nabla l_n(\hat{\beta}_{J,n})\right\|^2 \leq c_0 n^{-1}(\log n)^{1+2\varepsilon}\right\}$. We have*

$$\sum_n P\left(E_n^c\right) < \infty \tag{18}$$

for some proper choice of c_0.

Proof. Set s_0 in $\delta(n)$ to that in Lemma 2. By Taylor's formula and the boundedness of x_t's and Θ, there exists $c_5 > 0$ such that for all t,

$$\sup_{\beta\in B_{\delta(n)}(\beta^*)}\left|\sigma\left(x_t^{'}\beta^*\right) - \sigma\left(x_t^{'}\beta\right)\right| \leq c_5\delta(n). \tag{19}$$

Let $E_{4,n} = \left\{\sup_{v:\|v\|=1} n^{-1}\left|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta^*\right)\right)x_t^{'}v_i\right| \leq c_7\delta(n)\right\}$. Some algebraic manipulation, (14), a proper choice of c_7, and

$$\sum_n P(E_{4,n}^c) < \infty. \tag{20}$$

On $E_{4,n} \cap \left\{\left\|\hat{\beta}_{J,n} - \beta^*\right\| \leq \delta(n)\right\}$, by (19) and the boundedness of x_t, there exists some $c_6 > 0$ such that

$$\begin{aligned}
\left\|\nabla l_n(\hat{\beta}_{J,n})\right\|^2 &= \frac{1}{n^2}\left\|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\hat{\beta}_{J,n}\right)\right)x_t\right\|^2 \\
&\leq \frac{1}{n^2}\left\|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta^*\right)\right)x_t\right\|^2 \\
&\quad + \frac{1}{n^2}\sup_{\beta\in B_{\delta(n)}(\beta^*)} 2\left\|\sum_{t=1}^{n}\left(y_t - \sigma\left(x_t^{'}\beta^*\right)\right)x_t\right\| \\
&\qquad \times \left\|\sum_{t=1}^{n}\left(\sigma\left(x_t^{'}\beta^*\right) - \sigma\left(x_t^{'}\beta\right)\right)x_t\right\| \\
&\quad + \frac{1}{n^2}\sup_{\beta\in B_{\delta(n)}(\beta^*)}\left\|\sum_{t=1}^{n}\left(\sigma\left(x_t^{'}\beta^*\right) - \sigma\left(x_t^{'}\beta\right)\right)x_t\right\|^2 \\
&\leq c_6\frac{(\log n)^{1+2\varepsilon}}{n}.
\end{aligned} \tag{21}$$

By Lemma 2, (20), and (21), we have finished the proof. □

Proof of Theorem 3. Notation is simplified; see the notation notice of Lemma 2, 3. Since $p < \infty$, essentially we need (22) and (23). For $p \geq k$ such that $\beta_k^* \neq 0$,

$$\lim_{m\to\infty} P\left(\cap_{n\geq m}\left\{\mathrm{IC}_n(J_p - \{k\}) > \mathrm{IC}_n(J_p)\right\}\right) = 1; \tag{22}$$

and for $p \geq k$ such that $\beta_k^* = 0$,

$$\lim_{m\to\infty} P\left(\cap_{n\geq m}\left\{\mathrm{IC}_n(J_p - \{k\}) \leq \mathrm{IC}_n(J_p)\right\}\right) = 1. \tag{23}$$

Denote $J_p - \{k\}$ by $J_{p,k}$. For (22), we note that $\mathrm{IC}_n(J_p - \{k\}) > \mathrm{IC}_n(J_p)$; Taylor's formula implies for some $\beta_c \in B_{\left\|\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right\|}(\hat{\beta}_{J_p,n})$,

$$\begin{aligned}
R_n &< l_n(\hat{\beta}_{J_{p,k},n}) - l_n(\hat{\beta}_{J_p,n}) \\
&= -\nabla l_n(\hat{\beta}_{J_p,n})\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right) \\
&\quad - \frac{1}{2}\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right)^{'}\nabla^2 l_n(\beta_c)\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right).
\end{aligned} \tag{24}$$

The second term on the RHS of (24) can be bounded from below. By the boundedness of x_t and Θ, there exists some $C > 0$ such that for all large n, on $\mathcal{B}_n \cap E_{1,n}$,

$$\begin{aligned}
&-\frac{1}{2}\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right)^{'}\nabla^2 l_n(\beta_c)\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right) \\
&\quad \geq \frac{1}{2}\inf_{\beta\in\Theta}\lambda_{\min}\left(-\nabla^2 l_n(\beta)\right)\left\|\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right\|^2 \\
&\quad \geq |\beta_k|C.
\end{aligned} \tag{25}$$

Since $\nabla l_n(\hat{\beta}_{J_p,n})$ is a vector of zeros, (25) says

$$(24) \text{ is true on } \mathcal{B}_n \cap E_{1,n}, \text{ for all large } n. \tag{26}$$

To get (22), we use (26) and Lemma 1, 2.

(23) is tighter than the other; $\mathrm{IC}_n(J_p - \{k\}) \leq \mathrm{IC}_n(J_p)$ implies

$$R_n \geq l_n(\hat{\beta}_{J_{p,k},n}) - l_n(\hat{\beta}_{J_p,n}). \tag{27}$$

We need another Taylor's expansion:

$$\begin{aligned} \nabla l_n(\hat{\beta}_{J_{p,k},n})^{'} &= \nabla l_n(\hat{\beta}_{J_p,n})^{'} + \nabla^2 l_n(\beta_c)\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right) \\ &= \nabla^2 l_n(\beta_c)\left(\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\right), \end{aligned} \tag{28}$$

for some $\beta_c \in B_{\|\hat{\beta}_{J_p,n} - \hat{\beta}_{J_{p,k},n}\|}(\hat{\beta}_{J_p,n})$. By convexity of l_n and (28), we have that there exists some $C > 0$ such that for all large n, on $E_{1,n} \cap E_n$,

$$\begin{aligned} l_n(\hat{\beta}_{J_{p,k},n}) - l_n(\hat{\beta}_{J_p,n}) &\leq \nabla l_n(\hat{\beta}_{J_{p,k},n})\left(\hat{\beta}_{J_{p,k},n} - \hat{\beta}_{J_p,n}\right) \\ &\leq \left[\inf_{\beta \in \Theta} \lambda_{\min}\left(-\nabla^2 l_n(\beta)\right)\right]^{-1} \left\|\nabla l_n(\hat{\beta}_{J_{p,k},n})\right\|^2 \\ &\leq C\frac{(\log n)^{1+2\varepsilon}}{n}. \end{aligned} \tag{29}$$

By this, we have

$$\begin{aligned} &(27) \text{ is true on } E_{1,n} \cap E_n, \text{ for all large } n \\ &\text{and a proper choice of } R_n = O\left(\frac{(\log n)^{1+2\varepsilon}}{n}\right). \end{aligned} \tag{30}$$

To get (23), we use (30) and Lemma 1, 3, and the Borel-Cantelli Lemma (note that $J_p^* \subset J_{p,k}$ in this case). □