User Response Based Recommendations: A Local Angle Approach

Veeraruna Kavitha, Devanand R, Parmod Kumar, and Eitan Altman

***Abstract*—In this work, we propose novel recommendation policies, which are based on user-generated responses. These can be applied in variety of recommendation systems (RS), in particular are very useful in the context of anonymous users. These are users without any history and those who are not comfortable in leaving behind any information. Our objective is to satisfy the users requirement at the earliest, using the responses given by the later to the previous recommendations of the same session. The earlier the user is satisfied, the more is the profit to the content provider (CP). The proposed approach is different from the traditional recommendation schemes, as the recommendations are neither based on the history of the usage of the items nor on the history of the user(s) that used the item previously.**

We first derived optimal policies in the continuous Euclidean space and translated the same to the space of discrete items. In the optimal policy, the recommendations at the same time step are at 180 degrees from each other, while are at 90 degrees with respect to the ones at the previous time step. In the context of discrete items, notion of distance based on similarity measure is well known, however notion of angle is not known. We proposed the notion of local angle in the space of discrete items and then translated the optimal policies. Using simulation results, we demonstrated that the notion of local angle improved the average hitting time performance. The proposed recommendation policies can be used in discrete as well as continuous space based applications. For example, these policies can be used by video item providers, online shopping portals etc. They can also be applied in context of service and rescue robot navigation.

I. Introduction

Recommendation systems (RS) became an active research area after the research on collaborative filtering in mid-1990s, [5]-[8]. Over the past decade recommendation engines have become quietly ubiquitous, and used for various purposes, e.g. e-commerce, social and professional networks, such as YouTube, Pandora, Amazon etc., [6]. Even after plenty of work done by the industry and academia to develop new methods for recommender systems in the last decade, this research area still has lots of scope to explore practical applications ([9]).

In this work, we propose novel recommendation policies, which are based on user-generated responses. The objective is to satisfy the users requirement at earliest using the responses of the same user to the previous recommendations of the same session. The proposed recommendation policies are well suited to both, discrete as well as continuous space based applications, e.g. video item providers, online shopping portals, service and rescue robot navigation (here the radar signals take the role of user responses), etc.

Our approach can also help cold start problems ([4], [3], [2]), in particular a new user problem. Most of these problems utilize the history of items ([4]). In [3] authors consider demographic approach to classify new users. In [1], [2] authors consider prompting the new users for the purpose of learning them and for subsequent improved person based recommendations. They discuss the set of items which are optimal in certain information theoretic approach to learn the user. This is the nearest work related to our approach. However our work differs majorly in the following sense. We learn the user, with respect to the item of the current interest, and along the way maximize the hitting probability in the current session itself.

One can explore the proposed approach even in the existence of history and we would consider this in future. A hybrid scheme may perform better. The users interest can vary drastically from one session to another. For example, a user might visit an online portal to buy a book at one time, while the same user might be interested in mobile phones the next time. Users demanding similar items can have drastically different preferences. But the interest of the user during the same session would be consistent and the partial information about the same is available via the user responses. This method utilizes this information to construct optimal policies.

II. System Dynamics

We consider a large (finite) database $\mathcal{S}$, each item of which is specified by F number of defining features. For example, for a music video, the singer, composer, instruments etc. are some of the describing features. The similarities/dis-similarities between the items are obtained by comparing these features. A new user without any history is interested in one of the items referred by v_{ref}, which is unknown to the Content Provider (CP). We assume that the v_{ref} is equally likely to be any one of items available with CP. The user specifies its interest by initial search query and the system generates a series of recommendations based on this search query. User is satisfied if any one of the suggested recommendations is close to v_{ref}, i.e., with at least $\overline{F}$ number of matching features and then the CP derives benefit. The user starts a session with CP using an item X_0 (obtained after a search query), only if it has at least more than $\underline{F}$ similar features with $0 \leq \underline{F} \leq \overline{F}$. The case wherein user always starts the session is obtained when $\underline{F} = 0$. The CP displays a list of M recommendations (call $\mathbf{a}_1 = (a_{1,1} \cdots, a_{1,M})$, a vector of length M) while displaying the item X_0. The user, if not satisfied with X_0, explores one of the suggestions (call it X_1) from among $\mathbf{a}_1$. CP while displaying item X_1 also suggests new recommendation list $\mathbf{a}_2$. The user, if not satisfied with X_1, chooses one among $\mathbf{a}_2$ and this continues. That is, the user navigates through the recommendations of the CP and this the user does for maximum T number of steps. At every step, the user chooses one of the items recommended by the CP with maximum number of features in common with v_{ref}. That is,

$$X_k = \arg\max_m d(v_{ref}, a_{k,m}), \quad (1)$$

where $d(v_{ref}, a_{k,m})$ is the similarity measure based distance between items v_{ref} and $a_{k,m}$, proportional to the number of matching features between them.

In case the item recommended via initial search/subsequent recommendations is interesting to the user (matching features more than $\overline{F}$), the CP benefits and we assume the benefit is proportional to the search time. The faster the user is satisfied, the more is the benefit to the CP. For example in case of video content, the user can spend the remaining time (leftover time till T) watching the video and CP's profit is proportional to the watched time. Alternatively, if the user does not come across a satisfactory item in the T steps, then the user quits the session without any benefit/profit to the CP. Our aim is to maximize the profit of the CP or equivalently minimize the time taken to suggest a satisfactory item.

III. Modelling

We denote the space of all possible items by $\mathcal{S}$ and v_{ref} is one among $\mathcal{S}$. Our analysis is also applicable to the case when $\mathcal{S}$ is an Euclidean subset. We also consider the spaces with L^1 metric (where the distance between two points/items is $d(v_1, v_2) := \sum_i |v_{1_i}, v_{2_i}|$). We could handle this generalization without extra effort and this is useful in various other contexts. More importantly, we derive recommendation policies for discrete content space by translating appropriate optimal policies of a continuous space.

The user's interest v_{ref} is uniformly distributed among the items of $\mathcal{S}$. That is the probability that v_{ref} lies in a subset $\Lambda \subset \mathcal{S}$ is given by:

$$P(v_{ref} \in \Lambda) = \frac{\mu(\Lambda)}{\mu(\mathcal{S})},$$

where μ is the Lebesgue measure (length, area respectively in one and two dimensional spaces) in the continuous case and is the cardinality (the number of elements in a set) when $\mathcal{S}$ is discrete. The subsets Λ are Borel sets in continuous case while are any subset of $\mathcal{S}$ (which is finite) in the discrete case. We only consider finite sets when $\mathcal{S}$ is discrete.

We use similarity measure to represent the approximate notion of distance in discrete content space. We consider binary features, i.e., each feature can be either 1 or 0. Most of the results can be extended to other cases. The distance between two items $v_1 = (v_{1_i})_{i \le F}$ and $v_2 = (v_{2_i})_{i \le F}$ is defined as

$$d(v_1, v_2) = \frac{\text{\# of mismatching features}}{F} = \frac{\sum_{i=1}^{F} |v_{1_i} - v_{2_i}|}{F}.$$

In continuous spaces, d is either Euclidean or L^1 distance. One can now define the balls around an item/point v by:

$$\mathcal{B}(v, \underline{r}) := \{x \in \mathcal{S} : d(x, v) < \underline{r}\}.$$

Thus the user's interest lies in any item inside the ball $\mathcal{B}(v_{ref}, \underline{r})$ where $\underline{r}$:= 1 - $(\overline{F}/F)$ while the initial search X_0 lies inside $\mathcal{B}(v_{ref}, \overline{R})$ where $\overline{R}$ = 1 - $(\underline{F}/F)$. Recall that v_{ref} is unknown to CP, however, it knows that $v_{ref} \in \mathcal{B}(X_0, \overline{R})$. Let τ represent the first time a recommendation hits the inner circle $\underline{r}$ circle $\mathcal{B}(v_{ref}, \underline{r})$, i.e. (using (1)),

$$\begin{aligned} \tau &:= \inf_{k \ge 1} \{a_{k,i} \in \mathcal{B}(v_{ref}, \underline{r}), \text{ for some } i \le M\} \\ &= \inf_{k \ge 1} \{X_k \in \mathcal{B}(v_{ref}, \underline{r})\}. \end{aligned}$$

Our aim is to minimize the expected value of the hitting time, $E[\tau]$, or maximizing the time spent in the system, i.e. $E[T-\tau]$,

$$\min_\pi E[\tau] \equiv \max_\pi E[T - \tau] \quad (2)$$

We are presenting our results with $M = 2$. Some of the results can easily be extended for $M > 2$ and we briefly discuss the same at later stages.

IV. Analysis and Optimal Policies

This problem is a fixed horizon problem, and the time step is represented by the time epoch at which the user has chosen a new item to explore. The recommendations depend upon the user response to the previously recommended items. We describe the precise problem formulation below. The policy would be a pair of recommendations (actions), one for each time step k as below.

$$\pi = \{(a_{1,1}, a_{1,2}), (a_{2,1}, a_{2,2}), \cdots, (a_{T,1}, a_{T,2})\} \quad (3)$$

These policies depend upon appropriate state of the system. The state of the system at time k is given by (X_k, v_{ref}), where X_k defined in equation (1) is the choice of user at time k and v_{ref} is the interest of the user. Since, v_{ref} is not directly available to the CP, we consider the alternate state $Z_k = (X_k, B_k)$, inspired by Partially Observable MDP (POMDP) framework. Here belief random variable B_k is the conditional distribution of v_{ref} given history H_k, defined by:

$$H_k = \{(X_0, B_0), (X_1, B_1), ...(X_{k-1}, B_{k-1}), X_k, \mathbf{a_1}, ...\mathbf{a_k}\}.$$

The belief only improves with time, in the sense for all k, B_k is concentrated on certain sets $\mathcal{A}_k \subseteq \mathcal{A}_0$, in the following manner. The proof is provided in Appendix.

***Lemma** 1:* The belief random variable B_k at any time step k depends only upon the previous belief B_{k-1}, current recommendation $\mathbf{a_k}$ and current user choice X_k. For each k, $B_k \sim \mathcal{U}(\mathcal{A}_k)$, which implies that B_k is uniformly distributed over a subset $\mathcal{A}_k \subset \mathcal{S}$. Also the subsets $\{\mathcal{A}_k\}_{k \le T}$ are nested:

$$\mathcal{S} = \mathcal{A}_0 \supset \mathcal{A}_1 \cdots \supset \mathcal{A}_{T-1} \supset \mathcal{A}_T. \qquad \square$$

Thus the sequence of states (X_k, B_k) is a controlled Markov Chain, controlled by sequence of state dependent actions/policy, π. The required cost $E[T - \tau]$ with any policy

π and initial state $X_0 = x$ equals (details in Appendix)[1]

$$J(x,\pi) = \sum_{k=0}^{T} P(T-\tau > k) = \sum_{k=0}^{T-1} P(v_{ref} \in \cup_{l=0}^{k} \mathcal{B}(X_l, \underline{r})), \quad (4)$$

and we are interested in maximizing the above cost.

Nested notations: Before we proceed further we digress little to introduce nested notations which are important to state the results. Let $\mathcal{A}_1^0 := \mathcal{A}_0 = \mathcal{S}$, be the area of the first belief. Given a policy π, the recommendations of step $k = 1$ are given by $a_{1,i}$, $i = 1, 2$. We denote them by a_i^1 ($i = 1, 2$) and the corresponding (split) belief areas by $\mathcal{A}_i^1$, where

$$\mathcal{A}_1^1 = \mathcal{A}_{1,1}(\mathcal{A}_0) = \{v_{ref} \in \mathcal{A}_0 : d(v_{ref}, a_{1,1}) \leq d(v_{ref}, a_{1,2})\},$$
$$\mathcal{A}_2^1 = \mathcal{A}_{1,2}(\mathcal{A}_0) = \{v_{ref} \in \mathcal{A}_0 : d(v_{ref}, a_{1,2}) \leq d(v_{ref}, a_{1,1})\}.$$

Note here that $\mathcal{A}_1 = \sum_i \mathcal{A}_i^1 1_{\{X_1 = a_i^1\}}$. At time step $k = 2$, the two recommendations depend upon the choice X_1 of the user and there are in total 4 (2 pairs) recommendations available (see Fig. 1). Let a_i^2 with $i = 1, 2, 3, 4$ represent these recommendations and let $\mathcal{A}_i^2$ represent the corresponding split belief areas in the following order:

$$a_1^2 = a_{2,1}(\mathcal{A}_1^1),\ a_2^2 = a_{2,2}(\mathcal{A}_1^1),$$
$$a_3^2 = a_{2,1}(\mathcal{A}_2^1), \text{ and } a_4^2 = a_{2,2}(\mathcal{A}_2^1).$$

In the above $a_{2,1}(\mathcal{A}_1^1)$ implies the first recommendation at step $k = 2$, when $X_1 = a_1^1$. Continuing this way we define 2^k recommendations $\{a_j^k\}_{1 \leq j \leq 2^k}$ and the corresponding belief areas $\{\mathcal{A}_j^k\}_{1 \leq j \leq 2^k}$ for time step k. Define $\mathcal{B}_i^k := \mathcal{B}(a_i^k, \underline{r})$ to represent the desired ball around the recommendation a_i^k.

The areas and the recommendations so defined depend upon the policy π, but given a policy π one can determine all of them. We have a set of points $\mathcal{Q}_\pi = \{a_i^k\}_{k \leq T, i \leq 2^k}$, for any π, such that recommendation at any time k and for any state (X_k, B_k) belongs to $\mathcal{Q}_\pi$.

Using the above notations we state the first result, Theorem 1. The proof is provided in Appendix.

***Theorem** 1:* For any policy π, and for any k, the sets $\{\mathcal{A}_j^k\}_{j \leq 2^k}$ form a 2^k- partition of $\mathcal{S}$. Further 2^k partition is a finer division of 2^{k-1} partition, i.e., $\mathcal{A}_i^{k-1} = \mathcal{A}_{2i-1}^k \cup \mathcal{A}_{2i}^k$ for each i, k. We then have:

$$E[T-\tau](\pi) \overset{b}{\leq} \sum_{k=0}^{T-1} (T-k) \sum_{j=1}^{2^k} \frac{\mu(\mathcal{B}_j^k \cap \mathcal{A}_j^k)}{\mu(\mathcal{A}_0)}. \quad \square (5)$$

It is easy to see that each and every term of the summation representing the upper bound of equation (5) can be further upper bounded by $\min\{1, |\mathcal{B}|\}$ where $|\mathcal{B}| := \mu(\mathcal{B}(0, \underline{r}))$. Further the upper bound b can be satisfied with equality if all the balls $\{\mathcal{B}_j^k\}_{k,j \leq 2^k}$ are disjoint (see Appendix). Using this, one can obtain optimal policies, if we show that a particular policy simultaneously achieves all the upper bounds of equation (5).

In case the initial area $\mathcal{S}$ is large enough to contain at least $(2^T - 2)$ disjoint $\underline{r}$ radius balls, there is a possibility that one can obtain the optimal policy. This is exactly achieved in the corollary given below (which is easy to verify):

Corrollary 1: For any policy π,

$$E[T-\tau](\pi) \leq \frac{\sum_{k=0}^{T-1} (T-k) \sum_{j=1}^{2^k} \mu(\mathcal{B}_j^k)}{\mu(\mathcal{A}_0)} = \frac{(2^{T+1} - T - 2)|\mathcal{B}|}{\mu(\mathcal{A}_0)}.$$

If $\mathcal{A}_0$ contains at least $(2^T - 2)$ disjoint[2] $\underline{r}$ radius balls and if there exists a π^* which ensures $\mathcal{B}_i^k \subset \mathcal{A}_i^k$ for each and every (k, i), and if further the inequality (b) in equation (5) is satisfied with equality, then π^* is an optimal policy:

$$E[T-\tau](\pi^*) \geq E[T-\tau](\pi) \text{ for any } \pi. \qquad \square$$

For example, consider a one dimensional finite sorted set with cardinality based distance and we are interested in one particular item ($\underline{r}$ is small so that, $\mathcal{B}(v_{ref}, \underline{r}) = \{v_{ref}\}$). One can show that the well known binary search method achieves the upper bound of Corollary 1, and hence is an optimal policy. On the other hand if $\mathcal{A}_0$ is small and is itself contained in one ball $\mathcal{B}(\mathbf{a}, \underline{r})$ for some $\mathbf{a}$, then $E[T-\tau]$ can be upper bounded by T:

Corrollary 2: For any policy π, equation (4) can directly be upper bounded as $P(A) \leq 1$ for any A and hence we get,

$$E[T-\tau](\pi) \leq T.$$

If $\mathcal{A}_0 \subset \mathcal{B}(\mathbf{a}, \underline{r})$ for some $\mathbf{a}$ then π^* with $\pi_{k,i}^* = \mathbf{a}$ for all k, i becomes optimal policy achieving the upper bound T. $\square$

The theorem, corollaries are applicable even if we consider the cost from $k > 1$ and replace $\mathcal{A}_0$ with $\mathcal{A}_{k-1}$. The upper bound of Corollary 1 would be $\frac{(2^{k+1}-k-2)|\mathcal{B}|}{\mu(\mathcal{A}_{k-1})}$. While that in Corollary 2 would be $(T-k)$.

A. Optimal policies in Continuous space

L^1 **metric :** We first consider the case with L^1 metric. In this case any ball $\mathcal{B}(\mathbf{a}, r)$ is a rhombus, by rotating becomes a square. So, user interest v_{ref} is uniformly distributed in a square of dimension, $\sqrt{2}\,\overline{R}$, with centre as the initial recommendation X_0. The user is satisfied with any item which lies inside a square of dimension $\sqrt{2}\underline{r}$ with v_{ref} as the centre. If $(2^T - 2) \leq \frac{\overline{R}^2}{\underline{r}^2}$, we have sufficient disjoint balls and Corollary 1 is applicable. With $R := \sqrt{2}\,\overline{R}/2 = \overline{R}/\sqrt{2}$, one can easily verify that an optimal policy π^* (which provides the required disjoint balls) is given by:

[1] Some of these details (including the statements of Theorem 1 and the Corollaries) in the first version submitted to GI 2015 have errors and we corrected the same in this technical report.

[2] We need 2 balls at step 1, 4 at step 2 and 2^k at step k and so on up to time $k = T-1$ (Figure 1) and thus we need a total of $\sum_{k=1}^{T-1} 2^k = 2^T - 2$ disjoint balls.

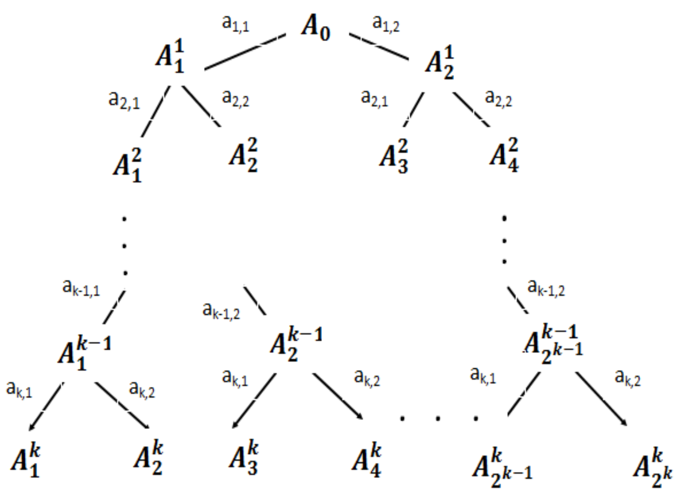
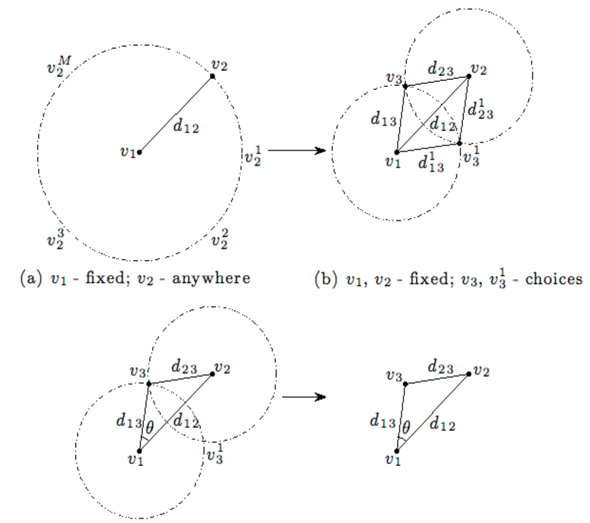


Fig. 1. Nested Notations



(a) v_1 - fixed; v_2 - anywhere (b) v_1, v_2 - fixed; v_3, v_3^1 - choices

(c) v_1, v_2 - fixed; v_3 - chosen (d) Local angle θ for v_1, v_2 and v_3

Fig. 2. Notion of 'Local' Angle.

$$
\begin{aligned}
&a^*_{1,1} = (\tfrac{R}{2}, 0), && a^*_{1,2} = (-\tfrac{R}{2}, 0),\\
&a^*_{2,1} = X_1 + (0, \tfrac{R}{2}), && a^*_{2,2} = X_1 + (0, -\tfrac{R}{2})\\
&a^*_{3,1} = X_2 + (\tfrac{R}{4}, 0), && a^*_{3,2} = X_2 + (-\tfrac{R}{4}, 0),\\
&\vdots\\
&a^*_{2k-1,1} = X_{2k-1} + (0, \tfrac{R}{2^{2k-2}}),\\
&\qquad\qquad a^*_{2k-1,2} = X_{2k-1} + (0, -\tfrac{R}{2^{2k-2}}).\\
&a^*_{2k,1} = X_{2k} + (\tfrac{R}{2^{2k-2}}, 0),\\
&\qquad\qquad a^*_{2k,2} = X_{2k} + (-\tfrac{R}{2^{2k-2}}, 0) \text{ for all } k,
\end{aligned} \tag{6}
$$

where user choice at any time step $k \geq 1$ is given by:

$$
\begin{aligned}
X_k \;=\; & a^*_{k,1} 1_{\{d(a^*_{k,1}, v_{ref}) < d(a^*_{k,2}, v_{ref})\}}\\
& + a^*_{k,2} 1_{\{d(a^*_{k,1}, v_{ref}) > d(a^*_{k,2}, v_{ref})\}}.
\end{aligned} \tag{7}
$$

L^1 **metric with** $(2^T - 2) > \overline{R}^2/\underline{r}^2$**:** Basic idea is to obtain the optimal policy using Corollary 1 till k^* where

$$k^* = \arg\max_k \left\{ (2^k - 2) \leq \frac{\overline{R}^2}{\underline{r}^2} \right\}$$

and then using Corollary 2 for the time steps from k^*+1 till T. The exact details are as below for the case when $\overline{R}/\underline{r}$ is an appropriate power of 2 such that $2^{k^*} - 2 = \overline{R}^2/\underline{r}^2$. One can give similar construction even otherwise. But some minor details need to be considered.

Note that we exactly have $(2^{k^*} - 2)$ disjoint balls and hence one can upper bound all the terms till k^* by $|\mathcal{B}|$ as in Corollary 1. Let $a^*_{k,i}$ be as defined in equation (6) for all i and for all $k < k^*$. At k^* all the remaining areas $\left\{\mathcal{A}_i^{k^*}\right\}_{i \leq 2^{k^*}}$ are already of size $\underline{r}$. As in Corollary 2, define for any $k > k^*$ and i

$$a^*_{k,i} = X_{k-1} = X_{k^*}.$$

One can easily verify that for this policy,

$$E[T - \tau] = \frac{(2^{k^*+1} - k^* - 2)|\mathcal{B}|}{\mu(\mathcal{A}_0)} + (T - k^*).$$

This $E[T-\tau]$ is obtained by summing upper bounds as in the two corollaries and hence the policy defined is optimal. From the above,

$$E[\tau] = k^* - \frac{(2^{k^*+1} - k^* - 2)|\mathcal{B}|}{\mu(\mathcal{A}_0)}.$$

Note that $E[\tau]$ is strictly less than k^*, which implies that the user satisfies in an average time, less than k^*.

L^1 **metric with 4 recommendations, i.e., with** $M = 4$ **:** The above analysis can easily be extended. We will have 4^k partitions in Theorem 1 and the optimal policy for L^1 metric is ($r\angle\theta$ is representation in polar coordinates):

$$a^*_{k-1,i} = X_{k-1} + \frac{R}{2^{k-2}} \angle((i-1)90^\circ + 45^\circ) \text{ for all } i \leq 4, k.$$

L^2 **metric** Here user interest v_{ref} is uniformly distributed in a circle of radius $\overline{R}$, with centre as X_0 while the user is satisfied with any item inside a circle of radius $\underline{r}$ with v_{ref} as the centre. The largest square contained in the bigger circle is of size $\sqrt{2}\,\overline{R}$ while the smallest square containing the smaller circle is of size $2\underline{r}$. Let $R := 2^{k^*} 2\underline{r}$ with $k^* := \max_k\{2^k 2\underline{r} \leq \sqrt{2}\,\overline{R}\}$. If $2^T - 2 < R^2/\underline{r}^2$ then again we have sufficient number of disjoint balls to use Corollary 1 and the optimal policy is exactly similar to that described in (6).

B. Implementation of continuous optimal policies

We refer optimal policy of equation (6) by π^*_{90}, to indicate that 90° plays major role as explained below.

Policy π^*_{90}**:** From (6) at any time step k, CP recommends two items $(a_{k,1}, a_{k,2})$, equidistant $(R/2^{k/2})$ from X_{k-1} while maintaining the following angular separations:

$$\angle\{\overrightarrow{X_k a_{k,1}}, \overrightarrow{X_k a_{k,2}}\} = 180^\circ \text{ while } \angle\{\overrightarrow{X_k a_{k,1}}, \overrightarrow{X_{k-1} X_k}\} = 90^\circ$$

where $\overrightarrow{xy}$ implies the line joining points $\mathbf{x}, \mathbf{y}$ and $\angle\{line1,\, line2\}$ represents the angle between the two lines.

One can obtain the estimate of $E[\tau]$, using the iterative procedure described in Algorithm 1. Repeat it for many samples and compute the same mean of τ to obtain an estimate $\hat{E}[\tau]$.

Algorithm 1: Policy π^*_{90} Algorithm (for each sample)

Initialize: Generate v_{ref}, X_0 randomly (uniformly)
Step k**.i:**, CP provides two recommendation $a^*_{k,i}$, $i = 1, 2$ as given by (6).
Step k**.ii:** User choses the best recommendation according to (7) and returns X_k.
If $k > T$ or $X_k \in \mathcal{B}(v_{ref}, \underline{r})$ **Then** Exit with $\tau = k$
Else return to **Step** k**.i** with $k = k + 1$

These results in continuous space are of independent interest. For example, they can be used in robotics: in rescue robot, robot navigation systems etc. In future we would like to explore this angle further.

Policy π_{180}**:** The policy π^*_{90} needs 90 degrees, and we will notice that the implementation of 90 degrees (after translation to discrete space) requires complicated logic. Hence we propose another policy π_{180} which only ensures 180 degree separation between the two recommendations ($a_{k,1}$ and $a_{k,2}$), but the angular separation with respect to the previous user choice X_{k-1} is chosen randomly, i.e, now

$$\angle\{\overrightarrow{X_k a_{k,1}}, \overrightarrow{X_{k-1} X_k}\} = \text{ random.}$$

Below, we study the loss of performance with this policy and one needs to chose an appropriate policy based on their own tradeoff of complexity versus performance.

Simulation results on Continuous space

We would like to recall that maximizing $E[T-\tau]$ is equivalent to minimizing $E[\tau]$. However we are interested in $E[\tau]$ and hence are plotting the estimate of $E[\tau]$ in all the simulations. We compare the $E[\tau]$ performance of policy π_{180} in comparison with that of π^*_{90}, using both L^1 as well as L^2 metrics. With large values of $\underline{r}$, the loss of performance is negligible with L^2 distance while it is significant with L^1 distance (see Tables I and II). But with smaller values of $\underline{r}$, the losses are more even with L^2 case. However the losses are significant with L^1 metic for all cases. In this case, again, the losses increase with the decreases in $\underline{r}$. This implies: a) in continuous space with Euclidean distance one may use less complicated π_{180} in place of π^*_{90}, when $\underline{r}$ is not very small; a) in all other spaces, it may be important to ensure 90 degrees separation.

$\underline{r}$	$E[\tau](\pi_{90})$	$E[\tau](\pi_{180})$
10	6.88	36.8
15	5.6	26.85
20	4.78	20.0
30	3.41	11.56
50	2.18	4.68
60	1.8	3.18
70	1.65	2.30
80	1.46	1.82

TABLE I
$E[\tau]$ in Continuous L^1 space with $\overline{R} = 100$

$\underline{r}$	$E[\tau](\pi_{90})$	$E[\tau](\pi_{180})$
10	6.25	28.15
15	4.86	19.25
20	4.08	13.37
30	2.72	7.15
50	1.73	2.74
60	1.52	1.90
70	1.41	1.54
80	1.26	1.30

TABLE II
$E[\tau]$ in Continuous L^2 space with $\overline{R} = 100$

V. Discrete Database and Notion of Local Angle

We consider binary database with F features, similarity based distance and cardinality based measure. A ball $\mathcal{B}(v, r)$ here includes all those items which match in more than $F(1-r)$ features with v, e.g., $\mathcal{B}((10), 0.5) = \{(10)\}$, $\mathcal{B}((10), 1) = \mathcal{S}$.

A. Optimal policies in Discrete space

We again use Corollary 1 to obtain optimal policies in some example scenarios. One can easily verify the following.

Case I: $F = 7$**,** $T = 2$ **and** $\underline{r} = 2/7$**:** Optimal $\mathcal{Q}_{\pi^*}$ is

$$a^1_1 = 1111111,\quad a^1_2 = 0000000,\qquad a^2_1 = 0011111,$$
$$a^2_2 = 1111100,\quad a^2_3 = 1100000 \quad\text{and}\quad a^2_4 = 0000011.$$

Case II: $F = 7$**,** $T = 2$ **and** $\underline{r} = 3/7$**:** Optimal $\mathcal{Q}_{\pi^*}$ is

$$a^1_1 = 1111111,\quad a^1_2 = 0000000,\qquad a^2_1 = 0001111,$$
$$a^2_2 = 111100,\quad a^2_3 = 1110000 \quad\text{and}\quad a^2_4 = 0000111.$$

Case III: $F = 9$**,** $T = 2$ **and** $\underline{r} = 4/9$**:** Optimal $\mathcal{Q}_{\pi^*}$ is

$$a^1_1 = 111111000,\quad a^1_2 = 000000111,\qquad a^2_1 = 001111001,$$
$$a^2_2 = 110001110,\quad a^2_3 = 110000110 \text{ and } \quad a^2_4 = 001110001.$$

One can continue this way, however we obtain optimal policies in a restrictive manner. Our aim is to obtain policies for discrete data base which are applicable for general scenarios. Further we would like iterative algorithms, which can easily handle large data sets and tracking (v_{ref} can be drifting too). The idea is to translate the continuous policies to the discrete space, by introducing a notion of local angle.

In this section, we also introduce/describe another policy referred to as π_R and show that the corresponding recommendations converge towards the satisfactory item with probability one. We compare the translated π_{90} with π_R and also study a hybrid policy. We verify the functionality using simulations.

Approximate notion of distance using similarity measure is already discussed in section III. We begin with the notion of angle and followed by the translation of the policies.

B. Notion of Local Angle

Distance between items defined via similarity measure is drastically different from that in continuous space. Consider a centre in a continuous space and points on a given circumference. There are uncountable such points. The points might be at equidistant from the centre, but their inter distances can vary significantly based on their angular separations. It is difficult to achieve such a notion in discrete space in a *consistent* manner, especially when the number of equidistant points is large. The number of points at equidistance from a reference point is finite in discrete space, nevertheless grows with the distance from the centre. *Hence we define a notion of local angle, which is applicable for small radius balls.*

We begin with three items (v_1, v_2 and v_3) and define the notion of angle using their inter distances, d_{12}, d_{23} and d_{13}. Fix the position of item v_1, then v_2 can be anywhere on the circumference of $\mathcal{B}(v_1, d_{12})$ (see Fig. 2). Fix one of these points as v_2 and now draw two circles, with respective centres as v_1 and v_2 and radii as d_{13} and d_{23}. Then item v_3 can be at one of the two intersecting points as shown (see Fig. 2). Choose any one of the two points as v_3 and form a triangle to obtain the required angle. For example, θ in Fig. 2 represents the 'local' angle between lines joining items v_1-v_2 and v_1-v_3. Note that, this angle is independent of the chosen points and depends only upon their inter distances, d_{12}, d_{23} and d_{13}.

180-degrees: An angular separation close to 180 degrees is achieved if the inter distances between the pair of points (equidistant from a reference point) is maximized. Given n items which are at the same distance from a fixed point, chose two among those points whose inter distance is maximum, i.e., a pair with maximum number of mis-match features.

90-degrees: An angular separation close to 90 degrees is achieved if the inter distances between the points satisfy the Pythagorus theorem (to the best extent possible). We achieve this by performing a mini optimization (8) at every step.

C. *Policy π_{90}, obtained by translation:*

This is obtained by translating the continuous π^*_{90} policy. From (6), subsequent recommendation are given at diminishing distances $\{\nu_1, \nu_2, \nu_3 \cdots\} = \{R/2, R/2, R/4 \cdots\}$ while the two recommendations at the same time step are maintained at 180 degrees with respect to each other. Distances in term of corresponding similarity measures are given by: $\phi_k = \min\{1, \lfloor \nu_k \rfloor\}$, where we set $R = F/2$. We also attempted simulations with other possible sequences of step sizes $\{\nu_k\}$ and found these to be the best.

We also require that the lines $\overrightarrow{X_k a_{k,1}}$ and $\overrightarrow{X_{k-1} X_k}$ are at 90 degrees. We achieve this by ensuring that the inter distances satisfy Pythagorus theorem, i.e., we minimize

$$\left(\sqrt{d(X_k, a_{k,1})^2 + d(X_{k-1}, X_k)^2} - d(a_{k,i}, X_{k-1})\right)^2$$

or equivalently minimize

$$(\hbar_k - d(a_{k,i}, X_{k-1}))^2 \text{ where hypotenuse } \hbar_k := \sqrt{\phi_k^2 + \phi_{k-1}^2}.$$

We also need to ensure $d(X_k, a_{k,1})$ is close to ϕ_k. Let $\mathcal{K}_k$ represent the set of indices where X_k matches with X_{k-1}:

$$\mathcal{K}_k = \{j : X_{k_j} = X_{k-1_j}\} \text{ and note that } |\mathcal{K}_k| = F - \phi_k.$$

In the above $|.|$ (when the argument is a set) implies cardinality. Our aim is to chose a pair of numbers (n, m) optimally, where n represents the number of matching features between $a_{k,i}$ and X_{k-1} in the $\mathcal{K}_k$ positions and m represents the same among $\mathcal{K}_k^c$ positions. With such a choice of (n, m), $a_{k,i}$ matches with X_k in n among $\mathcal{K}_k$ positions and in $|\mathcal{K}_k^c| - m$ among $\mathcal{K}_k^c$ positions . Thus given (n, m) we have

$$\begin{aligned} d(a_{k,i}, X_{k-1}) &= F - (n + m) := g_1(n, m) \text{ and} \\ d(a_{k,i}, X_k) &= F - (n + |\mathcal{K}_k^c| - m) := g_2(n, m). \end{aligned}$$

Let $l_k := F - \phi_k$. Now we propose to chose n^* and m^* which minimize the following joint cost (one term for 90 degrees and another for ϕ_k distance):

$$\min_{\begin{Bmatrix}(n,m): m \le \phi_{k-1} \\ \& \ n \le l_{k-1}\end{Bmatrix}} (\phi_k - g_1(n, m))^2 + w(\hbar_k - g_2(n, m))^2. \quad (8)$$

In the above w is introduced to consider the tradeoff between the two costs. To ensure 180 degree separation between the two ($i = 1, 2$) recommendations, n^* indices corresponding to the two recommendations are chosen from two disjoint sets of $\mathcal{K}_k$. In case $2n^* > l_{k-1}$, we consider $l_{k-1} - n^*$ and repeat similar construction, but now with inverted bits. A similar procedure is followed with m^* and $\mathcal{K}_k^c$. This construction ensures maximum possible distance between $a_{k,1}$ and $a_{k,2}$, which implies an angular separation close to 180 degrees. In policy π_{90} at any time step k, CP recommends two items to the user, which are generated by the Algorithm 2.

D. *Random Policy π_R:*

Policy: At any time k, one random index L among F features is chosen. First recommendation $a_{k,1}$ is same as X_k but at index L, it is set to 0. The second recommendation $a_{k,2}$ is similar except that at L it is set 1 and user selects the best match.

Analysis: Let α_k represent the number of mis-matching features between user choice X_k and v_{ref}. If the position L is one among the α_k mis-matching indices, then it improves the number of matching features, i.e., $\alpha_{k+1} = \alpha_k - 1$. If it is one among $F - \alpha_k$ (already matching features) then there is no improvement in α_{k+1}. Thus the sequence α_k follows a random walk, with correlated drift sequence $W_k \in \{0, -1\}$

$$\alpha_{k+1} = \alpha_k + W_{k+1},$$

with the distribution given by:

$$P(W_{k+1} = 0) = 1 - \frac{\alpha_k}{F} \text{ and } P(W_{k+1} = -1) = \frac{\alpha_k}{F}$$

When $\alpha_k > 0$, average drift $E[W_{k+1}] < 0$, one can show the random walk α_k converges to 0, i.e., the policy reaches satisfaction with probability one.

E. *Policy π_{180} :*

Policy: The step sizes remain same as is in the previous policy. At any time k, l_k indices are chosen randomly. In the first l_k indices, $a_{k,1}$ equals inverted value of the corresponding indices in X_k while it matches with the later in all other positions. Second recommendation $a_{k,2}$ is constructed similarly but using another set of l_k random indices, excluding the first l_k indices. This exclusion ensures that the distance between $a_{k,1}$ and $a_{k,2}$

Algorithm 2: Policy π_{90} Algorithm (Discrete)

Input: X_{k-1}, X_k, $\mathcal{K}_k$ and (n^*, m^*) from (8)
Step k**.i:** $\mathcal{I}$: permuted array of indices of $\mathcal{K}_k$
Step k**.ii: If** $n^* < |\mathcal{K}_k|/2$ **Then**
Define $\mathcal{I}_1$ and $\mathcal{I}_2$ as 1:n^* & n^*+1:2n^* indices of $\mathcal{I}$;
Do $a_{k,1}(i) = X_{k,1}(i) \forall i \in \mathcal{I}_1$ and
$a_{k,1}(i) \neq X_{k,1}(i) \forall i \in \mathcal{I} - \mathcal{I}_1$.
Do $a_{k,2}(i) = X_{k,1}(i) \forall i \in \mathcal{I}_2$ and
$a_{k,2}(i) \neq X_{k,1}(i) \forall i \in \mathcal{I} - \mathcal{I}_2$.
Else
Define $\mathcal{I}_1^c$ and $\mathcal{I}_2^c$ as 1:$(|\mathcal{K}_k| - n^*)$ &
$(|\mathcal{K}_k| - n^*)$+1:2$(|\mathcal{K}_k| - n^*)$ indices of $\mathcal{I}$;
Do $a_{k,1}(i) \neq X_{k,1}(i) \forall i \in \mathcal{I}_1^c$ and
$a_{k,1}(i) = X_{k,1}(i) \forall i \in \mathcal{I} - \mathcal{I}_1^c$. Similarly,
Do $a_{k,2}(i) \neq X_{k,1}(i) \forall i \in \mathcal{I}_2^c$ and
$a_{k,2_i} = X_{k,1_i} \forall i \in \mathcal{I} - \mathcal{I}_2^c$.
Step k**.iii:** $\mathcal{J}$: permuted array of indices of $\mathcal{K}_k^c$
Step k**.iv: If** $m^* < |\mathcal{K}_k^c|/2$ **Then**
Define $\mathcal{J}_1$ and $\mathcal{J}_2$ as 1:m^* & m^*+1:2m^* indices of $\mathcal{J}$;
Do $a_{k,1}(j) = X_{k-1,1}(j) \forall j \in \mathcal{J}_1$ and
$a_{k,1}(j) \neq X_{k-1,1}(j) \forall j \in \mathcal{J} - \mathcal{J}_1$.
Do $a_{k,2}(j) = X_{k-1,1}(j) \forall j \in \mathcal{J}_2$ and
$a_{k,2}(j) \neq X_{k-1,1}(j) \forall j \in \mathcal{J} - \mathcal{J}_2$.
Else
Define $\mathcal{J}_1^c$ and $\mathcal{J}_2^c$ as 1:$(|\mathcal{J}_k| - m^*)$ &
$(|\mathcal{J}_k| - m^*)$+1:2$(|\mathcal{J}_k| - m^*)$ indices of $\mathcal{J}$;
Do $a_{k,1}(j) \neq X_{k-1,1}(j) \forall j \in \mathcal{J}_1^c$ and
$a_{k,1}(j) = X_{k-1,1}(j) \forall j \in \mathcal{J} - \mathcal{J}_1$. Similarly,
Do $a_{k,2}(j) \neq X_{k-1,1}(j) \forall j \in \mathcal{J}_2^c$ and
$a_{k,2_j} = X_{k-1,1_j} \forall j \in \mathcal{J} - \mathcal{J}_2^c$.
End

is maximum possible ($2l_k$ mismatches) one which implies a degree close to 180 degrees. In case $l_k > F/2$, we consider $F - l_k$ and repeat similar construction, but now with inverted bits.

Analysis: Once again we write the evolution of number of mis-matching features as a random walk:

$$\alpha_{k+1} = \alpha_k + W_{k+1},$$

One can analyze this policy in a similar way as done with π_P. When $l_k = 1$, drift W_{k+1} takes value +1, if both the positions (in both the recommendations) are from among $F - \alpha_k$ matching positions. Then the matchings (with respect to v_{ref}) of both recommendations would be one less than corresponding value with X_k. On the other hand, if one of the two positions is from α_k positions then the matchings of the corresponding recommendation improves by one, i.e., $W_{k+1} = -1$. So,

$$\begin{aligned} P(W_{k+1} = 1) &= \frac{F - \alpha_k}{F} \frac{F - \alpha_k - 1}{F - 1} \\ &= \left(1 - \frac{\alpha_k}{F}\right)\left(1 - \frac{\alpha_k}{F - 1}\right), \\ P(W_{k+1} = -1) &= 1 - P(W_{k+1} = 1) \text{ when } l_k = 1. \end{aligned}$$

	$E[\tau]$ for different policies		
$\overline{F}$	$E[\tau](\pi_{90})$	$E[\tau](\pi_R)$	$E[\tau](\pi_{hybrid})$
60 (120)	9.8 (24.47)	14.43 (33.43)	10.22 (24.53)
65 (130)	23.37 (51.99)	27.8 (59.57)	23.25 (52.33)
70 (140)	39.89 (84.95)	42.89 (90.66)	39.06 (82.38)
75 (150)	60.7 (130.13)	60.53 (126.66)	57.06 (118.82)
80 (160)	91.57 (190.82)	82.68 (170.07)	79.68 (163.16)
85 (170)	134.67 (284.15)	110.79 (228.29)	107.27 (219.89)
90 (180)	214.35 (456.76)	149.42 (308.13)	146.22 (298.73)

TABLE III
DISCRETE POLICIES WITH $F = 100$ (200)

When $l_k = 2$ for some k the drift W_k can take either one of the 5 values $\{-2, -1, 0, 1, 2\}$. Drift $W_k = 2$ if all the four positions are from the previous matching positions and then the matching of the best also reduces by 2. If two of the positions are matching and two are not then we lose one matching in each recommendation and hence the best also looses one matching. For example with $l_k = 2$,

$$\begin{aligned} P(W_{k+1} = 2) &= \frac{(F - \alpha_k)!(F-4)!}{(F - \alpha - 4)!F!} \\ P(W_{k+1} = -2) & \\ = P(&\text{ In one of the recommendations both the} \\ &\text{positions are from mis-matching positions}). \end{aligned}$$

One can continue this way and show that $\alpha_k \to 0$, i.e., that X_k converges to v_{ref} with probability 1. We need to use dominating technique to complete the proof and we are currently working towards that.

F. Simulation results: Discrete space

We now study the various discrete policies via simulations with $F = 100$ and 200. The results are tabulated in Tables III and IV.

	$E[\tau]$ for Policy π_{180}	
$\overline{F}$	Policy π_{180}	Policy $\pi_{180hybrid}$
60 (120)	10.21 (97.23)	10.16 (39.91)
65 (130)	31.68 (491.66)	27.02 (434.46)
70 (140)	71.91 (690.55)	65.79 (685.11)
75 (150)	266.33 (2571.72)	257.96 (2652.58)
80(160)	3385.59 (T)	3304.24 (T)
85 (170)	T (T)	7423.06 (T)
90 (180)	T (T)	T (T)

TABLE IV
POLICY π_{180} WITH $F = 100$ (200)

We notice from the Table III that π_{90}^* performs superior to that of π_R for $\overline{F}$ almost up to 70 (140). However there is a degradation in the performance with larger values of $\overline{F}$. This is anticipated as with smaller distances the optimization (8) may not be effective because of fewer choices in the feasibility region. Hence we also consider a hybrid policy which used π_{90} in the beginning of the session and π_R during the later part. The performance of this policy is presented in the last column of the table, and it out performs all other policies.

In Table IV we study π_{180}. Simulation results show that the performance of Policy π_{180} is significantly inferior as compared to Policy π_{90}, π_R, and π_{hybrid}. Moreover, it fails when higher number of features are required to match (see Table IV). This again reinforces that without 90 degree logic, the performance is inferior. The degradation/losses increase significantly as larger levels of satisfaction (smaller $\underline{r}$ or larger $\overline{F}$) are demanded. Compare Tables IV and III to see the degradation of π_{180}. The losses are more even with a hybrid policy, consisting of π_{90} in the beginning and $\pi_{180hybrid}$ in the later stages, whose performance is given in the last column of Table IV.

VI. Conclusions

We presented and demonstrated novel recommendation policies, which are based on user-generated responses. Unlike the traditional recommendation schemes, the recommendations are neither based on the history of the usage of the items nor on the history of the user(s). These can be applied in variety of recommendation systems (RS), in particular are very useful in the context of anonymous users (cold start problems). But their bigger advantage is that they exploit the responses of the same user in the same session.

We proposed a notion of local angle in the context of discrete data base. Simulation based results clearly show that the notion of local angle improved the average hitting time performance. The proposed recommendation policies can be used in discrete as well as continuous space based applications, e.g. video item providers, online shopping portals, service and rescue robot navigation, etc. These are preliminary results and we would like to obtain theoretical performance of proposed policies.

When users with similar interests are sufficiently different from each other and if the same user is interested in drastically different items in different sessions the traditional history based methods may not perform well, while this approach can perform better. In future, we would like to test our policies against the other well known approaches in such scenarios and propose hybrid methods, if required.

Acknowledgements

Discussions with Prof. Sandeep Juneja and Rahul Vaze, TIFR, Mumbai, India have been very useful.

References

[1] A. M. Rashid, I. Albert, D. Cosley, S. K. Lam, S. McNee, J. A. Konstan, and J. Riedl. “Getting to know you: Learning new user preferences in recommender systems.” In Proceedings of the 2002 International Conference on Intelligent User Interfaces, San Francisco, CA, Feb. 2002, 127-134.

[2] Rashid, Al Mamunur, George Karypis, and John Riedl. “Learning preferences of new users in recommender systems: an information theoretic approach.” ACM SIGKDD Explorations Newsletter 10.2 (2008): 90-100.

[3] Blerina Lika, Kostas Kolomvatsos, and Stathes Hadjiefthymiades. “Facing the cold start problem in recommender systems.” Expert Systems with Applications 41 (2014) 2065-2073.

[4] M. Stritt, K. H. L. Tso, and L. S. Thieme, Attribute aware anonymous recommender systems, in Advances in Data Analysis, R. Decker and H.-J. Lenz, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, January 15, 2014, pp. 497504.

[5] Hill, Will and Stead, Larry and Rosenstein, Mark and Furnas, George, Recommending and Evaluating Choices in a Virtual Community of Use. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’95), Denver, Colorado, USA, 194-201.

[6] Lev Grossman, ”How Computers Know What We Want Before We Do.” Available at: http://item.time.com/time/magazine/article/0,9171, 1992403,00.html

[7] G. Adomavicius and A. Tuzhilin. Toward the Next Generation of Recommender Systems: A Survey of the State of the Art and Possible Extensions, IEEE Trans. Knowl. Data Eng, 17 (6), 2005, pp. 734-749.

[8] J. Gaillard, M. El-Beze, E. Altman and E. Ethis. “Flash reactivity : adaptative models in recommender systems,” International Conference on Data Mining (DMIN), WORLDCOMP, (2013).

[9] Adomavicius, Gediminas, and Alexander Tuzhilin. ”Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions.” Knowledge and Data Engineering, IEEE Transactions on 17.6 (2005): 734-749.

Appendix: Proofs

The proofs are given in the next page.

Proof of Lemma 1: Given that $B_0 = \mathcal{U}(\mathcal{A}_0)$ is the initial belief. If $\mathbf{a_1} = (a_{1,1}, a_{1,2})$ are the recommended items at time step k=1, user choses an item closer to the desired item, v_{ref} (recall it is unknown to the system) and hence

$$\begin{aligned} X_1 &= a_{1,1}1_{\{v_{ref}\in\mathcal{A}_{11}\}} + a_{1,2}1_{\{_{v_{ref}\in\mathcal{A}_{12}}\}}, \text{ where} \\ \mathcal{A}_{11} &= \{v_{ref} \in \mathcal{A}_0 \mid d(v_{ref}, X_{11}) < d(v_{ref}, X_{12})\} \text{ and} \\ \mathcal{A}_{12} &= \{v_{ref} \in \mathcal{A}_0 \mid d(v_{ref}, X_{12}) < d(v_{ref}, X_{11})\}. \end{aligned}$$

By definition B_1 is the conditional distribution of v_{ref} given H_1 i.e. v_{ref}/H_1. For any subset Λ (Borel if continuous case)

$$\begin{aligned} &P(B_1 \in \Lambda) = P(v_{ref} \in \Lambda / v_{ref} \in \mathcal{A}_0, B_0, \mathbf{a_1}) \\ &\quad = \frac{\mu(\Lambda \cap \mathcal{A}_{11})}{\mu(\mathcal{A}_{11})} 1_{\{X_1 = a_{1,1}\}} + \frac{\mu(\Lambda \cap \mathcal{A}_{12})}{\mu(\mathcal{A}_{12})} 1_{\{X_1 = a_{1,2}\}}. \end{aligned}$$

The equality in the above equation is because, v_{ref} lies in $\mathcal{A}_{11}$ when $X_1 = a_{1,1}$. Similar is the case when $X_1 = a_{1,2}$. Thus belief B_1 will be concentrated either on $\mathcal{A}_{11}$ or on $\mathcal{A}_{12}$ based only on X_k, $\mathbf{a_1}$ and B_0. And from equation (9) it is a uniform random variable:

$$B_1 \sim \mathcal{U}(\mathcal{A}_1) \text{ where } \mathcal{A}_1 = \mathcal{A}_{11} 1_{\{X_1 = a_{1,1}\}} + \mathcal{A}_{12} 1_{\{X_1 = a_{1,2}\}}.$$

Following similar logic, at any time step k we will have

$$B_k \sim \mathcal{U}(\mathcal{A}_k), \text{ with, } \mathcal{A}_k = \mathcal{A}_{k,1} 1_{\{X_k = a_{k,1}\}} + \mathcal{A}_{k,2} 1_{\{X_k = a_{k,2}\}}. \ \square$$

Proof of Theorem 1: We have

$$\begin{aligned} E[T-\tau] &= \sum_{k=0}^{T} P(T-\tau > k) = \sum_{k=0}^{T} P(\tau < T-k) \\ &= \sum_{k=0}^{T} P(\tau < k) = \sum_{k=1}^{T} P(\tau < k) \qquad \text{because} \quad P(\tau < 0) = 0 \\ &= \sum_{k=0}^{T-1} P(\tau \le k) = \sum_{k=0}^{T-1} P(v_{ref} \in \cup_{l=0}^{k} \mathcal{B}(X_l, \underline{r})) \\ &\le \sum_{k=0}^{T-1} \sum_{l=0}^{k} E\Big[\mu(\mathcal{B}(X_l, \underline{r}))\Big] = \sum_{k=0}^{T-1} \Big[(T-k) E\Big[\mu(\mathcal{B}(X_k, \underline{r}))\Big]\Big] \\ &\le \sum_{k=0}^{T-1} \Big[(T-k) E\Big[\mu(\mathcal{B}(X_k, \underline{r}))\Big]\Big]. \end{aligned}$$

It is easy to see that $E\Big[\mu(\mathcal{B}(X_k, \underline{r}))\Big] = E\Big[v_{ref} \in \mathcal{B}(X_k, \underline{r})\Big]$ and hence

$$E[T-\tau] \le \sum_{k=0}^{T-1} (T-k) E\Big[v_{ref} \in \mathcal{B}(X_k, \underline{r})\Big]. \tag{9}$$

For any set C, point $\mathbf{a}$ and any time k, it is easy to verify ($B_k \sim \mathcal{U}(\mathcal{A}_k)$ implies $v_{ref} \in \mathcal{A}_k$) the following

$$P(v_{ref} \in \mathcal{B}(\mathbf{a}, \underline{r}) | B_k) = P(v_{ref} \in \mathcal{B}(\mathbf{a}, \underline{r}) | v_{ref} \in \mathcal{A}_k) = \frac{\mu(\mathcal{B}(\mathbf{a}, \underline{r} \cap \mathcal{A}_k)}{\mu(\mathcal{A}_k)}.$$

We will use this logic repeatedly in the following. In addition if we condition on X_k we have,

$$\begin{aligned} P(v_{ref} \in \mathcal{B}(X_k, \underline{r})) &= E[P(v_{ref} \in \mathcal{B}(X_k, \underline{r}) / X_k, B_k)] \\ &= E\left[\frac{\mu(\mathcal{B}(X_k, \underline{r}) \cap \mathcal{A}_k)}{\mu(\mathcal{A}_k)}\right]. \end{aligned} \tag{10}$$

Further when $X_k = a_{k-1,i}$, i.e., when user choses i-th recommendation at step k, it implies the new belief is concentrated on $\mathcal{A}_{k-1,i}$, i.e., that $\mathcal{A}_k = \mathcal{A}_{k-1,i}$. So, for any $k > 1$

$$\begin{aligned} E[v_{ref} \in \mathcal{B}(X_k, \underline{R})] &= E\left[E\left[\frac{\mu(\mathcal{B}(X_k, \underline{R}) \cap \mathcal{A}_k)}{\mu(\mathcal{A}_k)} \left(1_{X_k=a_{k-1,1}} + 1_{X_k=a_{k-1,2}}\right)\middle| B_{k-1}\right]\right] \\ &= E\left[E\left[1_{X_k=a_{k-1,1}} \frac{\mu(\mathcal{B}(a_{k-1,1}, \underline{r}) \cap \mathcal{A}_{k-1,1})}{\mu(\mathcal{A}_{k-1,1})} + 1_{X_k=a_{k-1,2}} \frac{\mu(\mathcal{B}(a_{k-1,2}, \underline{r}) \cap \mathcal{A}_{k-1,2})}{\mu(\mathcal{A}_{k-1,2})}\middle| B_{k-1}\right]\right] \\ &= E\left[P(X_k = a_{k-1,1}|B_{k-1}) \frac{\mu(\mathcal{B}(a_{k-1,1}, \underline{r}) \cap \mathcal{A}_{k-1,1})}{\mu(\mathcal{A}_{k-1,1})} + P(X_k = a_{k-1,2}|B_{k-1}) \frac{\mu(\mathcal{B}(a_{k-1,2}, \underline{r}) \cap \mathcal{A}_{k-1,2})}{\mu(\mathcal{A}_{k-1,2})}\right] \\ &= E_{\mathcal{A}_{k-1}}\left[\frac{\mu(\mathcal{B}(a_{k-1,1}, \underline{r}) \cap \mathcal{A}_{k-1,1}) + \mu(\mathcal{B}(a_{k-1,2}, \underline{r}) \cap \mathcal{A}_{k-1,2})}{\mu(\mathcal{A}_{k-1})}\right] \end{aligned} \tag{11}$$

Note in the above that the recommendations $a_{k-1,i}$, $i = 1, 2$ depend upon previous belief B_{k-1}, but is a constant given the later as we are considering only pure policies (see Figure 1). Continuing in a similar way we get that:

$$\begin{aligned} E[v_{ref} \in \mathcal{B}(X_k, \underline{r})] &= \sum_{j=1}^{2^k} E\left[\frac{\mu(\mathcal{B}(a_j^k, \underline{r}) \cap \mathcal{A}_j^k)}{\mu(\mathcal{A}_0)}\right] \\ &= \sum_{j=1}^{2^k} \frac{\mu(\mathcal{B}(a_j^k, \underline{r}) \cap \mathcal{A}_j^k)}{\mu(\mathcal{A}_0)}. \end{aligned} \tag{12}$$

Now from equations (9) and (12), we get

$$E[T - \tau] \leq \sum_{k=0}^{T-1} (T - k) \sum_{j=1}^{2^k} \frac{\mu(\mathcal{B}(a_j^k, \underline{r}) \cap \mathcal{A}_j^k)}{\mu(\mathcal{A}_0)}. \qquad \square$$