Long Term Recommender Benchmarking for Mobile Shopping List Applications using Markov Chains

Sandro Schopfer
Chair of Information Management
ETH Zurich, Switzerland
sandro.schopfer@ethz.ch

Thorben Keller
Cosibon AG
St.Gallen, Switzerland
thorben.keller@cosibon.com

ABSTRACT

This paper presents a method to estimate the performance and success rate of a recommender system for digital shopping lists. The list contains a number of items that are allowed to occupy three different states (*to be purchased*, *purchased* and *deleted*) as a function of time. Using Markov chains, the probability distribution function over time can be estimated for each state, and thus, the probability that a recommendation is deleted from the list can be used to benchmark a recommender on its endurance and performance. An experimental set up is described that allows to test the presented method in an actual mobile application. The application of the method will allow to benchmark a variety of recommenders. An outlook is given on how the presented method can be used iteratively to support a recommender in finding the user's favourite items/products.

1. INTRODUCTION

Shopping lists have served for many decades as a tool to keep track of items to be purchased in physical stores. Consequently, shopping lists have been subject to a vast amount of scientific studies [2], [3] and [4].

With further advancement of mobile computing, the traditional hand-written shopping list has been digitalized and implemented as mobile applications. Many applications with different degrees of functionality have been deployed for all major mobile platforms. However, at the core they share the same functionality represented by one or more lists consisting of items which are displayed to and modified by the user in one of three possible states (*to be purchased*, *purchased* or *deleted*). Such a shopping list can potentially serve as a commercial platform which is realized as a brand recommendation system. Usually, recommendations on a mobile platform or web store occur based on popular items that are suggested to the user only once during a shopping session and rely heavily on purchase data. If the recommended item has been bought, the recommendation is considered to be successful. However, the items of a shopping list remain for some period of time while cyclically changing the status between *to be purchased* and *purchased* and eventually terminate in the state *deleted*. Under these circumstances, the success rate of a recommended product can be continuously evaluated given a certain type of recommender system. Furthermore, the durability horizon and therewith the probability that a recommended item will be purchased repetitively can be characterized. Recommendation applied to digital shopping list is unique in one point: no purchase data from the retailer is needed to build a recommender and to estimate its performance.

The present paper describes a methodology and an experimental set up which allows to benchmark different kinds of recommender systems with respect to their success rate. A recommender system can be benchmarked based on the probability that a recommended item is still actively used in the list after some time it has been suggested to the user.

RecSys 2014 Poster Proceedings, October 6-10, 2014, Foster City, Silicon Valley, USA.

2. METHODOLOGY

This section presents a method that allows to determine the performance of a recommendation. A straight forward solution to this problem is to track whether a recommended product has been accepted and consequently added to the list (in state *to be purchased*). However, this study aims to investigate the long term effect and therewith the durability of a recommended item. To that end, this section presents a model for long term analysis of a recommended item using Markov chains.

A shopping list is represented as a vector $\boldsymbol{\mathcal{I}}(t) = (\mathcal{I}_1(t), \mathcal{I}_2(t), \ldots, \mathcal{I}_n(t))^T$ with n items, where each item $\mathcal{I}_k(t)$ corresponds to a stochastic process continuous in time t with finite state space $S = \{-1, 0, 1\}$ which corresponds to *deleted*, *purchased* and *to be purchased*, respectively.

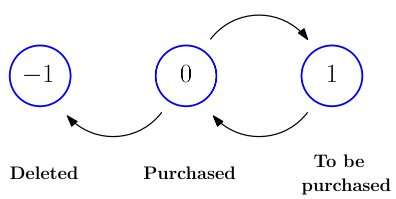

Figure 1: State space of a dynamic shopping list

These definitions allow the application of the theory and formalism of time continuous Markov processes with three discrete states for each item. The state space of the system with its possible transitions is visualized in Figure 1.

The probability that an item k occupies a state at time t is formally summarized in the vector

$$\begin{aligned}\boldsymbol{\pi}_k(t) &= (\pi_k^{(1)}, \pi_k^{(2)}, \pi_k^{(3)})^T \\ &= (P\left[\mathcal{I}_k(t)=-1\right], P\left[\mathcal{I}_k(t)=0\right], P\left[\mathcal{I}_k(t)=1\right])^T\end{aligned}$$

We define that every recommended item will be initialized in the state 1 *to be purchased* with $\boldsymbol{\pi}_k(t=t_k^0)=(0,0,1)^T$ where t_k^0 represents the point in time when item k is recommended. The occupancy probability over time for each state is governed by the Master Equation (1) with the solution (2)

$$\frac{d}{dt}\boldsymbol{\pi}_k = \boldsymbol{\pi}_k \mathbf{Q} \quad (1)$$

$$\boldsymbol{\pi}_k(t) = \boldsymbol{\pi}_k(t_k^0)\cdot \exp\left((t-t_k^0)\mathbf{Q}\right) \quad (2)$$

where $\mathbf{Q}$ describes the transition rate matrix and represents the probability that the process changes from state i to state j per infinitesimal time interval dt. In this context $\mathbf{Q}$ is not given, but can be estimated using a trajectory $I_k(t)$ which is a sample of $\mathcal{I}_k(t)$ as shown in Figure (2).

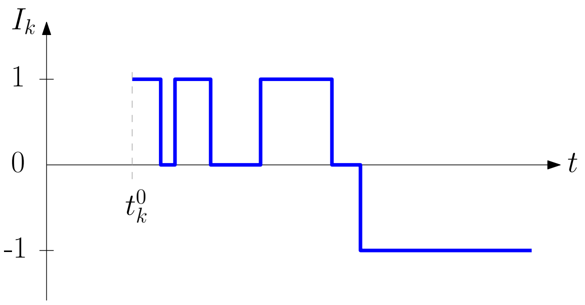

Figure 2: Visualization of a possible trajectory of a recommended item

Once the transition rate matrix is known, the probability distribution $\boldsymbol{\pi}_k$ for all states at all times $t \geq t_k^0$ can be computed using the solution of the Master Equation (2). Of particular interest is the probability that the process of an item occupies the deleted state as this yields a direct feedback on how likely a recommended product will be rejected after some time the item has been recommended to the user at $t=t_k^0$ as illustrated in Figure (3).

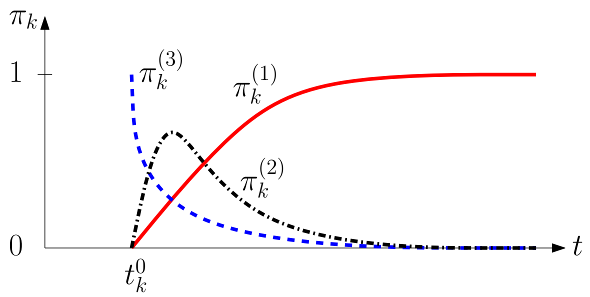

Figure 3: Distribution of the probability for state occupation over time

The probability that a recommended item ends up in the *deleted* state -1 is given by the first component of Equation (2), which is $\pi_k^{(1)}(t)$. The probability of rejecting a recommended item $\pi_k^{(1)}(t)$ can be used as a benchmark score for different recommenders, users and observed period Δt_k.

3. EXPERIMENTAL SETUP

A generic shopping list application – currently under development – will be exposed to a broader audience to generate item data. Grocery items are chosen as experimental context, but the methodology may be applied to other types of items as well. A master data set that maps electronic article numbers (EAN) to product names and categories serves as a basis for this application. Thus, the user can add/delete items in accordance with the master data set and change the status $I_k(t)$ of each item, depending on wheter the item needs to be purchased or not.

Using a selection of recommender systems, alternative and/or complementary items may be recommended. In reference with Figure (1), a recommended product k may be placed in to state 1 *to be purchased* at $t=t_k^0$ if accepted by the user. In case of acceptance, item k cyclically populates through the states $I_k(t)$ as illustrated in Figure (2). Each item and action of a user will be logged to a database allowing straightforward insights into the item dynamics using the methodology presented in the previous chapter.

4. EVALUATION & OUTLOOK

4.1 Recommender Benchmarking

As a benchmarking score we define the probability of rejecting a recommended item $\pi_k^{(1)}$ at a specified $\Delta t_k = t - t_k^0$. In general this score will depend on the user, the recommender system (implicitly the item) used and the time Δt_k that passed after the item has been recommended to the user. Hence, the experiment will reveal how successful a recommender system performs during its "lifetime" for different users, products, and periods of time Δt_k. The lower the benchmark score, the better the performance of the recommender.

4.2 Outlook: Iterative Recommendation

It is assumed that within the context of a shopping list, an ideal recommender is characterized with infinite lifetime which implies that the probability that the process occupies state -1 is below a critical value. This fact may be used to rank all used recommenders [1] during operation across different users to help them to find their favourite products that reside in the list for long periods of time. This may be obtained by minimizing $\pi_k^{(1)}(t)$, and successively replacing a rejected recommender until the user finds stationary items/products that reside over long periods of time in the list.

5. REFERENCES

[1] P. Nurmi et al. Promotionrank: Ranking and recommending grocery product promotions using personal shopping lists. *ACM Transactions on Interactive Intelligent Systems*, 4(1), 2014.

[2] A. Thomas and A. Garland. Supermarket shopping lists: their effect on consumer expenditure. *International journal of distribution management*, 22(2), 1993.

[3] A. Thomas and A. Garland. Grocery shopping: list and non-list usage. *Marketing Intelligence and Planning*, 22(6), 2004.

[4] F. J. Webster and F. Pechmann. A replication of the "shopping list" study. *Journal of Marketing*, 34(2), 1970.