Personalizable and Interactive Sequence Recommender System

Fan Du
University of Maryland
College Park, MD, USA
fan@cs.umd.edu

Sana Malik
Adobe Research
San Jose, CA, USA
sana.malik@adobe.com

Georgios Theocharous
Adobe Research
San Jose, CA, USA
theochar@adobe.com

Eunyee Koh
Adobe Research
San Jose, CA, USA
eunyee@adobe.com

CHI'18 Extended Abstracts, April 21–26, 2018, Montreal, QC, Canada.
ACM 978-1-4503-5621-3/18/04.
https://doi.org/10.1145/3170427.3188506

Abstract
Sequence recommender systems assist people in making decisions, such as which product to purchase and what places to visit on vacation. Despite their ubiquity, most sequence recommender systems are black boxes and do not offer justifications for their recommendations or provide user controls for steering the algorithm. In this paper, we design and develop an interactive sequence recommender system (SeRIES) prototype that uses visualizations to explain and justify the recommendations and provides controls so that users may personalize the recommendations. We conducted a user study comparing SeRIES to a black-box system with 12 participants using real visitor trajectory data in Melbourne and show that SeRIES users are more informed about how the recommendations are generated, more confident in following the recommendations, and more engaged in the decision process.

Author Keywords
Explainable smart systems; sequence recommendations.

ACM Classification Keywords
H.5.2 [User Interfaces]: Graphical user interfaces (GUI)

Introduction
With the development of mobile devices, electronic communication, and sensors, event sequence data are being

collected everywhere from social network activities and online clickstreams, to electronic health records and student academic activities. Sequence recommender systems have been designed to assist people in making decisions, such as choosing what movies to watch next based on the history of other viewers or which places to visit based on the trajectories of past visitors.

Many sequence recommender systems are black boxes and do not justify their recommendations nor allow users to steer the algorithm. Users often have limited knowledge about the underlying processes behind the recommendation, lowering their confidence in the system and discouraging the use of the recommendation. Further, users may not know what data and criteria were used to generate the recommendation, which may introduce unintentional bias by the system. While such black-box techniques have successful applications in entertainment scenarios, previous research found that users want to be more engaged when making more important decisions [3].

We present a personalizable and interactive sequence recommender system (SeRIES) that uses visualizations to explain the decision process and justify its results. It also provides controls and guidance to help users personalize the recommended action plans. We developed a prototype and conducted a user study with 12 participants to compare SeRIES to a black-box system. The study results show that SeRIES users are more informed about how the results are generated, more confident in following the recommendations, and more engaged in the decision process. Our direct contributions are:

- The design and implementation of an interactive sequence recommender system, SeRIES, which combines machine learning and visualization to generate, explain, and personalize recommendations of event sequences,
- A controlled user study with 12 participants measuring increased human understanding and performance when using SeRIES versus a black-box sequence recommender system in a travel planning scenario.

Related Work

Visualizations have been developed to improve the transparency of machine learning models. Alsallakh et al. [1] visualize the performance of classification models for analytics. Prospector [4] provides interactive dependence diagnostics to show how features affect a predictive model. TasteWeights [2] uses an interactive interface to explain the item recommendation process and elicit end users' preferences to improve the relevance of the results.

In this work, we focus on models for generating sequence recommendations. The closest related paper describes MDPvis [5], a suite of visualizations and controls for a category of sequence recommendation algorithms based on Markov Decision Process (MDP). While MDPVis is designed for machine learning practitioners to tune their models, our work focuses on helping end users understand how the recommendations are generated and enable them to interactively personalize the recommendations. To the best of our knowledge, we are the first to apply visualizations to sequence recommender systems for end users.

Description of SeRIES

User Interface

SeRIES consists of three views (Figure 1) to provide sequence recommendations, explain the recommendation generation process, and allow users personalize the recommended action plans.

Recommendation view (left column): The recommended sequence is displayed as a map and a list. Each circle rep-

Figure 1 (cont'd). The *Complex* version of SeRIES: (a) map context view, (b) recommendations, (c) user preference controls, and (d) overview of archived trajectories. This figure illustrates a real dataset of visitor trajectories in Melbourne. Points of interest (POIs) are categorized by their themes such as transport, shopping, and entertainment. In the *Simple* version, only (a) and (b) are visible and only the "recommended plan" is shown.

Our initial design displayed the recommendations (b) in a single list ranked by expected experience and provided three personalized recommendations. However, our pilot users found the list confusing since it was difficult to keep track of the changes as they adjust the preferences controls and to compare plans against each other. Our final design categorizes the recommendations and only shows one plan in each category.

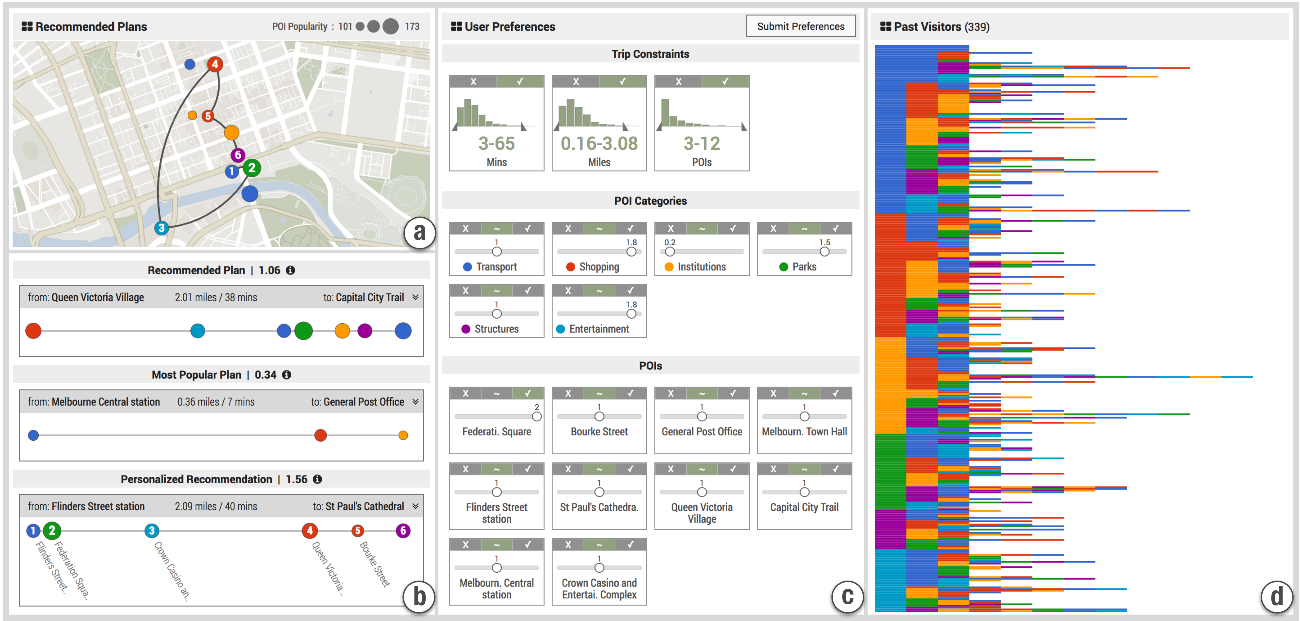

Figure 1: SeRIES combines machine learning with an interactive interface to explain and personalize recommendations. Continued on the left.

resents a point of interest (POI), where the size encodes the popularity of the place and the color indicates the type of the place. Three types of recommendations are provided, including (1) the machine-recommended plan with optimal expected experience based on all archived trajectories, (2) the most popular plan (by frequency) found in archived trajectories, and (3) the personalized recommendation.

User preference view (center): Personalization controls are displayed in three groups: trip constraints, POI categories, and specific POIs. Trip constraints are defined based on the duration, distance, and number of POIs of each archived trajectory. POI categories are automatically extracted from the archived trajectories (e.g., shopping or parks). Each control is represented in a rectangle showing its name and contextual information (i.e., the distribution of all archived trajectories), along with controls for tolerance range and weight.

Overview of archived trajectories (right): Archived trajectories are sorted by the type of POIs at each step and displayed in a compact list to provide an overview. Each row represents a trajectory and each column is a step. Zooming and panning interactions are provided for users to explore individuals or a group.

Other configurations: Simpler configurations may be preferred by intermittent users. Our prototype allows applica-

Study Task

"Imagine that you are visiting Melbourne. You will be asked to use two different user interfaces to make a plan for your one-day trip. We encourage you to really care about your trip and there is no time limit. Data from previous travellers will be used to generate recommendations."

Hypotheses

H1: Users will be more likely to follow the recommendation when using *Complex* than *Simple*.

H2: Users will be more confident in the recommended plan's experience when using *Complex* than *Simple*.

H3: Users will spend more time and perform more refinements when using *Complex* than *Simple*.

H4: Users will give higher ratings for ease of learning and ease of use for *Simple* than *Complex*.

tion designers to configure the visibility of the interface components to provide different levels of controls and details. In the user study, we created a *Simple* version that provides no user controls or details for the results, emulating a black-box interface. Only the "recommended plan" is shown for users to review (Figure 1a,b).

Recommendation Algorithm

Markov decision processes (MDPs) are widely used in applications for solving sequential decision problems (e.g., navigating a robot). Our implementation was based on the model introduced by Theocharous et al. [6]. We briefly describe the model and introduce our extensions for supporting user interactions.

Sequence modeling: The sequences were modeled using a probabilistic suffix tree (PST), which takes into account a visitor's path so far to suggest the next location. Each node in a PST encodes a frequent suffix $X = (s_1, s_2...s_t)$ and is associated with a probability distribution of the next places $P(s_{t+1}|X)$. The PST also compresses the input event sequences to accelerate computation.

Markov decision process: An MDP model can be computed directly from the PST where the states are nodes of the PST and the state transition probability is derived from the longest paths in the PST.

Thompson sampling: The last step is to find the optimal policies. Our implementation uses Thompson Sampling [8] to choose actions in real time to maximize the expected experience, as calculated by the "reward" on each state (provided in the dataset).

Supporting user interactions: In the original implementation [6], the reward function for computing the expected experience is defined as $r(x) = r(x_n)$, where x is a suffix available in the tree and x_n is the last symbol of x. Our system extends this definition by introducing a weighting factor w to represent users' preferences: $r(x) = w(x_n) \cdot r(x_n)$. By modifying only the reward function instead of the underlying models, we are able to provide faster feedback to the user based on the changes they make.

Evaluation

We conducted a within-subjects controlled study to compare the two interface configurations (*Simple* and *Complex*) to investigate the effects of controls and details on users' confidence and engagement in using the system.

The participants were 12 university students (10 males; 10 aged 25-34, and 1 each aged 18-24 and 35-44). All enjoyed travelling but none had visited Melbourne. No participant had prior experience with SeRIES. Each participant received a $10 gift card. The study was performed on a laptop computer with a 15.4-inch display.

Dataset

We used the YFCC100M dataset [7], which contains photos and videos from Yahoo! Flickr including meta information such as the time and location of the media. We extracted location sequences and narrowed the dataset to the 10 most popular POIs. After preprocessing and removing loops, we had 1,399 user trajectories and 10 POIs. Each trajectory on average consisted of 5 locations.

Procedure

Each session lasted about 60 minutes, including 5 minutes for general training and the study task overview (left). For each treatment, the participants were shown a brief tutorial (5 minutes) covering the interface components and operations. Participants used the interface to plan their trip and when satisfied with the recommendation, they clicked a "finish" button. They were encouraged to think aloud.

Q1: How easy was it to learn the interface? *(1=very difficult, 7=very easy)*

Q2: How easy was it to use the interface? *(1=very difficult, 7=very easy)*

Q3: Do you agree that the interface informed you about how the recommendations were made? *(1=strongly disagree, 7=strongly agree)*

Q4: How confident are you that you will follow the recommended plan in your trip? *(1=not confident at all, 7=very confident)*

Q5: How confident are you that the recommended plan will provide a good experience? *(1=not confident at all, 7=very confident)*

Table 1: Questions in the user satisfaction questionnaire using a 7-point Likert scale.

After each treatment, participants then completed a questionnaire using a 7-point Likert scale (Table 1). Interface order was counterbalanced and participants were allowed to adjust the ratings they gave for the previous version. The study system recorded task completion times and numbers of result refinements. After using both versions, participants were debriefed to collect feedback.

Results

We used Wilcoxon test to compare questionnaire ratings and T-tests for task completion times and numbers of result refinements, with a significance level of 0.01.

Questionnaire: As reported in Figure 2a, *Simple* (M=7.00 in Q1 and Q2) was rated easier to learn and easier to use than *Complex* (M=5.92 in Q1 and M=6.00 in Q2), supporting **H4**. In Q3, the participants felt more informed about how the recommendations were made when using *Complex* (M=6.17) than *Simple* (M=1.58). The ratings in Q4 showed that the participants were more confident to follow the recommendation when using *Complex* (M=5.42) than *Simple* (M=3.75), which supported **H1**. In particular, all participants preferred to follow the personalized recommendation compared to the generic recommended plan and the most popular plan. In Q5, *Complex* had a higher confidence rating for the trip experience (M=5.50) than *Simple* (M=3.50), supporting **H2**. The differences between the ratings of all questions were significant.

Completion time: On average, the participants spent 2.80 (SD=1.61) minutes on the *Simple* version and 10.38 (SD=3.86) minutes on *Complex* (Figure 2b), which was a significant increase of 108%, supporting **H3**.

Result refinement: On average, the participants made 20 result refinements (SD=9.22) using the *Complex* version (Figure 2c). No refinements were made using the *Simple* version, as it was not possible. In every case, the personalized recommendation differed from the original recommendation, indicating that personalization was necessary for and welcome by users.

Preference and Feedback

Transparency: When using *Simple*, all participants felt uninformed about how the recommendations were generated (Q3). For example, one said *"I have no idea how the plan was generated. The plan looks random to me"* and another asked *"is the plan created by hand?"* When using *Complex*, 11 out of 12 participants were able to identify the key factors considered by the recommendation algorithm. One participant explained, *"I understand that the recommendation is based on the popularity of different places, my preferences, and the trajectories of other visitors."* Another added that *"I do not have the knowledge to judge the algorithms, but knowing the high-level rules and logic is enough."* One participant had difficulty understanding the recommendation said he felt overwhelmed by the visualizations, in particular, the trajectory overview (Figure 1d).

Confidence: Overall, the participants lacked confidence when using *Simple* (Q4 and Q5). One explained, *"I am not sure how the plan was created. It would be a bad idea to follow a random plan."* Several emphasized the needs of personalization: *"I want to visit the Cathedral but it is not in the plan," "I want to avoid the Casino,"* or *"I want a shorter trip."* Participants appreciated the controls provided by *Complex*: *"The personalized plan is closer to my preferences. Most of my needs are satisfied"* and *"This Complex interface is more like my helper because it follows my controls. Simple was bossy."* Additional features were requested by the participants, including manually reordering the places, fixing the start and end locations, and integrating the controls into the map.

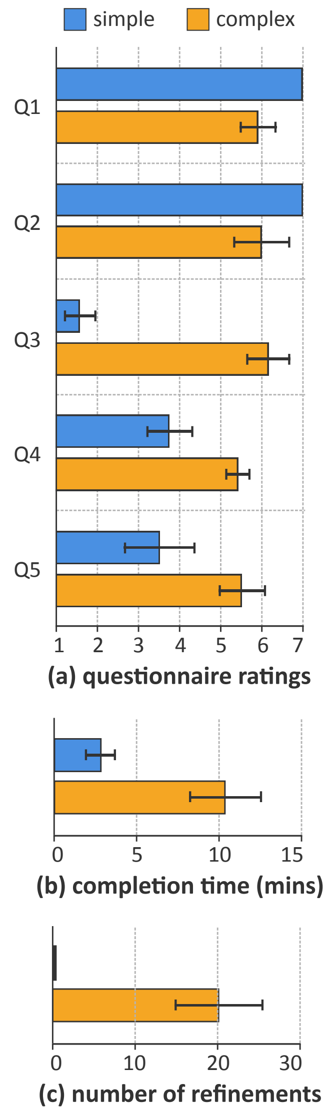

Figure 2: Study results (error bars show 95% confidence intervals).

Ease of learning and use: All participants agreed the *Simple* interface was very easy to learn and use but they felt disappointed due to the lack of controls and transparency. One commented *"I tried to click around but nothing happened"* and another added that *"I need more information to decide if this is a good recommendation."* All of them said they preferred to use *Complex* in real life. One explained: *"It takes some time to learn the interface but it is useful once you become familiar with it."* He suggested creating a moderate version with simpler visualizations for beginners.

User strategies: Most participants only briefly explored *Simple*. When using *Complex*, the participants typically tried different controls and carefully inspected the results. Some started by setting their preferences and then reviewed the personalized recommendation to see if it satisfied their needs. Others were less clear about their needs at the beginning and began exploring the generic recommended plan and archived trajectories, identifying places they wanted to visit or avoid, and then using the preference control. At the end, many compared their personalized recommendation against the generic recommended plan.

Conclusion and Future Work

This paper introduced an interactive sequence recommender system prototype that provides visualizations and user controls. Our user study indicated that the visualizations and controls were capable of informing users about reasons behind a recommendation, increasing users' confidence in following the recommendations, and engaging users in the decision making process. In future studies, we will explore alternative designs to improve the usability of the interface and generalize the prototype to other application domains such as recommending medical treatments, students' academic plans, and marketing strategies.

REFERENCES

1. B. Alsallakh, A. Hanbury, H. Hauser, S. Miksch, and A. Rauber. 2014. Visual methods for analyzing probabilistic classification data. *IEEE Trans. Vis. Comput. Graph.* 20, 12 (2014), 1703–1712.
2. S. Bostandjiev, J. O'Donovan, and T. Höllerer. 2012. TasteWeights: A visual interactive hybrid recommender system. In *Proc. of the ACM Conf. on Rec. Sys.* 35–42.
3. F. Du, C. Plaisant, N. Spring, and B. Shneiderman. 2017. Finding similar people to guide life choices: Challenge, design, and evaluation. In *Proc. of SIGCHI Conf. on Human Factors in Comp. Sys.* 5498–5509.
4. J. Krause, A. Perer, and K. Ng. 2016. Interacting with predictions: Visual inspection of black-box machine learning models. In *Proc. of SIGCHI Conf. on Human Factors in Comp. Sys.* 5686–5697.
5. S. McGregor, H. Buckingham, T.G. Dietterich, R. Houtman, C. Montgomery, and R. Metoyer. 2016. Interactive visualization for testing Markov Decision Processes: MDPvis. *J of Vis. Lang. & Comp.* (2016).
6. G. Theocharous, N. Vlassis, and Z. Wen. 2017. An interactive points of interest guidance system. In *Proc. Int. Conf. Intell. User Interfaces Companion*. 49–52.
7. B. Thomee, D.A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L. Li. 2016. YFCC100M: The new data in multimedia research. *Commun. ACM* 59, 2 (2016), 64–73.
8. W.R. Thompson. 1933. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika* 25, 3/4 (1933), 285–294.