

# Sequential Recommendations in IoT scenarios with a Generalized User Behaviour Model

David Massimo and Francesco Ricci

Free University of Bozen-Bolzano, Bolzano, Italy  
damassimo@inf.unibz.it, fricci@unibz.it

**Abstract.** Recommender Systems (RSs) are typically used to support users in finding web content of their interest. We consider here an alternative scenario: to support human decision making in the physical world. In particular, we focus on Internet of Things (IoT), where, for instance, users exploration of a sensor enabled city can be tracked and the knowledge of their choices (visit to points of interest, POIs) can be used to generate recommendations for not yet visited POIs. We leverage two distinct components: a generalised user behavioural model and a complementary recommender system; here recommendations can deviate from the usual approach to directly use the learned behaviour model and suggest the most likely actions the user will take next. We also propose techniques for simulating user behaviour and analysing the collective dynamics of a population of users. Moreover, we tackle the lack of data produced by interactions of users with IoT augmented areas, by designing a simulator that can be used to collect user preferences and monitor their decisions.

## 1 Introduction

RSs are techniques and tools that support human decision making by identifying items or actions relevant to a user [10]. Since users' preferences and behaviour may also be influenced by contextual factors, such as, the items they previously experienced, context-aware RSs have been introduced [1]. Moreover, in order to leverage the knowledge derived from the order in which the user has consumed items, pattern-discovery [6, 2, 9] and reinforcement learning [11, 7] approaches have also been proposed. The first approach extracts common patterns from users' behaviour logs and learns a predictive model of the next user action. The latter generates recommendations by using the optimal choice model (policy) of the user. Their common feature is to recommend items predicted by the user learnt behaviour, i.e., they suggest the user's predicted next choice. Moreover, while the first approach can only suggest items that have been already observed (new item problem), the second assumes that the system knows the utility the

user receives from actions, even though, in practice, users rarely provide explicit feedback (e.g., ratings).

We believe that these solutions tend to generate recommendations that lack novelty and therefore tend to be not interesting for the user (especially the first one, since it suffers from the cold-start problem). We therefore propose a recommendation technique that leverages a more general and explainable user behavioural model, which is learnt by observing users that are “similar” to the target one, and decouples behaviour learning from recommendations generation, so that, recommendations can deviate from the strategy to recommend the most likely next action of the user. We see this as a prerequisite to design novel and more compelling recommendation strategies to the user. In order to learn the generalised user behaviour (choice policy) we exploit the observation of the choices of a population that acts in physical areas. Observations are acquired via Internet of Things (IoT) technologies, i.e., sensors networks that can register users’ activities (e.g., movements) and environment conditions (e.g., weather) [3]. We also propose to reuse the predicted behavioural model to simulate the collective behaviour of a population with alternative environment conditions. Moreover, since testing IoT solutions in real life can be expensive (development and infrastructure cost) and suitable datasets for RS research in IoT augmented areas are scarcely available, we propose a simulation environment for generating realistic data logs of user/object interactions.

## 2 Research Progress Up To Date

In order to learn the generalised user behaviour, we have used Inverse Reinforcement Learning (IRL)[8], which has been already successfully applied to economics and robotics to learn an observed agent behaviour, but not yet to RSs. IRL, compared with other techniques for learning/predicting action sequences, such as, recurrent neural networks, has two distinguish aspects. Firstly, it learns to which extent items and context features are “liked” by a user (contribute to the acquired reward), whereas the other mentioned techniques do not. Secondly, it is able to learn even from small samples.

IRL is used by us to learn the sequential decision making behaviour of a group of users by leveraging observations of their actions sequences, i.e., without knowing the users’ reward (e.g. rating feedback). For instance, in the POI visit scenario the observations are the temporally ordered attraction-visit actions. Moreover, IRL assumes that the user’s reward (utility) for a state (e.g., visiting a POI) is a linear function  $r = \vartheta^T \cdot \varphi$  (reward function), where  $\varphi$  is the state feature vector and  $\vartheta$  is the (unknown) user preference vector for the state features. State features can describe the location or the visit context of the user, e.g., the crowdedness level of the POI. From the reward function  $r$ , by assuming that a user chooses actions that maximise her reward, it can be derived the user (optimal) action-selection policy, i.e., the policy that given the current user’s state tells her to act so that her expected utility is maximized.

IRL typically learns the user reward function from a set of observed user’s actions sequences. Since in many cases user observed actions are few, and noisy, a better model of the user behaviour can be learned by grouping (e.g., clustering) similar users (e.g., the users sharing the same visit goals) and then by learning user behaviour from grouped observations (generalised user behaviour of the group).

We have implemented this approach to behaviour learning in two case studies in the tourism domain: the first one in an indoor environment (i.e., a museum) [4] and the second one in an outdoor environment, i.e., a tourist area (under development). Our initial results show that IRL is capable of learning a group of users’ reward for actions and their action-selection policy even from small and noisy datasets.

After the generalised user behaviour model of a group (group action-selection policy) is learnt, action recommendations for a user can be generated by considering the observed sequence of her actions, as follows. If there are just a few observations (e.g., the sequence of actions performed by the user so far), then the generalised user behaviour (predictive model) of the group the user belongs to is used to suggest the optimal action that this user should do after the last visited POI. Conversely, in case there are more action observations for a user, recommendations can be generated by aggregating the group generalised user behaviour with an individual user preference model (built from the user observations). In practice, the two actions’ rankings derived from the generalised user behaviour and the user preference model, are aggregated and the suggested action is the top one in this ranking. For instance, if a commuter is understood to like arts because he mostly visited museums (individual user preference model) and a close-by exhibition is estimated to be an optimal choice for group members, then the system could recommend it to her.

We are currently studying an approach to group users’ actions sequences according to a common “semantic” structure that motivates the resulting bundles. Grouping is done by first representing action sequences according their state features and then by performing clustering (i.e., topic modelling). Groups formed by following the described approach are matching distinguished alternative tourist types and their differences can be further illustrated by the diverse reward functions that are learnt for the groups.

The proposed techniques are going to be implemented in a mobile application that exploits an IoT infrastructure to support tourists’ decision making by enabling them to discover new places or shops in an Italian alpine valley.

As mentioned above, since testing of IoT solutions is expensive and suitable datasets are missing, in order to bootstrap the application, we have developed a simulation tool that allows individuals or groups to experience a simulated itinerary and visit to POIs [5]. Collected data will be used to initially assess the proposed behaviour learning and recommendation approaches. Evaluations will be performed in both on-line (e.g., A/B testing, questionnaires) and off-line (e.g., algorithms performance) settings.

Finally, we intend to use the learnt behavioural model in order to generate synthetic action sequences of a population under conditions of interest (e.g., the opening of a new road or of a new mall). In this way the global system dynamics and the recommender performance can be evaluated even in novel IoT configurations, further reducing the testing costs.

## Acknowledgments

The research described in this paper is supported by the project Suggesto Market Space, funded by the Autonomous Province of Trento with Ectrl Solutions and Fondazione Bruno Kessler.

## References

1. Adomavicius, G., Tuzhilin, A.: Context-Aware Recommender Systems, pp. 217–253. Springer US, Boston, MA (2011)
2. Jannach, D., Kamehkhosh, L., Lerche, L.: Leveraging multi-dimensional user models for personalized next-track music recommendation. In: Proceedings of the Symposium on Applied Computing. pp. 1635–1642. SAC '17, ACM, New York, NY, USA (2017)
3. Li, S., Xu, L.D., Zhao, S.: The internet of things: a survey. *Information Systems Frontiers* **17**(2), 243–259 (2015)
4. Massimo, D., Elahi, M., Ricci, F.: Learning User Preferences by Observing User-Items Interactions in an IoT Augmented Space. In: 2017 Conf. on User Modelling, Adaptation and Personalization. ACM (2017)
5. Massimo, D., Not, E., Ricci, F.: User behaviour analysis in a simulated iot augmented space. In: Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion. IUI '18 Companion, ACM, New York, NY, USA (2018)
6. Mobasher, B., H. Dao, T. Luo, Nakagawa, M.: Using Sequential and Non-Sequential Patterns in Predictive Web Usage Mining Tasks. 2002 IEEE International Conference on Data Mining, 2002. Proceedings. pp. 669–672 (2002)
7. Moling, O., Baltrunas, L., Ricci, F.: Optimal radio channel recommendations with explicit and implicit feedback. Proceedings of the 6th ACM conference on Recommender systems - RecSys '12 p. 75 (2012)
8. Ng, A.Y., Russell, S.: Algorithms for inverse reinforcement learning. In: 17th Int. Conf. on Machine Learning. pp. 663–670. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2000)
9. Palumbo, E., Rizzo, G., Baralis, E.: Predicting Your Next Stop-over from Location-based Social Network Data with Recurrent Neural Networks. In: RECSYS 2017, 2nd ACM International Workshop on Recommenders in Tourism (RecTour'17), CEUR Proceedings Vol. 1906, August 27-31, 2017, Como, Italy,. pp. 1–8 (2017)
10. Ricci, F., Rokach, L., Shapira, B.: Recommender Systems: Introduction and Challenges, pp. 1–34. Springer US, Boston, MA (2015)
11. Shani, G., Heckerman, D., Brafman, R.I.: An mdp-based recommender system. *Journal of Machine Learning Research* pp. 1265–1295 (2005)