The Coaching Scenario: Recommender Systems with a Long Term Goal. A Case Study in Changing Dietary Habits

Doctoral Consortium

Jules Vandeputte
UMR MIA-Paris, AgroParisTech, INRAe, Université Paris-Saclay
Paris, France
jules.vandeputte@inrae.fr

ABSTRACT

This PhD work explores the way automated recommender systems can be built to help users develop healthier food consumption habits. The main focus is on developing methods of recommendation that allow long-term modifications in user's consumption habits and lasting changes in user's behaviour. We proposed a recommendation scenario were the user and the recommender can be seen as two agents interacting with each other. We also proposed a reinforcement learning formalism of the recommendation problem faced by the recommender system, as well as a choice criterion for recommendation and heuristics derived from that optimal criterion to guide recommendation.

KEYWORDS

Reinforcement Learning; Long-term Recommendation.

ACM Reference Format:

Jules Vandeputte. 2022. The Coaching Scenario: Recommender Systems with a Long Term Goal. A Case Study in Changing Dietary Habits: Doctoral Consortium. In *Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems : Doctoral Consortium (AAMAS 2022), Online, May 9–13, 2022*, IFAAMAS, 3 pages.

1 INTRODUCTION

Nowadays, humans choices are increasingly made in interaction with a machine or an algorithm. Indeed, as the quantity of information grows, and so the number of possible choices, it becomes more difficult for humans to make choices. Looking at the huge amount of products provided on *amazon.com*, or videos on *YouTube*, one can feel completely lost when looking for the best decision to take. The way this immense amount of data makes the choice barely impossible for a human being is called *information overload.* In order to overcome this problem, filtering systems were introduced, referred to as recommender systems. These are used to provide their users a short-list of items that are the most relevant for the user at the given time. There exists a wide variety of such recommender systems algorithms, that can be applied in numerous domains.

The PhD summarized here aims at using the tools of automated recommendation to help a user to modify his/her consumption habits towards a goal he/she pursues. The objective of this PhD is to explore the way to build a recommender system leading efficiently a user to modify his/her behaviour toward a given goal. In order to do so, we propose a model of the user-system interaction, we explore the space of recommendation policies, and the characteristics that lead to efficient recommendation. We also aim at investigating the effect of the specified goal on the quality of the recommendation, and the efficiency of learning recommendation policies. We study this problem in the particular context of dietary food recommendation.

First we explore how to model the interaction between the user and the recommender system. As we focus on helping a user to modify his/her behaviour, we investigate the effect of recommender systems on user behaviour. It appears that these effects are of two different kinds. On one hand, the immediate effects of recommendation. When one is given a recommendation by a system, this may impact his/her behaviour : if the recommended item is chosen, the recommender has lead to a one-off behaviour modification from the user. On the other hand, when interacting on the long term with a recommender system, the user's habits and even preferences may shift to new-ones [4]. This shift can be, if not induced, influenced by the items the system has showed to the user, by recommending them. One kind of long term modification that has been measured is known as *filter bubbles*, where the recommender induces a narrowing in diversity of user chosen items [7]. In this case the behaviour change is seen as an adverse effect of the recommendation. However throughout this work, we consider the user behaviour change as the goal of the recommendation as evoked in [6]. We call this recommendation task *coaching*, as it aims to help a user adopt a new behaviour by recommending slight changes in the user choices similar to a tennis coach giving his/her student recommendations to improve his/her game, rather than complete menus or dietary plans like existing healthy food recommender systems ([10], [1]).

2 FORMALIZATION

To study coaching, we provide a model of the user-coach interaction based on a long term, repeated interaction, to maximise the long term effects of recommendation. We model the interaction as an iterative two players game, between a user agent and a coach agent.

Model of the user U. We characterize a user by three components.

(1) A probability distribution over the set of items that may change over time: Π_t. This expresses the *preferences* of the user, and dictates the behavior of U.
(2) A matrix $\mathbf{M}_t$: $\mathcal{I} \times \mathcal{I} \to [0, 1]$ of which each element $m^t_{i,j}$ expresses the probability that U *accepts a suggestion* of change from an item $i \in \mathcal{I}$ to an item $j \in \mathcal{I}$ at time t, and can be called the substitutability from item i to item j [2]. If the suggestion is not accepted, U stays with his/her choice.

Proc. of the 21st International Conference on Autonomous Agents and Multiagent Systems : Doctoral Consortium (AAMAS 2022), P. Faliszewski, V. Mascardi, C. Pelachaud, M.E. Taylor (eds.), May 9–13, 2022, Online.

(3) A *propensity to modify* Π_t when a suggestion of change $(i \to j)$ by the coach C has been accepted by U. In our model, we consider changes of preferences of the form:

$$\begin{cases} \pi_{t+1}(i) &= (1-\lambda)\,\pi_t(i) \\ \pi_{t+1}(j) &= \pi_t(j) + \lambda\,\pi_t(i) \end{cases} \quad (1)$$

where $\lambda \in [0, 1]$. This formula guarantees that, if Π_t is a probability distribution, then so is Π_{t+1}. One can consider λ as a parameter that controls or characterizes the learning rate of the user. A value $\lambda = 0$ means that no learning takes place, while the closer to 1 the value of λ, the larger the effect of accepting a recommendation $i \to j$ of C.

Model of the coach C. The coach C examines the choice i made by U at time t and computes the expected gain $G(i, j)$ of each possible substitution $i \to j$ (including "no substitution": $i \to i$). C chooses the recommendation $i \to c_t(i)$ according to:

$$c_t(i) = \operatorname*{ArgMax}_{j \in \mathcal{I}} G(i, j)$$

The expected gain $G(i, j)$ depends on the respective quality $s(i)$ and $s(j)$ of i and j (e.g., their nutritional quality) and on other parameters depending on the recommending strategy used by C.

The coach C may maintain an estimate $\widehat{\mathbf{M}}_t$ of the acceptability matrix $\mathbf{M}_t$ of the user, an estimate $\widehat{\Pi}_t$ of Π_t and an estimate of λ, the learning rate that characterizes U.

Accordingly, we formulate the coaching scenario as an *iterated two-player game* between the user U and the coach C.

(1) U makes an item proposal, for example i, using his/her vector of preferences Π_t.
(2) C analyzes U's proposal, and suggests, if judged useful, a modified proposal j, using his/her knowledge of the value of the items through the score function $s(\cdot)$, and his/her estimation of U's ability to accept the proposal.
(3) U accepts or rejects the substitution proposal provided by C.
  - If U accepts C's proposal (replacing item i by item j), he/she modifies the preference vector Π_t according to his/her learning capacity, so as to propose the recommended item more frequently in the future.
  - Otherwise, U does not modify the preference vector.

We explore the question of recommendation's efficiency under the objective of behaviour change. To do so, it is necessary to define the efficiency of a recommendation, regarding the given goal of coaching. First we show that coaching, unlike classical recommendation tasks, leads to a compromise for the recommender agent between acceptability, and quality of the recommendation given the pursued goal. As in some health-aware recommender systems, like [5], the coach agent needs to take into account both information. This shows the importance of measuring the recommendation's value under the pursued goal. To do so, we introduce the concept of score, that, given a user agent and its current habits, measures the quality of these habits considering the goal. In the case of dietary food recommendation, one can find numerous nutritional scores that measure the dietary quality of eating habits ([8], [3]). Given a score, we propose a performance metric of the coaching, based on the idea of budget : under a given limited budget of interaction (ie : the number of times the coach agent can make a recommendation to the user agent) the coach agent has to maximize the score gain of the user agent. This was inspired by performance metrics found in teacher-student literature [11]. We propose a formalization of the coach recommendation task as a kind of reinforcement learning problem as defined in [9], with a environment that changes over time, where the coaching agent is rewarded proportionally to the actual score gain permitted by the recommendation.

3 FIRST RESULTS

Once we defined a performance metric, is therefore possible to compare different recommendation policies. Using the reinforcement learning formalism we proposed, we extract thanks to Bellman equations an optimal criterion for recommendation. From this theoretical criterion, we derive different recommendation policies, with their own characteristics. In particular, we can derive adaptive policies, whose recommendation will depend on the user as well as non-adaptive ones. But also myopic and non-myopic strategies. Non-myopic policies will take into account not only the immediate gain of score, but also the future possible gains. To test the importance of such characteristics, we then experimentally test these policies on our user model using a real-world food behaviour data-set. Results show that it is worth it to adapt to user characteristics and use non-myopic strategies,which aim for long-term gains, when the number of interactions becomes large.

4 PERSPECTIVES

After investigating the effect of the policies characteristics on the efficiency of recommendation, we are currently working on the effects of the way the coaching goal is specified. We showed that an interesting way to specify the goal is to use a score function. But one can consider other way to specify it, and even considering only score functions, their diversity leads to questions about the effect of their characteristics on recommendation policies and their relative efficiency. In the food domain, it exists various dietary scores with diverse characteristics. Some depend only on the items, and do not take into account any notion of context [8]. On the other hand, it exists scores that take context into account [3]. This context can be either an immediate context, or a temporal one, considering the consumption history of the user. In this PhD we aim at investigating the effect of such characteristics on recommendation policies. Particularly, we are interested in considering scores with a temporal context. This kind of scores seems to induce changes in the way we can formalize the coaching problem as a reinforcement learning problem. If confirmed, this would lead to a different optimal recommendation criterion and then to new policies, able to handle the temporal nature of context. In either case, we strongly believe that the study of recommendation policies and their relative performance under a score considering a temporal context is an interesting research question, as in the food domain such score exists. The work we provide in this PhD on the recommender systems we call coaching systems is strongly linked, in our approach and experiments, to the dietary recommendation task. However, the scope of the proposed method goes far beyond this use case. Indeed, the formalism proposed for coaching systems is general, and this PhD also aims at showing how coaching can be useful for other domains.

REFERENCES

[1] Agapito, G., Simeoni, M., Calabrese, B., Caré, I., Lamprinoudi, T., Guzzi, P.H., Pujia, A., Fuiano, G., Cannataro, M.: Dietos: A dietary recommender system for chronic diseases monitoring and management. Computer methods and programs in biomedicine **153**, 93–104 (2018)

[2] Akkoyunlu, S., Manfredotti, C., Cornuéjols, A., Darcel, N., Delaere, F.: Investigating substitutability of food items in consumption data. In: Second International Workshop on Health Recommender Systems co-located with ACM RecSys. vol. 5 (2017)

[3] Chaltiel, D., Adjibade, M., Deschamps, V., Touvier, M., Hercberg, S., Julia, C., Kesse-Guyot, E.: Programme national nutrition santé–guidelines score 2 (pnns-gs2): development and validation of a diet quality score reflecting the 2017 french dietary guidelines. British Journal of Nutrition **122**(3), 331–342 (2019)

[4] Chaney, A.J., Stewart, B.M., Engelhardt, B.E.: How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In: Proceedings of the 12th ACM Conference on Recommender Systems. pp. 224–232 (2018)

[5] Ge, M., Ricci, F., Massimo, D.: Health-aware food recommender system. In: Proceedings of the 9th ACM Conference on Recommender Systems. pp. 333–334 (2015)

[6] Liang, Y.: Recommender system for developing new preferences and goals. In: Proceedings of the 13th ACM Conference on Recommender Systems. pp. 611–615 (2019)

[7] Nguyen, T.T., Hui, P.M., Harper, F.M., Terveen, L., Konstan, J.A.: Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd international conference on World wide web. pp. 677–686 (2014)

[8] Rayner, M.: Nutrient profiling for regulatory purposes. Proceedings of the Nutrition Society **76**(3), 230–236 (2017)

[9] Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)

[10] Toledo, R.Y., Alzahrani, A.A., Martinez, L.: A food recommender system considering nutritional information and user preferences. IEEE Access **7**, 96695–96711 (2019)

[11] Torrey, L., Taylor, M.: Teaching on a budget: Agents advising agents in reinforcement learning. In: Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems. pp. 1053–1060 (2013)