Adaptive Sequential Recommendation Using Context Trees

Fei Mi,   Boi Faltings

Artificial Intelligence Lab, EPFL

Lausanne, Switzerland

firstname.lastname@epfl.ch

Abstract

Machine learning is often used to acquire knowledge in domains that undergo frequent changes, such as networks, social media, or markets. These frequent changes pose a challenge to most machine learning methods as they have difficulty adapting. So my thesis topic focus on adaptive machine learning models. At the first step, we consider a forum content recommender system for massive open online courses (MOOCs) as an example of an application where recommendations have to adapt to new items and evolving user preferences. We formalize the recommendation problem as a sequence prediction problem and compare different recommendation methods, including a new method called context tree (CT). The results show that the CT recommender performs much better than other methods. We analyze the reasons for this and demonstrate that it is because of better adaptation to changes in the domain.

1 Introduction

With the increased availability of data, machine learning has become the method of choice for knowledge acquisition in intelligent systems. However, data and the knowledge derived from it have a timeliness, such that in a dynamic environment not all of the knowledge acquired in the past remains valid. Therefore, machine learning models should acquire new knowledge incrementally and adapt to the dynamic environments. Dynamic environments have been considered in artificial intelligence in the field of robotics. Today, many intelligent systems deal with dynamic environments: information on websites, social networks, and stock markets changes dynamically every day.

In such dynamically changing environments, knowledge also must adapt to the changes. Many statistical machine learning techniques interpolate between input data and thus their models can adapt only slowly to new situations. As an example, we consider recommender systems where user preferences, as well as the set of available items, change dynamically over time. However, techniques such as collaborative filtering build an increasingly complex model of users and items, and this model only evolves slowly. Thus,

when an item is superseded by a newer model, it takes time for the recommendation to adapt. Another perspective sees recommendation as a dynamic machine learning problem. Recommender systems use the set of items a user liked in the past as an indication of her preferences. For example, in item-based collaborative filtering, a user is recommended items based on the items she has consumed before. Thus, the recommendation can be understood as the task of predicting the next items in a sequence of items liked by that user. In this paper, we use algorithms for sequence prediction based on variable-order Markov models for recommendation. More specifically, we use the context tree method developed by [Willems *et al.*, 1995] and proposed for news recommendation by [Garcin *et al.*, 2013; 2014].

2 Related Work

Typical recommender systems adopt a static view of the recommendation process and treat it as a prediction problem over all historical preference data. We argue that it is more appropriate to view the problem of generating recommendations as a sequential decision problem.

The most well-known class of recommender system is based on collaborative filtering (CF). Several attempts have been made to incorporate temporal components into the collaborative filtering setting to model users' drifting preferences over time [Ding and Li, 2005; Koren, 2010; Yuan *et al.*, 2013]. Markov models are applied to recommender systems so that they can learn the transition functions over items. [Shani *et al.*, 2002] viewed the problem of generating recommendations as a sequential decision problem and they considered a finite mixture of Markov models with fixed weights.

3 Methodology

We focus on a class of recommender systems based on context trees [Willems *et al.*, 1995]. These trees are used to estimate variable-order Markov models (VMM) which were originally applied to lossless data compression and then applied to various discrete sequence prediction tasks. Recently, [Garcin *et al.*, 2013; 2014] successfully applied CTs to a recommender system and conducted online evaluations of different methods for news recommendation in a live environment.
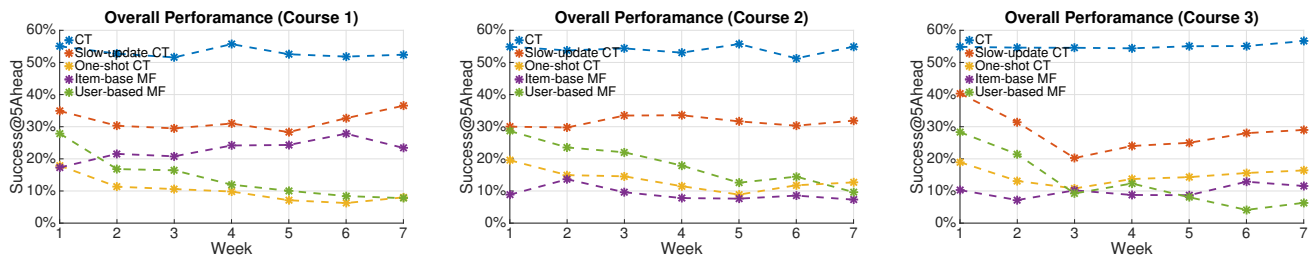
Figure 1: Overall performance comparison

Because of the sequential nature of item consumption in the application of recommender systems, the user's preference can be summarized by the last several items visited. When modeling the process as a k-order Markov process, it is not clear how to select the order k. A variable-order Markov model (VMM), like a context tree, alleviates this problem by using a context-dependent order. There are two key ideas behind applying CTs to recommender systems. First, the data structure is defined as a hierarchy of contexts, arranged in a tree such that a child node completely contains the context of its parents. Second, a local prediction model, called an expert, is assigned to to each context. For instance, a particular expert only gives predictions for users who have consumed a particular sequence of items.

The algorithm is truly online so that it adapts continuously to a dynamic environment. This class of algorithm can be applied elegantly to domains where adaptation and timeliness are of concern, mainly due to two advantages:

- The structure of the CT helps new items or paths to be recognized in new contexts, whereas old items can still be accessed in their contexts.

- The model parameter learning and recommendations generated are online.

4 Results

We mainly compared the proposed CT recommender against the traditional matrix factorization (MF) recommender, including user-base MF and item-based MF. As MF is not suitable for online recommendation due to the expensive model update, MF model is only updated periodically (week-by-week). To enable a fair comparison with matrix factorization techniques, we implement versions where the CT model is updated at fixed time intervals, equal to those of the MF model. In the "One-shot CT" version, we compute the CT recommendations for each user based on the data available at the time of the update, and the user then receives these same recommendations at every time step until the next update. This mirrors the conditions of user-based MF. To compare with item-based MF, in the "Slow-update CT" version we update the recommendations, but not the model, at each time point based on the information available at that time. Furthermore, we also consider the full CT algorithm with a continuously updated model.

The overall performances illustrated in Figure 1, for three MOOCs, show that CT recommender performs better than MF. We found, through empirical results, that what makes

the CT work better is that it adapts (recommends more fresh items) and furthermore that these recommendations for fresh items are precise. This finding shows the great potential of addressing adaptation of machine learning models while boosting performance.

5 Conclusion and Future Work

Till the current stage, we studied the adaptation of machine learning models from a sequences recommendation perspective. We focused on the application of recommender system and achieved both performance improvement and adaptation to new items using a new method called context tree. As a next step in the development of the current CT recommender, it would be highly interesting to design special structure to favor model adaptation or factorize the current fine-grained Markov chain in order to allow a mixture of different but similar paths, Besides, other applications of CTs and adaptation analysis are worth studying as well.

References

[Ding and Li, 2005] Yi Ding and Xue Li. Time weight collaborative filtering. In *CIKM*, pages 485–492. ACM, 2005.

[Garcin *et al.*, 2013] Florent Garcin, Christos Dimitrakakis, and Boi Faltings. Personalized news recommendation with context trees. In *RecSys*, pages 105–112. ACM, 2013.

[Garcin *et al.*, 2014] Florent Garcin, Boi Faltings, Olivier Donatsch, Ayar Alazzawi, Christophe Bruttin, and Amr Huber. Offline and online evaluation of news recommender systems at swissinfo.ch. In *RecSys*, pages 169–176. ACM, 2014.

[Koren, 2010] Yehuda Koren. Collaborative filtering with temporal dynamics. *Communications of the ACM*, 53(4):89–97, 2010.

[Shani *et al.*, 2002] Guy Shani, Ronen I Brafman, and David Heckerman. An MDP-based recommender system. In *UAI*, pages 453–460. Morgan Kaufmann Publishers Inc., 2002.

[Willems *et al.*, 1995] Frans MJ Willems, Yuri M Shtarkov, and Tjalling J Tjalkens. The context-tree weighting method: Basic properties. *IEEE Transactions on Information Theory*, 41(3):653–664, 1995.

[Yuan *et al.*, 2013] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *SIGIR*, pages 363–372. ACM, 2013.