



Managing Learners' Memory Strength in a POMDP-Based Learning Path Recommender System

Zhao Zhang^(✉), Armelle Brun^(✉), and Anne Boyer^(✉)

Université de Lorraine, CNRS Loria, Vandœuvre-lès-Nancy, France
{zhao.zhang, armelle.brun, anne.boyer}@loria.fr

Abstract. This paper views the learning path recommendation task as a sequential decision problem and considers Partially Observable Markov Decision Process (POMDP) as an adequate approach. This work proposes M-POMDP, a POMDP-based recommendation model that manages learners' memory strength, while limiting the increase in complexity and data required. M-POMDP has been evaluated on two real datasets.

Keywords: Learning path recommendation · POMDP · Memory strength

1 Introduction

Recommender systems (RS) in education help learners reach their learning goals, while keeping care of the recommendation adoption. The recommendation are sequences of resources, that maximise the probability the goal is reached. Such recommender systems are Learning Path Recommender Systems (LPRS). LPRS can be viewed as a sequential decision problem and approached by a Markov Decision Process (MDP). However, in the educational context some elements remain uncertain such as the learners' knowledge level or the motivation [2]. LPRS can thus be formulated as a POMDP. Although the learners' memory ability is an important factor, it is seldom considered in recommendation, generally at the cost of a high model complexity. We intend to manage it to promote the review of resources and foster long-term retention, with a limited complexity in a learning environment where no metadata about the resources is provided.

2 Related Work

Learning path recommender systems (LPRS) are designed to recommend a sequence of educational resources that contributes to reach a predefined goal. This goal can be the knowledge increase, minimization of learning time, etc. Associated models generally exploit the learners' past interactions with pedagogical resources.

Several approaches have been proposed to perform learning path recommendations, especially Markov-based algorithms, which are known to be good at dealing with this sequential problem. In the educational context, MDP and POMDP have shown to be relevant [2]. POMDPs compute a policy for selecting sequential actions when information may be unobserved. A POMDP consists of a tuple of at least 7 elements, among which the set of states, the set of actions, the observation probability, a transition and a reward function, beliefs, etc.

Memory is important in education and numerous studies in psychology have been interested in modeling human memory. They model how memory decays with time, through a forgetting curve or a half-life regression model [4]. [5] studies several forgetting curve models that incorporate human expertise: psychological and linguistic features, to predict the probability of word recall. One main limit of these works is their complexity and the large datasets they require.

3 Learners' Memory Strength in a POMDP-Based LPRS

We formulate LP recommendation as a POMDP when no content information about resources is available. An action is the act of accessing a resource. The set of actions is thus defined as the set of resources, as in [2]. A state s is defined by two simple attributes: s_{LP} represents the learner's history learning path and s_{KL} represents the estimated knowledge level of the learner. As resources are not indexed, we propose to represent the learner's knowledge for each resource in s_{LP} . To limit the complexity, we discretize the knowledge level [3]. Given that action a is taken in state s , the observation model $p(z|s, a)$ indicates the probability of observing z . The reward function is defined as $R(s, a) = r(s_{LP}, a) + r(s_{KL}, a)$. The transition function T models the possible effects of the actions on a state. $T(s, a, s') = P(s'|s, a) = P(s'_{LP}|s, a) \cdot P(s'_{KL}|s, a)$ manages both attributes of the state independently [3].

Given an learning path LP , two cases arise when estimating a learner's knowledge level for a resource. First, if the action points to an evaluation resource er (quiz, exam), the knowledge level $KL(LP, er)$ can be directly estimated from the grade obtained by the learner ($eval(LP, er)$). Second, if the action does not point to an evaluation resource, we assume that ($eval(LP, er)$) is an accurate indicator of knowledge level of all the resources that have been studied before er . So, the knowledge level of the current resource cr can be estimated from the evaluation resource that follows cr . We propose to apply a discount factor λ , the longer distant cr is from er , the lower the knowledge level for cr . We present the way we estimate the knowledge level on cr as: $KL(LP, cr) = round(\lambda^{dist(LP, cr, n_eval(LP, cr))} eval(LP, n_eval(LP, cr)))$ where $n_eval(LP, cr)$ is the next evaluation resource that follows cr in LP and $dist(LP, cr, n_eval(LP, cr))$ is the distance between cr and the corresponding next evaluation resource.

Managing Learners' Memory Strength. We propose Memory-based POMDP (M-POMDP) that manages the learners' memory strength to foster resource reviewing. M-POMDP is intended to limit the increase in complexity. M-POMDP stores learners' memory strength in the state, as an additional attribute,

under the form of a discretized attribute [4]. The corresponding attribute s_{NLT} is set as an array, where each element represents the number of times the corresponding resource in s_{LP} has been studied by the learner. It is used to evaluate the need of review of this resource. s_{NLT} is deterministically incremented each time the learner interacts with a resource. In line with the literature, when a learner has studied a resource MAX_{NLT} times, it does not need to be reviewed.

Reward Function. s_{NLT} is a supplement to s_{KL} and impacts the reward function. We propose to redefine the reward function as $R(s, a) = r(s_{LP}, a) + r(s_{KL}, a) + r(s_{NLT}, a)$ where $r(s_{NLT}, a)$ is the reward function that computes the reward based on s_{NLT} , defined as follows: if the NLT is increased, it gains a unit of reward u_{NLT} ; otherwise the reward is 0.

Transition Function. The transition function $T(s, a, s')$ is also impacted by s_{NLT} . It is evaluated by three independent sub-functions: $T(s, a, s') = P(s'_{LP}|s, a) \cdot P(s'_{KL}|s, a) \cdot P(s'_{NLT}|s, a)$. Since NLT is deterministic, $P(s'_{NLT}|a, s) = 1$, so the transition function remains unchanged.

This simple solution faces a limit: the time gap between two actions is not considered. Even if it could be simply stored as the last access date of each resource, this would be at the cost of a significant increase in the number of states due to the high number of possible values for this new attribute.

4 Experiments

Experiments are conducted on two real-world datasets: EOLE and the EdNet datasets. EOLE, described in [6], is a medium-sized dataset that contains 3,972 interactions from 104 learners on 39 resources. The median length of LP is 38 and the repetition rate is 0.30. About EdNet¹, it is a large dataset with a LP median length of 15 (twice lower than for EOLE) and a repetition rate of 0.22.

Evaluation Protocol. We propose to adopt a leaving one out cross validation. The interactions of each test learner in the review period are split into two. The first 50% form the elements which help to determine the initial state of the POMDP, the rest is used to evaluate the recommended LP. For the EdNet dataset, we select the last 50% of interactions of each learner as the test set.

Parameter Settings. The length of the history is set to the average length of the learning path in datasets ($N = 7$). The number of knowledge levels is set to $K = 4$ [3], $MAX_{NLT} = 3$ and $\lambda = 0.9$. The SARSOP solver [1] is used.

Evaluation Metrics. We use the well-known precision measure. To fit the sequential characteristics of our data, we redefine the “matched” resources from the upper part of the equation by the Longest Common Subsequence (*LCS*) between *RLP* and ground truth LP (*GTLP*). This updated precision is defined as $Precision = \frac{|LCS(RLP, GTLP)|}{|RLP|}$. Besides, we use precision of *SLLP* measure

¹ <https://github.com/riiid/ednet>.

(Similar Learners Learning Path) [6], noted $Prec_{SLLP}$. Based on [6], learners are split in three groups: Good (GL), Average (AL) and Promising (PL) Learners.

Table 1. Evaluation of POMDP and M-POMDP for EOLE and EdNet datasets

Dataset	Measures	POMDP				M-POMDP			
		Knowledge level				Knowledge level			
		ALL	GL	AL	PL	ALL	GL	AL	PL
EOLE	$Prec$	0.34	0.41	0.36	0.24	0.44	0.59	0.43	0.30
	$Prec_{SLLP}$	0.26	0.30	0.27	0.21	0.33	0.40	0.35	0.22
EdNet	$Prec$	0.20	0.14	0.31	0.15	0.29	0.27	0.38	0.23
	$Prec_{SLLP}$	0.07	0.1	0.11	0.02	0.10	0.09	0.15	0.05

Table 1 presents the values of $Prec$ and $Prec_{SLLP}$. Considering the baseline POMDP, $Prec$ decreases with the level of the group, which was expected. This confirms that POMDP recommends PL paths that are closer to those adopted by learners with a higher level, which is confirmed by $Prec_{SLLP}$. Considering M-POMDP on the entire set of learners $Prec$ and $Prec_{SLLP}$ are improved by similar rates and the quality of the recommendations if also increased for each group of learners. This confirms that the way M-POMDP manages learners' memory strength seems to be adequate.

We can see that the values on EdNet are lower than for EOLE, explained by the number of resources that is twice larger on EdNet; the average length of learners' learning path that is 3 times smaller than in EOLE. For M-POMDP, the increase on $Prec$ and $Prec_{SLLP}$ for the entire set of learners are similar. Considering each group of learners, $Prec$ is improved significantly for each group. $Prec_{SLLP}$ remains stable with M-POMDP.

From these experiments, we can conclude that the simple way M-POMDP to manages learners' memory strength is adequate and fits medium size datasets.

5 Conclusion and Perspectives

This work focused on the learning path recommendation task through POMDP. We have designed a model that manage learners' memory strength with a limited increase in complexity, validated experimentally. As a future work, we intend to incorporate additional information in the model, whether they are teacher expertise or from data.

References

1. Kurniawati, H., Hsu, D., Lee, W.S.: SARSOP: efficient point-based POMDP planning by approximating optimally reachable belief spaces. In: Robotics: Science and Systems (2008)

2. Rafferty, A.N., Brunskill, E., Griffiths, T.L., Shafto, P.: Faster teaching via POMDP planning. *Cogn. Sci.* **40**(6), 1290–1332 (2016)
3. Ramachandran, A., Sebo, S.S., Scassellati, B.: Personalized robot tutoring using the assistive tutor POMDP (at-POMDP). In: *AAAI Conference*, vol. 33, pp. 8050–8057 (2019)
4. Settles, B., Meeder, B.: A trainable spaced repetition model for language learning. In: *Proceedings of the 54th Annual Meeting of the ACL*, pp. 1848–1858 (2016)
5. Zaidi, A., Caines, A., Moore, R., Buttery, P., Rice, A.: Adaptive forgetting curves for spaced repetition language learning. In: *Proceedings of AIED Conference*, pp. 358–363 (2020)
6. Zhang, Z., Brun, A., Boyer, A.: New measures for offline evaluation of learning path recommenders. In: Alario-Hoyos, C., Rodríguez-Triana, M.J., Scheffel, M., Arnedillo-Sánchez, I., Dennerlein, S.M. (eds.) *EC-TEL 2020. LNCS*, vol. 12315, pp. 259–273. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-57717-9_19