



Adaptive Learning Material Recommendation in Online Language Education

Shuhan Wang¹(✉), Hao Wu², Ji Hun Kim¹, and Erik Andersen¹

¹ Department of Computer Science, Cornell University, Ithaca, USA
{sw788, jk2227, ela63}@cornell.edu

² Department of Computer Science, George Washington University, Washington, D.C., USA
fqq11679@gmail.com

Abstract. In online language education, it is challenging to recommend learning materials that match the student's knowledge since we typically lack information about the difficulty of materials and the abilities of each student. We propose a refined hierarchical structure to model vocabulary knowledge in a corpus and introduce an adaptive algorithm to recommend reading texts for online language learners. We evaluated our approach with a Japanese learning tool, finding that adding adaptivity into material recommendation significantly increased engagement.

1 Introduction

Engaging students with personalized content in online language learning presents two key challenges. First, we must prepare a corpus of learning materials that are organized by difficulty. Although we would like to utilize materials collected from the Internet, it is prohibitively expensive to ask experts to measure the difficulty of those materials. Second, we must assess each student's competency level and recommend content that is appropriate for that student. Most existing content recommender systems for language learning are designed for formal learning scenarios and make recommendations based on standardized pre-assessment results. However, these systems may not scale easily to informal learning scenarios such as online learning, where we usually do not have accurate and standardized information of a student's prior knowledge.

Existing assessment and recommendation systems [1, 3, 5] generally use unidimensional measurements for student ability and content difficulty, which is incomprehensive [2]. Ideally, a unified system could multidimensionally evaluate a student's ability and the relative difficulty of learning materials in order to prepare future lessons for that student, without requiring prior information from the student or significant expert labor. Moreover, previous work on multidimensional knowledge structuring for grammar knowledge uses strict constraints to specify the relative difficulty between two texts [8]. However, this does not

scale to teaching vocabulary with a large online corpus since these strict constraints yield too few edges in the structure. To this end, we propose the *fuzzy partial ordering graph*, a refined hierarchical knowledge structure with relaxed constraints, which significantly increases the density of the knowledge structure.

We also present a material recommender system for online language learning that incorporates adaptive knowledge assessment. It collects authentic and up-to-date learning materials from the Internet and organizes them with a fuzzy partial ordering graph. It also uses a probabilistic function to balance assessment and recommendation throughout the learning process in order to improve student engagement. We evaluated our approach through *JRec*, an online Japanese language learning tool that recommends appropriate reading texts from the Internet based on the student's prior knowledge. Our user study demonstrates that our adaptive recommendation system led users to read 62.5% more texts than a non-adaptive recommendation version. This suggests that our multidimensional assessment can improve engagement in material recommendation.

2 Approach

Fuzzy Partial Ordering Graphs. In order to multidimensionally assess a student's knowledge and make recommendation accordingly, we need to measure the difficulty of each learning material and organize the corpus into a hierarchical structure. In our model, a reading text t_1 is considered *fuzzily harder than* another text t_2 if t_1 covers a *majority* of vocabulary words in t_2 . This also implies that students who understand t_1 will also be able to understand t_2 . Based on this fuzzy partial ordering, we model the vocabulary knowledge within a corpus of texts using a *fuzzy partial ordering graph*, in which each node denotes a text, and a directed edge from t_1 to t_2 indicates t_1 is fuzzily harder than t_2 .

This model improves our previous work in hierarchical knowledge structures [8] by increasing the number of partial ordering edges within the structure (the density). This previous work was based on a strict partial ordering, meaning that there is an edge from t_1 to t_2 only if t_1 covers *all* knowledge in t_2 . This strict partial ordering works well for grammar learning but may not scale well to vocabulary, since it is not common in an authentic corpus that a text covers all vocabulary knowledge of another text. Consequently, the strict partial ordering yields a vocabulary-based knowledge structure that is too sparse. The fuzzy partial ordering, however, addresses this issue by increasing the number of edges in the vocabulary-based knowledge structure to make it dense enough for assessment and recommendation.

To avoid unacceptable loss of confidence in our fuzzy partial orderings, we conducted a series of case studies in our corpus of 4,269 Japanese texts. We selected the fuzzy parameter $\alpha = 0.8$, meaning that t_1 is fuzzily harder than t_2 if t_1 covers at least 80% of the vocabulary words in t_2 . The fuzzy partial ordering graph with $\alpha = 0.8$ has 71% more edges than the strict version.

Adaptive Learning Material Recommendation. Based on the fuzzy partial ordering graph, we seek to build a recommender system that carefully balances the trade-off between assessment and recommendation: in order for recommendations to be appropriate, the system needs to accurately assess each student; however, excessive assessment can potentially harm engagement because students might need to respond to too many problems that are far outside of their comfort zone. Our heuristics for assessment and recommendation are:

The Assessment Heuristic: Select the problem that maximizes the *expected* amount of information gained on the student’s prior knowledge. Formally, the assessment heuristic selects the problem s^* such that:

$$s^* = \arg \max_s [p_s n_s^+ + (1 - p_s) n_s^-] \quad (1)$$

where p_s indicates the probability that the student can solve s . If the student can solve s , n_s^+ represents how many problems we know that he/she can solve. Otherwise, if the student cannot solve s , n_s^- represents how many problems we know that he/she cannot solve. Both n_s^+ and n_s^- include s itself and exclude the problems we already know the student can/cannot solve before presenting s . The probability p_s can be estimated: $p_s = N^+ / (N^+ + N^-)$, where N^+ and N^- denote the number of presented problems that the student can/cannot solve.

The Recommendation Heuristic: Select the problem that is directly harder than some problem that the student can solve. This heuristic is based on Vygotsky’s Zone of Proximal Development (ZPD) theory [7].

Since we believe that students are more engaged while solving a problem relevant to their experience, if there are multiple problems satisfying this requirement, pick the one that is *most relevant* to the student prior knowledge. Practically, the relevance is measured as the number of edges from that problem’s node to any solvable problem’s node in the fuzzy partial ordering graph.

Balancing Assessment and Recommendation: Our system uses a probabilistic function to balance the assessment and recommendation heuristics. To select the next problem, our system chooses the assessment heuristic with probability $p = \#Prob/M$ and chooses the recommendation heuristic with probability $1 - p$. Here $\#Prob$ represents the number of the problems that the student has experienced, regardless of whether he/she has solved those problems. M is a pre-set parameter that controls how fast our system transitions from assessment-favoring to recommendation-favoring. It also indicates that our system will always choose the recommendation heuristic after the student has experienced M problems.

3 Evaluation of Adaptive Recommendation

We evaluate our adaptive learning material recommender system in *JRec* (Fig. 1), a Japanese reading text recommendation tool. Our corpus of 380 articles was collected from NHK Easy [4], a Japanese news website for language learners. In order to accommodate beginners, our tool split those articles into 4,267 sentences and paragraphs so that students do not have to read the whole article.

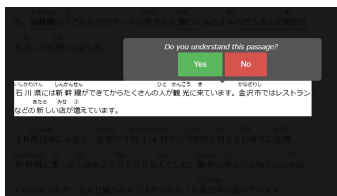


Fig. 1. Screenshot of *JRec*, which draws texts from NHK Easy [4].

Table 1. Wilcoxon Rank-sum tests for all pairs of our four groups.

Comparison	Results
A.R. vs N.R.	$p = .035, Z = 2.109$
A.O. vs A.R.	$p = .766, Z = 0.298$
A.O. vs N.R.	$p = .022, Z = 2.287$
Rand vs N.R.	$p = .547, Z = 0.603$
A.O. vs Rand	$p = .294, Z = 1.049$
A.R. vs Rand	$p = .389, Z = 0.861$

Afterwards, it analyzed the hierarchical structure of vocabulary knowledge in the corpus and built a fuzzy partial ordering graph. When using this tool, users are directed to an NHK Easy webpage, read a recommended text (a paragraph or a sentence), and respond whether or not they understand it. Our tool highlights the recommended text and grays out the rest of the webpage. We recruited 368 users from the Japanese Learning Sub-reddit [6].

Adding Adaptivity Improved Engagement Significantly. We tested four different versions: (1) adaptive recommendation (which balances recommendation and assessment using $M = 50$) and (2) non-adaptive recommendation (with no assessment incorporated), as well as (3) assessment-only and (4) random selection as additional baselines. We particularly wanted to see if adaptive recommendation is more engaging than non-adaptive recommendation, since this would demonstrate that adaptive assessment can enhance learning material recommendation.

In order to measure engagement, we recorded the number of texts each user read before leaving. 131 randomly selected users used adaptive recommendation (A.R.), 91 users used non-adaptive recommendation (N.R.), 115 users used assessment-only (A.O.) and 31 users used the random algorithm (Rand.). Users were assigned to these conditions at a ratio of 3:3:3:1, respectively, but the tool only recorded when a user responded to a text and some users may have quit before responding to the first problem. As a result, the number of recorded users in each group differs somewhat from the expected ratio.

Since our data was not normally distributed, we ran Wilcoxon Rank-sum tests for all pairs of the four groups (Table 1). We observed that the median user in the adaptive recommendation group ($Median = 13$) read 62.5% more texts than those in the non-adaptive recommendation group ($Median = 8$), and the difference between these two groups was statistically significant ($p = .035$), which indicates that adaptive recommendation led users to read more texts than non-adaptive recommendation. In addition, the median user in the assessment-only group read 12 texts, which was also significantly more than that in the non-adaptive recommendation group ($p = .022$). The median user in the random group read 8 texts and we did not find a statistically significant difference

compared to the other three groups, possibly because the random group had too few users. Overall, our results show that incorporating adaptive assessment can significantly enhance learning material recommendation in online learning.

Acknowledgements. This material is based upon work supported by the National Science Foundation under Grant No. IIS-1657176.

References

1. Chen, C.M., Hsu, S.H., Li, Y.L., Peng, C.J.: Personalized intelligent m-learning system for supporting effective English learning. In: IEEE International Conference on Systems, Man and Cybernetics, SMC 2006, vol. 6, pp. 4898–4903. IEEE (2006)
2. Falmagne, J.-C., Cosyn, E., Doignon, J.-P., Thiéry, N.: The assessment of knowledge, in theory and in practice. In: Missaoui, R., Schmidt, J. (eds.) ICFCFA 2006. LNCS (LNAI), vol. 3874, pp. 61–79. Springer, Heidelberg (2006). https://doi.org/10.1007/11671404_4
3. Lord, F.M.: Applications of Item Response Theory to Practical Testing Problems. Routledge, London (1980)
4. NHK: NEWS WEB EASY (2019). www3.nhk.or.jp/news/easy/
5. Rasch, G.: Probabilistic models for some intelligence and attainment tests. ERIC (1993)
6. Reddit: Learn Japanese (2019). <https://www.reddit.com/r/LearnJapanese/>
7. Vygotsky, L.S.: Mind in Society: The Development of Higher Psychological Processes. Harvard University Press, Cambridge (1980)
8. Wang, S., He, F., Andersen, E.: A unified framework for knowledge assessment and progression analysis and design. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 937–948. ACM (2017)