Recommendation System of Mobile Language Learning Applications:
Similarity versus Diversity in Learner Preference

Juyeong Song
Ewha Womans University
juyeong0910@ewhain.net

Hyo-Jeong So
Dept. of Educational Technology
Ewha Womans University
hyojeongso@ewha.ac.kr

Kisu Yang
Korea University
Willow4@korea.ac.kr

Hyeji Jang
Ewha Womans University
hyejijang@ewhain.net

ABSTRACT

Recognizing the potential of the preference-inconsistent recommendation systems (RS) for learning, this paper aims to examine two recommendation algorithms for mobile language learning applications: RS with similarity and RS with diversity. Diversity was measured through learning styles (how learners learn) and achievement goals (why learners learn). A total of 160 learners participated in the study for building learner profiles and the recommendation algorithm. Overall, our results with RSME indicate that both RS with similarity and RS with diversity performed better than the random recommendation.

Keywords

Recommendation system, Mobile language learning applications, Diversity

1. INTRODUCTION

Recommendation systems (RS) that suggest preferable items intend to help people make proper decisions in a flood of information and choices. In the field of education, researchers have developed RS to recommend learning resources based on ontology [22], learning styles [4], ranked lists of learning objectives [6], and learner preference [5]. Thus far, most educational RS have been based on the algorithm of learners' preferences, recommending items that learners are likely to prefer based on their previous data or learners with similar profiles. Cognitive dissonance theory [10] explains the popularity of this preference-consistent recommendation. When humans process information that does not conform to their existing thoughts and beliefs, they are in a state of cognitive dissonance and experience an uncomfortable psychological state. Hence, for psychological comfort, people tend to only pursue information and items that they prefer to avoid cognitive dissonance.

However, some scholars have started questioning the efficacy of preference-consistent recommendations, arguing that such recommendation methods may reinforce learners' confirmation bias and narrow the opportunity to discover new information and

experiences [3]. Researchers have proposed 'preference-inconsistent' recommendations that induce discovery learning and higher-order thinking by intentionally providing information that does not match learners' preferences or recommending information that contradicts learners' beliefs [19]. While the preference-inconsistent mechanism has been widely researched in political science and media study to minimize confirmation or information bias, its application in the educational RS is still in its infancy and lacks empirical data to support its efficacy.

Recognizing the potential of the preference-inconsistent recommendation for learning, this paper presents our work-in-progress research that aims to examine two recommendation algorithms for mobile language learning applications (apps): RS with similarity and RS with diversity. Here, 'RS with similarity' refers to the preference-consistent method that recommends the list of apps compatible with learners' learning styles and achievement goals. In contrast, 'RS with diversity' refers to the method that recommends the list of apps that do not match learners' current preferences but have the potential to help learners exposed to new learning styles and achievement goals. It should be noted that RS with diversity does not mean that the system recommends items that learners do not like. Instead, the method is intended to lead learners to discover new items that are likely to have learning efficacy in the future by extending learning experiences beyond a preferred comfort zone.

2. LITERATURE REVIEW

2.1 Recommendation Systems in Education

Recommendation systems have been widely used in the e-commerce area to suggest information and items for potential customers [18]. Recently, RS has been studied in the education sector for various purposes such as suggesting learning activities and courses based on learners' knowledge levels and preferences [16].

From a technical view, there are two main methods of RS: content-based filtering and collaborative filtering. Content-based filtering recommends items that are similar to what users already selected and liked in the past [15]. However, when there is a lack of data on previous behaviors by users, it is hard to recommend appropriate items that match users' preferences. Also, the attributes of items should be well-structured for the accuracy of content-based filtering. In contrast, collaborative filtering recommends items based on similar users' preferences [15]. It is possible to predict items that new users are likely to prefer even if there is a lack of available data

for the target user. Collaborative filtering, however, has a scalability issue. As the number of users or items increases, the complexity of computation becomes higher [11]. Moreover, user-based collaborative filtering systems tend to provide a low number of available items, making final recommendations biased toward popular items.

In the field of technology-enhanced learning (TEL), it has been suggested that there are two perspectives of enhancing filtering techniques: top-down and bottom-up [7]. The top-down perspective focuses on well-defined educational metadata whereas the bottom-up perspective collects user-generated data such as likes, tags, and ratings from learners. In this study, we focus on language learning with mobile applications, which are often used in informal learning contexts. As informal learning is self-directed and less structured with learning goals and time [7], it is difficult to divide applications into specific categories for educational metadata. Hence, this study focuses on the RS from the bottom-up perspective such as learner ratings of respective mobile apps.

2.2 Diversity in Learning Styles and Achievement Goals

Recently, there have been attempts to develop methods for recommend diverse and novel items beyond similarity measures. Described as a 'long-tail recommendation', this method recommends diverse and unpopular items in the non-mainstream area to mitigate confirmation bias and to enhance diversity in user experiences [23]. As for movie recommendations, it has been reported that the list of diverse movies positively affects viewers' satisfaction [14]. However, a long-tail recommendation also suffers from a low accuracy in the recommendation, which necessitates some tradeoff between diversity and accuracy [23].

For learning purposes, there are numerous ways to operationally define diversity. In this study, we focus on diversity in terms of learning styles (*how learners learn*) and achievement goals (*why learners learn*). These two constructs have been well researched and established as a theoretical framework to define learner characteristics in educational research.

First, learning styles refer to the ways learners prefer to receive and process information. According to Felder and Silverman [9], learning styles are divided into four categories: (a) processing (active/reflective), (b) perception (sensing/ intuitive), (c) input (visual/verbal), and (d) understanding (sequential/global). Many studies have been conducted on which learning styles are the basis for presenting effective learning methods. However, some scholars suggested two limitations about learning styles [13]. The first limitation is that the preferable way to learn and the actual effective and efficient learning method can be different. The second limitation is that the learning styles instrument showed low reliability [21]. While we are aware of such limitations, this study adopts learning styles as one of the learner-related variables for learner profiles based on the following studies. As for recommender systems, the Felder-Silverman Learning Style Model (FSLSM) has been widely used for two reasons [2, 17]. First, FSLSM can be implemented in a simple way to design personalized e-learning systems [2]. Second, the FSLSM model is appropriate to apply to mobile learning because the learning materials in mobile learning are composed of various formats such as video, text, audio, etc. [17].

Second, achievement goals refer to the aim and focus of one's actions concerning achievement. Elliot and McGregor proposed the 2X2 achievement goal framework in terms of definition (mastery vs. performance) and valence (approach focus vs. avoidance focus). The framework includes (a) mastery-approach goal, (b)

performance-approach goal, (c) mastery-avoidance goal, and (d) performance-avoidance goal [8]. When learners have the mastery-approach goal, they strive to master learning tasks with the emphasis on learning itself and self-improvement. In contrast, learners with the mastery-avoidance goal tend to avoid misunderstanding or the failure to master a task. Regarding performance, learners with the performance-approach goal concentrate on proving their ability relative to others. They may not be interested in mastering tasks, but doing better than others is more important. On the other, learners with the performance-avoidance goal tend to avoid performing more poorly than others do.

3. RECOMMENDATION SYSTEM

To construct our RS architecture (see Figure 1), three steps were performed: Step 1: constructing the database of RS, Step 2: data analysis, and Step 3: RS evaluation (Experiment). The whole process of the research was conducted with the IRB approval from the researchers' university.
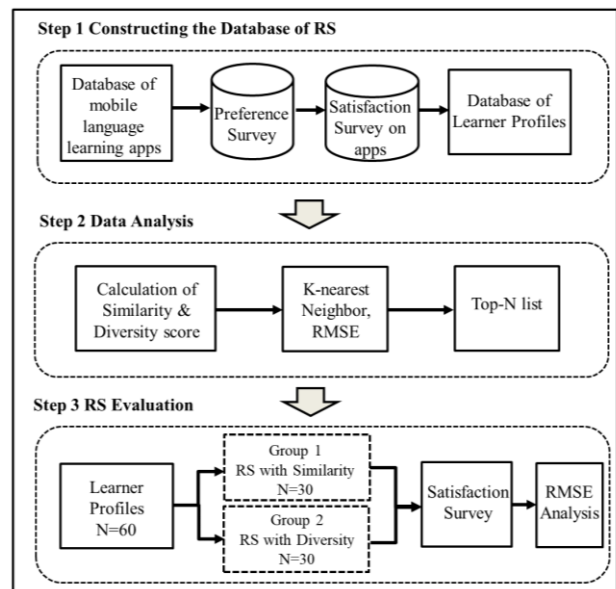

Figure 1. Proposed architecture and process

3.1 Step 1: Constructing the Database of RS

3.1.1 *Database of Mobile Language Learning apps*

In this study, we focused on recommending mobile language learning apps with artificial intelligence (AI) functions, which have been increasingly developed in the app market. For constructing the database of learner profiles, we chose 18 mobile apps that met five criteria: (1) an application for language learning, (2) an application with AI functions (e.g., voice recognition, adaptive learning system), (3) an application that the developer indicates the use of AI technologies, (4) an application used with no errors, and (5) an application locally available in Korea. Table 1 presents the final list of mobile language learning apps used in this study.

Table 1. List of mobile language learning apps with AI

ID	Application	ID	Application
app 1	Riid Tutor	app 10	Cake
app 2	Super Chinese	app 11	Youbot English Speaking

ID	Application	ID	Application
app 3	Plang	app 12	ELSA Speak: Accent Coach
app 4	Opic up	app 13	Bigple
app 5	Lingo Champ	app 14	Memrise
app 6	Easy Voca	app 15	Mondly
app 7	Say Voca	app 16	AI Tutor
app 8	Youbot Chinese	app 17	Rosetta Stone
app 9	Duolingo	app 18	Busuu

3.1.2 Preference and Satisfaction Survey

To construct the database of learner profiles, we recruited 100 adult learners. Among them, 82 learners fully completed the preference survey that include items about demographic information (e.g., gender, age, education, etc.), learning styles, and achievement goals. For learning styles, we used the ILS (Index of Learning style) instrument with 44 items [9]. For achievement goals, we used the 2X2 Achievement Goal Questionnaire-Revised with 24 items [20]. Then, the participants used four to six mobile language learning apps randomly assigned to them. Finally, they completed the satisfaction survey with 20 items on a 5-point Likert scale to evaluate each app they used. The satisfaction survey includes seven categories: (1) convenience, (2) personalization, (3) professionalism, (4) reliability, (5) appropriateness, (6) satisfaction, and (7) overall satisfaction. We refer to [1] to develop the items related to convenience, personalization, professionalism, and reliability and [12] for the items related to appropriateness of app use and satisfaction. We modified the items according to the context and purpose of this study.

3.1.3 Database of Learner Profiles

To construct the database of learner profiles, data from 82 learners who fully answered the surveys were analyzed based on their learning styles and achievement goals. We generated attribute values as shown in Figure 2. LS1, LS2, LS3, and LS4 refer to four dimensions of learning styles respectively: processing, perception, input, and understanding. LG1, LG2, LG3, and LG4 mean four achievement goals respectively: mastery-approach goal, mastery-avoidance goal, performance-approach goal, and performance-avoidance goal. Since each learner used 4 to 6 apps, the total data of learner profiles is 82*415 (the number of learners * the total number of applications used).

NO.	LS1	LS2	LS3	LS4	LG1	LG2	LG3	LG4
L1	1	0	1	0	4.67	3.00	4.00	1.67
L2	0	1	1	0	3.83	3.67	3.00	3.17
L3	0	0	1	0	5.00	3.00	3.17	1.67
L4	1	1	1	0	3.67	2.50	3.17	2.50
L5	0	1	1	0	4.83	4.17	4.67	1.67

Figure 2. Example of attribute values

3.2 Step 2: Data Analysis

Based on the survey results, we calculated the similarity score for RS with similarity and the diversity score for RS with diversity. The formula for calculating similarity and diversity scores for learner profiles is shown below. We measured the diversity score with Euclidean distance between two learners for learning styles and goals. For example, if the learning styles of learner A and Learner B are [1, 0, 1, 0] and [1, 1, 1, 0] respectively, the diversity score is 1. After calculation, we chose the best similarity measure and the number of K-neighbors.

$$sim(t, t')$$

$$= \sum_{j}^{2} \lambda_j \cdot sim_j(t, t')$$

$$= \lambda_1 \cdot sim_1(t, t') + \lambda_2 \cdot sim_2(t, t')$$

$$sim_i(t, t') = diversity\ score * -1$$

$$diversity\ score = Euclidean\ distance(t, t')$$

$$= \sqrt{\sum_{i=1}^{n} (t_i - t'_i)^2}$$

To evaluate the performance of RS, this research used Root Mean Squared Error (RMSE), which is widely used to evaluate collaborative filtering systems. From the postulation that users are satisfied with items recommended by diverse users, not just similar ones, RS with diversity returns the top N items sorted by the ascending order of the average scores of K dissimilar users. The diversity score is a negative value of the similarity score, which is the weighted sum of the learning styles similarity and the achievement goals similarity. In the following experiment in Step 3, we set K = 9 and the weights to each similarity $\{\lambda_1, \lambda_2\} = \{1.0, 1.5\}$ because the settings of K and λ show the lowest RMSE between diversity-based prediction and ground truths in training data (see Figure 3).
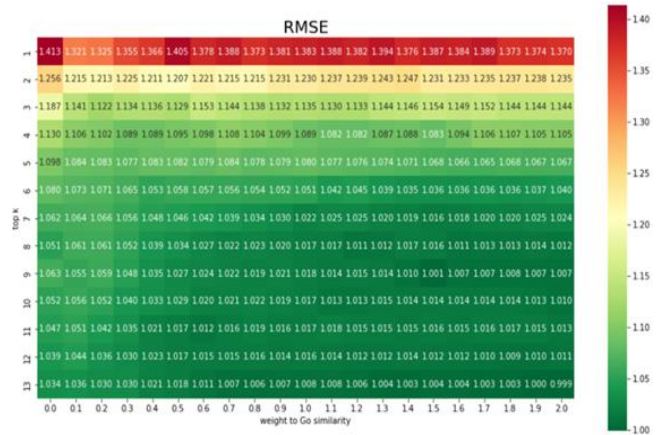

Figure 3. RMSE

3.3 Step 3: RS Evaluation

3.3.1 Experiment

For the experiment to compare the performance of RS with similarity and RS with diversity, we recruited 60 participants aged from 19 to 39 years old. They used the visualized RS (see Figure 4) to complete the same survey administered in Step 1. We analyzed the preference survey data to gauge their learning styles and achievement goals, and then made the recommendations of apps for each participant. We divided 60 participants into two groups with the consideration of their age, job, and gender. The first group was recommended with three apps based on the algorithm of RS with similarity. The second group was recommended with three apps based on the algorithm of RS with diversity. After using the recommended apps, the participants completed the satisfaction survey to evaluate each app they used.

3.3.2 Experimental Setup

For the experiment, we developed the mobile web-based system (see Figure 4) to simulate the real situation of using a RS for learning. The interface shows three buttons to guide learners to access

each page. In the first button 'What is my preference?', learners complete the preference survey about demographic information, learning styles, and achievement goals. Clicking the second button 'Checking recommended apps', learners see the list of recommended mobile apps for language learning. After using the recommended apps, learners move to the third button 'What is my satisfaction?' to complete the survey to indicate their satisfaction with the recommended apps.
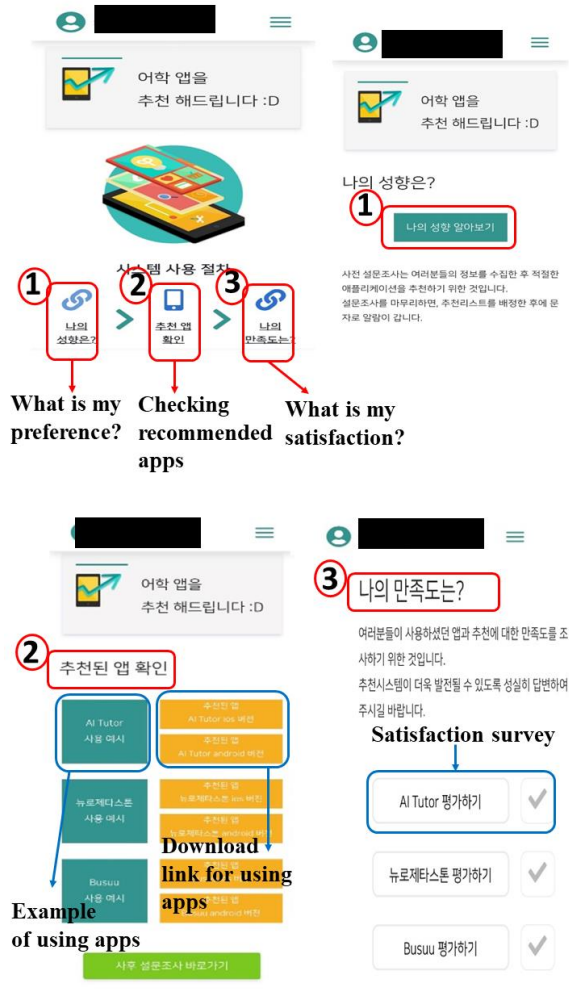

Figure 4. RS Visualization

3.3.3 Data Results

We evaluated our recommendation systems with RMSE representing the difference between the predicted and actual levels of learners' satisfaction. We chose this approach since an accuracy metric only covers the exactly matched values. We observed that RMSE values of similarity-based recommendation and diversity-based recommendation are 0.91 and 1.26 respectively. They are significantly smaller than 1.86 from the random prediction. With this, we interpret that the recommended method that considers how and why learners learn are likely to lead to higher learner satisfaction than the random recommendation. Further, RS with similarity showed a better performance than RS with diversity.

4. LIMITATIONS AND FUTURE WORK

The limitations of this research are as follows. First, the recommendation system was only evaluated by accuracy. Although it is important to predict items accurately, qualitative data such as interviews and proxy indicators (e.g., usage behaviors) should be used additionally to further test and improve the proposed RS. Second, since the data sample is rather small to build a robust model, the research findings have limited generalizability. The small sample size was due to the time and effort required for the participants to use the assigned apps and to provide valid evaluation data. In the future, we plan to recruit more users and expand the database of the apps to build a more robust model. Lastly, while this research compares two RS algorithms, developing a hybrid RS that considers the optimal level of both similarity and diversity is certainly a promising area for future research.

Despite these limitations, we believe that this research makes meaningful contributions to the existing body of literature on RS by providing empirical data that examine similarity versus diversity approaches.

5. ACKNOWLEDGMENTS

6. REFERENCES

[1] Baek, C. H. 2020. A study on satisfaction of artificial intelligence speaker users using quality evaluation method of artificial intelligence service. *Journal of Korean Service Management Society*. 21, 5 (Dec. 2020), 157-174. DOI= https://doi.org/10.15706/jksms.2020.21.5.007.

[2] Bourkoukou, O., El Bachari, E., & El Adnani, M. A. 2016. Personalized e-learning based on recommender system. *International journal of learning and teaching*. 2, 2 (Dec. 2016, 99-103. DOI=https://doi.org/10.18178/ijlt.2.2.99-103.

[3] Buder, J. and Schwind, C. 2012. Learning with personalized recommender systems: *A psychological view. Computers in Human Behavior*. 28, 1 (Jan. 2012), 207–216. DOI= https://doi.org/10.1016/j.chb.2011.09.002.

[4] Chen, H., Yin, C., Li, R., Rong, W., Xiong, Z., and David, B. 2019. Enhanced learning resource recommendation based on online learning style model. *Tsinghua Science and Technology*. 25, 3 (Oct. 2019), 348–356. DOI= https://doi.org/10.26599/TST.2019.9010014.

[5] Dascalu, M.-I., Bodea, C.-N., Mihailescu, M.N., Tanase, E.A., and Ordoñez de Pablos, P. 2016. Educational recommender systems and their application in lifelong learning. *Behaviour & information technology*. 35, 4 (Jan. 2016), 290–297. DOI= 10.1080/0144929X.2015.1128977.

[6] De Medio, C., Limongelli, C., Sciarrone, F., and Temperini, M. 2020. MoodleREC: A recommendation system for creating courses using the moodle e-learning platform. *Computers in Human Behavior*. 104 (Mar. 2020), 106168. DOI=https://doi.org/10.1016/j.chb.2019.106168.

[7] Drachsler, H., Hummel, H., and Koper, R. 2009. Identifying the goal, user model and conditions of recommender systems for formal and informal learning. *Journal of Digital Information*. 10, 2 (Jan. 2009), 4–24.

[8] Elliot, A. J. and McGregor, H. A. 2001. A 2 X 2 achievement goal framework. *Journal of Personality and Social Psychology.* 80, 3 (2001), 501-519.

[9] Felder, F.M. and Soloman, B.A. (1991), Index of Learning Styles. North Carolina State University. Web address: www.ncsu.edu/effective_teaching/ILSdir/ILS-a.htm.

[10] Festinger, L. 1954. A theory of social comparison processes. *Human relations.* 7, 2 (May. 1954), 117–140. DOI= https://doi.org/10.1177/001872675400700202.

[11] Jalili, M., Ahmadian, S., Izadi, M., Moradi, P., and Salehi, M. 2018. Evaluating collaborative filtering recommender algorithms: a survey. *IEEE access.* 6 (Nov. 2018), 74003–74024. DOI= https://doi.org/10.1109/ACCESS.2018.2883742.

[12] Jang, E., Park, Y., and Lim, K. 2012. Research on factors effecting on learners' satisfaction and purchasing intention of educational applications. *The Journal of the Korea Contents Association*. 12, 8 (Aug. 2012), 471-483. DOI= https://10.5392/JKCA.2012.12.08.471.

[13] Kirschner, P.A. 2017. Stop propagating the learning styles myth. *Computers & Education*. 106 (Mar. 2017), 166-171. DOI= https://doi.org/10.1016/j.compedu.2016.12.006.

[14] Nguyen, T.T., Maxwell Harper, F., Terveen, L., and Konstan, J.A. 2018. User personality and user satisfaction with recommender systems. *Information Systems Frontiers*. 20, 6 (Sep. 2018), 1173–1189. DOI= https://doi.org/10.1007/s10796-017-9782-y.

[15] Ricci, F., Rokach, L., and Shapira, B. 2015. Recommender systems: Introduction and challenges. In *Recommender systems handbook,* F. Ricci, L. Rokach, and B. Shapira, Eds. Springer. Boston, MA, 1–34. DOI= https://doi.org/10.1007/978-1-4899-7637-6_1.

[16] Santos, O.C. and Boticario, J.G. 2015. User-centred design and educational data mining support during the recommendations elicitation process in social online learning environments. *Expert Systems*. 32, 2 (Jul. 2015), 293–311. DOI= https://doi.org/10.1111/exsy.12041.

[17] Saryar, S., Kolekar, S.V., Pai, R.M., and Manohara Pai, M.M. 2019. Mobile learning recommender system based on learning styles. In *Soft computing and signal processing,* J. Wang, G. Reddy, V. Prasad, and V. Reddy, Eds. Springer. Singapore, SG, 299–312. DOI= https://doi.org/10.1007/978-981-13-3600-3_29.

[18] Schafer, J.B., Konstan, J., and Riedl, J. 1999. Recommender systems in e-commerce. In *Proceedings of the 1st ACM Conference on Electronic Commerce* (Colorado, USA, November 03-05, 1999). EC'99. ACM, New York, NY, 158–166.

[19] Schwind, C., Buder, J., Cress, U., and Hesse, F.W. 2012. Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education*. 58, 2 (Feb. 2012), 787–796. DOI= https://doi.org/10.1016/j.compedu.2011.10.003.

[20] Shin, S.-Y. 2007.*Identifying predictability of achievement goal orientation, reward structure and attributional feedbacks on learning achievement in e-PBL.* Doctoral Thesis**.,** Ewha womans university.

[21] Stahl, S. A. 1999. Different strokes for different folks? A critique of learning styles. *American Educator.* 23, 3 (1999). 98-107.

[22] Tarus, J.K., Niu, Z., and Mustafa, G. 2018. Knowledge-based recommendation: a review of ontology-based recommender systems for e-learning. *Artificial intelligence review*. 50, 1 (Jan. 2018), 21–48. DOI= https://doi.org/10.1007/s10462-017-9539-5.

[23] Wang, S., Gong, M., Li, H., and Yang, J. 2016. Multi-objective optimization for long tail recommendation. *Knowledge-Based Systems*. 104 (Jul. 2016), 145–155. DOI= https://doi.org/10.1016/j.knosys.2016.04.