Educational Technology Research and Development
DOI 10.1007/s11423-014-9344-7

Collaborative filtering for expansion of learner's background knowledge in online language learning: does "top-down" processing improve vocabulary proficiency?

Masanori Yamada · Satoshi Kitamura · Hideya Matsukawa · Tadashi Misono · Noriko Kitani · Yuhei Yamauchi

Abstract In recent years, collaborative filtering, a recommendation algorithm that incorporates a user's data such as interest, has received worldwide attention as an advanced learning support system. However, accurate recommendations along with a user's interest cannot be ideal as an effective learning environment. This study aims to develop and evaluate an online English vocabulary learning system using collaborative filtering that allows learners to learn English vocabulary while expanding their interests. The online learning environment recommends English news articles using information obtained from other users with similar interests. The learner then studies these recommended articles as a method of learning English. The results of a two-month experiment that compared this system to an earlier collaborative filtering system called ''GroupLens'' reveal that learners who used the collaborative filtering system developed in this study read various news articles and had significantly higher scores on topic-specific vocabulary tests than did those who used the previous system.

Keywords Recommendation system · Learning support · Language learning · Vocabulary learning

Introduction

As the amount of information available on the Internet is increasing at a tremendous rate, recommendation algorithms have become important tools for users seeking to retrieve focused information, and have thus been applied to various online services.

A recommendation algorithm is one that predicts a user's attributions, such as ability and preference, and suggests information that corresponds with these data. Collaborative filtering is one type of recommendation algorithm. It has been the focus of recent research, and it is used by various online services such as Amazon.com.

Collaborative filtering is a recommendation algorithm that predicts an active user's preferences and presents suggested material by using the preferences of users who appear similar to the active user according to their rating data. As indicated by previous research (e.g., Herlocker et al. 1999; Resnick et al. 1994), recommendation algorithms aim to determine a user's preferences and suggest optimized information according to that determination. Recently, tens of thousands of educational resources such as Open Course Ware (OCW) have become available to the world through the Internet. Therefore, interested parties around the world are examining educational resource recommendation algorithms. Researchers have been applying these resources to learning systems to clarify the effects of collaborative filtering on learning (e.g., Liang et al. 2006; Recker and Walker 2003). However, optimization according to learners' preferences is not always ideal in educational settings. This study aims to develop a collaborative filtering design optimized for learning and to evaluate its effects.

Related works

Background theory of language learning for the application of collaborative filtering

Along with the advancement of information communication technology, interest in using online learning environments for language learning has grown. Online learning environments can support learning outside the classroom, and in foreign language learning, they can increase opportunities for using foreign language resources. However, Sakai (2008) pointed out that they do not increase learners' time spent using these resources. Online learning technologies do help increase learning opportunities; however, the use of these environments depends on a learner's motivation, which online learning environments for foreign language learning should be designed to encourage. It would be advantageous to develop a foreign language learning system that allows learners to easily access online learning environments. Consideration should also be given to enabling learners to increase ''input,'' or foreign language information for learning (Krashen 1985), since it plays an important role in foreign language learning (Oxford 1990; Payne 2011). Payne (2011) suggested that input should be provided to pupils in natural language use settings.

Yamada et al. (2009) have developed an online English learning environment with a collaborative filtering system that recommends learning material according to a learner's interests. This system enables the study of English outside of school time. This learning environment allows learners to read English news articles that match their own interests, thereby motivating them to learn English. Motivation is a central factor in successful learning, and problems related to motivation often arise in online learning environments. One such problem is related to the learning materials themselves. In language learning, input information such as learning material has a strong effect on understanding content in the target language (Krashen 1985; Payne 2011) and enhancing motivation (Dörnyei 2001; Manolopoulou-Sergi 2004). Input here refers to written or spoken information in the target language that the learner is able to comprehend (e.g., Gass et al. 1998; Krashen 1985). It is difficult for an online learning environment to provide appropriate learning materials for each individual learner since this would require teachers to prepare materials that conform to each learner's prior knowledge and preferences. Collaborative filtering is one viable solution for this problem.

The basic idea behind recommender systems with collaborative filtering is that the results are based on predictions of a user's interests (e.g., Herlocker et al. 2004). High prediction accuracy is, of course, desirable. However, such a system of learning may cause bias within the learning content because learners will read only the English news articles that correspond with their interests. In other words, there will be less variety because only items in which learners have a strong interest will be recommended. Therefore, in the current study, learners will be required to understand diverse information in the context of their diverse background knowledge; however, this can also cause bias in terms of comprehension.

A learning environment that continuously recommends material that matches a user's interests is essential for maintaining the user's motivation to learn. However, in order to learn to use English in everyday situations, it is important to study the language without concentrating solely on one specific field. Building a learning environment that goes beyond the range of the learner's interests is desirable, as the expansion of background knowledge has been suggested to have a positive impact on learning. Especially in regards to listening and reading comprehension, background knowledge is an effective and important informational source for comprehension (Anderson and Lynch 1988; Lynch 2011; McNeil 2012). Specifically, comprehension proceeds as a combination of bottom-up and top-down processing. In bottom-up processing, comprehension is built from smaller linguistic units (phoneme, word) to larger units (clause, sentence, passage/paragraph). In top-down processing, comprehension follows predictions and inferences that emerge from the background knowledge of the listener (schema) and the context. Lynch (2011) explained that one of conditions to promote learners' comprehension of lectures in a foreign language is to recall the relevant background knowledge. Burgoyne et al. (2013) reported that children who were taught the relevant knowledge for reading material used this taught knowledge to answer questions, and this process promoted text comprehension in foreign language learning settings. A learning method that stresses background knowledge is adopted in top-down processing, which has a positive impact on learning (Field 1998). It is expected that reading articles on various topics promotes top-down processing, increasing various kinds of background knowledge. The suggestions and findings mentioned above seem to be reasonable from the view of educational technology research, which states that input information relevant to learners' background knowledge should be considered in designing instruction (e.g., Keller 2010).

Collaborative filtering works to optimize an exact fit with learners' preferences and interests; thus, it supports top-down processing for comprehension in foreign language learning. The learning objectives in foreign language learning, though, are sometimes dependent on topic; therefore, an exact fit with learner preferences and interests may not be effective. Vocabulary acquisition depends on the topics covered in the learning materials.

For example, when we read medical textbooks or news articles, we learn medical vocabulary. In order to learn a variety of vocabulary, it is better for learners to read learning materials on diverse topics.

However, Gilmore (2007) points out the danger of demotivating learners by teaching them foreign language vocabulary related to topics that lie completely outside their preferences and interests. Thus, collaborative filtering is required in order to recommend topics that are at least close to learners' preferences and interests. Such a system may help increase the variety of foreign language vocabulary to which learners are exposed and thus have a positive effect on the acquisition of such vocabulary.

Effects of collaborative filtering on learning performances

One recent educational movement seeks to provide global access to open learning materials such as OCW, thus making tens of thousands of learning materials available on the Internet. Learners have to find suitable learning materials from a mass of search results. However, it seems to be difficult for learners to find online learning resources (OLRs) that fit with their interests from materials such as OCW, news articles, academic papers, and blogs. Thus, a recommendation algorithm provides an appropriate learning environment that helps learners sift through the vast amount of OLRs on the Internet. Many researchers have conducted studies on the application of recommendation algorithms to learning support systems in technology-enhanced learning research areas. Hsu (2006) developed and evaluated an English learning system that recommends English news articles by using a score based on learning time and the association rule with text-mining. Hsu's research evaluated the rate at which the system's recommendations fit learners' interests; however, this research did not evaluate learning performance.

Bobadilla et al. (2009) evaluated the fit rate of the recommendations made by collaborative filtering by comparing performance test scores while taking into account learner performance levels. The researchers did not establish any cohesive effects of collaborative filtering and suggested that other variables be considered for the educational use of collaborative filtering. Manouselis et al. (2010) compared several methods of calculating the approximation, processing time, and covering rate of topics in collaborative filtering. Recker and Walker (2003) and Recker et al. (2003) evaluated the effects of a collaborative filtering system called ''Altered Vista'' on learning performance in essay education. When comparing Altered Vista with random recommendations, the satisfaction with recommendations and the essay test scores of learners using Altered Vista were higher than those of the learners who used random recommendations. Recker et al. (2003) suggested that learners using Altered Vista felt encouraged by finding other learners who had similar opinions. Previous research about collaborative filtering for educational use aimed to optimize the fit with learner attributions according to data such as preference and proficiency, and to evaluate the effect of its optimization.

However, Manouselis et al. (2010) pointed out that collaborative filtering using learner preference data is difficult to apply to educational and learning support systems; that is, optimization to fit learner preferences cannot be considered ''optimization'' in education and learning. In order for this to work in education and learning, it is necessary to adjust the range of strength of interests on items or variables for recommendation according to the problems that teachers want to solve. When the system recommends learning materials by considering the learners' level in one subject, this level is the variable for recommendation (Bobadilla et al. 2009). On the other hand, when the system recommends high-level learning materials, rating data from high-level learners is the variable for recommendation

(Liang et al. 2006). Ghauth and Abdullah (2010) also conducted comparative research in the context of information technology education. These authors compared a collaborative filtering algorithm with a content-based recommendation and no recommendation treatment. The results of their research reveal that the algorithm proposed in their research was the most effective on the performance test because of the model used for the algorithm. In cases in which learning depends on the topic of the learning material, recommendations of high-interest material may not be desirable. As mentioned in the previous section, in foreign language vocabulary learning, the acquisition of a varied vocabulary depends on the topics in the textbook or news articles being studied. This research suggests that increasing the fit rate of recommendations along with learners' interests cannot always work in educational and learning contexts.

In foreign language learning, it may be desirable to recommend topics that are close to, rather than exactly matching, learner preferences or interests. This study aims to develop and evaluate the effect of a foreign language learning support system that uses a collaborative filtering system to recommend news articles that are close to, rather than an exact fit with, learner preferences, in order to expand background knowledge to promote top-down processing.

Research questions

This study aims to develop and evaluate an English vocabulary learning support system with collaborative filtering that recommends English news articles on topics that are close to learners' preferences and interests, and to compare this system with the popular earlier collaborative filtering algorithm known as ‘‘GroupLens’’ (Resnick et al. 1994). This research aims to investigate the effects of the new collaborative filtering algorithm, which recommends English news article in which learners have some interest (not strong interest), on perceived appropriateness of recommendation and learning performance. There are two hypotheses as follows:

H1: Learners will read about more various topics recommended by the algorithm developed by this research.

H2: Learners' vocabulary proficiency will improve because the system developed in this research allows them to read more various topics than previous systems did.

In order to test these hypotheses, we developed four research purposes. This research aims to evaluate the effects of the new recommendation system according to the following four points.

(1) Is there a difference in the fit rate of recommendation between the collaborative filtering developed in this research and GroupLens? (for H1).
(2) Does this collaborative filtering promote diversity in topics of news articles? (for H1).
(3) Does the system tend to recommend specific topics in order to promote specific vocabulary? (for H2).
(4) Overall, does this system contribute to the improvement of vocabulary test scores? (for H2).

The first point aims to investigate the differences between the two systems among learners' perceptions of fit with their interests. The system developed in this study recommends news articles close to learners' preferences and interests, as mentioned above. If this system is superior to GroupLens from the viewpoints of both perception of fit and learning performance, this system may be more appropriate to foreign language vocabulary learning than GroupLens. We investigate this point using a questionnaire. The second and third points are the main areas of research interest. Recommending news articles close to learners' preferences and interests is expected to lead learners to read various categories of news topics, and learn other topic-specific vocabulary. We evaluate these points using system logs and topic-specific vocabulary tests. The last point aims to investigate the effects of this system on overall learning performance. If this system is more effective on learning performance overall than GroupLens, this system may be superior to GroupLens. This point is evaluated by a vocabulary test, using JACET, which will be explained later.

System

The online learning environment system developed in this study includes four functions: news article recommendations, vocabulary definitions, a marker function, and a comments function. Although this system uses the same vocabulary definitions, marker function, and comments function as that used in online learning environments developed by Yamada et al. (2009), a major change was made to the algorithm for the news article recommendation function.

The system's structure is shown in Fig. 1; its main screen is shown in Fig. 2; and its article details screen is shown in Figs. 3 and 4. This web application consists of two parts: client and server. In the client part, six functions are implicated: a marker function to support cognitive learning strategies, a dictionary, a word check function, comments for each news article, an evaluation of interest for each news article, and an evaluation history function that allows learners to see other learners' evaluation histories. The functions

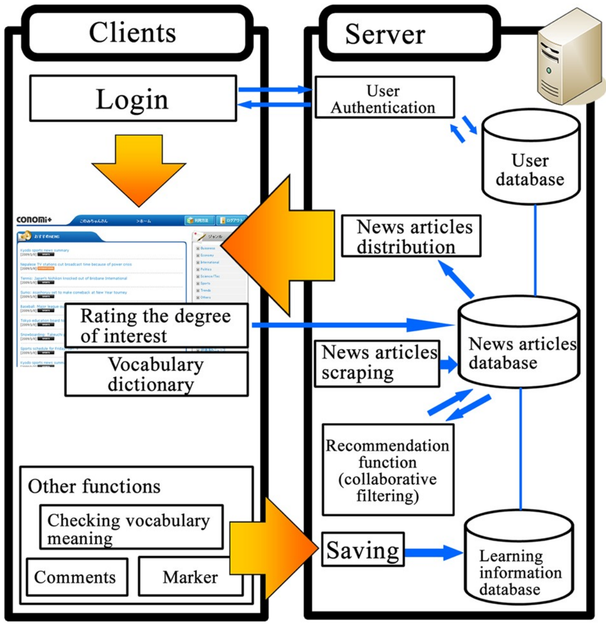

Fig. 1 System structure

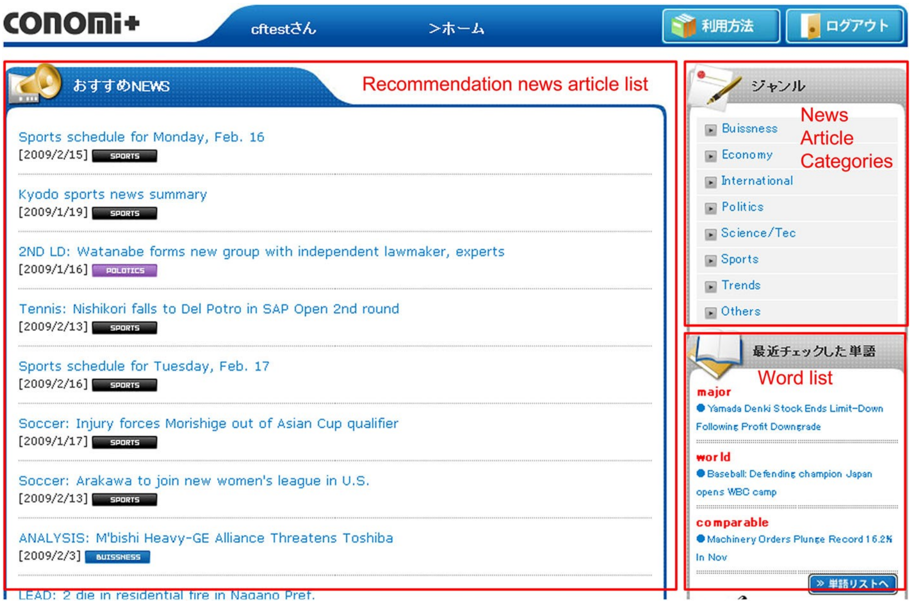

Fig. 2 Main screen of this system. The recommendation news article list displays the news articles' titles recommended by collaborative filtering

Fig. 3 Dictionary function

mentioned above were developed by AJAX (Asynchronous Java Script and XML). On the server side, five functions are implicated: user certification, news articles distribution, collaborative filtering, a user log function that collects what functions learners used and

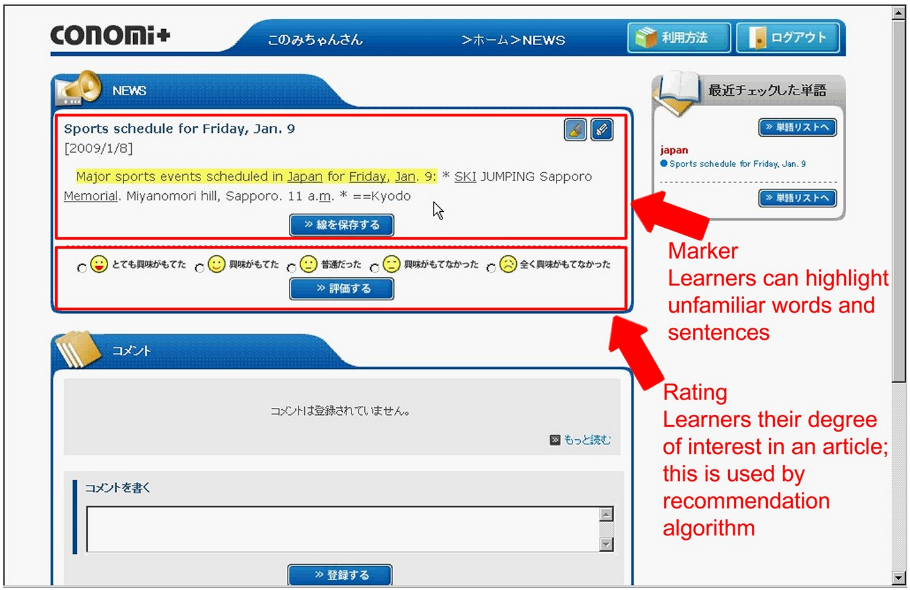

Fig. 4 Marker and rating functions

when, and a news scraping that regularly collects news article data from a news article company's website. These functions were developed by Java and PHP. The server environment consists of three server software applications: Apache Web Server 2.0 as the Web server, JBOSS 3.2.7 as the Application server, and PostgreSQL 8.1.4 as the Database server. The main functions are explained in detail below.

News article recommendation function

The news article recommendation function displays on the main screen a list of articles that are based on the learner's interests. A user-based collaborative filtering system is used for the recommendation algorithm. Yamada et al. (2009) used the algorithm proposed by Resnick et al. (1994), which is a basic user-based collaborative filtering algorithm used to motivate learners by providing appropriate learning materials. The system with added collaborative filtering predicts a user's preferences by analyzing of a similar user's preferences. A collaborative filtering algorithm first finds similar users by employing various methods, one of which uses the correlation rate between users in a preference pattern. Users who have a high correlation rate are regarded as similar to active users. This study, however, applied an alteration to this algorithm.

First, a group of users (user group X) that excludes the learner who receives recommendations is separated into two smaller groups depending on correlation of interests: user group A, which has a high absolute value of correlation coefficient with the learner, and user group B, which has a low absolute value of correlation coefficient with the learner. User group C, which has a high absolute value of correlation coefficient with a certain group of users in group A, is then extracted from user group B. The data for user group C are used for the recommendation method proposed by Resnick et al. (1994).

Making recommendations using only the data from user group A does not result in less accurate predictions than those made using all groups' data (Herlocker et al. 1999). On the other hand, using only the data from user group B will decrease accuracy.

To expand the range of interests, we focused on the sociological concept of transitivity. Transitivity states that if A and B have a strong close link, and A and C have a strong close link, then B and C tend to have a strong close link with similar thoughts, actions, and values (Granovetter 1973). Given this concept, we hypothesized that there is a high possibility that although all users' preferences are not similar, when three users are considered, the preferences of the first and second users will be similar and the preferences of the first and third users will be similar. Hence, the preferences of the second and third users will also be similar. Therefore, we decided to apply the above-mentioned algorithm. The concrete process of recommendation is explained in the following steps.

1: Each active user rates his or her interest in each news article.

2: The system makes data set ''Y_{ai}'' of news articles that are rated positively by both active user a and user i.

3: The system evaluates the similarity of interest in news articles between users a and i, according to the rank correlation coefficient q_{ai} using the rating data from new articles including dataset Y_{ai}

$$\rho_{ai} = \frac{\sum_{k \in Y_{ai}} \left(rank_{a,i} - \overline{rank_a}\right)\left(rank_{u,i} - \overline{rank_u}\right)}{\sigma_a * \sigma_u}$$

Herlocker et al. (1999) reported that Spearman's rank correlation coefficient was a higher threshold resulting in an accurate fit rate, thus improving the GroupLens algorithm. We also used a rank-correlation coefficient.

4: The system calculates rank correlation coefficients for all users using all users' rating data.

5: The system makes an active user dataset ''A,'' which has an absolute value of correlation coefficient over ''m,'' with active user a. The dataset A is a neighbor dataset, and represents user group A.

$$A=\{i \mid |\rho_{ai}| \geqq m\}$$

6: The system makes user(j) dataset ''B,'' which has an absolute value of correlation coefficient over ''c'' and under ''m'' (m > c), with active user a. The dataset B is a non-neighbor dataset, and represents user group B

$$B=\{j \mid m > |\rho_{aj}| \geqq c\}$$

7: The system makes a user(j) dataset Γ, which has an absolute value of correlation coefficient over ''m,'' with user i. The dataset Γ represents user group C.

$$\Gamma=\{j \mid |\rho_{ij}| \geqq m\}$$

8: The system makes a user(j) dataset Δ, included in both user datasets B and Γ.

$$\Delta=\{j \mid j \in B,\ j \in \Gamma\}$$

9: The system makes a user dataset ''Ex,'' which contains users in dataset Δ who have already rated their interest in news article x that active user a did not rate.

10: The system weighs rating scores on item x by users included in the dataset Ex by a correlation coefficient of rating scores between active user a and users included in the user dataset Ex.

11: The system predicts the rating score $\hat{s}_{aj}$ by active user a, using a weighted rating score calculated in step 10. The system calculates a predicted rating score by weighted average of rating score for item j using the following formula:

$$\widehat{S_{ai}} = \overline{S_a} + \frac{\sum_{i \in x_j} \rho_{ai} \left(S_{ij} - \overline{S_i'} \right)}{\sum_{i \in x_j} |\rho_{ai}|}$$

12: The system calculates the predicted rating score s_x on items that active users a did not rate, and makes a recommendation list that is based on s_x.

This study set 0 as c, and 0.3 as m. We decided these scores by distribution.

We expected that the system would make recommendations that match a learner's preferences to a certain extent but would then go beyond the boundaries of the learner's interests.

Vocabulary definition and marker functions

When a user places the cursor on an underlined word in a news article, its definition (first definition) is displayed. By clicking the ''detail'' button, the user can also view further detailed definitions of the word.

Table 1 Sample of words in each level of JACET 8000

Level	Sample word (three words in each level)
1	The
	And
	To
2	Audience
	Article
	Cultural
3	Interpret
	Disabled
	Significant
4	Senator
	Chapter
	Investment
5	Boom
	Respondent
	Hazard
6	Heap
	Accumulation
	Indulge
7	Exemplify

8	Arbitration Coarse Ensemble Mythology Inconsistency

The underlined words are determined on the basis of the results of an English vocabulary test taken by the user in advance. The words follow JACET 8000 (Ishikawa et al. 2003; Uemura and Ishikawa 2004), a vocabulary glossary created by the Japan Association of College English Teachers (JACET), which separates words into eight levels according to frequency of use in several corpuses (Uemura and Ishikawa 2004). JACET 8000 is a standardized vocabulary level measurement for English learners at the college level. Table 1 shows the sample of vocabularies for each level. Depending on the learner's test results, this system underlines and displays meanings of vocabulary that are listed at a higher JACET level than learner's JACET vocabulary level. Clicking the ''detail'' button enables learners to make lists of unknown words as well. When a learner finds an unknown word, he or she can create such a list by clicking the button on the vocabulary definition display. This system records the word and the news headers that learners have read.

In the article details screen, learners can mark any sentence or word. These marks can be saved and displayed whenever the same article is reopened. These functions were developed according to cognitive learning strategies, such as underlining and motivating with learning materials in which learners have an interest, that have been reported to be effective for successful learning experiences (Garcia and Pintrich 1994). This study first designed and developed marking and vocabulary definition as support functions for using cognitive learning strategies such as elaborating by highlighting and using a dictionary.

Comments and rating functions

This function allows learners to insert Japanese comments into the articles they read. With this function, a learner can share thoughts and opinions on the articles with other learners.

At the bottom of each article details screen, learners can rate the perceived interest in each news article. Collaborative filtering works using this rating data. Learners can find a list of all users who rated that article. When a user's name is clicked, other articles that he/she rated will be displayed. This function, related to interaction among learners, aims to motivate learners to understand content and to provide active feedback. White (2003) suggests that social learning support, such as feedback among learners, promotes self-regulated learning. After the active learner clicks the article reader icon, this system shows a list of the readers of each article as well as a list of news articles that readers have already read. Learners can also comment on each article. This function seems to reduce isolation and to encourage readers by making them aware of similar learners.

Experiment

Content of experiment

The purpose of this experiment was to observe how the alteration in the algorithm influences learning. To conduct a controlled experiment, the algorithm of Resnick et al. (1994), known as a basic user-based collaborative filtering algorithm, was used.

In order to minimize the probability of bias caused by individual differences, the two recommendation algorithms were applied randomly to learners. Hereafter, the group that studied English news articles recommended by the user-based collaborative filtering algorithm developed for this study will be called the experimental group, and the group that studied English news articles recommended by the basic user-based collaborative filtering algorithm will be called the control group. There were no differences between the control and experiment groups in terms of the function of the system except for the collaborative filtering algorithm.

English news articles used in the experiment

The English news articles used in this study were acquired from Kyodo News. The articles in Kyodo News include tags that separate them into different categories. After considering the number and content of news articles in a month, we identified seven categories: politics, economics, accidents, science and technology, sports, Asian issues, and society. No news items were tagged with the topic categories ‘‘Trends’’ or ‘‘Others.’’

Experiment procedure

The period of registration was 1 week. Learners were randomly assigned to the control group or the experimental group, and they began studying sometime during this period. During registration, learners took an online English vocabulary test based on JACET 8000, which consisted of 80 questions. As previously mentioned, this system chooses which vocabulary to display according to the results of this vocabulary test. This system asks learners to rate the degree of interest on distributed news topics because new articles cannot be recommended without learners’ interest data at the beginning of using collaborative filtering; this is referred to as a ‘‘cold start’’ problem (Drachsler et al. 2008). Thus, this system needs to collect learners’ interest data before collaborative filtering can work. After the learner registration, this system recommends news articles for 1 week in order to collect learners’ interest data, using the rating data with which learners rated new articles at the beginning (Content-based Recommendation). After 1 week, the algorithm was changed. For the next 2 weeks, both groups studied from recommendations made by the algorithm of Resnick et al. (1994). Then, for 1 month, the experimental group studied from recommendations made with the user-based collaborative filtering algorithm developed in this study, while the control group continued with recommendations made by the algorithm of Resnick et al. (1994). The week following the one-month experiment was designated for collecting data. During this time, learners completed a post-questionnaire and took an 80-item English vocabulary test based on JACET 8000; this was the same test they took before the experiment began. They also took an online 42-item English vocabulary test on each news category; the test included six characteristic words from the news articles of each category.

Subjects

First- to fourth-year undergraduates and graduate students from multiple universities and graduate schools participated in this experiment. There was a total of 374 registered members (the age range was 18–25; average 19.53). The native language of all subjects is Japanese. They had studied English since junior high school. They were required to select their TOEIC score level range. Subject number of each TOEIC level was displayed in Table 2. After the random division of subjects into two groups by the system, the number of subjects in the control group was 189 (male: 45, female: 130, no answer: 14, average age: 19.65), and the experiment group was 185 (male: 39, female: 125, no answer: 21, average age: 20.00). The final stage included 116 learners who had provided complete data (including pre-test, post-test, and questionnaire data).

Table 2 Subjects' TOEIC levels

Not take TOEIC	Less than 220	220–465	470–730	735–855	Over 860	Total
166	3	105	78	12	10	374

Table 3 Differences between the experimental and control groups in terms of their Toeic levels

	Not take TOEIC	Less than 470	Over 470	Total
Control group	19	16	10	45
Experimental group	17	17	19	53
Total	36	33	29	98

df = 2, $v^2 = 2.30$, $p = 0.317$, ns

As this study focused on vocabulary learning, learners who had a high vocabulary score at the starting point were excluded from the final analysis. Learners who attained less than 50 correct answers in the English vocabulary pre-test were included in the analysis; this score is less than one standard deviation from the overall average score. The number of learners included in the final analysis was 53 from the experimental group (15 males, 38 females) and 45 from the control group (15 males, 30 females).

In order to investigate the differences between the experimental and control groups in terms of their TOEIC levels, a Chi square test was conducted. Before the Chi square test, we condensed five TOEIC levels into two (less than or over 470), because we found cells containing 0. We found a non-significant difference between the experimental and control groups. Table 3 displays these results.

Data used in analysis

The results of the English vocabulary pre- and post-test based on JACET 8000 (80 questions), results of the category-based English vocabulary post-test (42 questions), and answers to the post-questionnaire about perceived fit rate and usability of both systems, were used in the analysis. Table 4 shows the questionnaire items.

Results

Analysis of a simulation test

We conducted a simulation test using dummy data in order to clarify the difference between the algorithm developed by this study and GroupLens in terms of predicted scores. The simulation program predicts the rating score of 20 items that 150 users will rate, using the rating score of eight items that were rated by 1,350 dummy users. We calculated the average product-moment correlation coefficient between the algorithm and GroupLens. The result of the calculation of the average correlation coefficient was 0.798 (S.D. 0.187), and a 95 % confidence interval was from 0.768 to 0.829. These results revealed the similarity between the two algorithms to some extent, but that the recommended items by these algorithms do not match perfectly. This means that these algorithms do not recommend the same items to the same active users.

Table 4 Post-questionnaire items

#	Items
1	The numbers of recommended news articles were too high for me
2	Rating my interest in each news article was intrusive for me
3	The function that displays the meaning of a word by pointing the cursor to an underlined word was helpful to read news articles
4	Marking words and sentences was easy to use for me
5	Comments from other learners were helpful for me to comprehend the content of news articles
6	This system was effective for the improvement of my English vocabulary proficiency

Analysis of access frequency

As described in section ''Target Group,'' about two thirds of participants did not continue to learn English vocabularies. First, we examined the difference in number of dropped-out learners between the experimental and control groups, including the subjects who had a pre-test score of over 50. Pearson's Chi square test was conducted for this analysis, and revealed no significant difference between the control and experiment groups in terms of discontinuation. This result is displayed in Table 5.

Analysis of usability

The only difference between the experiment and control groups in terms of the system function is the use of collaborative filtering algorithm. The results of the *t* test showed that there were no significant differences in usability of the system between the experiment and control groups. Table 6 displays the results of usability.

Analysis of response to news recommendations

To examine the influence of the recommendation algorithms on the relationship between the learners' interests and recommended English news articles, we analyzed the post-questionnaire data. The questionnaire included two fields: interesting news articles that were recommended and interesting news articles that were not recommended. For each field, learners selected an answer on a scale of five (1 = *Yes, very much* and 5 = *No, not at all*). After the rating of the second question was reversed, the ratings of the two questions were summed up in each group as the perceived fit rate of recommendation. Cronbach's alpha between the two variables is 0.72, and correlation is 0.56.

Table 5 The results of Pearson's Chi square test of the difference in discontinuation between the control and experiment groups

	Discontinuation	Accomplishment	Total
Experiment group	124	61	185
Control group	134	55	189
	258	116	374

$df = 1$, $\chi^2 = 0.66$, ns $p = .42$

Table 6 The results of a t-test of the difference in usability between the control and experiment groups

	Experimental group		Control group		*d.f.*	*t*
	M	*S.D.*	*M*	*S.D.*		
The numbers of recommended news articles were too much for me	2.79	1.01	2.98	.97	96.59	.92
Rating my interest in each news article was intrusive for me	3.79	.82	3.86	.76	97.27	.46
The function that displays the meanings of words by pointing the cursor to an underlined word was helpful to read news articles	1.79	.77	1.95	1.02	82.26	.88
Marking words and sentences was easy to use for me	2.47	1.17	2.73	1.40	87.67	.99
Comments from other learners were helpful for me to comprehend the content of news articles	2.51	1.20	2.64	.98	97.82	.61
This system was effective for the improvement of my English vocabulary proficiency	2.09	.74	2.20	.72	96.01	.71

Table 7 Perceived fit rate of recommendation in collaborative filtering developed in this study (experimental group) and grouplens (control group)

	Experimental group		Control group		df	t	d
	M	S.D.	M	S.D			
Perceived fit rate of recommendation	5.79	1.58	6.04	1.62	94.14	.79	.16
The first question	2.87	.91	3.11	.91	94.95	1.33	.27
The second question (reversed)	3.08	.88	3.07	.91	93.93	.06	.01

Table 8 Average number and standard deviation of access logs by learners in each group

Group	M	SD	t
Experiment group	18.40	12.80	$t(96) = 0.37$, ns
Control group	17.36	14.80	

We applied Welch's t test to the results of three items, namely the first question, second question (reversed), and summed up score of the two questions, in order to increase the validity of evaluation of perceived fit rate. The results showed no significant difference between the experimental and control groups in response to the three items. Hence, it can be inferred that making recommendations using the user-based collaborative filtering algorithm developed in this study does not significantly decrease the level of accuracy of the recommendations. This research aims to improve topic-specific vocabulary proficiency, keeping the same perceived fit rate as that of GroupLens. The perceived fit rates of this system and GroupLens are almost the same, despite the low rating score on the first point. This result is displayed in Table 7.

Analysis of utilization status

To examine the influence of the recommendation algorithm on the articles read by the learners, we analyzed the access logs and the numbers of news articles rated by the learners. The results are displayed in Table 8 for the access log and Fig. 5 for the number of rated news articles.

The average number of articles rated by the experimental group was 87.56, while the number rated by the control group was 88.93. There was no significant difference in the Welch's t test results (t (80.70) = .08, n.s.). The comparison of the average number of rated articles by category also showed no significant difference between the experimental and control groups. The comparison of standard deviations of the number of articles rated by category by both groups showed a significant difference in the sports and politics category ($p < .05$) and a weak tendency in the Asian issues category ($p < .10$). The results of the Welch's t-test and variance-comparison test are displayed in Tables 9 and 10.

The following conclusions can be made from this analysis. As the t-test results for the total and each category's numbers for learners' interest ratings in the news articles show, we did not confirm a significant difference between the experiment and control groups. These results indicated that the difference of the algorithm did not affect the total number of rated news articles. However, the results of the variance-comparison test revealed that subjects in the experiment group read various news articles on several topics because of the significant difference between the variance-comparison test results of each group. The algorithm developed in this study seems to reduce deviation of read news articles.

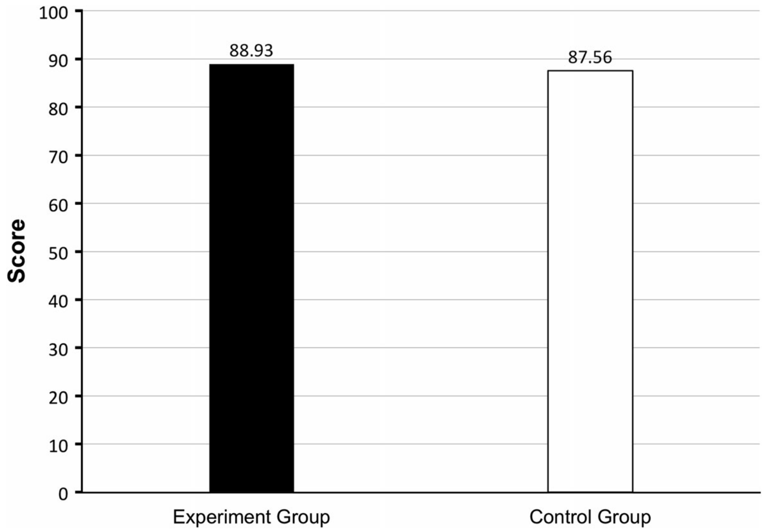

Fig. 5 Average number of recommended articles evaluated by learners

Table 9 Average number of recommended articles evaluated by learners (category)

	M				
	Experimental group	Control group	df	t.	Sig.
Politics	16.64	18.13	81.49	.35	
Economics	12.58	12.93	87.18	.13	
Accidents	9.75	8.18	91.32	.78	
Science and technology	6.04	4.71	97.18	1.00	
Sports	11.25	17.73	55.19	1.31	
Asian issues	8.79	8.96	83.55	.07	
Society	15.64	12.53	97.58	1.21	

Table 10 Standard deviation of recommended articles evaluated by learners (category)

	SD			
	Experimental group	Control group	F	Sig.
Politics	17.50	23.60	$F(44,52)=1.82$	*
Economics	11.10	13.45	$F(44,52)=1.47$	
Accidents	9.36	10.42	$F(44,52)=1.24$	
Science and technology	6.74	6.28	$F(44,52)=.87$	
Sports	11.89	31.37	$F(44,52)=6.96$	***
Asian issues	9.19	11.92	$F(44,52)=1.68$	
Society	13.28	12.05	$F(44,52)=.82$	

*** $p < 0.001$, * $p < 0.05$, $p < 0.1$

Analysis of study results

To examine the influence of the algorithms on the study results, we analyzed the scores of the category-based English vocabulary test and the JACET 8000 English vocabulary test.

Analysis of category-based English vocabulary test scores

To examine the bias in individual strengths, the category-based English vocabulary test scores were analyzed. This test consisted of 42 questions (six questions 9 seven categories). The number of correct answers was calculated for each category and compared to the average scores. There was a 5 %-level significant difference in the science and technology ($t(84.91) = 2.23$, $p < .05$, $d = .46$), politics ($t(91.53) = 2.12$, $p < .05$, $d = .43$), and accidents categories ($t(94.96) = 2.37$, $p < .05$, $d = .48$), as well as a tendency in the Asian issues category ($t(85.63) = 1.88$, $p < .10$, $d = .39$). Figure 6 shows these results.

In contrast, there was no significant difference between the experimental and control groups regarding the standard deviation of correct answers in each category.

These results show that the experimental group had overall higher scores in the English vocabulary test category when compared to the control group.

Analysis of JACET 8000 English vocabulary test scores

To examine the overall study results, we analyzed the scores of the JACET 8000 English vocabulary test displayed in Fig. 7. While there was no significant difference in scores on the JACET 8000 English vocabulary pre-test (the experimental group: $Mean = 39.09$, $SD = 7.79$; the control group: $Mean = 37.84$, $SD = 6.36$; $t(97.83) = .87$, n.s., $d = .17$), the average scores on the JACET 8000 English vocabulary post-test showed a significant difference (the experimental group: $Mean = 43.83$, $SD = 9.35$; the control group: $Mean = 40.07$, $SD = 8.85$; $t(96.87) = 2.04$, $p < .05$, $d = .41$). When the effect of scores on the pre-test was statistically controlled using a multiple regression analysis, the group treatment had only a marginal effect on the scores in the post-test ($b = 3.05$, $SE = 1.67$,

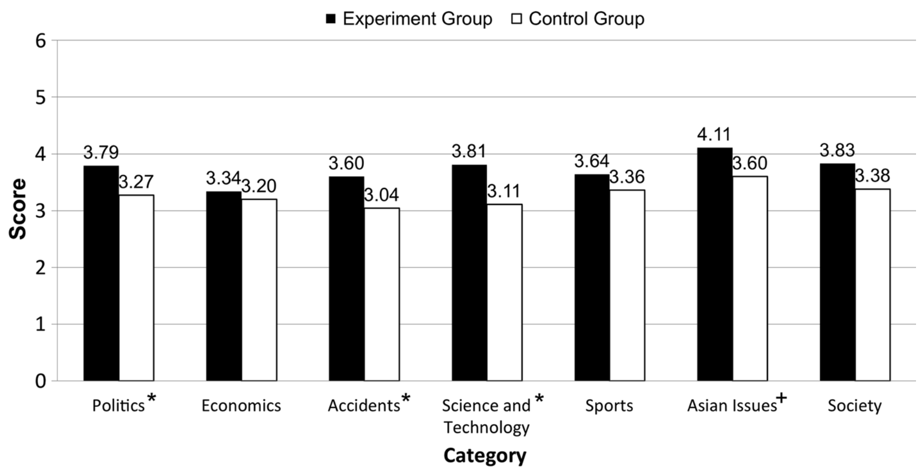

Fig. 6 Average score of the category-based English vocabulary test ($* p<0.05$, $^{+} p < 0.1$)

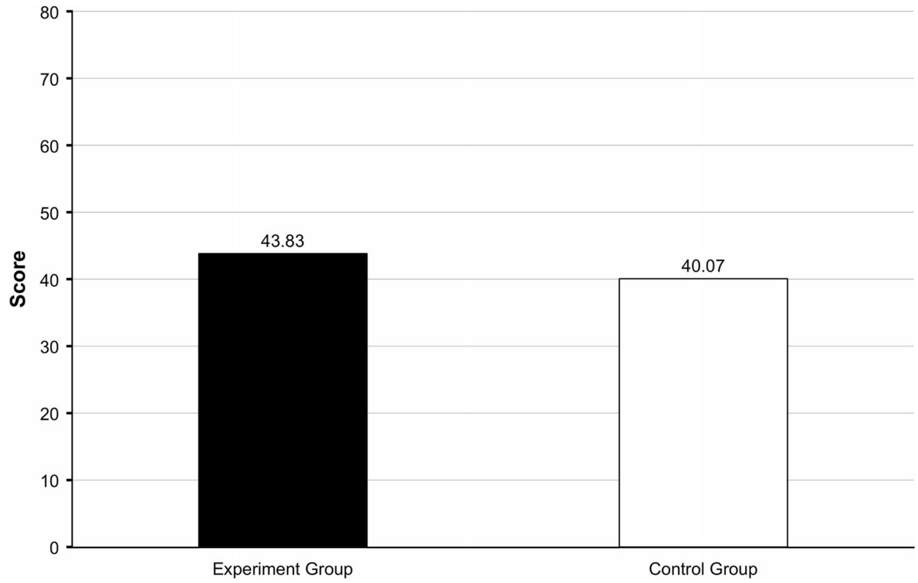

Fig. 7 Average score of JACET test ($p \searrow 0.05$)

$\Delta R^2 = .03$, $F(1, 95) = 3.34$, $p < .10$). Additionally, when the TOEIC score's effect was statistically controlled using a multiple regression analysis, the group treatment had a marginal effect on the scores in the post-test ($b = 3.16$, $SE = 1.83$, $\Delta R^2 = .03$, $F(1, 94) = 2.97$, $p < .10$). As the members in each group were selected randomly, the difference in the post-test can be considered to be caused by the difference in group treatment. These results show that the experimental group had a marginally higher learning performance than the control group.

Discussion

This study considered the effect of a new collaborative filtering system on learning as compared with previous collaborative filtering known as GroupLens. The researchers hypothesized that receiving recommendations from the collaborative filtering algorithm developed in this study allows learners to read a wide range of English news articles and possibly leads to higher learning results. In order to explain the effects of the collaborative filtering algorithm developed in this study, we posed four research questions:

(1) Is there a difference in the fit rate of recommendation between the collaborative filtering developed in this research and GroupLens?
(2) Does this collaborative filtering promote diversity in topics of news articles?
(3) Does the system tend to recommend specific topics in order to promote specific vocabulary acquisition?
(4) Overall, does this system contribute to the improvement of vocabulary test scores?

Concerning RQ1, the results of the comparison data by simulation indicated the difference of estimation of learner's interest by correlation analysis. However perceived fit of news article recommendations showed no significant difference between the two systems, indicating that the collaborative filtering developed in this study did not seem to recommend news articles in which learners had no interest. Significant differences between the control and experimental groups were not found in terms of the number of read news articles and access times. Considering the average correlation between the algorithm developed in this study and GroupLens, it seems too small of a difference to affect any change in learners' perceptions, though each algorithm recommends different items to learners. The difference in recommended items seems too small to impact their learning, considering that there is no significant difference in terms of frequency of access to the systems.

Concerning RQ2, the average numbers of rated articles in both groups displayed no significant differences. However, according to the results of the variance-comparison test, subjects in the experiment group read more various news articles than those in the control group did. The system distributed various topics of news articles, and learners seemed to have interest in the middle level, given the estimation results of learners' interests. Therefore, the subjects read various topics of news articles. In terms of RQs 3 and 4, the average score for several genres using the collaborative filtering system developed in this study is significantly higher than that of GroupLens for vocabulary proficiency. However, not only on the category-base test, but also on the overall vocabulary test, learners using the collaborative filtering developed in this study scored significantly higher than those using GroupLens. For the learning environment of this study, only the recommendation algorithms were different for the experimental group and control group; the interface was the same, and no assistance for their learning was provided without the system. This system recommends news articles close to but not exactly matching learners' preferences and interests, using similar learners' preference data. This function seems to allow learners to read news articles on various topics and acquire specific vocabulary, thus improving overall vocabulary proficiency. These results indicate that the collaborative filtering developed in this study contributes to the enhancement of vocabulary proficiency with the support of ''top-down'' processing (Anderson and Lynch 1988) through learning with news articles closely related to the learners' interests. The several previous research studies, which investigated the relationship between recommendation algorithms and learning performance, found positive effects on the improvement of learning performance. In English education settings, Hsu et al. (2013) developed a mobile-based language learning system with a recommendation algorithm that uses learners' proficiency and preference with a shared annotation function. The authors evaluated its effects on a performance test compared with a recommendation algorithm with an individual annotation function. The results reveal that the recommendation algorithm is effective on the learning performance, but a significant difference related to annotation type in the effects on learning performance was not found. It is difficult to compare these results with those found in the current study because the evaluation criteria were different. Nevertheless, both studies found a positive effect on the improvement of language learning performance.

A different point involves evaluating content-specific performance. Depending on the design of the recommendation algorithm, it is possible to recommend various topics close to learners' preferences and improve language learning performance. In order to investigate this point, this research evaluated the number of news articles and vocabulary proficiency in each topic. However, we did not find significant differences of standard deviation in several categories. The system used with the experiment group allows subjects to read various topics; however, this depends on the range of learners' interests. Future research should consider methods for estimating the range of learners' preferences in order for learners to find topics closer to their preferences.

Importantly, two thirds of all subjects dropped out, though there was no significant difference between the control and experiment groups in terms of the access log. Input information relevant to learners' background knowledge has a relationship with motivation (e.g., Keller 2010; Manolopoulou-Sergi 2004), as mentioned previously. The results of this study reveal that the effect of input relevant to learners' preferences is limited in cases where learners continue to learn.

As Manouselis et al. (2010) indicated, in order to understand ''interest'' in educational settings, discussion should focus on what kinds of ''variables'' influence learners' interests for a recommendation algorithm. This study revealed that recommendations that exactly match learners' interests are not always ideal in some educational settings, such as language learning. When new technology is used for educational purposes, it is important to consider how the technology was developed, how the educational and learning contexts were developed, and how the technology should be applied to the education field. Though this research focused on English vocabulary learning, the findings are applicable to other learning areas. Recent educational movements focus on learning analytics for OER such as the use of MOOC (Massive Open Online Courses). The common problems in OER environments are high drop-out ratios and satisfying learners (Irvine et al. 2013). In this situation, it is important to consider how to support learners' choice of courses or learning material from a massive pool of learning materials. In order to do that, it should be required to give learners the appropriate information for their courses. Recommendation algorithms can be applied to this research area. One of effective ways to give information related to courses is to provide recommendations, predicting learners' interests or previous knowledge, but also providing opportunities to take various courses and increase learners' curiosity and interests. Doing so would help improve learners' success in massive online learning environments. The recommendation algorithm developed in this study can contribute to supporting effective massive learning environments by incorporating recommendations that are based on accurate predictions of learners' interests. A recommendation algorithm such as GroupLens can be appropriate to support learning at the beginning phase, in order to motivate learners and store accurate data about learners' preferences. In the next step, the recommendation algorithm developed in this study can be applied to extend learners' interests by using their accurate preference data.

Recommendation algorithms such as collaborative filtering represent one type of personalization systems. Current personalization systems aim to suggest the information that individuals will want. Therefore, such a system should have a high fit rate with the degree of users' preferences. However, personalization is a system that constructs an information environment by using ''unknown information'' that resembles a user's prior knowledge.

Curiosity is "a form of cognitively induced deprivation that arises from the perception of a gap in knowledge or understanding" (Loewenstein 1994, p. 75). Several studies have also indicated that the strength level of curiosity determines the strength of learning motivation (Kang et al. 2009; Litman 2005; Rotgans and Schmidt 2014). Personalized information environments may not provide information that arouses users' curiosity because users may draw on their previous knowledge to find information that they comprehend. From the viewpoint of the social sciences, there is a concern that personalization limits the information environment not just in education, but in daily life as well (e.g., Sunstein 2001). The results of this research support this concern. In order to promote successful use of OER such as MOOCs, developing learners' curiosity seems to be one of key factors to support learners' selections of OER for their learning.

This study focused on the possibility that when data processing technologies are considered for educational purposes, ''optimization'' of data processing does not necessarily lead to ‘‘optimization’’ of learning. This study examined the effect of collaborative filtering in the area of language learning. The features of educational domains and contexts should be considered for the design and use of advanced technologies such as collaborative filtering. This study suggests the necessity of considering the ideal situation, process, and learning goal when endeavoring to apply information technologies to learning support systems.

Conclusion

This study aims to examine the effect of collaborative filtering that predicts the degree of learner interest in unread English news articles and compare it to that of general collaborative filtering, which aims to predict the accurate fit rate of learner interest. The results of this research reveal that the system developed in this study allowed learners to read various genres of English news articles and to improve their vocabulary proficiency even though there was no difference between the collaborative filtering developed in this study and the general filtering system in the perceived fit rate of recommendations, despite the difference in the simulation test between the algorithm developed in this study and GroupLens. From the view of learning behavior, there was no significant difference between the control and experiment group in terms of the access log. The algorithm developed in this research did not seem to de-motivate subjects to read English news articles, according to the results of this research.

However, the perceived fit rate in both systems did not seem to be high, because of the small data set (about 300 learners) needed in order for collaborative filtering to work more effectively. It is necessary to increase the number of participants in future research for more precise recommendations. Mixing other personalization methods such as the ephemeral personalization method with this algorithm could be effective in improving the perceived fit rate of recommendation (Schafer et al. 2001). Recommending articles that users who read the same article previously read is one ephemeral personalization method. Article content analysis using text-mining and co-occurrence of words in news articles to calculate the similarity between news articles is one possible way to improve the perceived fit rate of recommendation. Learners’ proficiency levels also should be considered. In this research, we focused on vocabulary proficiency levels, although using other parameters concerning learners’ proficiency levels such as the length of news articles could also improve the perceived fit rate in terms of learners’ proficiency levels.

The analysis of the number of read new articles should be conducted using an accurate number of news articles. In this research, the system did not trace the number of clicks on news URLs, but the number of rated news articles. However, rating the interest in news articles can be a reliable type of data that indicates ‘‘subject read news articles.’’ It would be difficult to regard ‘‘clicking the URL of each news article’’ as meaning ‘‘reading news article.’’

Other psychological factors such as motivation should also be considered in future research. This study focused on learners’ preferences in order to design and develop an effective collaborative filtering algorithm for learning. However, psychological factors also play an important role in the use of learning systems. As mentioned above, two thirds of all subjects dropped out, which seemed to promote motivation. In future research, we should investigate the relationship between psychological factors, system use, and learning performance.

Another limitation of this study involves the gender issue in language learning. Many previous studies have indicated that gender differences have an influence on learning styles (e.g., Allakbari and Tazik 2011; Grenfell and Harris 2012), and learning performance (e.g., Cochran et al. 2010). Learners' gender is one of factors that affect learning; therefore, this factor should be considered in future studies on language learning. Privacy should be considered in the future work. Preference data is among the most important data regarding one's personal life. The systems used in this research used anonymous names; system design should take privacy protection into consideration by remaining disconnected from external systems to avoid potential personal information leaks.

We will continue to investigate the effects of this algorithm on learning in various learning settings and take into consideration the limitation mentioned above.

Acknowledgments This research was conducted through the Benesse Department of Educational Advanced Technology (BEAT) and managed by the University of Tokyo as a collaborative research project with Benesse Corporation. This work was partly supported by JSPS KAKENHI Grant Number 21300302.

References

Allakbari, M., & Tazik, K. (2011). On the relationship between gender and perceptual language learning styles: the case of Iranian academic EFL learners. *Educational Psychology: An International Journal of Experimental Educational Psychology, 31*(6), 657–674. doi:10.1080/01443410.2011.592275.

Anderson, A., & Lynch, T. (1988). *Listening*. Oxford: Oxford University Press.

Bobadilla, J., Serradilla, F., & Hernando, A. (2009). Collaborative filtering adapted to recommender systems of e-learning. *Knowledge-Based Systems, 22*(4), 261–265. doi:10.1016/j.knosys.2009.01.008.

Burgoyne, K., Whiteley, H. E., & Hutchinson, J. M. (2013). The role of background knowledge in text comprehension for children learning English as an additional language. *Journal of Research in Reading, 36*(2), 132–148. doi:10.1111/j.1467-9817.2011.01493.x.

Cochran, J. L., McCallum, R. S., & Bell, S. M. (2010). Three A's: How Do Attributions, Attitudes, and Aptitude Contribute to Foreign Language Learning? *Foreign Language Annals, 43*, 566–582. doi:10.1111/j.1944-9720.2010.01102.x.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511667343.

Drachsler, H., Hummel, H. G., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model. *International Journal of Learning Technology, 3*(4), 404–423. doi:10.1504/IJLT.2008.019376.

Field, J. (1998). Skills and strategies: Towards a new methodology for listening. *ELT Journal, 52*(2), 110–118. doi:10.1093/elt/52.2.110.

Garcia, T., & Pintrich, P. R. (1994). Regulating motivation and cognition in the classroom: The role of self-schemas and self-regulatory strategies. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulation of learning and performance: Issues and educational applications* (pp. 127–154). Hillsdale: Lawrence Erlbaum Associates.

Gass, S., Mackey, A., & Pica, T. (1998). The role of input and interaction in second language acquisition. *The Modern Language Journal, 82*(3), 299–305. doi:10.1111/j.1540-4781.1998.tb01206.x.

Ghauth, K. I., & Abdullah, N. A. (2010). Learning materials recommendation using good learners' ratings and content-based filtering. *Educational Technology Research and Development, 58*, 711–727. doi:10.1007/s11423-010-9155-4.

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching, 40*(2), 97–118. doi:10.1017/S0261444807004144.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380. doi:10.2307/202051.

Grenfell, M., & Harris, V. (2012). Making a difference in language learning: the role of sociocultural factors and of learning strategy instruction. *The Curriculum Journal, 24*(1), 121–152. doi:10.1080/09585176.2012.744326.

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, 230–237. doi: 10.1145/312624.312682.

Herlocker, J. L., Konstan, J. A., Riedl, J., & Terveen, L. G. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems, 22*(1), 5–53. doi:10.1145/963770.963772.

Hsu, M.-H. (2006). A personalized English learning recommender system for ESL students. *Expert Systems with Applications, 34*, 683–688. doi:10.1016/j.eswa.2006.10.004.

Hsu, C.-K., Hwang, G.-J., & Chang, C.-K. (2013). A personalized recommendation-based mobile learning approach to improving the reading performance of EFL students. *Computers and Education, 63*, 327–336. doi:10.1016/j.compedu.2012.12.004.

Irvine, V., Code, J., & Richards L. (2013). Realigning Higher Education for the 21st-Century Learner through Multi-Access Learning, *MERLOT Journal of Online Learning and Teaching, 9*(2), 172–186. Accessed from http://jolt.merlot.org/vol9no2/irvine_0613.pdf.

Ishikawa, S., Murata, M., Tono, Y., Uemura, T., Mochizuki, M., Shimizu, S., et al. (2003). *JACET8000: A Word List Constructed Using a Scientific Method and its Applications to Language Teaching and Learning, a paper presented at The Third ASIALEX Biennial International Conference*. Urayasu: Meikai University. 2003.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G., McClure, S. M., Wang, J. T.-Y., et al. (2009). The wick in the candle of learning—Epistemic curiosity activates reward circuitry and enhances memory. *Psychological Science, 20*(8), 963–973. doi:10.1111/j.1467-9280.2009.02402.x.

Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. New York, NY: Springer. doi:10.1007/978-1-4419-1250-3.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Harlow: London House.

Liang, G., Weining, K., & Junzhou, L. (2006). Courseware recommendation in e-learning system. In W. Liu, Q. Li, & R. W. H. Lau (Eds.), *International Conference on Web-based Learning 2006, Lecture Notes in Computer Science 4181* (pp. 10–24). doi: 10.1007/11925293_2.

Litman, J. A. (2005). Curiosity and the pleasures of learning: Wanting and liking new information. *Cognition and Emotion, 19*(6), 793–814. doi:10.1080/02699930541000101.

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75–98. doi:10.1037/0033-2909.116.1.75.

Lynch, T. (2011). Academic Listening in the 21st Century: Reviewing a Decade of Research. *Journal of English for Academic Purposes, 10*(2), 79–88. doi:10.1016/j.jeap.2011.03.001.

Manolopoulou-Sergi, E. (2004). Motivation within the information processing model of foreign language learning. *System, 32*, 427–441. doi:10.1016/j.system.2004.04.002.

Manouselis, N., Vuorikari, R., & Van Assche, F. (2010). Collaborative recommendation of e-learning resources: An experimental investigation. *Journal of Computer Assisted learning, 26*, 227–242. doi:10.1111/j.1365-2729.2010.00362.x.

McNeil, L. (2012). Extending the compensatory model of second language reading. *System, 40*(1), 64–76. doi:10.1016/j.system.2012.01.011.

Oxford, R. (1990). *Language learning strategies: what every teacher should know*. Boston: Heinle & Heinle Publishers.

Payne, M. (2011). Exploring Stephen Krashen's 'i ? 1^0 acquisition model in the classroom. *Linguistics and Education, 22*(4), 419–429. doi:10.1016/j.linged.2011.07.002.

Recker, M. M., & Walker, A. (2003). Supporting ''Word-of-Mouth'' social networks through collaborative information filtering. *Journal of Interactive Learning Research, 14*(1), 79–98.

Recker, M. M., Walker, A., & Lawless, K. (2003). What do you recommend? Implementation and analyses of collaborative information filtering of web resources for education. *Instructional Science, 31*, 299–316. doi:10.1023/A:1024686010318.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM conference on computer supported cooperative work*, 175–186. doi: 10.1145/192844.192905.

Rotgans, J. I., & Schmidt, H. G. (2014). Situational interest and learning: Thirst for knowledge. *Learning and Instruction, 32*, 37–50. doi:10.1016/j.learninstruc.2014.01.002.

Sakai, S. (2008). A study on the relationship between developing learner autonomy and ICT in ELT. *Journal of Multimedia Aided Education Research, 6*(1), 46–56.

Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). E-Commerce recommendation applications. *Data Mining and Knowledge Discovery, 5*(1–2), 115–153. doi:10.1023/A:1009804230409.

Sunstein, C. (2001). *Republic.com*. Princeton, NJ: Princeton University Press.

Uemura, T., & Ishikawa, S. (2004). JACET 8000 and Asia TEFL vocabulary initiative. *Journal of Asia TEFL, 1*(1), 333–347.

White, C. (2003). *Language learning in distance education*. Cambridge: Cambridge University Press.

Yamada, M., Kitamura, S., Miyahara, S., Yamauchi, Y. (2009). Vocabulary Learning Environment with Collaborative Filtering for Support of Self-regulated Learning, In J.D. Velásquez. S. A. Ríos, R. J. Howlett & L.C. Jain (Eds.), *Knowledge-Based and Intelligent Information and Engineering Systems, Lecture Notes in Computer Science 5712* (pp. 523–530). doi: 10.1007/978-3-642-04592-9_65.

Masanori Yamada is associate professor in Faculty of Arts and Science at Kyushu University. He is engaged in research and development of computer-mediated communication systems and multimedia language teaching materials. He worked for NTT Comware Corporation where he was engaged in the research on the technology related to moving images from 2000 to 2003. He was a project assistant professor in the Benesse department of Educational Advanced Technology at the Interfaculty Initiative in Information Studies, the University of Tokyo since 2006 to 2010. He received M.A. and Ph.D in Human System Science from Tokyo Institute of Technology in 2005 and 2008 respectively. He is also a recipient of the outstanding young researcher award from Information Processing Society of Japan (2003), and Japan Society for Educational Technology, and best article award from the The Computer-Assisted Language Instruction Consortium (CALICO).

Satoshi Kitamura is currently an associate professor in the Faculty of Communication Studies at Tokyo Keizai University. Before joining in the faculty, he was a project assistant professor in the Benesse department of Educational Advanced Technology at the Interfaculty Initiative in Information Studies, the University of Tokyo. His research interests focus on media environment and human information behavior.

Hideya Matsukawa was born in 1977. He graduated from Osaka University where he earned a Ph.D degree at the graduate program. Dr. Matsukawa's research centers on the development of various systems: a liaison system between a kindergarten and a child's home, LMS for online courses and other educational systems using an electronic bulletin board, and analysis of information involved in those systems using data and text mining technology. He also engages in the research for distance education. He is a recipient of the Best Research Paper Award from the Japanese Society for Information and Systems in Education, and also from the Japanese Society for Educational Technology. He also received the Best Presentation Award from the Japanese Society for Educational Technology.

Tadashi Misono is associate professor in Faculty of Education at Shimane University. He is engaged in research of mathematics education and educational technology. He received Ph.D in Human System Science from Tokyo Institute of Technology in 2009.

Noriko Kitani is researcher in mathematical education department at Benesse Corporation. She had been the coordinator of Benesse department of Educational Advanced Technology project from 2009 to 2010. She was engaged in the research project of the development of mobile-based learning system. She was graduated from the faculty of liberal arts at Tsuda Collage. She received M.A. in engineering from Hosei University.

Yuhei Yamauchi is associate professor, Interfaculty Initiative in Information Studies, The University of Tokyo. He has been engaged in the research that is combined with fieldworks with regard to designing educational environments that make use of information technology. He is also a recipient of the Outstanding Young Researcher Award and the Best Research Paper Award from the Japan Society for Educational Technology.