academicresearchJournals

Vol. 3(1), pp. 7-23, January 2015
DOI: 10.14662/IJALIS2014.046
Copy © right 2015
Author(s) retain the copyright of this article
ISSN: 2360-7858
http://www.academicresearchjournals.org/IJALIS/Index.htm

*International Journal of Academic Library and Information Science*

*Full Length Research*

# Book Recommendation Using Machine Learning Methods Based on Library Loan Records and Bibliographic Information

**Keita Tsuji[1)*)], Fuyuki Yoshikane[2)], Sho Sato[3)], and Hiroshi Itsumura[4)]**

[1, 2, 4] Faculty of Library, Information and Media Science, University of Tsukuba, Kasuga, Tsukuba City, Ibaraki-ken 305-8550, Japan, Phone & Fax: +81-29-859-{1428 [1)], 1346 [2)], 1274 [4)]}
[3] Faculty of Social Studies, Doshisha University, Karasuma Higashi-iru, Imadegawa-dori, Kamigyo-ku, Kyoto 602-8580, Japan. E-mail: min2fly@gmail.com. Phone & Fax: +75-251-3454
[1]*Corresponding author's E-mail: keita@slis.tsukuba.ac.jp

Accepted 30 December 2014

**In this paper, we propose a method to recommend Japanese books to university students through machine learning modules based on several features, including library loan records. We determine the most effective method among the ones that used (a) a support vector machine (SVM), (b) a random forest, and (c) Adaboost. Furthermore, we assess the most effective combination of relevant features among (1) the association rules derived from library loan records, (2) book titles, (3) Nippon Decimal Classification (NDC) categories, (4) publication years, and (5) frequencies with which books were borrowed. We conducted an experiment involving 60 subjects who were students at T University. The books recommended by our candidate methods as well as the loan records used were obtained from the T University library. The results showed that books recommended by the method that employs an SVM based on features (2), (3), and (5) were rated most favorably by subjects. The method outperforms previous book recommendation techniques, such as that proposed by Tsuji et al. (2013), and is comparable in recommendation performance to the website Amazon.co.jp. The results obtained were independent of student grades, NDC categories, and the publication years of books.**

**Keywords—** Book Recommendation; Recommender System; Library Loan Records; Support Vector Machine (SVM); Random Forest; Adaboost

**Cite This Article As**: Tsuji K, Yoshikane F, Sato S, Itsumura H (2015). Book Recommendation Using Machine Learning Methods Based on Library Loan Records and Bibliographic Information. Inter. J. Acad. Lib. Info. Sci. 3(1): 7-23.

## INTRODUCTION

In this paper, we propose a method to recommend Japanese books to university students through machine learning modules based on several features, including library loan records. We determine the most effective method among ones that use (a) a support vector machine (SVM), (b) a random forest, and (c) Adaboost.

Furthermore, we assess the most effective combination of relevant features among (1) the association rules derived from library loan records, (2) book titles, (3) Nippon Decimal Classification (NDC) categories, (4) publication years, and (5) frequencies with which books were borrowed.

The recommendation of books to students by university libraries through Online Public Access Catalogs (OPACs) is important in two ways: (a) the recommended books might help improve students' knowledge of the relevant area and advance their research by providing new information, and (b) the recommended books might draw students to the university library, where they can seek the assistance of other students as well as professional librarians, e.g., in learning commons, where they can study on their own or in groups (Beagle et al., 2006). Being drawn to libraries hence creates numerous educational opportunities and experiences for students.

Despite its importance, scant research has been conducted on book recommendation by university libraries. Chen & Chen (2007), Luo et al. (2009), and Shirgaonkar et al. (2010) have all proposed book recommendation methods without revealing the details of their respective proposals. On the other hand, Whitney & Schiff (2006), Harada (2009), Harada & Masuda (2010), Tsuji et al. (2012), Tsuji et al. (2013), and Chen (2013) proposed relatively detailed methods and tested them through experiments. In this paper, we will focus on the method proposed by Tsuji et al. (2013) for the following reasons. First, Chen (2013) focused on the location of users in libraries in an approach that was markedly different from others in the area. Second, Tsuji et al. (2012) showed that their method is more effective than those proposed by Harada (2009) and Harada & Masuda (2010), and is similar to the method suggested by Whitney & Schiff (2006). Third, Tsuji et al. (2013) showed that their method was more effective than their previous proposal (Tsuji et al., 2012). Therefore, the method proposed by Tsuji et al. (2013) is the most effective book recommendation system available at present. Hence, if we can develop a method that outperforms theirs, it will be beneficial for university libraries as well as students. In this paper, we propose such a method.

As we will subsequently explain in greater detail, Tsuji et al. (2013) proposed a method that uses SVMs based on NDC matches, similarity of words contained in book titles, and association rules derived from library loan records. However, their method has the following three shortcomings:

(A) The method only uses three features for book recommendation – i.e., NDC matches, book titles, and association rules – and ignores two other promising features, i.e., the publication years of books and the frequencies at which they have been borrowed. The former is an important feature because it is possible for relatively new books to be evaluated more favorably by library users. With regard to the latter, books that have been frequently borrowed might be more useful to students than ones that have not. Using them as features for book recommendation might improve the effectiveness.

(B) Tsuji et al. (2013) did not use a machine learning method other than the SVM. As we subsequently explain, some studies have reported that a random forest ensemble learning method is more effective than the SVM, and that Adaboost is comparable in performance to random forest. If we adopt random forest or Adaboost, we might be able to outperform Tsuji et al.'s method (2013).

(C) Tsuji et al. (2013) showed that book recommendation by Amazon.co.jp (henceforth "Amazon") is more effective than their method, and mentioned the possibility to ask Amazon to help recommend books in the university OPACs. However, it is preferable for university libraries to develop book recommendation methods of their own and not rely on other organizations or private companies. Furthermore, Amazon's recommendation method is not available to the public (although a brief outline of it was reported by Linden et al. (2003))[1]. Thus, an open-source recommendation method for university librariesis more advantageous and preferable.

In light of the above ideas, in this paper, we (1) examine the effectiveness of adopting new features for machine learning, (2) assess the performance of other machine learning methods, such as random forest, and (3) compare the effectiveness of our consequent book recommendation method with that of Amazon.

This paper is structured as follows: we survey related research in the next section. We then explain our proposed method and the data used for our experiment. Finally, we report the results of our experiment and discuss their implications.

## Related Studies

In this section, we survey research related to (1) book recommendation by university libraries and (2) machine learning for classification.

### Book Recommendation by University Libraries

Scant research has been conducted on book recommendation systems for libraries. Whitney & Schiff (2006) proposed a recommendation method that uses a weighted graph model similar to the association rule method. Chen & Chen (2007), Luo et al. (2009), and Shirgaonkar et al. (2010) proposed various recommendation methods but did not conduct

[1] The recommendation method reported was quite similar to the method that uses association rules.

experiments to assess their effectiveness.

Harada (2009) and Harada & Masuda (2010) proposed a method that applies collaborative filtering to library loan records, and conducted experiments to determine the parameters. Tsuji et al. (2012) compared the effectiveness of (1) collaborative filtering, (2) association rules based on library loan records, and (3) the Amazon recommendation system, and showed that association rules were more effective than collaborative filtering for book recommendation. However, their results showed that Amazon's recommendation system is more effective than one that employs association rules. Tsuji et al. (2013) examined the effectiveness of book recommendation using SVM based on (a) NDC matches, (b) similarity among book titles, (c) association rules, and (d) abstracts of books described in the BOOK Data base (maintained by the Japan Book Publishers' Association). They showed that an SVM based on (a), (b), and (c) is more effective than that using only association rules, i.e., the method proposed in Tsuji et al. (2012). However, Amazon's recommendation system yet again proved to be more effective than their method.

Chen (2013) proposed a method that detects users' locations inside libraries and recommend books in close proximity to them. Apart from location information, a simple cosine measure between users' queries and book titles was used to recommend books.

## Machine Learning for Classification

Automatic classification can be directly used for recommendation, i.e., by classifying items into ones that should be recommended and ones that should not. In the following, we summarize the studies that compared (1) Adaboost with SVM or random forest, and (2) random forest with SVM.

(1) Breiman (2001) compared the accuracy of random forest with that of Adaboost. He used 20 data sets, the number of features of which ranged from six to 256, and concluded that the two algorithms are comparable. Chan et al. (2001) also showed that the accuracies of SVM and Adaboost are comparable. The sample they used was a land cover map and the number of features used was 24.Chan & Paelinckx (2008) compared the accuracies of random forest and Adaboost to classify ecotopes of land. They used 21-126 bands of airborne hyperspectral imagery as features and found that the accuracy of random forest was almost identical to that of Adaboost. Ishii et al. (2010) proposed a method that uses rough set reducts and *k*-nearest neighbor for automatic classification, and compared its performance to that of SVM and Adaboost. Their results showed that the performance of SVM and Adaboost is comparable. The sample they used consisted of documents from the Reuters news wire. Khan et al. (2010) compared the accuracies of SVM (polynomial kernel), random forest, Adaboost, and other methods for skin detection. They used three features and concluded that random forest is the most accurate learning algorithm. Miao & Heaton (2010) compared the accuracy of random forest with that of Adaboost for ecosystem classification. They used 66 features and concluded that Adaboost is more accurate than random forest.

(2) Longjun et al. (2011) compared the accuracies of SVM and random forest to predict the characteristics of blasting vibration. The number of features and the training data used were nine and 93, respectively. They found that the average predicted error ratio of SVM was lower than the results obtained by random forest. Nitze et al. (2012) compared the accuracies of SVM (radial basis function (RBF) and polynomial kernels), random forest, and neural networks in classifying crop types in satellite images. The number of features used was 10-40 depending on the images. They concluded that SVM with the RBF kernel is the most accurate algorithm. Jia et al. (2013) compared the accuracies of SVM and random forest to predict β-hair pin motifs in proteins. They concluded that random forest was more accurate than SVM. Liu et al. (2013) introduced SVM, random forest, and neural networks to an electronic tongue and, using seven features, compared their accuracy in recognizing the types and brands of orange beverages and Chinese vinegars. They concluded that random forest was the most accurate algorithm. Hasan et al. (2014) compared the accuracies of SVM and random forest in detecting computer network intrusion. They concluded that SVM is slightly more accurate but more time-consuming. The number of features of their dataset was 41.

## Data

In this section, we introduce the data for our experiment, i.e., (1) library loan records and recommended books, (2) subjects, and a book that the subject would like to borrow at the time, and (3) training data for machine learning.

### Library Loan Records and Recommended Books

For this study, we used 2,293,642 loan records at the T University library. The checkout dates of the books ranged from January 2, 2006 to March 31, 2012. Of these records, 999,630 involved books checked out by undergraduate students, 1,294,012 books had been checked out by graduate students and faculty members, and 30,776 books were checked by other types of patrons. A total of 477,668 titles of books were borrowed by 44,571 users. The number of baskets – i.e., the sets of books borrowed together – was 821,771. The same data were used by Tsuji et al. (2013).

For books to recommend, we used 643,676 books in the T University Library (again, the same data were used by Tsuji et al. (2013)). Many books in the T University Library written in English or other non-Japanese languages are not assigned NDC categories, and we did not choose such books for recommendation. This is why the 643,676 books used for our experiment are only part of the total number of books at the T University library.We limited the books for recommendation to those with NDC categories because these have been shown to be effective for recommendation (Tsuji et al., 2013).

### Subjects and a Book the Subject would like to Borrow

We gathered two sets of subjects: (A) one for the main experiment, and (B) one for additional experiments concerning the effectiveness of association rules. In (A), 40 students majoring in Library and Information Science (LIS) at T University participated as subjects in our experiment. They consisted of 15 graduate students, 11 senior-year undergraduate students, and 14 sophomore undergraduate students. For the sake of convenience, we will call these three groups are different in “grades.” For (B), we chose 20 students majoring in LIS at T University, consisting of 10 graduate students and 10 senior-year undergraduate students.

Subjects were asked to provide the title (and other bibliographical information, if necessary) of a book that they would like to borrow from the T University library for research or studying[2] (henceforth “LBOOK” for brevity). This information was used to recommend books based on association rules, similarities between the titles, matches between the NDC categories, and the Amazon recommendation system.

### Training Data

We used as training data the 358 pairs of books that Tsuji et al. (2013) employed for their experiment. These books were composed of two kinds of data obtained as follows:

(1) Tsuji et al. (2012) asked 33 subjects to each provide an LBOOK. Based on the books, they recommended 460 books by using an association rule, and asked the subjects to assess their degree of interest in the recommended books using the criteria used in this study (to be described later). In this way, they obtained 460 pairs of books assessed by the subjects. In this study, the pairs where the recommended book was rated as “2: Very Interested” and “0: Not Interested” were considered positive examples and negative examples, respectively, and were used as part of the training data. The number of positive and negative examples was 59 and 127, respectively.

(2) Tsuji et al. (2013) asked 16 subjects to each provide an LBOOK. By using SVM based on association rules, similarities in book titles, NDC matches, etc., they recommended books to subjects and obtained their evaluation. In the same manner as in (1), they obtained 52 positive examples and 120 negative examples.

We combined the data mentioned from (1) and (2) and used them as training data.[3] The total number of pairs is 358 (=59+127+52+120). The numbers of pairs of positive and negative examples were 111 and 247, respectively.

## RECOMMENDATION METHOD

We adopted the SVM, random forest, and Adaboost machine learning methods. As mentioned above, some studies in the area have concluded that random forest is the most accurate learning method, whereas others have favored SVM or Adaboost. We will compare the accuracies of these methods in order to determine the one most suited to book recommendation.

The features we used for machine learning were as follows:

(a) Confidence and support of association rules (we will explain the details later) between candidate books and the LBOOKs extracted from the library loan records
(b) Similarities of words contained in titles between candidate books and the LBOOKs
(c) Matches/mismatches in the NDC category of candidate books and the LBOOKs
(d) Publication years of the candidate books
(e) Frequencies at which each of the candidate books has been borrowed (henceforth “loan frequencies”)

We selected six books with the highest probabilities to belong to positive examples for SVM, random forest, and Adaboost based on the following combinations of the

[2] Due to the limitation of our bibliographic data, we asked the subjects not to choose books published in and after 2013.

[3] Actually, Tsuji et al. (2013) used data mentioned in (1) as training data as well. Therefore, the training data used by us and Tsuji et al. (2013) are identical.

features described above: (a)(b)(c), (a)(b)(c)(d), (a)(b)(c)(e), and (a)(b)(c)(d)(e). We then recommended these books to the subjects. Therefore, we recommended to each subject six (books)× 3 (machine learning methods) × 4 (combination of features) = 72 books. There were several duplicate results among the recommendation methods, because of which the actual number of recommended books is lower than 72 and different for each subject.

Henceforth, "SVM-APL," "SVM-AxL," "SVM-APx," and "SVM-Axx" represent the recommendation by SVM based on the combination of features (a)(b)(c)(d)(e), (a)(b)(c)(e), (a)(b)(c)(d), and (a)(b)(c), respectively. "RnF" and "Ada" represent recommendations by random forest and Adaboost, respectively, and APL, AxL, APx, and Axx for each of these represent features analogous to those in the case of SVM. For instance, "RnF-AxL" represents the recommendation by random forest based on features (a)(b)(c)(e).

In addition to machine learning, we recommended six books whose NDCs were the same as those of the LBOOKs, and whose titles were the most similar (i.e., the values of (b) were the highest. Henceforth, we represent this method as "Ndc Title"). The purpose was to show that methods that use machine learning are more effective than simpler methods that do not. Furthermore, we recommended books based on Amazon's system to show the effectiveness of our method. In the following subsections, we introduce (1) the details of machine learning, (2) the features used for machine learning, (3) the Amazon recommendation system, and (4) our method for evaluation.

## Machine Learning Methods

We used the following three machine learning methods.

### SVM

An SVM is a representative machine learning method for automatic classification. It chooses several training samples as "support vectors," which maximize the margin between positive and negative examples in the training data. We used LIBSVM ver. 3.12 in our experiment and adopted the L1 soft-margin SVM "C-Support Vector Classification" and the Radial Basis Function (RBF) kernel. We used the easy.py script to obtain the optimal parameters C (the margin parameter that determines the generalization capability) and γ (the parameter that determines the influence of a training instance). We found that the best combinations of C,γ, and the cross-validation accuracies of SVM-APL, SVM-AxL, SVM-APx, and SVM-Axx were ($3.3\times10^{4}$, $7.8\times10^{-3}$, 73.7%), ($8.2\times10^{3}$, $7.8\times10^{-3}$, 73.5%), ($3.1\times10^{-2}$, $1.2\times10^{-4}$, 72.9%), and ($5.0\times10^{-1}$, $5.0\times10^{-1}$, 72.1%), respectively. We also used the-b option to express the probability that a book interests the subject (more precisely, the probability that a book belongs to the positive examples in the training data that had interested past subjects).

### Random Forest

Random forest is a machine learning method for classification that generates several decision trees based on the training data, and classifies the test data by a majority vote of the decision trees (Breiman, 2001). For random forest, we used the "random Forest" package of R, a free software environment for statistical computing and graphics. We first used the "tune RF" function to tune the "mtry" parameter (the number of features from among which the best split was chosen at each node) based on the training data. We found that the optimal "mtry" was "2" for RnF-APL, RnF-AxL, RnF-APx, and RnF-Axx and their out-of-bag (OOB) error rates were 23.5%, 27.1%, 27.4%, and 28.8%, respectively. We then used the "random Forest" function to learn the testing data, and the "predict" function to classify the data and display their probabilities. Furthermore, we used the "importance" function to calculate the degree of importance of each feature.

### Adaboost

Adaboost (short for "Adaptive Boosting") generates several decision trees and assigns each tree a weight based on its ability to correctly classify training samples (Friedman et al., 2000). It classifies the test data by a weighted majority vote (i.e., a decision tree that classifies training samples more accurately has more votes than one that classifies them less accurately). For Adaboost, we used the "ada" package in R. We used the "ada" function for the module to learn the testing data, and the "predict" function to classify them and display their probabilities. We chose discrete boosting and carried out 50 boosting iterations. These options are the default settings of the "ada" function.

## Features for Machine Learning

We used the following five features for machine learning: confidence and support of association rules, similarities in titles, NDC matches, publication years, and loan frequencies.

### Confidence and Support of Association Rules

Books that have been frequently borrowed through

LBOOKs are favorably evaluated. In order to recommend such books, we calculated the confidence and support of the association rules (Agrawal et al., 1993). When a user borrows n books $S_i$ (i = 1, ..., n) at a given time, we call the set {$S_1$, …, $S_n$} a "transaction." If we extract two subsets X and Y from this set (X∩Y=φ), we can say that when a user borrows books of set X also borrows books of set Y. We represent such rule as X → Y, where X and Y are called antecedent and consequent, respectively. For instance, if a user borrows three books, A, B, and C, this transaction can be represented as {A, B, C}. From this transaction, we can extract the rule A → {B, C}, which means "the user who borrows book A also borrows books B and C."Using all transactions carried out by all users, we thus extracted rules that were frequently observed and, in this sense, were useful and reliable. Confidence is defined as the ratio of the number of transactions that contain X and Y to the number of transactions that only contain X. Support is defined as the ratio of the number of transactions containing members of X and Y to the number of all transactions.

### Similarities of Words Contained in Book Titles

Books whose titles were similar to those of one of the LBOOKs are favorably evaluated. In order to recommend such books, we calculated the similarities between titles as follows: (1) The titles of all books in T University library and the LBOOKs are divided into words using the Japanese morphological analyzer MeCab ver. 0.994. (2) Single nouns and two-noun sequences were extracted from strings of words. (3) Vectors consisting of term frequency–inverse document frequency (tf–idf) from the single nouns and two-noun sequences were created for each book. The tf–idf of a single noun or a two-noun sequence S concerning book A is defined as the "frequency of S in the title of book A" ×the natural logarithm of" (the number of all candidate recommendations, i.e., 643,676) / (the number of books whose titles contain S)." (4) The similarity between the titles of the two books is defined as the cosine measure of their corresponding vectors.

### Matches/Mismatches in the NDC Categories

Books whose NDCs are similar to those of one of the LBOOKs are favorably evaluated. In order to recommend such books, we adopted matches/mismatches in the NDC categories as a feature for machine learning. Like Dewey decimal classification, each NDC category is represented by three numbers representing classes, divisions and sections, respectively. For instance, "324" represents a civil law book, where "3" indicates the social sciences class, division "2" indicates "laws," and section "4" indicates "civil law." We assigned to each book "1" as feature value if its NDC category matched the section level of an LBOOK; if not, we assigned it the value"0."[4]

### Publication Years

Recently published books are more favorably evaluated than older ones. In order to recommend such books, we adopted the year of publication of a book as a feature. More concretely, we assigned to each book a natural logarithm of 2014- its publication year. For instance, we assigned ln(6) to a book that was published in 2008.[5] We attempted to directly use publication years, but using the logarithmic value seemed to be more effective in our pilot examination. Confirming this observation by questioning subjects is postponed until future research.

### Loan Frequencies

Books that have been frequently borrowed by students and teachers belonging to the same faculty as a relevant subject in our experiment are favorably evaluated. In order to recommend such books, we adopted as a feature the loan frequencies of books. More concretely, we first counted the number of times (N) that a book was borrowed by students and faculty belonging to LIS from January 2, 2006 to March 31, 2012. We then assigned to each book the value ln(N/Y+1), where Y is defined as follows: (a) 2013 - the publication year of the book, if the book was published in or after 2006, and (b) seven, if the book was published before 2006. For instance, we assigned ln(20/5+1)=ln(5) to a book published in 2008 and borrowed 20 times during the above-mentioned time period. This is the natural logarithm of a rough approximation of the number of times the book was borrowed per year.[67] We attempted to directly use N/Y, but its natural logarithm

[4] Examining the effectiveness of assigning "1" when only the classes match, and when the classes and divisions both match, will be left for future research.

[5] For books whose years of publication were not recorded in the bibliographical data maintained by T University library (it happened sometimes when books were very old), we assigned them the value ln(50).

[6] We assume that a copy of a book was purchased by T University Library shortly its publication.

[7] We should have used frequency per month, although the result would have been similar.

seemed to be more effective in our pilot examination. Confirming this observation by questioning subjects is postponed until future research.

### Amazon

When we click a book on the Amazon website, it recommends other books with the caption "Customers who bought this item also bought ...."Therefore, we first located LBOOKs on Amazon, and then sequentially chose books from its recommendation list, starting from the left,[8] and checked the OPAC of the T University library to see if it had the recommended books. We thus collected six books recommended by Amazon and available in T University library, and recommended them to the subjects (for brevity, we will call these books "those recommended by Amazon").

Unfortunately, the number of books that Amazon recommends for each book chosen by the user is not infinite. Hence, for many LBOOKs, we were unable to recommend six books. In such cases, we recommended to our subjects the books recommended by Amazon. This is why the total number of books recommended by Amazon is lower than 240 (six books×40 subjects), as we show in the next section.

### Evaluation Method

The bibliographical information for the recommended books was shown to the subjects, who were then asked to describe their level of interest in each book using the following five-point scale that was also used by Tsuji et al. (2013): "2: Very Interested," "1: Interested," "0: Not Interested," "x: Have No Idea," and "A: Have Already Bought or Read."

## RESULTS AND DISCUSSION

In this section, we show the recommendation results for different methods in relation to probabilities, grades, and NDCs. We also discuss the reasons for the results by focusing on the years of publication of books, their loan frequencies, association rules, and random forests. All the results were obtained from subject group (A) except a part of those described in subsection 5.8, which were obtained from subject group (B).

### Overall Results

Table 1 shows the results for the book recommendation methods tested in our experiment. We can see, for instance, that 240 books were recommended by SVM-APL, of which 67 were rated "2: Very Interested" by subjects, accounting for 27.9% (= 67/240 × 100) of the recommended books. On the other hand, 150 books were recommended by Amazon, of which 64 were rated "1: Interested," accounting for 42.7% (= 64/150 × 100) of all recommendations.

If "A: Already Bought or Read," "2: Very Interested," and "1: Interested" are considered to be positive evaluations – as they were inTsuji et al. (2013) – the method (excluding Amazon) with the highest percentage of positive evaluations (henceforth "PPE" for brevity) is SVM-AxL, with 80.0% of its recommendations receiving a positive evaluation (= (16 + 65 + 111)/240). This is higher than 71.3% for SVM-Axx (= (7 + 51 + 113)/240), and the difference of population percentages calculated by a statistical Z-test is significant at 0.05. Tsuji et al. (2013) had examined several book recommendation methods and had concluded that SVM-Axx is the most effective. In contrast, our method, which adds loan frequencies as a featureto SVM, was found to be more effective than that proposed by Tsuji et al. (2013). This is among the main findings of this paper.

The PPE of SVM-AxL (=80.0%) is also higher than those of both random forest and Adaboost. Some past studies have reported that random forest is more effective than SVM (Khan et al., 2010, Longjun et al., 2011, Jia et al., 2013, and Liu et al., 2013) and that the effectiveness of Adaboost is comparable to that of SVM (Chan & Paelinckx, 2008, and Ishii et al., 2010). However, SVM was more effective than either method in our experiment. If we adopt the features used in our experiment for book recommendation in each of the above experiments, SVM might outperform the other methods. This is another main finding of this paper, and will be discussed presently.

The PPE of SVM-APL which uses year of publication as a feature for SVM was 77.5% (= (16 + 67 + 103)/240. It was not significantly different from that of SVM-AxL (=80.0%) at 0.05. The publication year might not contribute to improve recommendation performance. We will later examine this point.

Recommendations made by Amazon's book recommendation system recorded an 82.7% PPE (= (20 + 40 + 64)/150). While this is higher than the PPE for SVM-AxL, the difference is not significant at 0.05. Note

[8] The number of recommended books depends on the PC window size (at least when we conducted our experiment in February 2014 with Windows 7 and 8) and if we enlarge the window, Amazon adds another recommended book to the right side of the window. Therefore, we regarded that Amazon recommended books in the left more highly than those in the right.

**Table 1.** Overall results

| | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0: Not Insterested | x: Have No Idea | Total |
|---|---|---|---|---|---|---|---|
| SVM-APL | 77.5 | 16 ( 6.7 ) | 67 ( 27.9 ) | 103 ( 42.9 ) | 51 ( 21.3 ) | 3 ( 1.3 ) | 240 |
| SVM-AxL | 80.0 | 16 ( 6.7 ) | 65 ( 27.1 ) | 111 ( 46.3 ) | 44 ( 18.3 ) | 4 ( 1.7 ) | 240 |
| SVM-APx | 70.8 | 3 ( 1.3 ) | 57 ( 23.8 ) | 110 ( 45.8 ) | 68 ( 28.3 ) | 2 ( 0.8 ) | 240 |
| SVM-Axx | 71.3 | 7 ( 2.9 ) | 51 ( 21.3 ) | 113 ( 47.1 ) | 60 ( 25.0 ) | 9 ( 3.8 ) | 240 |
| RnF-APL | 63.3 | 8 ( 3.3 ) | 44 ( 18.3 ) | 100 ( 41.7 ) | 84 ( 35.0 ) | 4 ( 1.7 ) | 240 |
| RnF-AxL | 57.1 | 6 ( 2.5 ) | 48 ( 20.0 ) | 83 ( 34.6 ) | 100 ( 41.7 ) | 3 ( 1.3 ) | 240 |
| RnF-APx | 55.4 | 1 ( 0.4 ) | 50 ( 20.8 ) | 82 ( 34.2 ) | 103 ( 42.9 ) | 4 ( 1.7 ) | 240 |
| RnF-Axx | 57.9 | 4 ( 1.7 ) | 43 ( 17.9 ) | 92 ( 38.3 ) | 92 ( 38.3 ) | 9 ( 3.8 ) | 240 |
| Ada-APL | 63.8 | 5 ( 2.1 ) | 50 ( 20.8 ) | 98 ( 40.8 ) | 80 ( 33.3 ) | 7 ( 2.9 ) | 240 |
| Ada-AxL | 65.8 | 10 ( 4.2 ) | 48 ( 20.0 ) | 100 ( 41.7 ) | 75 ( 31.3 ) | 7 ( 2.9 ) | 240 |
| Ada-APx | 63.8 | 3 ( 1.3 ) | 45 ( 18.8 ) | 105 ( 43.8 ) | 84 ( 35.0 ) | 3 ( 1.3 ) | 240 |
| Ada-Axx | 61.7 | 6 ( 2.5 ) | 48 ( 20.0 ) | 94 ( 39.2 ) | 84 ( 35.0 ) | 8 ( 3.3 ) | 240 |
| NdcTitle | 73.3 | 6 ( 2.5 ) | 54 ( 22.5 ) | 116 ( 48.3 ) | 52 ( 21.7 ) | 12 ( 5.0 ) | 240 |
| Amazon | 82.7 | 20 ( 13.3 ) | 40 ( 26.7 ) | 64 ( 42.7 ) | 24 ( 16.0 ) | 2 ( 1.3 ) | 150 |

that SVM-AxL recommended 240 books while Amazon recommended only 150. If the PPE for the two methods is similar, the method that can recommend more books is preferable. In this sense, SVM-AxL is more effective than Amazon. We will elaborate upon this point below.

The PPE of “Ndc Title” was 73.3% (= (6 + 54 + 116)/240). As previously mentioned, Ndc Title does not use machine learning and makes recommendations based only on NDC matches and similarities in book titles. Table 1 shows that machine learning methods, especially SVM-AxL, are more effective than this simpler method.

### Results Concerning Probabilities

SVM, random forest and Adaboost assign to each book the probability that it interests the subject. If we only recommend books that are likely to interest subjects, our results will change. Based on this idea, we divided the recommended books into three groups: books whose probabilities are (1) greater than or equal to 0.9, (2) no less than 0.8 and less than 0.9, and (3) less than 0.8. The results are shown in Tables 2, 3 and 4, respectively.

We see in Table 2 that the PPE for 146 books recommended by SVM-AxL was 84.2%. We know that the PPE for 150 books recommended by Amazon was 82.7% (Table 1). Therefore, the accuracy of SVM-AxL is comparable to that of Amazon when we recommend books whose probabilities of interesting the subject are greater than or equal to 0.9.

The PPE for SVM-APL was slightly higher than that for SVM-AxL. However, the number of books recommended was only 83, which is far fewer than the number of books recommended by SVM-AxL. With regard to other methods, such as RnF-APL and RnF-AxL, either the number of books recommended was small or the PPE was low for each of them.

### Results Concerning Grades

The results were compiled by dividing the subjects into sophomore undergrad students, senior-year undergraduate students, and graduate students, as shown in Tables 5, 6 and 7, respectively. As the tables show, the PPE values for sophomores and senior undergraduate students were the highest for SVM-AxL at 91.7% and 72.7%, respectively (excluding Amazon). For graduate students, Ndc Title was the most effective method with 75.6% (excluding Amazon). However, SVM-AxL was in second place with a recommendation accuracy of 74.4% and the difference was not statistically significant. Therefore, SVM-AxL seems to be most effective in general regardless of the subjects’ “grade.” At the very least, it is highly effective for sophomore undergraduate students.

### Results Concerning NDCs

We divided the recommended books according to the NDCs of the aforementioned LBOOKs and examined the results. The PPEs for SVM-AxL were the highest of all recommendation methods (excluding Amazon) when the NDCs of the LBOOKs were 0, 1, 2, 4, 5, 6, 7, 8, and 9. When the NDC of a book was 3 (which indicates social sciences), the PPE for SVM-AxL was 81.9%, which followed that for SVM-APL (84.7%). However, the difference was small.

### Results Concerning the Length of Titles

If the titles of the LBOOKs were too short or too long, their topics might not have been adequately represented by the

**Table 2.** Results for books whose probabilities are 0.9 or more

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| 0.9〜 | SVM-APL | 85.5 | 4 ( 4.8 ) | 29 ( 34.9 ) | 38 ( 45.8 ) | 12 ( 14.5 ) | 83 |
| | SVM-AxL | 84.2 | 7 ( 4.8 ) | 44 ( 30.1 ) | 72 ( 49.3 ) | 23 ( 15.8 ) | 146 |
| | SVM-APx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |
| | SVM-Axx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |
| | RnF-APL | 81.3 | 0 ( 0.0 ) | 6 ( 18.8 ) | 20 ( 62.5 ) | 6 ( 18.8 ) | 32 |
| | RnF-AxL | 65.0 | 3 ( 2.9 ) | 27 ( 26.2 ) | 37 ( 35.9 ) | 36 ( 35.0 ) | 103 |
| | RnF-APx | 51.9 | 0 ( 0.0 ) | 11 ( 20.4 ) | 17 ( 31.5 ) | 26 ( 48.1 ) | 54 |
| | RnF-Axx | 70.4 | 1 ( 3.7 ) | 6 ( 22.2 ) | 12 ( 44.4 ) | 8 ( 29.6 ) | 27 |
| | Ada-APL | 64.5 | 0 ( 0.0 ) | 8 ( 25.8 ) | 12 ( 38.7 ) | 11 ( 35.5 ) | 31 |
| | Ada-AxL | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |
| | Ada-APx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |
| | Ada-Axx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |

**Table 3**. Results for books whose probabilities are between 0.8 and 0.9

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| 0.8〜0.9 | SVM-APL | 79.5 | 4 ( 9.1 ) | 16 ( 36.4 ) | 15 ( 34.1 ) | 9 ( 20.5 ) | 44 |
| | SVM-AxL | 73.0 | 2 ( 5.4 ) | 9 ( 24.3 ) | 16 ( 43.2 ) | 10 ( 27.0 ) | 37 |
| | SVM-APx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |
| | SVM-Axx | 64.3 | 1 ( 3.6 ) | 3 ( 10.7 ) | 14 ( 50.0 ) | 10 ( 35.7 ) | 28 |
| | RnF-APL | 62.4 | 8 ( 4.1 ) | 37 ( 19.1 ) | 76 ( 39.2 ) | 73 ( 37.6 ) | 194 |
| | RnF-AxL | 50.4 | 2 ( 1.6 ) | 20 ( 15.7 ) | 42 ( 33.1 ) | 63 ( 49.6 ) | 127 |
| | RnF-APx | 62.5 | 1 ( 0.7 ) | 36 ( 23.7 ) | 58 ( 38.2 ) | 57 ( 37.5 ) | 152 |
| | RnF-Axx | 54.5 | 0 ( 0.0 ) | 1 ( 9.1 ) | 5 ( 45.5 ) | 5 ( 45.5 ) | 11 |
| | Ada-APL | 67.0 | 5 ( 2.7 ) | 39 ( 21.1 ) | 80 ( 43.2 ) | 61 ( 33.0 ) | 185 |
| | Ada-AxL | 69.7 | 8 ( 4.0 ) | 43 ( 21.4 ) | 89 ( 44.3 ) | 61 ( 30.3 ) | 201 |
| | Ada-APx | 67.2 | 1 ( 0.6 ) | 36 ( 20.3 ) | 82 ( 46.3 ) | 58 ( 32.8 ) | 177 |
| | Ada-Axx | — | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 ( — ) | 0 |

**Table 4.** Results for books whose probabilities are less than 0.8

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| 〜0.8 | SVM-APL | 70.8 | 8 ( 7.1 ) | 22 ( 19.5 ) | 50 ( 44.2 ) | 33 ( 29.2 ) | 113 |
| | SVM-AxL | 73.7 | 7 ( 12.3 ) | 12 ( 21.1 ) | 23 ( 40.4 ) | 15 ( 26.3 ) | 57 |
| | SVM-APx | 70.8 | 3 ( 1.3 ) | 57 ( 23.8 ) | 110 ( 45.8 ) | 70 ( 29.2 ) | 240 |
| | SVM-Axx | 72.2 | 6 ( 2.8 ) | 48 ( 22.6 ) | 99 ( 46.7 ) | 59 ( 27.8 ) | 212 |
| | RnF-APL | 35.7 | 0 ( 0.0 ) | 1 ( 7.1 ) | 4 ( 28.6 ) | 9 ( 64.3 ) | 14 |
| | RnF-AxL | 60.0 | 1 ( 10.0 ) | 1 ( 10.0 ) | 4 ( 40.0 ) | 4 ( 40.0 ) | 10 |
| | RnF-APx | 29.4 | 0 ( 0.0 ) | 3 ( 8.8 ) | 7 ( 20.6 ) | 24 ( 70.6 ) | 34 |
| | RnF-Axx | 56.4 | 3 ( 1.5 ) | 36 ( 17.8 ) | 75 ( 37.1 ) | 88 ( 43.6 ) | 202 |
| | Ada-APL | 37.5 | 0 ( 0.0 ) | 3 ( 12.5 ) | 6 ( 25.0 ) | 15 ( 62.5 ) | 24 |
| | Ada-AxL | 46.2 | 2 ( 5.1 ) | 5 ( 12.8 ) | 11 ( 28.2 ) | 21 ( 53.8 ) | 39 |
| | Ada-APx | 54.0 | 2 ( 3.2 ) | 9 ( 14.3 ) | 23 ( 36.5 ) | 29 ( 46.0 ) | 63 |
| | Ada-Axx | 61.7 | 6 ( 2.5 ) | 48 ( 20.0 ) | 94 ( 39.2 ) | 92 ( 38.3 ) | 240 |

term vector. In light of this, we classified the books recommended by SVM-AxL according to the number of nouns in the titles of the LBOOKs. The results are shown in Table 8, where we can see that the PPE of SVM-AxL does not depend on the number of nouns, and that it can make effective recommendation seven if the title of an

**Table 5**. Results for sophomore undergraduate students

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Sophomore Undergraduate Students | SVM-APL | 88.1 | 3 ( 3.6 ) | 32 ( 38.1 ) | 39 ( 46.4 ) | 10 ( 11.9 ) | 84 |
| | SVM-AxL | 91.7 | 2 ( 2.4 ) | 32 ( 38.1 ) | 43 ( 51.2 ) | 7 ( 8.3 ) | 84 |
| | SVM-APx | 79.8 | 0 ( 0.0 ) | 28 ( 33.3 ) | 39 ( 46.4 ) | 17 ( 20.2 ) | 84 |
| | SVM-Axx | 71.4 | 1 ( 1.2 ) | 24 ( 28.6 ) | 35 ( 41.7 ) | 24 ( 28.6 ) | 84 |
| | RnF-APL | 71.4 | 1 ( 1.2 ) | 26 ( 31.0 ) | 33 ( 39.3 ) | 24 ( 28.6 ) | 84 |
| | RnF-AxL | 65.5 | 1 ( 1.2 ) | 25 ( 29.8 ) | 29 ( 34.5 ) | 29 ( 34.5 ) | 84 |
| | RnF-APx | 60.7 | 0 ( 0.0 ) | 24 ( 28.6 ) | 27 ( 32.1 ) | 33 ( 39.3 ) | 84 |
| | RnF-Axx | 67.9 | 1 ( 1.2 ) | 19 ( 22.6 ) | 37 ( 44.0 ) | 27 ( 32.1 ) | 84 |
| | Ada-APL | 73.8 | 1 ( 1.2 ) | 21 ( 25.0 ) | 40 ( 47.6 ) | 22 ( 26.2 ) | 84 |
| | Ada-AxL | 76.2 | 1 ( 1.2 ) | 25 ( 29.8 ) | 38 ( 45.2 ) | 20 ( 23.8 ) | 84 |
| | Ada-APx | 67.9 | 1 ( 1.2 ) | 22 ( 26.2 ) | 34 ( 40.5 ) | 27 ( 32.1 ) | 84 |
| | Ada-Axx | 64.3 | 2 ( 2.4 ) | 18 ( 21.4 ) | 34 ( 40.5 ) | 30 ( 35.7 ) | 84 |
| | NdcTitle | 72.6 | 0 ( 0.0 ) | 25 ( 29.8 ) | 36 ( 42.9 ) | 23 ( 27.4 ) | 84 |
| | Amazon | 81.8 | 4 ( 7.3 ) | 14 ( 25.5 ) | 27 ( 49.1 ) | 10 ( 18.2 ) | 55 |

**Table 6.** Results for senior-year undergraduate students

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Senior-year Undergraduate Students | SVM-APL | 71.2 | 1 ( 1.5 ) | 19 ( 28.8 ) | 27 ( 40.9 ) | 19 ( 28.8 ) | 66 |
| | SVM-AxL | 72.7 | 1 ( 1.5 ) | 20 ( 30.3 ) | 27 ( 40.9 ) | 18 ( 27.3 ) | 66 |
| | SVM-APx | 71.2 | 2 ( 3.0 ) | 18 ( 27.3 ) | 27 ( 40.9 ) | 19 ( 28.8 ) | 66 |
| | SVM-Axx | 66.7 | 2 ( 3.0 ) | 14 ( 21.2 ) | 28 ( 42.4 ) | 22 ( 33.3 ) | 66 |
| | RnF-APL | 62.1 | 0 ( 0.0 ) | 12 ( 18.2 ) | 29 ( 43.9 ) | 25 ( 37.9 ) | 66 |
| | RnF-AxL | 65.2 | 0 ( 0.0 ) | 14 ( 21.2 ) | 29 ( 43.9 ) | 23 ( 34.8 ) | 66 |
| | RnF-APx | 54.5 | 0 ( 0.0 ) | 13 ( 19.7 ) | 23 ( 34.8 ) | 30 ( 45.5 ) | 66 |
| | RnF-Axx | 53.0 | 2 ( 3.0 ) | 13 ( 19.7 ) | 20 ( 30.3 ) | 31 ( 47.0 ) | 66 |
| | Ada-APL | 57.6 | 1 ( 1.5 ) | 16 ( 24.2 ) | 21 ( 31.8 ) | 28 ( 42.4 ) | 66 |
| | Ada-AxL | 69.7 | 2 ( 3.0 ) | 13 ( 19.7 ) | 31 ( 47.0 ) | 20 ( 30.3 ) | 66 |
| | Ada-APx | 63.6 | 0 ( 0.0 ) | 10 ( 15.2 ) | 32 ( 48.5 ) | 24 ( 36.4 ) | 66 |
| | Ada-Axx | 60.6 | 2 ( 3.0 ) | 16 ( 24.2 ) | 22 ( 33.3 ) | 26 ( 39.4 ) | 66 |
| | NdcTitle | 71.2 | 2 ( 3.0 ) | 16 ( 24.2 ) | 29 ( 43.9 ) | 19 ( 28.8 ) | 66 |
| | Amazon | 83.8 | 7 ( 18.9 ) | 10 ( 27.0 ) | 14 ( 37.8 ) | 6 ( 16.2 ) | 37 |

**Table 7.** Results for graduate students

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Graduate Students | SVM-APL | 72.2 | 12 ( 13.3 ) | 16 ( 17.8 ) | 37 ( 41.1 ) | 25 ( 27.8 ) | 90 |
| | SVM-AxL | 74.4 | 13 ( 14.4 ) | 13 ( 14.4 ) | 41 ( 45.6 ) | 23 ( 25.6 ) | 90 |
| | SVM-APx | 62.2 | 1 ( 1.1 ) | 11 ( 12.2 ) | 44 ( 48.9 ) | 34 ( 37.8 ) | 90 |
| | SVM-Axx | 74.4 | 4 ( 4.4 ) | 13 ( 14.4 ) | 50 ( 55.6 ) | 23 ( 25.6 ) | 90 |
| | RnF-APL | 56.7 | 7 ( 7.8 ) | 6 ( 6.7 ) | 38 ( 42.2 ) | 39 ( 43.3 ) | 90 |
| | RnF-AxL | 43.3 | 5 ( 5.6 ) | 9 ( 10.0 ) | 25 ( 27.8 ) | 51 ( 56.7 ) | 90 |
| | RnF-APx | 51.1 | 1 ( 1.1 ) | 13 ( 14.4 ) | 32 ( 35.6 ) | 44 ( 48.9 ) | 90 |
| | RnF-Axx | 52.2 | 1 ( 1.1 ) | 11 ( 12.2 ) | 35 ( 38.9 ) | 43 ( 47.8 ) | 90 |
| | Ada-APL | 58.9 | 3 ( 3.3 ) | 13 ( 14.4 ) | 37 ( 41.1 ) | 37 ( 41.1 ) | 90 |
| | Ada-AxL | 53.3 | 7 ( 7.8 ) | 10 ( 11.1 ) | 31 ( 34.4 ) | 42 ( 46.7 ) | 90 |
| | Ada-APx | 60.0 | 2 ( 2.2 ) | 13 ( 14.4 ) | 39 ( 43.3 ) | 36 ( 40.0 ) | 90 |
| | Ada-Axx | 60.0 | 2 ( 2.2 ) | 14 ( 15.6 ) | 38 ( 42.2 ) | 36 ( 40.0 ) | 90 |
| | NdcTitle | 75.6 | 4 ( 4.4 ) | 13 ( 14.4 ) | 51 ( 56.7 ) | 22 ( 24.4 ) | 90 |
| | Amazon | 82.8 | 9 ( 15.5 ) | 16 ( 27.6 ) | 23 ( 39.7 ) | 10 ( 17.2 ) | 58 |

**Table 8**. Results of SVM-AxL by the number of nouns in LBOOKs

| Number of Nouns | Recommended Books | Positively Evaluated Books | PPE |
|---|---|---|---|
| 1 | 6 | 5 | 83.3 |
| 2 | 18 | 16 | 88.9 |
| 3 | 18 | 17 | 94.4 |
| 4 | 30 | 21 | 70.0 |
| 5 | 30 | 23 | 76.7 |
| 6 | 42 | 34 | 81.0 |
| 7 | 36 | 27 | 75.0 |
| 8-11 | 30 | 25 | 83.3 |
| 12+ | 30 | 24 | 80.0 |
| Total | 240 | 192 | 80.0 |

LBOOK is short or too long.

We also classified the books recommended by SVM-AxL based on the number of nouns in their titles. The results are shown in Table 9. We can see that books containing more than four nouns in the title record a significantly lower PPE. Therefore, if we only recommend books whose titles are not too long (more concretely, whose titles contain two to four nouns), the PPE could be higher.

## Results Concerning Publication Years

To observe the effect of publication year on the results, we divided the results according to the years of publication of (1) LBOOKs and (2) books recommended by SVM-AxL. For (1), we divided the recommended books into two groups: those whose LBOOKs were published before 2005 (120 books), and those whose LBOOKs were published in or after 2005 (also 120 books). The results are shown in Tables 10 and 11. We can see that the PPE was the highest for SVM-AxL (excluding Amazon in Table 11).

The accuracy results of a classification according to that of (2) are shown in Table 12. We can see that recommended books that were published before 1990 were less favorably evaluated than those in or after 1990. Therefore, if we only recommend books published in or after 1990, the PPE would be higher.

The averages of publication years (i.e., natural logarithm of (2014 - publication year)) of books recommended by each method are shown in Table 13. We can see that the average of the publication years of the books that were recommended by SVM-AxL and were positively evaluated was 2.42, which was larger than those of SVM-APL, SVM-APx, RnF-APL, RnF-APx, Ada-APL, and Ada-APx (i.e., 2.17, 1.75, 2.14, 2.10, 2.13, and 2.11, respectively). Therefore, SVM-AxL recommended “older” books than these methods did. We expected relatively newer books to be more positively evaluated than older ones. However, SVM-AxL was found to have recommended older books and received high PPE.

**Table 9.** Results of SVM-AxL by the number of nouns in titles of recommended books

| Number of Nouns | Recommended Books | Positively Evaluated Books | PPE |
|---|---|---|---|
| 1 | 9 | 7 | 77.8 |
| 2 | 35 | 30 | 85.7 |
| 3 | 60 | 52 | 86.7 |
| 4 | 51 | 43 | 84.3 |
| 5 | 17 | 11 | 64.7 |
| 6 | 16 | 12 | 75.0 |
| 7 | 17 | 12 | 70.6 |
| 8+ | 35 | 25 | 71.4 |
| Total | 240 | 192 | 80.0 |

## Results Concerning Loan Frequencies

To observe the effectiveness of incorporating loan frequencies, we investigated the averages of loan frequencies of the recommended books. The results are shown in Table 14. [9]For instance, we can see that the average loan frequencies of books recommended by SVM-AxL and evaluated as “2: Very Interested” by the subjects was 1.08. By comparing the results obtained by two methods, one of which uses loan frequencies as a feature and the other does not (all other conditions being equal) – such as the pairs {SVM-AxL, SVM-Axx} and {RnF-APL, RnF-APx} – we can see that the former recommends books whose loan frequencies are high. For instance, while the average of the loan frequencies of books recommended by SVM-AxL and was positively evaluated by the subjects was 1.06, that of the loan frequencies of books recommended by SVM-Axx was 0.33. In general, SVM-AxL recommends books with higher loan frequencies than either random forest or Adaboost. The superior performance of SVM-AxL can be attributed to this difference. On the other hand, Amazon recommended books whose loan frequencies were lower than the loan frequencies of those recommended by SVM-AxL. The reason why Amazon attains a high PPE even though the loan frequencies of its recommended books are low will be investigated in future research.

[9] Due to a lack of recent loan records, we could not calculate the loan frequencies of 75 books recommended by Amazon. Therefore, we used the loan frequencies of the remaining 75 (= 150 - 75) books for Amazon.

**Table 9**. Results for books published before 2005

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| ～2004 | SVM-APL | 76.7 | 4 ( 3.3 ) | 35 ( 29.2 ) | 53 ( 44.2 ) | 28 ( 23.3 ) | 120 |
| | SVM-AxL | 78.3 | 2 ( 1.7 ) | 33 ( 27.5 ) | 59 ( 49.2 ) | 26 ( 21.7 ) | 120 |
| | SVM-APx | 70.0 | 1 ( 0.8 ) | 23 ( 19.2 ) | 60 ( 50.0 ) | 36 ( 30.0 ) | 120 |
| | SVM-Axx | 74.2 | 1 ( 0.8 ) | 27 ( 22.5 ) | 61 ( 50.8 ) | 31 ( 25.8 ) | 120 |
| | RnF-APL | 61.7 | 6 ( 5.0 ) | 20 ( 16.7 ) | 48 ( 40.0 ) | 46 ( 38.3 ) | 120 |
| | RnF-AxL | 55.8 | 2 ( 1.7 ) | 29 ( 24.2 ) | 36 ( 30.0 ) | 53 ( 44.2 ) | 120 |
| | RnF-APx | 55.8 | 1 ( 0.8 ) | 26 ( 21.7 ) | 40 ( 33.3 ) | 53 ( 44.2 ) | 120 |
| | RnF-Axx | 59.2 | 0 ( 0.0 ) | 22 ( 18.3 ) | 49 ( 40.8 ) | 49 ( 40.8 ) | 120 |
| | Ada-APL | 65.0 | 2 ( 1.7 ) | 25 ( 20.8 ) | 51 ( 42.5 ) | 42 ( 35.0 ) | 120 |
| | Ada-AxL | 65.0 | 5 ( 4.2 ) | 33 ( 27.5 ) | 40 ( 33.3 ) | 42 ( 35.0 ) | 120 |
| | Ada-APx | 62.5 | 2 ( 1.7 ) | 21 ( 17.5 ) | 52 ( 43.3 ) | 45 ( 37.5 ) | 120 |
| | Ada-Axx | 65.8 | 2 ( 1.7 ) | 24 ( 20.0 ) | 53 ( 44.2 ) | 41 ( 34.2 ) | 120 |
| | NdcTitle | 78.3 | 1 ( 0.8 ) | 28 ( 23.3 ) | 65 ( 54.2 ) | 26 ( 21.7 ) | 120 |
| | Amazon | 76.1 | 6 ( 9.0 ) | 19 ( 28.4 ) | 26 ( 38.8 ) | 16 ( 23.9 ) | 67 |

**Table 10.** Results for books published inorafter 2005

| | | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| 2005～ | SVM-APL | 78.3 | 12 ( 10.0 ) | 32 ( 26.7 ) | 50 ( 41.7 ) | 26 ( 21.7 ) | 120 |
| | SVM-AxL | 81.7 | 14 ( 11.7 ) | 32 ( 26.7 ) | 52 ( 43.3 ) | 22 ( 18.3 ) | 120 |
| | SVM-APx | 71.7 | 2 ( 1.7 ) | 34 ( 28.3 ) | 50 ( 41.7 ) | 34 ( 28.3 ) | 120 |
| | SVM-Axx | 68.3 | 6 ( 5.0 ) | 24 ( 20.0 ) | 52 ( 43.3 ) | 38 ( 31.7 ) | 120 |
| | RnF-APL | 65.0 | 2 ( 1.7 ) | 24 ( 20.0 ) | 52 ( 43.3 ) | 42 ( 35.0 ) | 120 |
| | RnF-AxL | 58.3 | 4 ( 3.3 ) | 19 ( 15.8 ) | 47 ( 39.2 ) | 50 ( 41.7 ) | 120 |
| | RnF-APx | 55.0 | 0 ( 0.0 ) | 24 ( 20.0 ) | 42 ( 35.0 ) | 54 ( 45.0 ) | 120 |
| | RnF-Axx | 56.7 | 4 ( 3.3 ) | 21 ( 17.5 ) | 43 ( 35.8 ) | 52 ( 43.3 ) | 120 |
| | Ada-APL | 62.5 | 3 ( 2.5 ) | 25 ( 20.8 ) | 47 ( 39.2 ) | 45 ( 37.5 ) | 120 |
| | Ada-AxL | 66.7 | 5 ( 4.2 ) | 15 ( 12.5 ) | 60 ( 50.0 ) | 40 ( 33.3 ) | 120 |
| | Ada-APx | 65.0 | 1 ( 0.8 ) | 24 ( 20.0 ) | 53 ( 44.2 ) | 42 ( 35.0 ) | 120 |
| | Ada-Axx | 57.5 | 4 ( 3.3 ) | 24 ( 20.0 ) | 41 ( 34.2 ) | 51 ( 42.5 ) | 120 |
| | NdcTitle | 68.3 | 5 ( 4.2 ) | 26 ( 21.7 ) | 51 ( 42.5 ) | 38 ( 31.7 ) | 120 |
| | Amazon | 88.0 | 14 ( 16.9 ) | 21 ( 25.3 ) | 38 ( 45.8 ) | 10 ( 12.0 ) | 83 |

**Table 11.** Results of SVM-AxL by publication year

| Publication Year | Recommended Books | Positively Evaluated Books | Proportion |
|---|---|---|---|
| 1960-79 | 13 | 9 | 69.2 |
| 1980-89 | 23 | 17 | 73.9 |
| 1990-99 | 40 | 34 | 85.0 |
| 2000-14 | 164 | 132 | 80.5 |
| Total | 240 | 192 | 80.0 |

The average loan frequencies of books recommended by SVM-AxL divided by students' grades are shown in Table 15. We can see that the average loan frequencies of books that were positively evaluated by graduate students were lower than those positively evaluated by undergraduate students (0.88, 1.19, and 1.13 for graduate students, senior-year undergraduate, and sophomore undergraduates, respectively). The same can be said for "A: Already Bought or Read," "2: Very Interested," and "1: Interested." Graduate students are interested in relatively narrower areas of research than undergraduate students, and books intended for graduate studies may not be

**Table 12.** Averages of publication years of recommended books

| | Positively Evaluated | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested |
|---|---|---|---|---|---|
| SVM-APL | 2.17 | 2.42 | 2.21 | 2.10 | 2.03 |
| SVM-AxL | 2.42 | 2.33 | 2.34 | 2.49 | 2.55 |
| SVM-APx | 1.75 | 2.14 | 1.79 | 1.72 | 1.50 |
| SVM-Axx | 2.75 | 2.55 | 2.74 | 2.77 | 2.88 |
| RnF-APL | 2.14 | 2.24 | 2.15 | 2.12 | 2.17 |
| RnF-AxL | 2.68 | 2.55 | 2.70 | 2.67 | 2.65 |
| RnF-APx | 2.10 | 1.79 | 2.09 | 2.10 | 2.07 |
| RnF-Axx | 2.72 | 2.36 | 2.80 | 2.70 | 2.97 |
| Ada-APL | 2.13 | 2.23 | 2.16 | 2.12 | 2.23 |
| Ada-AxL | 2.54 | 2.46 | 2.57 | 2.53 | 2.72 |
| Ada-APx | 2.11 | 1.94 | 2.13 | 2.11 | 2.17 |
| Ada-Axx | 2.86 | 2.37 | 2.89 | 2.88 | 2.85 |
| NdcTitle | 2.76 | 2.65 | 2.79 | 2.74 | 2.96 |
| Amazon | 1.79 | 1.66 | 1.88 | 1.78 | 2.30 |

**Table 13**. Averages of loan frequencies

| | Positively Evaluated | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested |
|---|---|---|---|---|---|
| SVM-APL | 1.10 | 1.53 | 1.12 | 1.02 | 0.89 |
| SVM-AxL | 1.06 | 1.36 | 1.08 | 1.00 | 1.04 |
| SVM-APx | 0.35 | 0.83 | 0.38 | 0.31 | 0.18 |
| SVM-Axx | 0.33 | 0.82 | 0.29 | 0.32 | 0.18 |
| RnF-APL | 0.64 | 0.67 | 0.57 | 0.67 | 0.48 |
| RnF-AxL | 0.69 | 0.86 | 0.73 | 0.65 | 0.65 |
| RnF-APx | 0.36 | 0.29 | 0.35 | 0.37 | 0.19 |
| RnF-Axx | 0.27 | 1.02 | 0.28 | 0.24 | 0.14 |
| Ada-APL | 0.51 | 0.63 | 0.51 | 0.50 | 0.44 |
| Ada-AxL | 0.67 | 0.75 | 0.68 | 0.65 | 0.63 |
| Ada-APx | 0.41 | 0.89 | 0.34 | 0.42 | 0.23 |
| Ada-Axx | 0.27 | 0.86 | 0.22 | 0.25 | 0.17 |
| NdcTitle | 0.31 | 0.70 | 0.32 | 0.29 | 0.12 |
| Amazon | 0.50 | 0.89 | 0.59 | 0.34 | 0.18 |

**Table 14**. Averages of loan frequencies of books recommended by SVM-AxL divided by grades

| | Positively Evaluated | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested |
|---|---|---|---|---|---|
| Sophomore | 1.13 | 1.65 | 1.16 | 1.09 | 0.85 |
| Senior-year | 1.19 | 2.17 | 1.07 | 1.25 | 1.05 |
| Graduate | 0.88 | 1.25 | 0.91 | 0.75 | 1.09 |

frequently borrowed. As we previously mentioned, Ndc Title, which does not use loan records, outperformed SVM-AxL with regard to graduate students. Hence, methods of this sort might be better to recommend books to graduate students. Switching recommendation methods depending on whether the users are graduate students or undergraduate students might be an effective manner of attaining accurate results.

Finally, the degrees of importance of each feature in random forest are shown in Table 16. Random forest determined that loan frequencies are more important than both the confidence and support of association rules and NDC matches/mismatches.

### Results Concerning Association Rules

Among 2,880 books recommended by machine learning methods (= 12 methods i.e., SVM-APL, SVM-AxL, ..., Ada-APx and Ada-Axx) × 240 books), the books that composed association rules with LBOOKs (i.e., one of which was antecedent and the other of which was consequent of association rules) were only two. In other words, only two books had been borrowed together with LBOOKs in the past, and the other 2,878 books were not.

This result raises the question of whether the association rules are really effective and contribute to book recommendation. Tsuji et al. (2013) concluded that recommendation based on (a) NDC matches, (b) similarities in titles, and (c) confidence and support of association rules were more effective than that based merely on (a) and (b). However, the difference between the PPEs of the two kinds of methods was small and not statistically significant. Therefore, we conducted an additional experiment to examine the effectiveness of association rules. We asked for LBOOKs of subjects in group (B) and used each of the four methods – i.e., SVMs based on features (a)(b), (a)(b)(c), (a)(b)(d), and (a)(b)(c)(d), where (a), (b), (c), and (d) represent NDC matches, similarities in titles, confidence and support of association rules, and loan frequencies, respectively – to recommend six books. Henceforth, we represent SVMs based on (a)(b), (a)(b)(d), (a)(b)(c), and (a)(b)(c)(d) as SVM-NTxx, SVM-NTRx, SVM-NTxL, and SVM-NTRL, respectively. SVM-NTRL is identical to SVM-AxL.

The recommendation results are shown in Tables 17, 18, and 19. From these Tables, we can form the following conclusions :(1) The PPE of SVM-NTRx was slightly higher than that of SVM-NTxx (63.3% and 60.0% in Table 17, respectively), although the difference was not statistically significant. This result is consistent with that obtained by Tsuji et al. (2013). (2) The PPEs of SVM-NTxL and SVM-NTRL (=SVM-AxL) were identical (71.7% in Table 17), and (3) the PPE of SVM-NTxL was higher than that of SVM-NTRx (71.7% and 63.3%, respectively). Similar results were obtained when we divided the subjects by their grades (see Tables 18 and 19). Furthermore, we found that (4) the two sets of 120 books recommended by each of SVM-NTxL and SVM-NTRL were quite similar, i.e., they had 112 books in common. From (1), (2), (3), and (4), we can conclude that adding association rules to NDC matches and similarities in titles slightly improves recommendation performance. However, if we use loan frequencies as features along with them, association rules are not needed.

Incidentally, we also used Amazon’s recommendation system in our additional experiment. We can see in Table 17 that the PPE of SVM-NTxL was comparable to that of Amazon (i.e., 71.7% and 72.9%, respectively). Although we have not listed this in any of the tables, the PPE of SVM-NTxL for the top 70 books (the same number of books that Amazon recommended) according to probability was 72.9%, which is identical to the PPE of Amazon’s recommendation system. Finally, while Amazon had exhibited the best results for senior undergraduate students in Table 6, our methods outperformed it in the additional experiment (Table 18). From these results, we can conclude that SVM-NTxL is comparable to Amazon.

### Results Concerning Random Forests

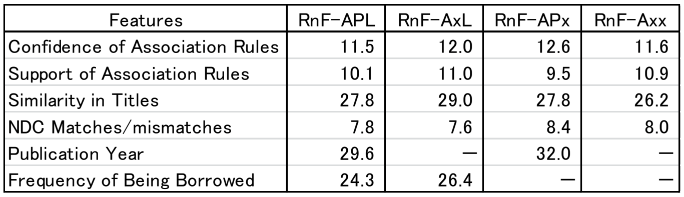
In our experiment, random forests performed poorly compared to SVM, although some past research had indicated the opposite. This can be attributed to the small number of features used for random forests. Nitze et al. (2012) showed that SVM was more accurate than random forest, but the difference depended on the number of features used. When they used fewer features, the difference became larger and random forest became much less effective than SVM.[10] As previously mentioned, Khan et al. (2010) reported that random forest based on three features was more effective than SVM. However, the kernel they used for SVM was polynomial. A polynomial kernel is smaller than the RBF kernel, and Nitze et al. (2012) reported that it was less effective than RBF. Therefore, if Khan et al. (2010) had used the RBF kernel, the result might have changed. Breiman (2001) claimed that a higher correlation between decision trees leads to higher values of generalization errors. We only used four to six features, which might have produced a higher correlation between trees and poor recommendation results.

## CONCLUSIONS

In this paper, we proposed a method to recommend Japanese books to university students through machine learning modules based on several features, including

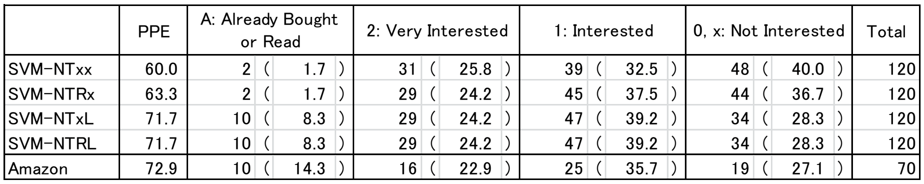
[10] For instance, when 40 features were used, the overall accuracies of SVM and random forest were 88.1% and 87.4%, respectively. However, when only 10 features were used, they were 55.8% and 68.6%, respectively.

**Table 15.** Importance of features concerning random forest



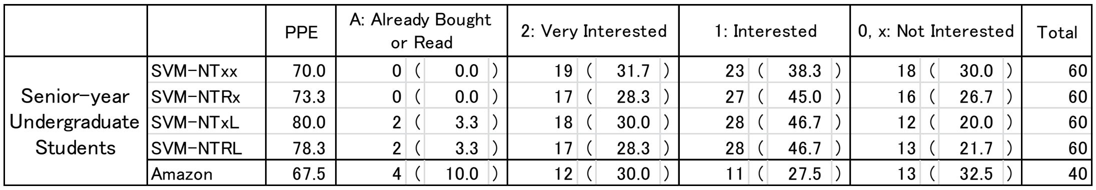
| Features | RnF-APL | RnF-AxL | RnF-APx | RnF-Axx |
|---|---|---|---|---|
| Confidence of Association Rules | 11.5 | 12.0 | 12.6 | 11.6 |
| Support of Association Rules | 10.1 | 11.0 | 9.5 | 10.9 |
| Similarity in Titles | 27.8 | 29.0 | 27.8 | 26.2 |
| NDC Matches/mismatches | 7.8 | 7.6 | 8.4 | 8.0 |
| Publication Year | 29.6 | — | 32.0 | — |
| Frequency of Being Borrowed | 24.3 | 26.4 | — | — |

**Table 16**. Results with/without association rules



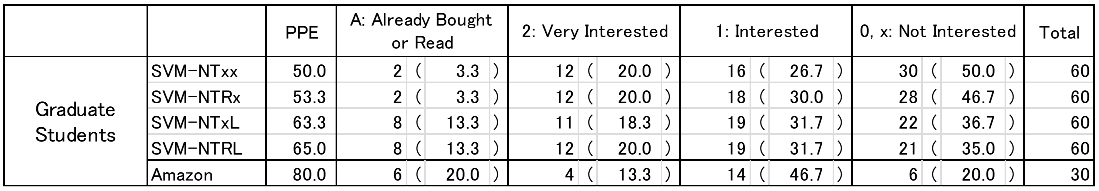
|  | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|
| SVM-NTxx | 60.0 | 2 ( 1.7 ) | 31 ( 25.8 ) | 39 ( 32.5 ) | 48 ( 40.0 ) | 120 |
| SVM-NTRx | 63.3 | 2 ( 1.7 ) | 29 ( 24.2 ) | 45 ( 37.5 ) | 44 ( 36.7 ) | 120 |
| SVM-NTxL | 71.7 | 10 ( 8.3 ) | 29 ( 24.2 ) | 47 ( 39.2 ) | 34 ( 28.3 ) | 120 |
| SVM-NTRL | 71.7 | 10 ( 8.3 ) | 29 ( 24.2 ) | 47 ( 39.2 ) | 34 ( 28.3 ) | 120 |
| Amazon | 72.9 | 10 ( 14.3 ) | 16 ( 22.9 ) | 25 ( 35.7 ) | 19 ( 27.1 ) | 70 |

**Table 17**. Results with/without association rules for senior-year undergraduate students



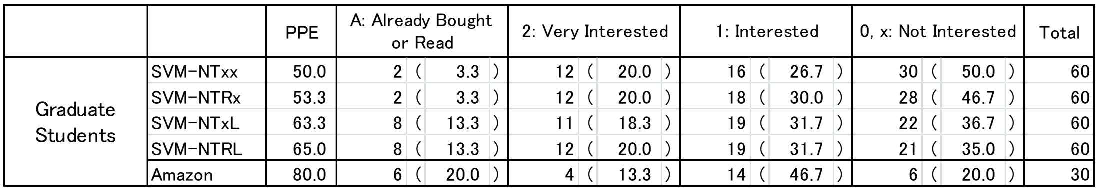
|  |  | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Senior-year Undergraduate Students | SVM-NTxx | 70.0 | 0 ( 0.0 ) | 19 ( 31.7 ) | 23 ( 38.3 ) | 18 ( 30.0 ) | 60 |
|  | SVM-NTRx | 73.3 | 0 ( 0.0 ) | 17 ( 28.3 ) | 27 ( 45.0 ) | 16 ( 26.7 ) | 60 |
|  | SVM-NTxL | 80.0 | 2 ( 3.3 ) | 18 ( 30.0 ) | 28 ( 46.7 ) | 12 ( 20.0 ) | 60 |
|  | SVM-NTRL | 78.3 | 2 ( 3.3 ) | 17 ( 28.3 ) | 28 ( 46.7 ) | 13 ( 21.7 ) | 60 |
|  | Amazon | 67.5 | 4 ( 10.0 ) | 12 ( 30.0 ) | 11 ( 27.5 ) | 13 ( 32.5 ) | 40 |

**Table 18.** Results with/without association rules for graduate students



|  |  | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Graduate Students | SVM-NTxx | 50.0 | 2 ( 3.3 ) | 12 ( 20.0 ) | 16 ( 26.7 ) | 30 ( 50.0 ) | 60 |
|  | SVM-NTRx | 53.3 | 2 ( 3.3 ) | 12 ( 20.0 ) | 18 ( 30.0 ) | 28 ( 46.7 ) | 60 |
|  | SVM-NTxL | 63.3 | 8 ( 13.3 ) | 11 ( 18.3 ) | 19 ( 31.7 ) | 22 ( 36.7 ) | 60 |
|  | SVM-NTRL | 65.0 | 8 ( 13.3 ) | 12 ( 20.0 ) | 19 ( 31.7 ) | 21 ( 35.0 ) | 60 |
|  | Amazon | 80.0 | 6 ( 20.0 ) | 4 ( 13.3 ) | 14 ( 46.7 ) | 6 ( 20.0 ) | 30 |

**Table 19**. Results with/without association rules for graduate students



|  |  | PPE | A: Already Bought or Read | 2: Very Interested | 1: Interested | 0, x: Not Interested | Total |
|---|---|---|---|---|---|---|---|
| Graduate Students | SVM-NTxx | 50.0 | 2 ( 3.3 ) | 12 ( 20.0 ) | 16 ( 26.7 ) | 30 ( 50.0 ) | 60 |
|  | SVM-NTRx | 53.3 | 2 ( 3.3 ) | 12 ( 20.0 ) | 18 ( 30.0 ) | 28 ( 46.7 ) | 60 |
|  | SVM-NTxL | 63.3 | 8 ( 13.3 ) | 11 ( 18.3 ) | 19 ( 31.7 ) | 22 ( 36.7 ) | 60 |
|  | SVM-NTRL | 65.0 | 8 ( 13.3 ) | 12 ( 20.0 ) | 19 ( 31.7 ) | 21 ( 35.0 ) | 60 |
|  | Amazon | 80.0 | 6 ( 20.0 ) | 4 ( 13.3 ) | 14 ( 46.7 ) | 6 ( 20.0 ) | 30 |

library loan records. We showed that the method that used SVM based on (1) matches/mismatches between NDC categories, (2) similarities in book titles, and (3) loan frequencies performed better than those that used (a) SVMs based on other combinations of features, (b) random forest and Adaboost with features identical to (1), and (c) the method proposed by Tsuji et al. (2013). Furthermore, the results of recommendations made by our proposed methods are comparable to those of Amazon's recommendation system.

In the future, we aim to examine (1) the effectiveness of incorporating into recommendation methods information specific to university students, such as courses in which they are enrolled, and (2) the effect of the size of the training data on the performance of different methods. Furthermore, subjects were only asked to indicate their "degree of interest" in books in our experiment, regardless of what they felt when they were "interested" in a particular book. For instance, subjects might have felt that a recommended book belonged to an unexpected novel field or that they knew a recommended book but had not yet read it. We would like to account for all the types of interests in order to develop methods for book recommendations that are suited to various kinds of interests.

## REFERENCES

Agrawal R. et al. (1993). Mining association rules between sets of items in large databases. Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, 207-216.

Beagle D. R. et al. (2006). *The Information Commons Handbook*. New York: Neal-Schuman Publishers, 247p.

Breiman L (2001). Random forests. Machine Learning, 45(1), 5-32.

Chan J. C. et al. (2001). Enhanced algorithm performance for land cover classification from remotely sensed data using bagging and boosting. IEEE Transactions of Geosciences and Remote Sensing, 39, 693-695.

Chan J, Paelinckx D (2008). Evaluation of random forest and Adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyper spectral imagery. Remote Sensing of Environment, 112(6):2999-3011.

Chen C (2013). An intelligent mobile location-aware book recommendation system that enhances problem-based learning in libraries. Interactive Learning Environments, 21(5): 469-495.

Chen C, Chen A (2007). Using data mining technology to provide a recommendation service in the digital library. The Electronic Library, 25(6), 711-724.

Friedman J. et al. (2000). Additive logistic regression: A statistical view of boosting. The Annals of Statistics, 28(2), 337-407.

Harada T (2009). The book recommendation system using library loan records. Digital Libraries, 36, 22-31 (text in Japanese).

Harada T, Masuda K (2010). A trial approach of weighting for library loan records for developing a book recommendation system. Digital Libraries, 38, 54-66 (text in Japanese).

Hasan AM et al. (2014). Support vector machine and random forest modeling for intrusion detection system (IDS). Journal of Intelligent Learning Systems and Applications, 6(1), 45-52.

Ishii N. et al. (2010). Classification by rough set reducts, Adaboost and SVM. Proceedings of the 11th ACIS International Conference on Software Engineering Artificial Intelligence Networking and Parallel/Distributed Computing (SNPD), 63-68.

Jia S. et al. (2013). The comparison between random forest and support vector machine algorithm for predicting β-hairpin motifs in proteins. Engineering, 5, 391-395.

Khan R. et al. (2010). Skin detection: A random forest approach. Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010), 4613-4616.

LIBSVM and easy.py <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.[Accessed: December30, 2014].

Linden, G. et al. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. IEEE Internet Computing, 7(1), 76-80.

Liu, M. et al. (2013).Comparison of random forest, support vector machine, and back propagation neural network for electronic tongue data classification: Application to the recognition of orange beverage and Chinese vinegar. Sensors and Actuators B: Chemical, 177, 970-980.

Longjun, D. et al. (2011). Comparisons of random forest and support vector machine for predicting blasting vibration characteristic parameters. Procedia Engineering, 26, 1772-1781.

Luo, Y. et al. (2009). A privacy-preserving book recommendation model based on multi-agent. Proceedings of the 2009 Second International Workshop on Computer Science and Engineering, 323-327.

MeCab<https://code.google.com/p/mecab/>.[Accessed: December30, 2014].

Miao X, Heaton JS (2010). A comparison of random forest and Adaboost tree in ecosystem classification in east Mojave Desert, Proceedings of the 18th International Conference on Geoinformatics, 1-6.

Nitze, I. et al. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. Proceedings of the Geographic Object-Based Image Analysis (GEOBIA 2012), 35-40.

R <http://www.r-project.org/>.[Accessed: December30, 2014].

Shirgaonkar, S. et al. (2010). Application of improved a priori in university library. International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), 535-540.

Tsuji K. et al. (2012).Use of library loan records for book recommendation. Proceedings of the 3rd IIAI International Conference on e-Services and Knowledge Management (IIAI ESKM 2012), 30-35.

Tsuji K. et al. (2013). Book recommendation based on library loan records and bibliographic information. Proceedings of the 3rd International Conference on Integrated Information (IC-ININFO 2013). 8p.(No Pagination).

Whitney C, Schiff L (2006). Melvil Recommender Project: Developing library recommendation services. D-Lib Magazine, 12(2).<http://www.dlib.org/dlib/december06/whitney/12whitney.html>. [Accessed: December30, 2014].