Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Evaluation of Difficulty Estimation for Learning Materials Recommendation

Yasuo MIYOSHI [a*], Kazuhiro SUZUKI [a], Ken-ichi SHIOTA [a] & Ryo OKAMOTO [a]
[a]*Faculty of Science, Kochi University, Japan*
*miyoshi@is.kochi-u.ac.jp

Abstract: The popular technology for the information recommendation of books or web pages is based on taste information from many users, but it is important to be based on difficulty and proficiency for the recommendation of learning materials. We have developed an algorithm to estimate difficulty of learning materials and proficiency of learners, for recommendation considering difficulty of learning materials. The algorithm uses only a bipartite learner-material graph that consists of the reader relations with materials and learners. In this paper, we describe how to make an accurate data to evaluate the estimated difficulty, and report about the result that evaluated the precision of our proposed algorithm.

Keywords: Personalized recommendation, learning materials, difficulty estimation

1. Introduction

Most of mainstream information recommendation systems use content-based filtering or collaborative filtering based on the past behaviors of users, and may use hybrid filtering that redeems their defects and capitalizes on their respective strengths. Both the filters are based on taste information from many users. In recommendation for learning materials, the recommender systems that based on taste information of users might recommend the materials too difficult to a user. They might recommend the materials too easy adversely. The recommender system for learning materials desires to recommend the materials that have suitable difficulty for users. Therefore we propose new filtering based on difficulty of learning materials and proficiency of users. As you can see in figure 1, we propose also a recommender system that adopts pipelined hybrid filtering. Our proposed system recommends the learning materials suitable for proficiency of users after having narrowed down candidate materials by users' taste information.

2. Related Work

Durand, Belacel and LaPlante (2013) proposed a learning path recommendation algorithm using graph theory based model. This approach focuses on ways to search for potential learning paths. We suggest that a new and good learning material would be hard to be recommended by affected by old learning paths. Ghauth and Abdullar (2010) and Guo, Erdt and Lee (2013) proposed a learning materials recommendation algorithm based on difficulty of materials. Their recommender systems ask to learners to specify a difficulty level. Among other examples, there is a difficulty based recommender system for recommending games on mobile phones (Skocir et al., 2012).

3. Algorithm to Estimate Difficulty and Proficiency

3.1 Summary of Estimation Algorithm

Our algorithm does not use contents of learning materials at all. It only use the reading relations that denote who read which materials. The reading relations are expressed in bipartite graph like figure 2, and we call it a learner-material graph. Even if learning materials may be acquired no contents, this algorithm can make it a target of recommendation by combining with collaborative filtering like figure 1 because it can estimate difficulty and proficiency without using the contents of the learning

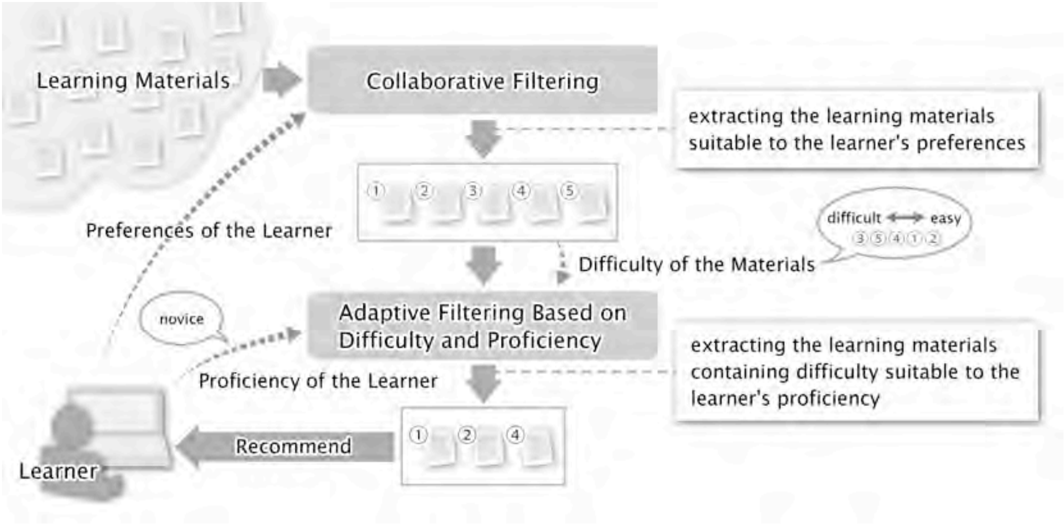

Figure 1. A Proposed Recommender System that Adopts Pipelined Hybrid Filtering.

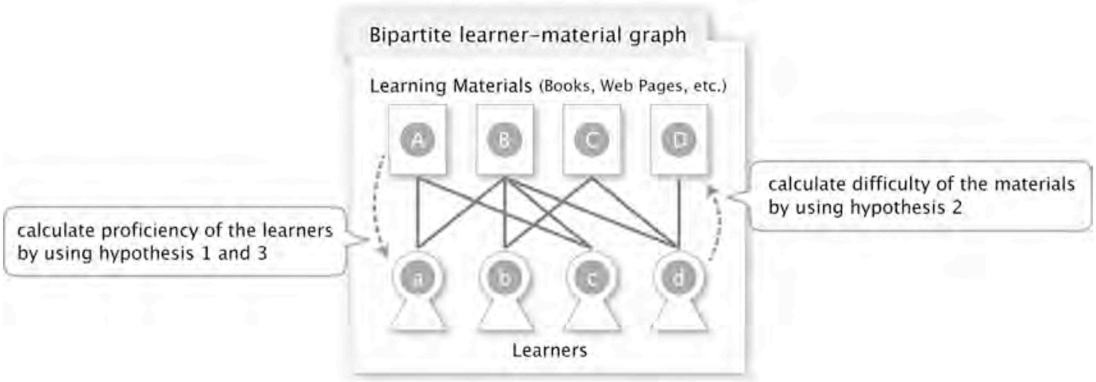

Figure 2. To Estimate Difficulty and Proficiency Using Learner-Material Graph.

materials. One of feature of this algorithm is to be able to simultaneously estimate difficulty of learning materials and proficiency of learners, but the estimation algorithms based on contents can estimate only difficulty. In addition, this algorithm is similar to PageRank (Page, Brin, Motwani and Winograd, 1999) and HITS (Kleinberg, 1999) that are popular webpage ranking algorithms, in concept to use a link structure of the graph network. The elemental algorithm performs based on the following hypotheses.

- Hypothesis 1: A learner reading a lot of learning materials of a domain should know a lot about the domain.
- Hypothesis 2: A learning material read by learners that have much knowledge should be difficult, and a material read by learners that do not know a lot should be easy.
- Hypothesis 3: A learner reading difficult learning materials should know a lot, and a learner reading easy materials should not have much knowledge.

At first, proficiency of learners is estimated by hypothesis 1. Next, based on hypothesis 2, difficulty of learning materials is calculated from proficiency of the learners. Then proficiency is recalculated from difficulty of the materials by hypothesis 3. Hypothesis 2 and 3 are influenced by each other. The calculations based on hypothesis 2 and 3 are repeated until the calculation result converges.

Hypothesis 1 and 3 are expressed as:

$$p_u = \sum_{i \in I_u} \dot{d}_i \tag{1}$$

where p_u is proficiency of the learner u and $\dot{d}_i$ is normalized difficulty of the learning material i and I_u is a set of learning materials read by the learner u. The initial value of $\dot{d}_i$ is 0.5. The difficulty value d_i of hypothesis 2 is written as:

$$d_i = \sum_{u \in U_i} \left(\dot{p}_u - 0.5 \right) \qquad (2)$$

where $\dot{p}_u$ is normalized proficiency of the learner u and U_i is a set of learners who read the material i. The normalization values of $\dot{d}_i$ and $\dot{p}_u$ are found by dividing the deviation value by 100.

3.2 Improve of the Algorithm

We will define difficulty of learning materials and proficiency of learners before we describe an improved algorithm. The premise to estimate difficulty of learning materials and proficiency of learners is for recommending the learning materials. Therefore, in this study, we define the term "difficulty of a learning material" as presupposed knowledge quantity necessary to get most knowledge with the learning material, and define the term "proficiency of a user" as knowledge quantity that the user has about the domain of the learning materials and the associated domain.

The obtained values of difficulty and proficiency from elemental algorithm are relativity, and then there are not the direct relations in each other's values. Therefore the values are hard to be handled to recommend learning materials, because the value of difficulty of the learning material suitable for a learner who has a proficiency value is unclear. In addition, the elemental algorithm does not consider the order of that a learner read learning materials. The algorithm improved from elemental algorithm considers the order and estimates how learners acquire knowledge, for improvement of precision. Specifically we add hypothesis 4 and revise hypothesis 2 and 3 because a learner that has much knowledge might read an easy material.

- Hypothesis 4: A learning material read by expert learners while they are beginners should be easy.
- Hypothesis 2': A learning material read by learners that have much knowledge should be difficult.
- Hypothesis 3': A learner reading difficult learning materials should know a lot.

Furthermore, the improved algorithm makes it clear that the difficulty value of the learning material suitable for a learner is near to the proficiency value of the learner because the improved algorithm uses a same unit of the values of difficulty and proficiency by changing methods of calculation and normalization. Figure 3 and 4 show the difference of the estimation methods between

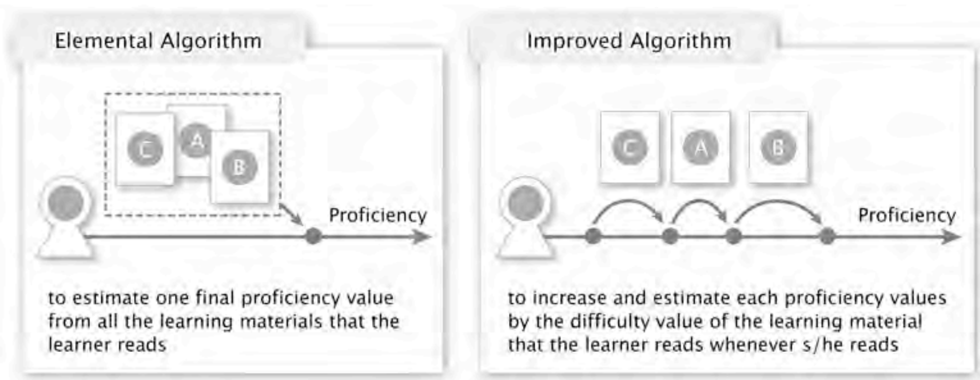

Figure 3. The Difference of the Methods for Proficiency Estimation.

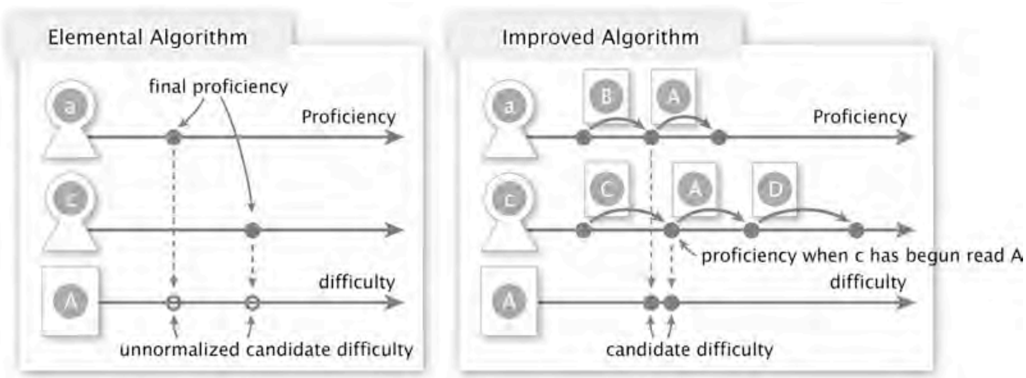

Figure 4. The Difference of the Methods for Difficulty Estimation.

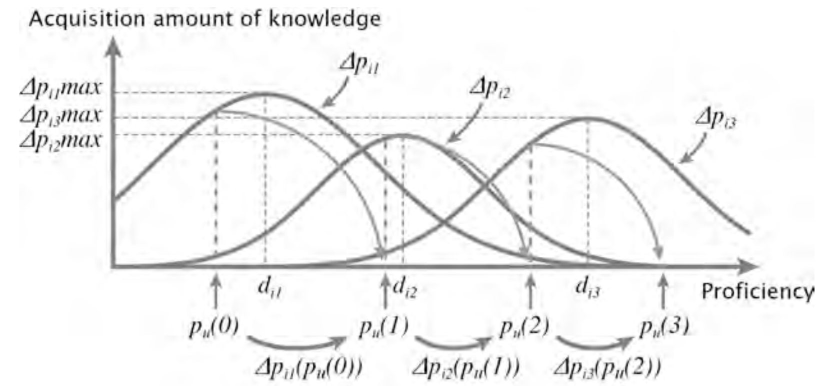

Figure 5. The Amount of Knowledge Acquired by Learner u

elemental algorithm and improved algorithm. The improved algorithm estimates the amount of knowledge acquired by the learner that reads a learning material. We describe the flows of estimation of the amount of acquisition knowledge by using figure 5. Learner u acquires knowledge from initial proficiency $p_u(0)$ to $p_u(3)$, when the learner reads in order of learning material $i1$, $i2$, $i3$. $p_u(n)$ is the proficiency value of learner u after having read the n-th learning material. Δp_i as the amount of knowledge acquired by reading learning material i depends on p_u at that time:

$$\Delta p_i = \Delta p_i \max \cdot \sigma_i^{-(p_u - d_i)/\Delta p_i \max} \tag{3}$$

where $\Delta p_i \max$ is the height of the curve and σ_i is the width of the curve. For example, quantity of knowledge $\Delta p_{i1}(p_u(0))$ is slightly lower than $\Delta p_i \max$ because $p_u(0)$ is lower than d_{i1}, and then $p_u(1)$ is given by $p_u(1) = p_u(0) + \Delta d_{p1}(p_u(0))$. The initial value of proficiency $p_u(0)$ is tentatively the difficulty value of the learning material that is read by u at the beginning in this study. However, $p_u(0)$ is the difficulty value of the next learning material if the next material is easy than first material. Therefore, although $p_u(0)$ is lower than d_{i1} in figure 5, $p_u(0)$ is d_{i1} accurately. Maximum quantity of acquirable knowledge $\Delta p_i \max$ and a spread of target scope σ_i are tentatively unity values, although the calculation methods of them are under consideration. As you can see in figure 4, a learning material gets some candidate difficulty values from the learners who read it. Most learners should have proficiency higher than or comparable with the difficulty value d_i, because d_i means quantity of knowledge necessary to read learning material i. Therefore, in this study, d_i is tentatively the middle value with the minimum and the median of the proficiency values of each learners.

4. Evaluation of Difficulty Estimation

We evaluated precision improvement between the elemental algorithm and the improved algorithm. The evaluation of estimation precision is relative evaluation. We compared the order of values estimated by the algorithm with the order of accurate data. We gathered and created the learner-material graph that represents the relations between web pages written about programming language C and the users that bookmarked them from social bookmark site “Hatena Bookmark” (http://b.hatena.ne.jp/) that is famous in Japan. The learners of the graph are the bookmark users, and the learning materials of the graph are the bookmarked web pages. Although our proposed algorithm is able to estimate not only difficulty of learning materials but also proficiency of learners, we evaluated only difficulty estimation because it is impossible to appreciate accurate proficiency of the bookmark site users. We would like to evaluate the learner proficiency estimation of our algorithm by letting the subject, whose accurate proficiency is known, use the bookmark site in the long term.

4.1 Accurate Difficulty Values of Learning Materials for Comparing with Estimated Data

We extracted the partial graph from the learner-material graph at random because to get accurate values of difficulty of all-bookmarked pages increases in cost. A scale of the learner-material network is 6,079 learners, 16,016 materials and 40,145 bookmarks. Their bookmarks have “C language” tag. We have decided the target page count of getting accurate value as 30 in consideration of workload of the subjects. We randomly selected 50 pages from the web pages that were bookmarked by over 30 users, because the pages may include inappropriate contents and have broken links. There were no pages including inappropriate contents, but 9 pages were not found or not available. Then we randomly reselected 30 pages from the 41 pages. All extracted web pages are technical contents about C programming and consist of blogs, news articles, reference pages, curated pages and so on.

We calculated accurate difficulty values from the results of the experiment evaluated by subjects that are learning C language. The subjects are 42 undergraduate students that learn information science, university freshmen are 13 people, second year students are 14 people and third year students are 15 people. The experiment was carried out in December 2013, and freshmen did not yet study a class of the C language at that point in time. Therefore we assume that the subjects consist of beginners and intermediate graders. The subjects firstly glanced through a web page for around 10 seconds and then evaluated in 5-point scale how much knew the contents of the page by oneself (Q1). The subjects secondly read the page and then evaluated in 4-point scale whether it was a useful page for oneself (Q2). The scales of Q1 are “not to know at all (0%)”, “know a little (25%)”, “know half (50%)”, “know most (75%)” and “know at all (100%)”. The scales of Q2 are “1) almost useless”, “2) useless a little”, “3) helpful a little”, “4) very helpful”. We confirmed that the self-evaluation of understanding became low as the young year students and the tendency of the frequency distribution of Q1 and Q2 became different, by analyzing the answers of all subjects for 30 pages. Therefore we calculated accurate difficulty values of learning materials from the answer data, using Item Response Theory (IRT). We used 1 parameter logistic model because we need only simple indicator of difficulty. The binary response data represents whether a subject is able to understand a learning material. Specifically, if answer in Q1 is 0% then the binary data is 0, but answer in Q2 also is 3) or 4) then it is 1. The probability of a correct response item j in the Rasch model is given by:

$$p_j(\theta) = \frac{1}{1 + \exp(-Da(\theta - b_j))} \tag{4}$$

where θ is the standing on the underlying trait and b_j is the difficulty of item j. The variable a is the parameter of the logistic curve. We estimated accurate difficulty from the binary response data with rasch function of R software package “ltm” (Rizopoulos, 2006). Table 1 shows the difficulty values of the learning materials in descending order of difficulty.

4.2 Results

We evaluated difficulty rankings precision by comparing accurate difficulty ranking, and table 2 shows result of improved algorithm. We applied some measures such as NDPM (Yao, 1995) and Spearman's rank correlation. NDPM measure will give a perfect score of 0 when estimated ranking completely agrees with accurate ranking, and will give a worst score of 1 when reversed estimated ranking completely agrees with accurate ranking. A score of 0.5 represents that there is no correlation between estimated ranking and accurate ranking. The Spearman's rank correlation coefficient would be near 1 when estimated ranking and accurate ranking have positive correlation. -1 is a negative correlation and 0 is no correlation. Therefore, as you can see in table 3, the difficulty ranking estimated by improved algorithm is better than elemental algorithm. We implemented simple content-based algorithm for evaluation. This algorithm estimates difficulty of a web page from difficulty of terms and frequency of terms about C language in the page. We defined difficulty of terms from the page number of index of the best-known book (Kernighan and Ritchie, 1988) in

Table 1: Accurate difficulty values of the partial learning materials.

Label	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Difficulty	1.19	0.71	0.49	0.39	0.19	0.00	0.00	-0.31	-0.41	-0.41
Label	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20
Difficulty	-0.51	-0.61	-0.61	-0.94	-1.32	-1.32	-1.47	-1.62	-1.80	-1.80
Label	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30
Difficulty	-2.00	-2.00	-2.23	-2.23	-2.51	-2.51	-2.51	-2.51	-2.89	-3.51

Table 2: Estimated difficulty ranking and values with the improved algorithm.

Label	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
Ranking	1	5	14	8	16	2	7	19	11	6
Difficulty	2.798	2.619	2.590	2.612	2.583	2.629	2.614	2.565	2.610	2.615
Label	M11	M12	M13	M14	M15	M16	M17	M18	M19	M20
Ranking	28	12	27	4	20	18	10	25	15	9
Difficulty	2.553	2.605	2.553	2.624	2.564	2.570	2.610	2.559	2.585	2.611
Label	M21	M22	M23	M24	M25	M26	M27	M28	M29	M30
Ranking	26	3	17	13	21	23	22	30	29	24
Difficulty	2.558	2.629	2.583	2.591	2.564	2.563	2.563	2.550	2.552	2.560

Table 3: Evaluation difficulty rankings by each ranking measures.

Metrics	Improved Algorithm	Elemental Algorithm	Content-Based Algorithm
NDPM	0.289	0.332	0.499
Spearman’s Correlation	0.576	0.468	-0.07

Japanese edition. We suppose that the reason to become worse estimation precision of content-based algorithm is that the algorithm is too simple.

5. Conclusions

We have described difficulty estimation algorithm to recommend learning materials. The evaluation result of difficulty estimation of our improved algorithm is better than elemental algorithm. We would like to improve the algorithm more and develop a learning material recommender system using the algorithm. Bookmark data of a social bookmark site is untrustworthy because a user may bookmark an unread page. We think that read pages should estimate difficulty and proficiency. Unread bookmarks are suitable to be re-recommended because a learner certainly has an interest in the bookmarked learning materials.

Acknowledgements

This work was supported by JSPS KAKENHI Grant Number 25330364.

References

Durand, G., Belacel, N. and LaPlante, F. (2013). Graph Theory Based Model for Learning Path Recommendation. *Elsevier Information Science*, 251, 10-21.

Ghauth, K. I. And Abdullar, N. A. (2010). Learning Materials Recommendation Using Good Learners’ Ratings and Content-Based Filtering. *Educational Technology Research and Development*, 58(6), 711-727.

Guo, G., Erdt, M. H. and Lee, B. S. (2013). A Hybrid Recommender System based on Material Concepts with Difficulty Levels. *Proceedings of the 21st International Conference on Computers in Education*, 90-96.

Kernighan, B. W., and Ritchie, D. M. (1988). The C programming language second edition. *Englewood Cliffs: prentice-Hall.*

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM), 46*(5), 604-632.

Page, L., Brin, S., Motwani, R. and Winograd, T. (1999). The PageRank Citation Ranking: Bringing Order to the Web. *Technical Report.* Stanford InfoLab.

Rizopoulos, D. (2006). ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses, *Journal of Statistical Software, 17(5), 1-25.*

Skocir, P., Marusic, L., Marusic, M. and Petric, A. (2012). The MARS – A Multi-Agent Recommendation System for Games on Mobile Phones. *Agent and Multi-Agent Systems: Technologies and Applications, LNCS 7327*, 104-113.

Yao, Y. Y. (1995). Measuring Retrieval Effectiveness Based on User Preference of Documents. *Journal of the American Society for Information Science, 46*(2), 133-145.