

Application of Social Computing to Collaborative Web Search

Chenn-Jung Huang^{1*}, Heng-Ming Chen², Shun-Chih Chang¹, Sheng-Yuan Chien¹

¹Department of Computer Science and Information Engineering, National Dong Hwa University, Taiwan

²Department of Electrical Engineering, National Dong Hwa University, Taiwan

Email: *cjhuang@mail.ndhu.edu.tw

Received 8 April 2015; accepted 13 September 2015; published 18 September 2015

Abstract

Recently, e-learning has been paid much attention in the area of education. However, it is difficult for the low-achievement students to find out the key points and keywords while searching online articles. They usually cannot accurately obtain the website information even after searching for large amount of data in the Internet. Meanwhile, these low-achievement students often lack of the related prior knowledge to determine if the website is useful. Accordingly, they select those websites based on their disorderly instincts. In this work, an intelligent collaborative web search assistance platform is proposed. A group grading module is presented to derive three parameters that are used to calculate the ranking of each website via the Support Vector Regression method. The effects of website ranking shorten the searching processes, and the learners can thus have more time to focus on comprehending the contents of the recommended website. The experimental results reveal that the proposed algorithm can effectively guide learners to search the appropriate website; accordingly, the target of self-learning assistance can be achieved and the learning performance of the students is enhanced.

Keywords

Cooperative/Collaborative Learning, Collaborative Information Seeking, Intelligent Tutoring Systems, Information Retrieval, Support Vector Regressions

1. Introduction

Collaborative information seeking (CIS) is a field of research that involves studying situations, motivations, and methods for people working in collaborative groups for information seeking projects, as well as building systems for supporting such activities. Researchers in the fields of command and control [1], education [2], and engineering [3], as well as research on technologies supporting collaborative information seeking [4] [5], study how people engage in information seeking in hopes of building more efficient and effective systems to support these activities. Besides, the literature [6] points out that information seeking generally happens in the context of a broader task and involves searching for, collecting, and analyzing information, as well as sense making and

*Corresponding author.

sharing. Morris [7] addresses that many tasks in both professional and casual settings can benefit from the ability to jointly search the websites with others.

World Wide Web (WWW) is a crucial learning environment for learners solving problems with a variety of network resources. However, learner's cognition load is heavy for they must filter useful knowledge out of diverse information. On the other hand, e-learning has been highly regarded in the area of education. However, it is difficult for the low-achievement students to find out the key points and keywords and obtain the website information while searching online articles in the Internet. In addition, these low-achievement students often lack of the related prior knowledge to determine if the website is useful. Accordingly, they select those websites based on their disorderly instincts. In view of this standpoint, a collaborative web search assistance platform is proposed to tackle above-mentioned problems. We expect this platform can effectively guide learners to search for the appropriate websites and achieve the purpose of self-learning assistance and enhance the students' learning performance. An experiment was conducted in a junior high school natural science course to examine the effectiveness of the collaborative web search assistance platform.

The remainder of the paper is organized as follows. The details of the proposed collaborative web search assistance platform are presented in Section 2. Section 3 addresses and discusses the experimental results. Conclusions and future work are set out in Section 4.

2. Architecture of the Collaborative Web Search Assistance Platform

The proposed platform consists of two major components, including the group grading module and the website ranking module. First, students issue their website search by typing in the keywords on the platform, and then the website ranking module is employed to evaluate the contents of the website. Notably, a well-known machine learning technique, Support Vector Regressions (SVRs), is adopted in the website ranking module to rank the website. Meanwhile, a group grading approach is adopted to derive three input parameters for the SVRs.

2.1. Group Grading Module

After a student typed a keyword on the platform, she/he is asked to press "Agree", "Strongly Agree", "Disagree", or "Strongly Disagree" buttons in order to evaluate the quality of the contents of the websites located by the platform. Notably, we also allowed students to press an "Uncertain" button on the Facebook-like platform, if they were not confident in terms of making a comment on the websites. The group grading module was then used to derive the required parameters for the SVRs to give a final evaluation of each websites after enough peers' comments were collected. In addition, the system will adjust the weights of the students that commented on the web sites based on their achievement. The detailed description of how the proposed group grading module operates is as follows.

Initialization: Each student's weight is first assigned by the teacher based on the student's past performance. The students' weights can be dynamically adjusted each time they comment on the websites.

Step 1: First, we sort the reliability of each student as follows,

$$w_1 \geq w_2 \geq \dots \geq w_M, \quad (1)$$

where M denotes the number of peers.

Step 2: Obtain the optimal combination of the students,

$$\hat{l} = \arg \max_l \left\{ \psi_0^l, \dots, \sum_{i=0}^{\lceil l/2 \rceil - 1} \psi_i^l, \dots, \sum_{i=0}^{\lceil K/2 \rceil - 1} \psi_i^K \right\}, \quad (2)$$

where K is the total number of students giving their opinions, and $\arg \max_l$ is the maximum of the elements in set \hat{l} . Each element in set \hat{l} can be expressed by,

$$\psi_0^l = \prod_{j=1}^l w_j, \quad (3)$$

$$\psi_1^l = \prod_{j=1}^l w_j + \sum_{n=1}^l \left\{ (1 - w_n) \cdot \prod_{j=1, j \neq n}^l w_j \right\}, \quad (4)$$

follows,

$$Imf = \frac{1}{\sum_{i=1}^M \beta_i} \sum_{i=1}^M \beta_i C_i \quad (12)$$

where C_i is a flag that is marked if the i th student gave a comment on the website. That is, $C_i = 1$ if i th student gave a comment on the website. Otherwise, C_i is set to zero. β_i denotes the weight of the selected keyword that is predetermined by the teacher.

2.2. Website Ranking Module

The website ranking module is used to compute the final ranking for the website comparing with other sample websites in the database. Support Vector Regression (SVR) [8] is used to implement this module. Three parameters for the SVR are three equations obtained from Equation (6), Equation (9), and Equation (12). SVR is a kind of supervised machine learning method that recognizes patterns and analyzes data; it is mostly used for classification and regression analysis. The major difference between the SVR and traditional regression techniques is that the SVR employs the structural risk minimization (SRM) approach, rather than the empirical risk minimization (ERM) approach typically adopted in statistical learning. The SRM attempts to minimize an upper threshold on the generalization rather than minimize the training error, and is expected to perform better than the traditional ERM approach. Furthermore, the SVR is a convex optimization, which guarantees that the local minimization is the unique minimization. In the recent literature, numerous researchers have adopted SVR to deal with classification and regression problems. Wu *et al.* [9] used SVR to predict the time spent on driving according to the speed of vehicles, traffic flow, and weather conditions. Users can handle the overall schedule more efficiently with this method. In addition, Liu *et al.* [10] compared three regression approaches, including SVR, Back-propagation Neural Network, and Partial Least Squares, to predict the Cold Modulus of Silicon Ceramic, and the results showed that SVR obtained better performance in terms of root mean square error than the other two methods.

3. Experimental Results

To verify the effectiveness of the proposed collaborative website search ranking mechanism, a total of 63 junior high school students participated in the study. One class related to natural science was assigned to be the experimental group in which the students were assisted by the website ranking mechanism, and the other was the control group in which the students used the ordinary search engine in the Internet. The experimental group included 31 students, while the control group had 32 students. After the teacher of a Natural Science course gave a detailed description of how to build a toy plane during a traditional classroom teaching activity, the students were asked to make a paper airplane using a plain piece of A4 size paper.

The students were asked to login to the platform after the cessation of classroom teaching activities. A question asking how to build an aircraft that would fly as long and as far as possible using a plain piece of A4 size paper was given to the students. The students were expected to surf on the Internet to address three main learning topics, including how to fly far, how to fly straight, and how to fly stably. The platform kept students' web search results in the database and evaluated the students' websites based on the students' comments. Each student was asked to take a pre-test and a post-test immediately before and after the learning activity, respectively.

We first compared the website ranking results between the proposed algorithm and that of the teacher. Notably, the evaluation results of the students' website search were also provided by the teacher after the experiment, and were then compared with those given by the proposed algorithm. The correction rate for the proposed algorithm was up to 88.27%.

We next compare the students' achievement before and after the learning activity, as illustrated in **Table 1**. Ten multiple choice questions were included in the pre-test and post-test, respectively, to verify students' learning performance. The questions were related to the basic information related to flying, such as how to make the paper airplane fly straight, far, and stably. The statistical results were obtained by running a t-test with the SPSS software package. As shown in **Table 1**, the average score received by the 31 pupils whose learning activity was supported by the proposed website search mechanism was significantly better than the average score before using the mechanism. The evidence of a highly significant correlation between pretest and posttest obtained in **Table 2** and the high significance level indicated by the results of the t-test in **Table 3** indicate that the learning

Table 1. Paired sample statistics.

	Mean	Standard deviation	Standard error mean
Pre-test	50.87	12.85	2.77
Post-test	64.32	16.25	3.22

Table 2. Paired sample correlations.

Correlation	Significance
0.526	0.003

Table 3. Paired sample test.

Mean Difference	t	df ^a	Sig. (2tailed)	95% C.I. ^b	
				Lower	Upper
-13.45	-5.15	30	0.000	-18.34	-8.03

^aDegrees of freedom; ^bConfidence interval.

effectiveness of the learners was indeed improved by the website search mechanism.

To verify whether the website search algorithm was helpful, two short questionnaires were given to students in the experimental group that included the questions “Is the website search system able to provide you with appropriate guidance?” and “Are you satisfied with the usage of the website search system?”. Among the 31 participants, 28 students believed that the web search assistance platform was useful and 29 students were satisfied with the usage of the web search assistance platform.

4. Conclusions

An intelligent web search assistance platform is proposed in this research. The proposed work is able to automatically evaluate whether or not the website the students search have addressed the related learning issues. The experimental results reveal that the online web search assistance platform proposed in this study effectively enhances the achievement of the students from a junior high school who are participating in learning activities related to natural science. In future work, we will consider incorporating students’ learning styles into the experiments, in order to provide adaptive feedback tailored toward students with different learning styles.

One limitation of this study is that the 31 students participate in this experiment to assess their learning performance before and after using the proposed system for a period lasting only one-month, due to the limitation of course time. We only used t-test to analyze students’ achievement before and after the conducted learning activity, and we will add more participants engaging in learning activity to explore the factors which may affect the learning performance via our proposed algorithm in future work. In addition, pre-existing personal friendships between the participants and how these relationships might affect learners’ social interactions have not been discussed in this study. Thus, further exploration of the impacts of social network interaction on enhancing collaboration and learning performance will be needed and considered in future work as well.

Acknowledgements

The authors would like to thank the National Science Council of Taiwan for financially supporting this research under Contract NSC 102-2511-S-259-010-MY3.

References

- [1] Prekop, P (2002) A Qualitative Study of Collaborative Information Seeking. *Journal of Documentation*, **58**, 533-547. <http://dx.doi.org/10.1108/00220410210441000>
- [2] Hyldega, J. (2006) Collaborative Information Behaviour—Exploring Kuhlthau’s Information Process Model in a Group-Based Educational Setting. *Information Processing & Management*, **42**, 276-298.

<http://dx.doi.org/10.1016/j.ipm.2004.06.013>

- [3] Fidel, R., Pejtersen, A.M., Cleal, B. and Bruce, H. (2004) A Multidimensional Approach to the Study of Human-Information Interaction: A case Study of Collaborative Information Retrieval. *Journal of the American Society for Information Science and Technology*, **55**, 939-953. <http://dx.doi.org/10.1002/asi.20041>
- [4] Lieberman, H., van Dyke, N. and Vivacqua, A. (1999) Let's Browse: A Collaborative Browsing Agent. *Knowledge-Based Systems*, **12**, 427-431. [http://dx.doi.org/10.1016/S0950-7051\(99\)00036-2](http://dx.doi.org/10.1016/S0950-7051(99)00036-2)
- [5] Romano, N.C., Roussinovm D., Nunamakerm J.F. and Chen, H. (1999) Collaborative Information Retrieval Environment: Integration of Information Retrieval with Group Support Systems. *Proceedings of the 32nd Hawaii International Conference on Systemsciences*, IEEE Computer Society Press, Los Alamitos. <http://dx.doi.org/10.1109/hicss.1999.772729>
- [6] Chirag, S., Robert, C. and Preben, H. (2014) Collaborative Information Seeking. *Computer*, **47**, 22-25.
- [7] Morris, M.R. (2008) A Survey of Collaborative Web Search Practices. *Proceedings of CHI 2008*, 1657-1660. <http://dx.doi.org/10.1145/1357054.1357312>
- [8] Drucker, H., Chris, J., Burges, C., Kaufman, L., Smola, A. and Vapnik, V. (1997) Support Vector Regression Machines. *Advances in Neural Information Processing Systems*, **9**, 155-161.
- [9] Wu, C.H., Ho, J.M. and Lee, D.T. (2004) Travel-Time Prediction with Support Vector Regression. *IEEE Transactions on Intelligent Transportation Systems*, **5**, 276-281. <http://dx.doi.org/10.1109/TITS.2004.837813>
- [10] Liu, X., Lu, W.C., Jin, S.G., Li, Y.W. and Chen, N.Y. (2006) Support Vector Regression Applied to Materials Optimization of Sialon Ceramics. *Chemometrics and Intelligent Laboratory Systems*, **82**, 8-14.