

Subject Recommendation Using Ontology for Computer Science ACM Curricula

M. M. Lotfy^{1,*}, A. A. Salama¹, H. A. El-Ghareeb², M. A. El-dosuky²

¹ Department of Mathematics and Computer Science, Faculty of Science, Port Said University, Egypt.

² Department of Information Systems, Faculty of Computers and Information Sciences, Mansoura University, Egypt.

Received: 01 November 2013; Accepted: 04 January 2014

Abstract

Recommender systems are needed to find subject items of one's interest. We review recommender systems and recommendation methods. We propose a subject personalization framework based on adaptive hypermedia for Computer Science ACM Curricula. We extend Hermes framework with subject recommendation functionality. We combine TF-IDF term extraction method with cosine similarity measure. Specialization and standard subject database are incorporated into the knowledgebase. Based on the performed evaluation, we conclude that semantic recommender systems in general outperform traditional recommenders systems with respect to accuracy, precision, and recall, and that the proposed recommender has a better F-measure than existing semantic recommenders.

Keywords: Ontology; Semantics-Based recommendation; Heuristics
©Martin Science Publishing. All Rights Reserved.

1. Introduction

Recommender systems are needed to find subject items of one's interest. Challenges in building recommender systems can be classified as those concerning the user, and those concerning the algorithms used [1]. Different models are proposed [2] to deal with the missing or incorrect data from subject recording measurements. Other challenges have a trade-off between them such as the perfect databases size and the cold-start problem. The cold-start problem can be solved by using information about the user's previous experience to calculate similarity measures to recommend new files [3]. Challenges about user compliance can benefit from many suggested strategies[4].

* Corresponding author.

E-mail address: dev.magdy2011@gmail.com (M. M. Lotfy).

Section 2 reviews the previous attempts in building subject recommenders and recommendation approaches. Section 3 presents our solution and the evaluation of the proposed framework. We conclude in Section 4 with plans for future work.

2. Previous Work

The State-of-the-Art for building recommender systems is shown in a recent papaer [5]. Many successful recommendation systems are proposed for news ([6], [7]) , nutrition planning ([8], [9], [10], [11], [12]) . The personalized recommendation technique in recommender systems, one of the most important tools of personal service in websites, makes great significance in Internet marketing activities of e-Commerce([13], [14]).

Recommendation system is a significant part of e-learning systems for personalization and recommendation of appropriate materials to the learner [15]. In the context of e-learning recommender systems, the users with greater knowledge have greater weight in the calculation of the recommendations than the users with less knowledge [16]. Recommender system can be used for predicting student performance [17].

There are four types of recommender approaches: content-based, semantics-based, collaborative filtering, and hybrid [22], but we restrict our discussion to the first two only. Content-based recommenders make use of Term Frequency-Inverse Document Frequency (TF-IDF)[18] and cosine similarity to compare the similarity between documents. Semantics is concerned only with concepts, and employing approaches such as concept equivalence [7], binary cosine[7], Jaccard [19], and semantic relatedness [20]. Next section shows how these approaches can be implemented.

3. Proposed Framework

The proposed framework is shown in fig. 1.

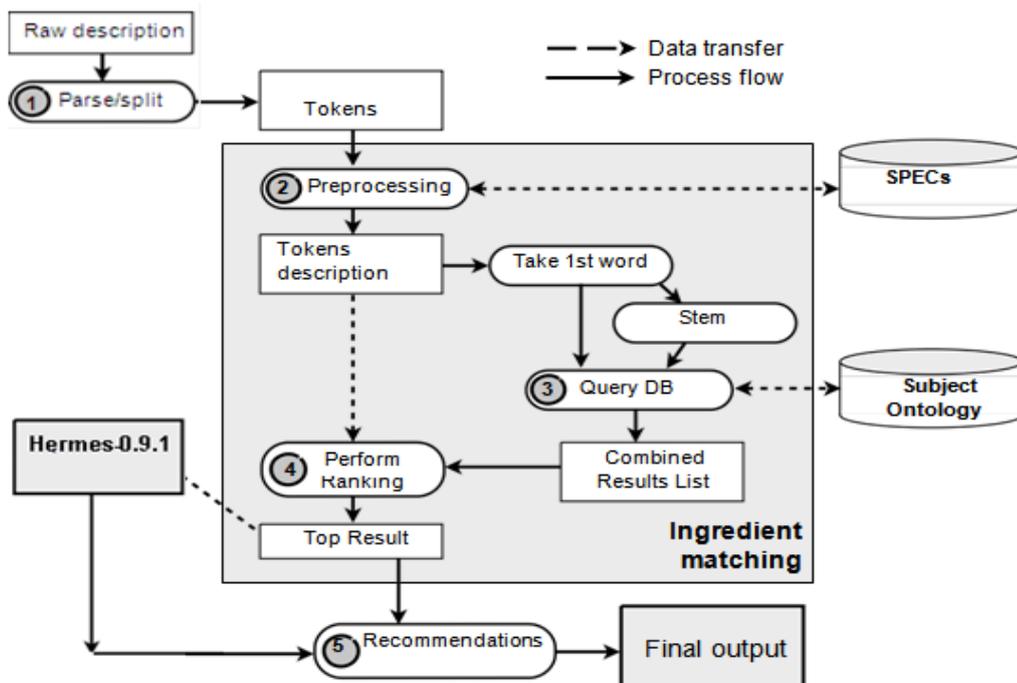


Figure 1. The proposed framework

The first step is to take the raw description directly from the user or from his profile. Stop words are removed, followed by stemming words back to the root and removing punctuation and converting to lower case. The next stage is to match the description or the output of the rule to the domain ontology. User profile is constructed by calculating TF-IDF values for each term. We determine the term frequency (TF) $f_{i,j}$ for a term t_i within an recipe a_j :

$$tf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

dividing $n_{i,j}$, the number of occurrences of term t_i in file by a_j , the total number of terms in the document. Then the inverse document frequency (IDF):

$$idf_i = \log \frac{|A|}{|\{a: t_i \in a\}|} \tag{2}$$

dividing the total number of subject items by the number of subject items containing term t_i . The final value is computed by multiplying TF and IDF:

$$tfidf_{i,j} = tf_i \times idf_i \tag{3}$$

Semantic measures benefit from the ontology that is defined by a set of concepts:

$$C = \{c_1, c_2, c_3, \dots, c_n\} \tag{4}$$

The subject file can be defined by a set of p concepts:

$$A = \{c_1^a, c_2^a, c_3^a, \dots, c_p^a\} \tag{5}$$

The user profile, U , consists of q concepts found in the subject items read by the user:

$$U = \{c_1^u, c_2^u, c_3^u, \dots, c_q^u\} \tag{6}$$

The similarity between a subject file and the user profile can be computed by:

$$Similarity(U, A) = \begin{cases} 1 & \text{if } |U \cap A| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

We can employ binary cosine to compute the similarity:

$$B(U, A) = \frac{|U \cap A|}{|U| \times |A|} \tag{8}$$

by dividing the number of concepts in the intersection of the user profile and the unread subject file by the product of the number of concepts in respectively U and A .

Similarly, Jaccard computes the similarity between two sets of concepts:

$$J(U, A) = \frac{|U \cap A|}{|U \cup A|} \tag{9}$$

Semantic neighborhood of c_i is all concepts directly related to c_i including c_i :

$$N(c_i) = \{c_1^i, c_2^i, c_3^i, \dots, c_n^i\} \tag{10}$$

A subject item a_k , which consists of m concepts is described as the following set:

$$A_k = \{c_1^k, c_2^k, c_3^k, \dots, c_m^k\} \tag{11}$$

To compare two new items n_i and n_j , a vector can be created:

$$V_l = (\langle c_1^l, w_1^l \rangle, \dots, \langle c_p^l, w_p^l \rangle) \quad l \in \{i, j\} \tag{12}$$

where w_i is the weight of c_i . The similarity between subject items a_i and a_j is :

$$\text{SemRel}(a_i, a_j) = \cos(V_i, V_j) = \frac{V_i \cdot V_j}{\|V_i\| \times \|V_j\|} \in [0,1] \tag{13}$$

The proposed framework is implemented in PHP. It allows the user to formulate queries and execute them to retrieve relevant subject items. We use the approach applied to adaptive hypermedia [21] and Hermes framework[6]. We extend Hermes with subject recommendation functionality. It utilizes OWL[23] for representing the ontology.

Performed tests are based on a subject corpus of 145 courses extracted from the Computer Science ACM Curricula [24] as shown in Table 1.

Table 1. Subject database

Group	No. of courses	Group	No. of courses
Algorithms	11	Artificial intelligence	11
ComputerArchitecture	10	Networking	9
Computation	3	Operating Systems	14
Discrete math	6	Problem formulation	8
Graphic	12	Programming languages	11
HCI	10	Software engineering	14
Information Models	15	Special topics	11

We have used 5 users with different but well-defined interests in our experiments. An example of a user interest is “algorithms”. Each user has manually rated the subject items as relevant or non-relevant for his interest. For each user we split the subject items corpus in two different sets: 60% of the subject items are the training set and 40% of the subject items are the test set. Recommenders compute the similarity between the subject items and previously

computed user profile. If the computed similarity value is higher than a predefined cut-off value the subject item is recommended and ignored otherwise.

Evaluating the recommenders is done by measuring accuracy, precision, recall, specificity, and F-measure. This is done by calculating a confusion matrix for each user. Fig. 2 shows the results of the evaluations.

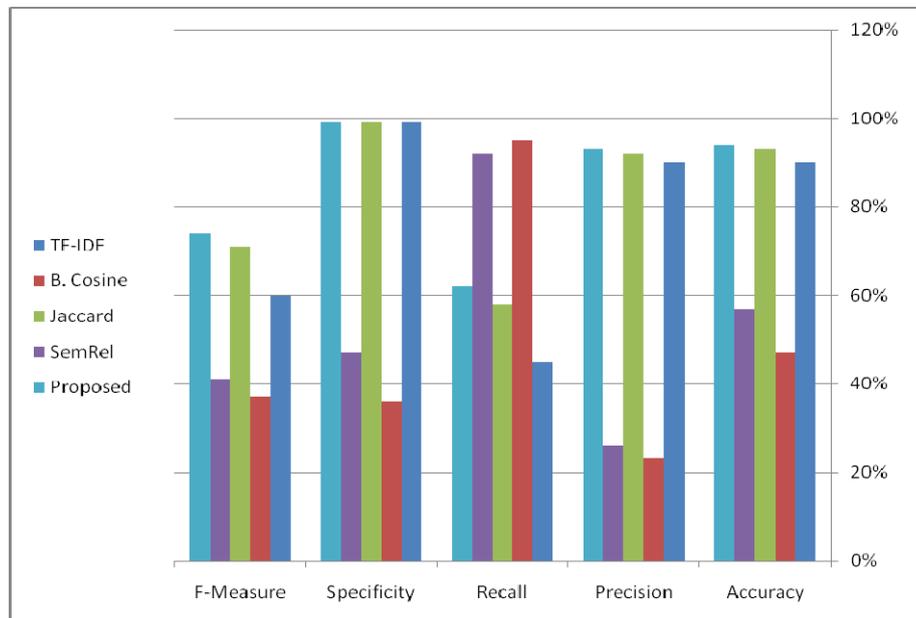


Figure 2. Evaluation results

The best recommenders for accuracy is the proposed framework, for precision is the proposed framework, for recall is binary cosine, for specificity are TF-IDF, Jaccard, and the proposed framework, and for F-measure is the proposed framework. The proposed algorithm scores well on accuracy as it makes relatively small amount of errors for both recommended subject as well as discarded subject items. For precision, the proposed algorithm scores the best for precision as most recommended subject items are relevant. The good results for recall obtained by the concept equivalence are due to the optimistic nature of the algorithm: any subject item which involves previously viewed concepts is recommended. TF-IDF, Jaccard, and the proposed framework score well on specificity as these algorithms do not recommend most of the non-relevant subject items.

4. Conclusion and Future Work

We propose a subject personalization framework based on adaptive hypermedia for Computer Science ACM Curricula. We extend Hermes framework with subject recommendation functionality. We combine TF-IDF term extraction method with cosine similarity measure. Specialization and standard subject database are incorporated into the knowledgebase. Performed tests are based on a subject corpus of 145 courses extracted from the Computer Science ACM Curricula.

Based on the performed evaluation, we conclude that semantic recommender systems in general outperform traditional recommenders systems with respect to accuracy, precision, and

recall, and that the proposed recommender has a better F-measure than existing semantic recommenders.

In the future we plan to extend the querying language by defining its grammar, and applying it for extracting deep knowledge from subject ontology.

Another possible research direction relates to the advanced traditional weighting schemes that other than TF-IDF such as logarithmic TF functions [25]. Another research direction is the considered similarity function. We would like to evaluate alternatives for cosine similarity as Lnu.ltu [26] which seem to remove some of the cosine similarity bias favoring long documents over short documents.

Also, we would like to incorporate other services such as opinion mining [27]. As additional further work we would like to consider other types of available subjects in the ACM Curricula.

References

- [1] Stefanie Mika, "Challenges for Nutrition Recommender Systems," *Proceedings of the 2nd Workshop on Context Aware Intel. Assistance*, Berlin, Germany, pp. 25-33, 2011.
- [2] R. H. Keogh and I. R. White, "Allowing for never and episodic consumers when correcting for error in food record measurements of dietary intake", *Biostatistics*, 2011.
- [3] Y. Van Pinxteren, G. Geleijnse and P. Kamsteeg, "Deriving a recipe similarity measure for recommending healthful meals", *Proc. of the 16th international conference on Intelligent user interfaces*, New York, NY, USA, pp. 105-114, 2011.
- [4] M. H. Becker and L. A. Maiman, "Strategies for enhancing patient compliance", *Journal of Community Health*, vol. 6, no. 2, pp. 113-135, 1980.
- [5] J. Bobadilla, F. Ortega, A. Hernando and A. Gutiérrez, "Recommender systems survey", *Knowledge-Based Systems*, vol. 46, pp. 109-132, 2013.
- [6] J. Borsje, L. Levering and F. Frasincar, "Hermes: a Semantic Web-Based News Decision Support System", *In 23rd Annual ACM Symposium on Applied Computing*, pp. 2415-2420, 2008.
- [7] W. IJntema, F. Goossen, F. Frasincar and F. Hogenboom, "Ontology-Based News Recommendation", *In EDBT/ICDT International Workshop on Business Intelligence and the Web*, 2010.
- [8] P. M. Powers and L. W. Hoover, "Calculating the nutrient composition of recipes with computers," *J. Am. Diet. Assoc.*, vol. 89, pp. 224-232, 1989.
- [9] K. Hammond, "Chef: A model of case-based planning", in *Proceedings of the National Conference on AI*, 1986.
- [10] J. Freyne and S. Berkovsky, "Intelligent food planning: personalized recipe recommendation", *Proceedings of the 15th international conference on Intelligent user interfaces*, New York, NY, USA, pp. 321-324, 2010.
- [11] J. Freyne and S. Berkovsky, "Intelligent food planning: personalized recipe recommendation," *Proceedings of the 15th international conference on Intelligent user interfaces*, New York, NY, USA, pp. 321-324, 2010.
- [12] M. A. El-Dosuky, Ahmed EL-Bassiouny, Taher Hamza, Magdy Rashad, "Food Recommendation Using Ontology and Heuristics", *CCIS*, vol. 322, pp. 423-429, 2012.
- [13] J. Ben Schafer, Joseph Konstan, and John Riedl, "Recommender systems in e-commerce", *Proceedings of the 1st ACM conference on Electronic commerce*, ACM, New York, NY, USA.

- [14] Kangning Wei, Jinghua Huang, Shaohong Fu, “A Survey of E-Commerce Recommender Systems”, *International Conference on Service Systems and Service Management*, pp.1-5, 2007.
- [15] Mojtaba Salehi, Isa Nakhai Kmalabadi, “A Hybrid Attribute-based Recommender System for E-learning Material Recommendation”, *IERI Procedia*, vol. 2, pp. 565-570, 2012.
- [16] J. Bobadilla, F. Serradilla, A. Hernando and MovieLens, “Collaborative filtering adapted to recommender systems of e-learning”, *Knowledge-Based Systems*, vol. 22, no. 4, pp. 261-265, 2009.
- [17] Nguyen Thai-Nghe, Lucas Drumond, Artus Krohn-Grimberghe and Lars Schmidt-Thieme, “Recommender system for predicting student performance”, *Procedia Computer Science*, vol. 1, no. 2, pp. 2811-2819, 2010.
- [18] G. Salton and C. Buckley, “Term-Weighting Approaches in Automatic Text Retrieval”, *Information Processing and Management*, vol. 24, no.5, pp. 513-523, 1988.
- [19] Jaccard, P., “Étude Comparative de la Distribution Florale dans une Portion des Alpes et des Jura”, *Bulletin del la Société Vaudoise des Sciences Naturelles*, vol. 37, pp. 547-579, 1901.
- [20] F. Getahun, J. Tekli, R. Chbeir, M. Viviani and K. Yetongnon, “Relating RSS News/Items”, *In 9th International Conference on Web Engineering*, pp. 442-452, 2009.
- [21] P. D. Bra, A. T. M. Aerts, G. J. Houben and H. Wu, “Making General-Purpose Adaptive Hypermedia Work”, *In World Conference on the WWW and Internet*, pp. 117-123, 2000.
- [22] G. Adomavicius and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions”, *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, 2005.
- [23] S. Bechhofer, F. V. Harmelen, J. Hendler, I. Horrocks, D. L. McGuinness and P. F. Patel-Schneider, “OWL Web Ontology Language Reference W3C Recommendation”, 10 February, 2004.
- [24] <http://www.acm.org/education/curricula-recommendations>, accessed December 10th, 2013.
- [25] C. Buckley, J. Allan and G. Salton, “Automatic Routing and Retrieval Using Smart: TREC-2”, *Information Porcessing and Management*, vol. 31, no. 3, pp. 315-326, 1995.
- [26] A. Singhal, C. Buckley and M. Mitra, “Pivoted Document Length Normalization”, *In 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 21-29, 1996.
- [27] M. A. El-Dosuky, M. Z. Rashad, T. T. Hamza, A.H. EL-Bassiouny, “Robopinion: Opinion Mining Framework Inspired by Autonomous Robot Navigation”, CoRR cs.CL/arXiv:1209.0249, 2012.