ASEE's 123rd Annual • Conference & Exposition • New Orleans, LA • June 26-29, 2016

New Orleans JAZZED about Engineering Education

Paper ID #15891

User-Based Collaborative Filtering Recommender Systems Approach in Industrial Engineering Curriculum Design and Review Process

Dr. Ebisa Wollega, Colorado State University - Pueblo

Ebisa D. Wollega, Ph.D. is an Assistant Professor of Engineering at Colorado State University-Pueblo. He received his Ph.D. degree in industrial engineering from the University of Oklahoma. His research interest areas include stochastic systems modeling and optimization, big data analytics, large scale optimization algorithms, and engineering education. His email is ebisa.wollega@csupueblo.edu and his web page is http://ceeps.csupueblo.edu/Engineering/FacultyandStaff/Pages/EbisaWollega.aspx.

Mr. Vitor Ambrosio Winckler, Colorado State University - Pueblo

Vitor Ambrosio Winckler is an undergraduate student of Mechatronics Engineering program at the Polytechnic School of Sao Paulo University, Brazil. He is participating in the Brazil Scientific Mobility Program with the Colorado State University – Pueblo as the host university. He worked for Siemens in Brazil where he was recognized for two awards. His research interests include artificial intelligence and big data analytics.

User-based Collaborative Filtering Recommender Systems Approach in Industrial Engineering Curriculum Design and Review Process

Abstract

Industrial engineering curriculum is relatively very sensitive to changes in industry needs compared to other engineering disciplines because of its structure. The effectiveness of the curriculum design and review process depends on the variety and the volume of input data. Industrial engineering educators usually collect data from students, alumni, and industry stakeholders. With the availability of massive online data, mining relevant information and combining the information with the data collected from the traditional data sources would improve the efficacy of the industrial engineering curriculum design and review process. In this paper, we propose online job posting data as additional relevant information that can be integrated to the curriculum design and review process. We describe the adaptation of a user-based collaborative filtering recommender systems algorithm to analyze the online data and to convert the data into relevant information that can be used as input to the process. An undergraduate industrial engineering Operations Planning and Control course case study was used to illustrate the adaptation of the algorithm. Some of the topics taught in the course were searched on websites that advertise jobs and tallied. A professor who is familiar with the topics also provided expert judgments with regard to the relevance of the topics to industry needs. Both data sets were used as inputs to the algorithm. The experimental results show that some of the topics are highly correlated with the expert judgment than others; these topics would be given more emphasis than the less correlated topics during the curriculum design and review process. Analysis of new topics that did not receive expert judgments is also presented. The method proposed in this paper plays a great role in continuous curriculum review process as massive data sets can be extracted from online sources and processed within short time window. The industrial engineering educators can make use of more of the online data as input to curriculum design and review process to improve the efficiency of the process. This paper can also lead engineering educators to possibly explore the contribution of massive online data as an input to curriculum design and review process instead of simply relying on the traditional data sources.

1.0. Introduction

The effectiveness of engineering curriculum design and review process is a major part of engineering education as engineering curriculum is highly affected by technological enhancements and stakeholders' requirement changes. Industrial engineering curriculum is relatively very sensitive to the changes compared to other engineering disciplines because of its composition. It covers a wide range of areas including systems design and improvement, data analytics, process optimization, energy utilization, and human factors. The sources and the types of input data used in the industrial engineering curriculum design and review process significantly affect the responsiveness of the curriculum to the technological changes. The data are generally gathered from students, alumni, and industry stakeholders. Additional data from online sources has not been fully exploited by industrial engineering educators. As we have observed in recent years, the use of massive online data has improved the decision making accuracy of many companies such as Microsoft, Facebook, Amazon, and Google. We believe that using online data such as industrial engineering related job posting data would also improve the efficiency of the industrial engineering curriculum design and review process.

This paper presents the use of the job posting data in industrial engineering curriculum design and review process. We illustrate how the user-based collaborative filtering recommender systems algorithm can utilize the job posting data to evaluate the content of one of the undergraduate level industrial engineering courses, Operations Planning and Control (OPC) course.

Recommender systems have been widely used in an online shopping to match users with items[4]. It is common to see Google and Amazon recommending you an item similar to what you have searched or purchased before. These recommendations are generated using three main techniques: collaborative filtering, content based filtering, and knowledge based filtering algorithms[13]. Collaborative filtering groups users according to their preferences[1], then provides recommendations for an active member of the group using the preferences of the remaining members of the group[6]. Content based filtering associates items and recommends users the items that they have experienced before[10]. Knowledge-based filtering uses knowledge about users and products to generate recommendations, and to indicate what types of products meet the users' requirements[9].

Collaborative filtering algorithm is the commonly used type of recommender system algorithm in online shopping[2]. However, one can choose from any of the three recommender systems depending on the availability of data and the characteristics of the items needed[5]. Researchers have used collaborative filtering algorithm in e-commerce, organizational knowledge management, and education[11]. In education, collaborative filtering recommender systems are applied to help intelligent tutoring systems, effective e-learning, computer managed instructions, internet based training, and online course recommendations. In this paper, the user-based type collaborative filtering recommender systems algorithm is adapted. The implementation procedure of the algorithm is presented in the following section.

2.0. Methodology

The first step in user-based collaborative filtering algorithm implementation is to identify the items and the users. The items selected for our analysis were the topics covered in OPC course. The online job posting websites and a professor who teaches the course were chosen as the users. However, anyone who is involved in the course content design and review process can play the role of the professor.

Five websites where industrial engineering related jobs are posted were selected. These are Indeed.com, CareerBuilder.com, US.jobs, AmericasJobExchange.com, and Dice.com. The selection of the websites was based on their ratings. The first two were top rated in 2015. The last three received lower ratings in the same year.

The course topics selected were Forecasting, Inventory Control, Production Scheduling, Capacity Planning, Cycle Time, and Lean. Three other topics were also added later on in the discussion section. The topics were selected from *Factory Physics* (3rd ed.) by Hall and Spearman[3] which is used as a textbook in many industrial engineering schools.

The topics were searched on all the five websites in six U.S. states between January 18, 2016 and January 22, 2016. The states were Colorado, Oklahoma, Texas, Kansas, Arizona, and Nebraska. The selection of the states was through informal conversations with students that are enrolled in the OPC course at Colorado State University-Pueblo. The students would like to work as industrial engineers in one of the six states.

Online text mining technique was used to extract the topics from the job descriptions of the websites. The data were stored in a database. The topics frequency data on the websites were the particular interest for this paper. The appearance of the topics in the job descriptions of the websites on a given day was counted. The counts were normalized to unity across the topics for all the websites. The professor also provided expert rating for each topic. The rating here is defined as the portion of the semester hour each topic may require assuming that each topic would be included in the course.

After the data were collected, the user-based collaborative filtering recommender systems algorithm was run to find similarities between the users (the websites and the professor). The similarities between the websites were computed to determine their nearest neighbors. Cosine similarity measure Pearson's correlation coefficient formula shown below was used to measure the similarities[6,8]:

$$sim\,(x,y) = \frac{\sum_{s\in S_{xy}} r_{x,y} r_{y,s}}{\sqrt{\sum_{s\in S_{xy}} r_{x,s}^2}\sqrt{\sum_{s\in S_{xy}} r_{x,s}^2}},$$

where $r_{x,s}$ is the rating of user x on topic s and $r_{y,s}$ is the rating of user y on topic s, $s_{x,y}$ indicates the items that users x and y co-experienced. Then rating is computed by a weighted average of the nearest neighbors using the following formula[7,8]:

$$r_{x,s} = \bar{r}_x + \frac{\sum_{y\in S_{xy}} (r_{y,s} - \bar{r}_x) sim(x,y)}{\sum_{y\in S_{xy}} sim(x,y)},$$

where, $\bar{r}_x$ is the average rating of user x. Two assumptions were used in the algorithm: (1) the professor was interested in relating the frequency of the topics on the websites to the OPC course, and (2) the frequency of the topics on the websites reflected the technological and stakeholders' need changes and the professor agreed with the existence of the changes[12].

The OPC course case study is presented in the next section to illustrate the adaptation of the user-based collaborative filtering recommender systems algorithm. The reader should note that a larger data set than presented in this paper can be processed using the algorithm proposed.

3.0. Results and Discussions

Fourteen thousand pages of job postings data related to industrial engineering were extracted and used in the analysis. The topics frequency data for each website are presented in Table 1 below.

Production Scheduling appeared 35 times on Indeed.com, 12 times on CareerBuilder.com, 7 times on US.jobs, and 8 times on Dice.com, and so on.

Table 1. Topics frequency data on websites

	Website				
Topics	Indeed.com	CareerBuilder.com	US.jobs	AmericasJobExchange.com	Dice.com
Production Scheduling	35	12	7	7	8
Capacity Planning	16	9	16	8	97
Inventory Control	202	98	86	44	124
Lean	239	98	162	68	260
Forecasting	44	30	49	19	46
Cycle Time	7	6	10	9	0

The Pearson's correlation coefficient matrix in Table 2 is computed for the websites and the topics data presented in Table 1. The correlation coefficients show that the presence of the topics on the websites is strongly correlated to each other since all the values are close to one. The topics frequency progression on the websites is very similar.

Table 2. Correlation coefficient matrix between job posting websites

	Indeed.com	CareerBuilder.com	US.jobs	AmericasJobExchange.com
Indeed.com				
CareerBuilder.com	0.988			
US.jobs	0.936	0.912		
AmericasJobExchange.com	0.971	0.950	0.991	
Dice.com	0.861	0.817	0.929	0.913

Now given the pairwise Pearson's correlation coefficient of the websites in Table 2, we compare every website topics rating to the professor's rating. The comparison provides the similarities of the frequency of the topics on the websites with the professor's ratings. The professor gives expert rating for each topic. In the previous section, rating was defined as the portion of the semester course hours each topic may require. Five different scenarios of the professor's ratings were created and named as the time allocation (A1, A2, A3, A4, and A5). The time allocation was normalized to unity and the corresponding website similarities (Sim1, Sim2, Sim3, Sim4, Sim5) were computed for each scenario. The results are shown in Table 3 below.

Table 3. Topics time allocation and the corresponding similarity results

A1	Forecasting: 0.10 Inventory Control: 0.10		Production Scheduling: 0.10 Capacity Planning:0.10		Cycle Time:0.05 Lean:0.10
	Dice.com	Indeed.com	US.jobs	AmericasJobExchange.com	CareerBuilder.com
Sim1	0.45	0.40	0.36	0.33	0.29
A2	Forecasting: 0.20 Inventory Control: 0.20		Production Scheduling: 0.10 Capacity Planning:0.10		Cycle Time:0.05 Lean:0.10
	Dice.com	Indeed.com	US.jobs	AmericasJobExchange.com	CareerBuilder.com
Sim2	0.12	0.34	0.26	0.26	0.77
A3	Forecasting: 0.20 Inventory Control: 0.30		Production Scheduling: 0.10 Capacity Planning:0.10		Cycle Time:0.05 Lean:0.10
	Dice.com	Indeed.com	US.jobs	AmericasJobExchange.com	CareerBuilder.com
Sim3	0.16	0.46	0.28	0.33	0.94
A4	Forecasting: 0.20		Production Scheduling: 0.20		Cycle Time:0.05

	Inventory Control: 0.20		Capacity Planning:0.05		Lean:0.20
	Dice.com	Indeed.com	US.jobs	AmericasJobExchange.com	CareerBuilder.com
Sim4	0.32	0.59	0.54	0.54	0.43
A5	Forecasting: 0.20 Inventory Control: 0.20		Production Scheduling: 0.20 Capacity Planning:0.10		Cycle Time:0.05 Lean:0.10
	Dice.com	Indeed.com	US.jobs	AmericasJobExchange.com	CareerBuilder.com
Sim5	-0.13	0.15	0.00	0.01	0.57

The topics time allocation A1 has more similarity with the frequency of the topics on Dice.com (Sim1 = 0.45) than the other websites; however, A1 has lesser similarity with the frequency of the topics on CareerBuilder.com (Sim1 = 0.29). Under this scenario, the professor would prefer to use the data extracted form Dice.com than the remaining websites during the OPC course design and review process. Similarly, the time allocation A2 is better aligned with the CareerBuilder.com (Sim2 = 0.77) than the rest of the websites. Now the professor would prefer to extract and use data from the CareerBuilder.com website than the rest of the websites. For the time allocation scenario A5 we can see that there is actually a dissimilarity of the topics presence on Dice.com (-0.13). This scenario indicates that Dice.com data would not be a wise choice for the OPC course design and review process.

The reason why the professor is looking for a particular website here is that a job may be advertised across multiple websites at a time. Aggregating the data may exaggerate the ratings of the topics. For example, company X may post a job description that contains Production Scheduling on all the five websites. Summing up the topic across all the websites would give an equivalent rating of frequency five while the topic should have received a rating equivalent to frequency one.

The professor may also need to explore more topics that s/he has not yet provided a rating for. For example, the professor needs time allocation recommendations for the Throughput Control, Shop Floor Control, and Bottleneck Planning. Given the frequency of the topics on the websites as shown in Table 4, s/he needs to determine the OPC course semester time allocation for each of the new topics. In this case, a time allocation recommendation procedure that uses the websites weighted data is necessary as described below.

Table 4. New topics appearance frequency data from job posting websites

	Indeed.com	CareerBuilder.com	US.jobs	AmericasJobExchange.com	Dice.com
Throughput Control	19	7	10	8	2
Shop Floor Control	8	12	2	1	5
Bottleneck Planning	5	2	0	0	1

The topics are rated by generating a weighted score that ranks the websites. The normalized frequency of the new topics (topics in Table 4) on each website is multiplied by how similar they are to the professor’s time allocation for the previous topics (topics in Table 3). The product of the two is cumulated across all the websites. Then the cumulative is divided by the summation of the professor’s similarities across the websites to recommend time allocations for the new topics. The recommended results are given in Table 5 below. The professor’s similarity data in the second column of Table 5 is taken from the time allocation A1 scenario presented in Table 3.

Table 5. New topics time allocation recommendation

	Sim	Throughput	(Sim)*(Throughput)	Shop Floor	(Sim)*(Shop Floor)	Bottleneck	(Sim)*(Bottleneck)
Indeed.com	0.45	0.079	0.036	0.033	0.015	0.021	0.009
CareerBuilder.com	0.40	0.029	0.012	0.050	0.020	0.008	0.003
US.jobs	0.36	0.041	0.015	0.008	0.003	0.000	0.000
AmericasJobE.com	0.33	0.033	0.011	0.004	0.001	0.000	0.000
Dice.com	0.29	0.008	0.002	0.021	0.006	0.004	0.001
Total			0.075		0.045		0.014
Sim sum			1.830		1.830		1.830
Total/Sim sum			0.041		0.025		0.008

One can see from Table 5 that the recommended time allocation for Throughput Control = 0.041, Shop Floor Control = 0.025, and Bottleneck Planning = 0.008. The professor will decide which one of the three topics to include in the OPC course, or how much time to allocate to each of the new topics should all the three topics were included in the course.

4.0. Conclusion

This paper presents an adaptation of the user-based collaborative filtering recommender systems algorithm that can be used in industrial engineering curriculum design and review process. A case study for one of the industrial engineering courses was presented to illustrate the implementation. The case study experimental results indicate two major contributions of the algorithm in curriculum design and review process: (1) the algorithm can suggest what type of course topics require more emphasis than others based on their frequency of appearance on job posting websites, and (2) the algorithm can suggest ratings of new topics based on past experiences as shown in Table 5. It was assumed that the frequencies of the topics indicate the extent at which the industry and stakeholders' technological requirements change. The adaptation of the algorithm would improve the efficiency of industrial engineering curriculum design and review process for two main reasons: (1) massive data sets can be extracted and analyzed within a very short time window, and (2) continuous curriculum review process can be done through the automation of the algorithm. As a future research direction, it would be interesting to study the impact of each of the input data sources on the efficiency of the industrial engineering curriculum design and review process. It would also be crucial to build a database and keep the progression of the course topics appearance on job posting websites for a wide range of the industrial engineering courses. This could provide industrial engineering educators with real-time information and historical trends associated with the technological changes of a topic in a course.

Bibliography

[1]. Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction*, *4*(2), 81-173.

[2]. Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*, *22*(1), 5-53.

[3]. Hopp, W.J. and Spearman, M.L. (2011). *Factory Physics* (3rd ed.). Long Grove, IL: Waveland Press, Inc.

[4]. Jannach, D., & Friedrich, G. (2011, July). Tutorial: Recommender Systems. In *Proceedings of the International Joint Conference on Artificial Intelligence, Barcelona*.

[5]. Melville, P., & Sindhwani, V. (2010). Recommender systems. In *Encyclopedia of machine learning* (pp. 829-838). Springer US.

[6]. Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to recommender systems handbook* (pp. 1-35). Springer US.

[7]. Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In *The adaptive web* (pp. 291-324). Springer Berlin Heidelberg.

[8]. Segran, T. (2007). *Programming Collective Intelligence: Building Smart Web 2.0 Applications.* Sebastopol, CA: O'Reilly Media, Inc

[9]. Trewin, S. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Science: Volume 69-Supplement 32*, 180.

[10]. Van Meteren, R., & Van Someren, M. (2000, May). Using content-based filtering for recommendation. In *Proceedings of the Machine Learning in the New Information Age: MLnet/ECML2000 Workshop* (pp. 47-56).

[11]. Xiao, B., & Benbasat, I. (2007). E-commerce product recommendation agents: Use, characteristics, and impact. *Mis Quarterly*, *31*(1), 137-209.

[12]. Zhao, Z. D., & Shang, M. S. (2010, January). User-based collaborative-filtering recommendation algorithms on hadoop. In *Knowledge Discovery and Data Mining, 2010. WKDD'10. Third International Conference on* (pp. 478-481). IEEE.

[13]. Zuva, T., Ojo, S. O., Ngwira, S., & Zuva, K. (2012). A survey of recommender systems techniques, challenges and evaluation metrics. *International Journal of Emerging Technology and Advanced Engineering*.