UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

OER Recommendations to Support Career Development

Tavakoli, M.; Faraji, A.; Mol, S.T.; Kismihók, G.

DOI
10.1109/FIE44824.2020.9274175

Publication date
2020

Document Version
Final published version

Published in
2020 IEEE Frontiers in Education Conference (FIE)

License
Article 25fa Dutch Copyright Act (https://www.openaccess.nl/en/policies/open-access-in-dutch-copyright-law-taverne-amendment)

Link to publication

Citation for published version (APA):
Tavakoli, M., Faraji, A., Mol, S. T., & Kismihók, G. (2020). OER Recommendations to Support Career Development. In *2020 IEEE Frontiers in Education Conference (FIE): 21-24 Oct. 2020, Uppsala, Sweden* (Proceedings - Frontiers in Education Conference; Vol. 50). IEEE. https://doi.org/10.1109/FIE44824.2020.9274175

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:09 Nov 2026

OER Recommendations to Support Career Development

Mohammadreza Tavakoli
Leibniz Information Centre for Science and Technology
Hannover, Germany
reza.tavakoli@tib.eu

Ali Faraji
RWTH Aachen University
Aachen, Germany
abdolali.faraji@rwth-aachen.de

Stefan T. Mol
University of Amsterdam
Amsterdam, Netherlands
s.t.mol@uva.nl

Gábor Kismihók
Leibniz Information Centre for Science and Technology
Hannover, Germany
gabor.kismihok@tib.eu

***Abstract*—This *Work in Progress Research* paper departs from the recent, turbulent changes in global societies, forcing many citizens to re-skill themselves to (re)gain employment. Learners therefore need to be equipped with skills to be autonomous and strategic about their own skill development. Subsequently, high-quality, on-line, personalized educational content and services are also essential to serve this high demand for learning content. Open Educational Resources (OERs) have high potential to contribute to the mitigation of these problems, as they are available in a wide range of learning and occupational contexts globally. However, their applicability has been limited, due to low metadata quality and complex quality control. These issues resulted in a lack of personalised OER functions, like recommendation and search. Therefore, we suggest a novel, personalised OER recommendation method to match skill development targets with open learning content. This is done by: 1) using an OER quality prediction model based on metadata, OER properties, and content; 2) supporting learners to set individual skill targets based on actual labour market information, and 3) building a personalized OER recommender to help learners to master their skill targets. Accordingly, we built a prototype focusing on *Data Science* related jobs, and evaluated this prototype with 23 data scientists in different expertise levels. Pilot participants used our prototype for at least 30 minutes and commented on each of the recommended OERs. As a result, more than 400 recommendations were generated and *80.9%* of the recommendations were reported as useful.**

***Index Terms*—OER, Open Educational Resource, educational recommender system, labour market intelligence, lifelong learning, machine learning, text mining**

I. Introduction

In recent years, we have been facing with a quickly evolving labour market leading to several educational challenges caused by the gap between what educational institutions offer and what is actually required by labour market [1]–[3]. Furthermore, with the emergence of COVID-19 pandemic, and the collapse of a number of local and global industries, large numbers of employees need to re-skill themselves in order to regain employment. Therefore, updating individuals' skills and knowledge, to adapt to post COVID-19 labour markets, is a deep concern of our current societies, including educational decision makers, employers, job holders/seekers alike. To tackle this enormous challenge, an individual level self development plan based on dynamic labour market information and Open Educational Resources (OERs) could be critically important [4]. The advantage of OERs is that they are provided by experts from diverse disciplinary backgrounds (e.g., computer science, medicine, construction) operating in different contexts (e.g., languages, expertise levels, government, profit organizations), which is important for labour market driven education.

Previous efforts to build educational search and recommender systems using the rapidly growing amount of OERs [5]–[7] revealed the lack of high-quality OER metadata, and quality control [8]. These issues seriously limit the deployment of high-quality OER search and recommendation services [9], [10].

In this paper, therefore, we showcase our attempt to extend our personalised, labour market information based OER recommender [4] with OER quality prediction models [11] and enhanced metadata scoring [8]. Furthermore, we show the integration process of 7 different OER repositories, and our latest learner dashboard prototype focused on *Data Science* related jobs. Finally, we describe our efforts to validate our prototype using a sample of 23 data science subject matter experts (SMEs).

II. State of the Art

We can group the related work to date into two main categories: A) OER quality assessment, and B) OER recommender systems.

A. OER Quality Assessment

Due to the large amount of OERs being created and uploaded by people around the world, manual quality control of OERs is getting more and more complicated, if not impossible. In general, high-quality metadata is a mandatory property for providing data driven services [12]. In the area of OERs, low-quality metadata not only reduces the discoverability of OERs, but also has a strong negative effect on their usability [13]. Accordingly, some studies attempted to define metrics (e.g. completeness and consistency of metadata [14], provenance, and accuracy of metadata [15]) in order to assess the quality of OER metadata. To support this, [8] showed that a metadata scoring method can be used to define a model to predict OER quality.

Authorized licensed use limited to: Universiteit van Amsterdam. Downloaded on February 25,2021 at 15:09:47 UTC from IEEE Xplore. Restrictions apply.

Additionally, [16] provides a quality assessment framework based on existing developmental models in the area of e-learning, semantic-based methods, and NLP techniques. Moreover, in the case of open educational videos, quality can be predicted on the basis of video transcripts, popularity metrics (i.e. rate, likes, dislikes), and length [11].

B. OER Recommendation Systems

The literature on high-quality OER recommender systems is very limited [10]. However, in recent years, some studies built recommendation algorithms based on ontologies, linked data, and open source RDF data to leverage semantic content [6], [10]. As an example, [5] defined an ontology for learners, learning objects, and their environments in order to provide adaptive recommendations, based on similarities between object properties. [9] examined the Cold Start problem [17] in the case of new micro OERs, by defining rules based on recommended sequences of learning objects using existing ontologies. Furthermore, [4] built an OER recommendation system for learners in order to help them to achieve skill based learning objectives.

III. Data Collection

In this section, we explain the steps we followed to collect relevant and reliable data sources for our recommender prototype, focusing on data science related skills and jobs.

A. Labour Market Information

We applied the method proposed by [4] on 500 job vacancies, published in February and January 2020, which resulted in the extraction of the 16 most relevant and required skills for data science jobs. Additionally, we used the Wikipedia python API[1] and crawled Wikipedia content describing these 16 extracted skills.

B. OER Resources

Regarding the educational content, we collected more than 700 OERs from the following repositories: Youtube, MIT OpenCourseWare[2], Skills Commons[3], OER Commons[4], Wisc-Online[5], Khan Academy[6], and Wikipedia by using their APIs and/or search services. For each OER, we store the following fields (if available), in order to provide effective recommendations: source, title, description, target skill, level (Beginner, Intermediate, Advanced, Master), URL, length, transcription, view count, rating (or combination likes and dislikes), ranking position score (reverse of ranking position in the source repository search: $1/ranking_position_in_repository_search$), and transcription similarity with the target skill description (calculated by the cosine similarity between the transcription vector and the target skill description vector using pre-trained 300-dimensional Glove vectors [18]).

[1]https://pypi.org/project/wikipedia/
[2]https://ocw.mit.edu/
[3]https://www.skillscommons.org/
[4]https://www.oercommons.org/
[5]https://www.wisc-online.com/
[6]https://www.khanacademy.org/

IV. Methods

In this section, we detail the most important components of our OER recommender prototype. These components are responsible for OER quality control, recommendation generation, and the interaction with the learner.

A. OER Quality Predictor

The aim of this component is to predict the overall quality of the collected OERs from different repositories, and define an automatic quality control mechanism.

1) Quality Prediction Based on Properties: As a first step, we established a measure for OER-skill matching. For this measure we used procedures from [11][7] to define a random forest model that classifies OERs into two groups: *Fit for Achieving a Skill*, and *Not Fit for Achieving a Skill*. The model uses the following OER properties to execute this prediction: length, rate, transcription similarity with the target skill description, view count, and ranking position score.

Furthermore, we had to find a solution for missing values in our prediction model, since not all OERs contained all required features. For this problem, we built a number of alternative random forest models with various combinations of the above-mentioned features, and selected the one with the highest F1-score. It should be mentioned that we applied alternative models on the data-set defined by [11], and got a minimum accuracy of *77.2 %*. The feature-set was adjusted towards maximum accuracy, but in some cases we had to rely on models with fewer features (from all desired OER features) than optimal, which resulted in a slightly reduced accuracy.

2) Quality Prediction Based on Metadata: Previously we proposed a metadata scoring model based on 8,887 crawled OERs from *Skill Commons* [8][8]. This model divides OERs in two groups: *OERs with Manual Quality Control*, and *OERs Without Quality Control*. At the same time, we also showed that there is a tight relationship between OER metadata quality and OER quality control processes. This means, that the more quality controls an OER passes, the higher the probability of containing high-quality metadata is. However, due to the rapidly growing amount of OERs, manual quality control is getting more and more impossible [8]. As a consequence, we built a model that predicts OER quality based on their metadata with *F1-score* of *94.6%*. Therefore, in this paper we capitalized on this scoring and prediction model, not only to qualify metadata, but also to predict the quality of the OERs we collected form 7 distinct repositories..

3) Quality Control Based on the Defined Models: To apply an automatic quality control process on the collected OERs, we used the above-mentioned models in a probability mode, which shows the probability of the target class in the classifiers. Therefore, for each OER, based on their properties and metadata, we applied both of our prediction models and defined two float numbers between 0 and 1 to determine the

[7]With *F1-score* of *86.3%*; The data-sets and the implementation protocols are included in the paper.
[8]The data-sets and the implementation protocols are included in the paper.

Authorized licensed use limited to: Universiteit van Amsterdam. Downloaded on February 25,2021 at 15:09:47 UTC from IEEE Xplore. Restrictions apply.

OER quality scores. Finally, we removed OERs with quality scores less than 0.5 (altogether 184 OERs) from the list of potential OERs for recommendation.

B. Recommendation Generation

In order to build our OER recommender system, we created a 6-dimensional vector of X for each OER including:

- Normalised length
- Normalised rate
- Text similarity with target skill description
- Metadata based quality probability
- Property based quality probability
- Source (containing 7 binary values to identify the 7 OER repositories we used for this study)

Respectively, for each learner, we define a 6-dimensional vector P as a preference vector that contains a weight (between 0 and 1) for each parameter in X.

Moreover, to provide personalised recommendations for learners, we rely on the following features describing learners' context:

- Country
- City
- Job experience including skills, expertise levels in each skill (Beginner, Intermediate, Advanced, Master) and industry[9]
- Birth date
- Gender (optional)

The goal is to find the best weights (P vector) for each learner. Therefore, at learner registration, we initialise the P vector for each learner, based on preference vectors of similar learners in the learner's context (e.g. Country, City, Job experience). Subsequently, we use Gradient Descent to optimize the preference vector based on users' ratings by minimizing the following loss function:

$$LossFunction = \sum_{o=recommended_OERs} |(P * X_o) - Y_o| \quad (1)$$

where X_o is the 6-dimensional vector of an OER o and Y_o is the satisfaction rating of the learner for that particular OER o.

Finally, to recommend an OER to a learner *u* for a particular target skill *s*, the system retrieves all OERs with the level that learner *u* has in skill *s*, and uses cosine similarity to recommend the OER with the closest X vector to the user preference vector (P).

C. Learning Dashboard

In this section, we explain the features of the implemented learner dashboard prototype, focusing on data science related jobs, using the above-mentioned components[10]. As an illustration, Figure 1 shows the different components of our system and their interaction.

When a learner uses our recommender for the first time, (s)he has to go through a two step registration process, which is critical to initialise essential learner features for effective recommendations:

1) Profile Creation: Learners register in our system by adding their demographic features (e.g. country, city, and birth date). Subsequently, they can also add their previous job experience(s) to their profile. Based on their previous jobs, our system recommends related skills (based on our labour market analysis) for their profile. Learners also need to set their expertise level(s) for each skill.

2) Defining Goals: Once a learner profile is created, our system asks learners to set their desired job. Based on this target occupation, our system shows all skills, which are required for that job (in the desired geographical location of the learner: city, country). Learners can select their skill goals from a list, but also search and add additional skill targets to their profiles[11].

D. Educational Content Page

For each selected skill objective, our system recommends the most relevant OERs. Learners can click and navigate to the physical location of the OER, and proceed with the learning. After learning, users can rate their satisfaction with each OER (from 1: very dissatisfied to 5: very satisfied), request another OER on the same level of difficulty, or ask for OERs with a higher level of difficulty. Moreover, if dissatisfied with the recommendation, learners can either replace the recommended OERs with another recommendation, or mark the OER as irrelevant for their skill target and ask for a replacement OER[12].

1) Progress Page and Reports: At each stage, learners can monitor their progress towards their goals (target job and skills), and consult reports on finished OERs at each expertise level they passed. Moreover, they can retrieve reports on periodically (e.g. monthly) completed OERs, and also the status of their current recommendations (e.g. in progress, finished, changed)[13].

V. Validation

To validate our approach and the proposed prototype, we invited 23 experts (including 16 PhD students and 7 university instructors) in the area of *Data Science* (with minimum 1 year of related teaching experience and 3 years of related industrial experience) and asked them to work with our prototype for at least 30 minutes. It should be mentioned that participants had different expertise levels regarding the required skills for a data scientist job as they had various range of teaching and industrial experience. Phd students had related teaching experience between 1 year and 6 years and related industrial experience between 3 years to 10 years. These numbers for

[9]We defined industries based on https://en.wikipedia.org/wiki/Statistical_Classification_of_Economic_Activities_in_the_European_Community

[10]The interested reader can find a demo of the prototype from: https://github.com/rezatavakoli/FIE2020

[11]Screenshot of the goal definition page: https://github.com/rezatavakoli/FIE2020

[12]Screenshot of the educational content page: https://github.com/rezatavakoli/FIE2020

[13]Screenshot of the progress page: https://github.com/rezatavakoli/FIE2020

Authorized licensed use limited to: Universiteit van Amsterdam. Downloaded on February 25,2021 at 15:09:47 UTC from IEEE Xplore. Restrictions apply.

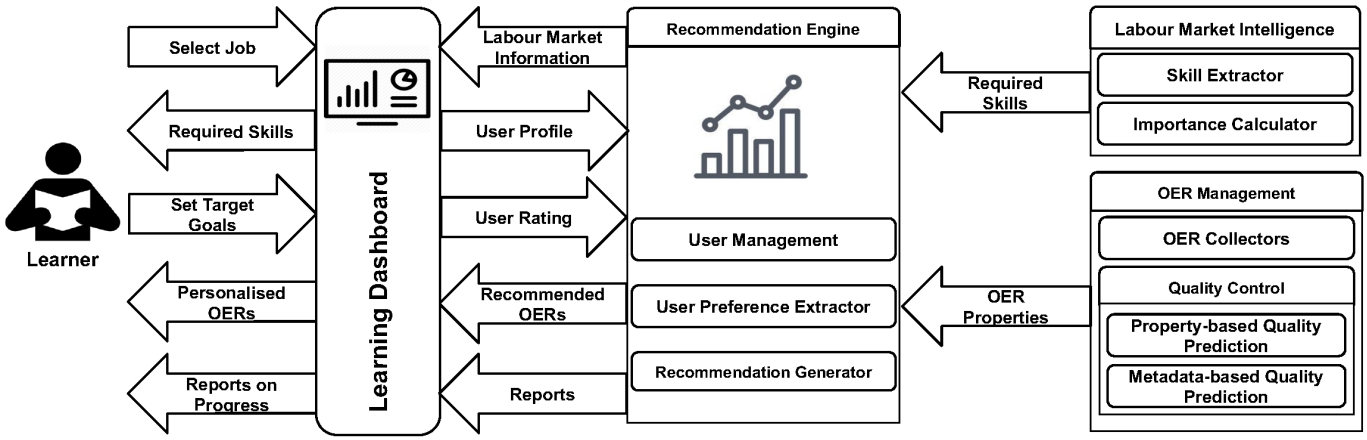

Fig. 1: Components of our Labour Market Driven Personalised OER Recommender System

TABLE I: Users Satisfaction Through Evaluation

Progress in Evaluation (%)	Very Dissatisfied (%)	Dissatisfied (%)	OK (%)	Satisfied (%)	Very Satisfied (%)	Report as Useful (%)
0-25	5	21	37	22	15	74
25-50	4	18	34	27	17	78
50-75	3	14	32	33	18	83
75-100	3	9	33	35	20	88
Total (0-100)	**≈4**	**≈15**	**≈34**	**≈29**	**≈18**	**≈81 (=80.9)**

instructors were between 7 to 16 years for teaching experience and 8 to 19 years for industrial experience. Accordingly, when experts registered in our system, 22% of skill expertise levels were set to Beginner, 26% were set to Intermediate, 31% were set to Advanced, and 21% were set to Master.

After creating their profile and setting their goals, participants started working with their learning dashboard, including their recommended OERs. Each participant generated at least 17, but maximum 20 recommendations. During the validation, altogether 415 OERs were recommended, and 336 of them (*80.9%*) were reported as useful (*OK*, *Satisfied*, or *Very Satisfied*). To illustrate how the quality of recommendations increased over time, we grouped all recommendations into four clusters. Each quarter contains a bit more than 100 recommendations, and Table I shows participants' satisfaction rate for each cluster (over time). The *Report as Useful* column reveals that the more the users work with our system, the more satisfactory recommendations they receive, since our system attempts to identify learner preferences and provide personalised recommendations. Moreover, based on our quality prediction and quality control process, more than 70% of the recommendations satisfied learners right from the beginning. As a consequence, we believe that this exercise revealed that our approach and the implemented prototype can potentially provide high-quality support to learners working towards labour market's demanded skills.

VI. Conclusion and Future Work

This study is one of the first steps to 1) empower learners to take control and responsibility for their own skill development, 2) improve skills on the basis of labour market information and personalised OER recommendation, and 3) progress the literature on OER quality control. As a long term vision, we expect that learners will show enhanced self-regulation, and spend less effort to find relevant, high-quality OERs. This approach also contributes to OER property identification accuracy, which is essential to increase OER (re)usability.

In this paper, we also demonstrated an OER recommender system prototype, which collects OERs from 7 different OER repositories, predicts their quality, and applies automatic quality control. Learners interact with the recommender through a dashboard, in which they can search for their desired job, display the list of required skills, set their level of expertise for each skill, set their learning context, and receive relevant OER recommendations accordingly. During their learning process, learners rate their satisfaction with recommendations, and update their learning preference vector. This strategy detects changes in learner profiles and fine tune the precision of recommendations.

For the next steps, we plan to add more OER repositories, extract more learner and OER properties, and use (quasi-)experimental designs for further developing and validating our prototype in a number of use cases. Moreover, feeding the users satisfaction back into our quality control cycle is among our important next steps.

References

[1] E. Colombo, F. Mercorio, and M. Mezzanzanica, "Applying machine learning tools on web vacancies for labour market and skill analysis," *Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence*, 2018.

[2] I. Wowczko, "Skills and vacancy analysis with data mining techniques," in *Informatics*, vol. 2, no. 4. Multidisciplinary Digital Publishing Institute, 2015, pp. 31–49.

Authorized licensed use limited to: Universiteit van Amsterdam. Downloaded on February 25,2021 at 15:09:47 UTC from IEEE Xplore. Restrictions apply.

[3] M. M. McGill, "Defining the expectation gap: a comparison of industry needs and existing game development curriculum," in *Proceedings of the 4th International Conference on Foundations of Digital Games*. ACM, 2009, pp. 129–136.
[4] M. Tavakoli, S. T. Mol, and G. Kismihók, "Labour market information driven, personalized, oer recommendation system for lifelong learners," *arXiv preprint arXiv:2005.07465*, 2020.
[5] S. Wan and Z. Niu, "An e-learning recommendation approach based on the self-organization of learning resource," *Knowledge-Based Systems*, vol. 160, pp. 71–87, 2018.
[6] G. Sun, T. Cui, G. Beydoun, S. Chen, F. Dong, D. Xu, and J. Shen, "Towards massive data and sparse data in adaptive micro open educational resource recommendation: a study on semantic knowledge base construction and cold start problem," *Sustainability*, vol. 9, no. 6, p. 898, 2017.
[7] J. Duffin, B. Muramatsu, and S. Henson Johnson, "Oer recommender: A recommendation system for open educational resources and the national science digital library," *White paper funded by the Andrew W. Mellon Foundation for the Folksemantic. org project*, 2007.
[8] M. Tavakoli, M. Elias, G. Kismihók, and S. Auer, "Quality prediction of open educational resources a metadata-based approach," *arXiv preprint arXiv:2005.10542*, 2020.
[9] G. Sun, T. Cui, D. Xu, J. Shen, and S. Chen, "A heuristic approach for new-item cold start problem in recommendation of micro open education resources," in *International conference on intelligent tutoring systems*. Springer, 2018, pp. 212–222.
[10] J. Chicaiza, N. Piedra, J. Lopez-Vargas, and E. Tovar-Caro, "Recommendation of open educational resources. an approach based on linked open data," in *Global Engineering Education Conference*. IEEE, 2017, pp. 1316–1321.
[11] M. Tavakoli, S. Hakimov, R. Ewerth, and G. Kismihók, "A recommender system for open educational videos based on skill requirements," *arXiv preprint arXiv:2005.10595*, 2020.
[12] P. Király and M. Büchler, "Measuring completeness as metadata quality metric in europeana," in *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 2018, pp. 2711–2720.
[13] S. Ushakova, "Usability of metadata standards for open educational resources," Oct 2015. [Online]. Available: https://hclemuseum.wordpress.com/2015/10/02/usability-of-metadata-standards-for-open-educational-resources/
[14] A. R. Pelaez and P. P. Alarcon, "Metadata quality assessment metrics into ocw repositories," in *Proceedings of the 2017 9th International Conference on Education Technology and Computers*. ACM, 2017, pp. 253–257.
[15] A. Romero-Pelaez, V. Segarra-Faggioni, and P. P. Alarcon, "Exploring the provenance and accuracy as metadata quality metrics in assessment resources of ocw repositories," in *Proceedings of the 10th International Conference on Education Technology and Computers*. ACM, 2018, pp. 292–296.
[16] A. Romero-Pelaez, V. Segarra-Faggioni, N. Piedra, and E. Tovar, "A proposal of quality assessment of oer based on emergent technology," in *2019 IEEE Global Engineering Education Conference (EDUCON)*. IEEE, 2019, pp. 1114–1119.
[17] X. N. Lam, T. Vu, T. D. Le, and A. D. Duong, "Addressing cold-start problem in recommendation systems," in *Proceedings of the 2nd international conference on Ubiquitous information management and communication*. ACM, 2008, pp. 208–211.
[18] J. Pennington, R. Socher, and C. D. Manning, "Glove: Global vectors for word representation," in *Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. [Online]. Available: http://www.aclweb.org/anthology/D14-1162

Authorized licensed use limited to: Universiteit van Amsterdam. Downloaded on February 25,2021 at 15:09:47 UTC from IEEE Xplore. Restrictions apply.