A statistical information system in support of job policies orientation

Adham Kahlawi, Francesca Giambona, Lucia Buzzigoli, Laura Grassini, Cristina Martelli

1. Introduction

One of the main issues in modern labour market governance is about connecting people with jobs (Martin, 2015; Varanasi, 2021); bridging the skills gap is on the agenda of many governs and institutions (Mohla, 2020; Ras, et al., 2017), and many efforts have been done, also at the European level, to address vocational training to the real needs of the different economic sectors. Under the European Blueprint Initiative, for instance, stakeholders work together in sector-specific partnerships, called alliances for sectoral cooperation for skills, which develop and implement strategies to address skills gaps in different sectors[1].

In this perspective, the availability of suitable data centred on skills needs in the labour market is strategic and addresses vocational training investment and lifelong learning politics. However, up to now, data sources are mainly organised around the concept of occupation which is too wide to orient politics and investments. In this work, we intend to use the recommendation systems approach to describe the skills more requested by the different occupations in order to improve the granularity of labour market description.

Born in the era of big data, recommendation systems are a family of information filtering procedures that help users make choices in an extremely rich and variable information context (for a brief, recent review, see Jariha and Jain, 2018). They can also be interpreted as methods of predicting whether a particular user will like a particular item based on its preference structure and characteristics. These methods are widely used in various fields: to suggest the purchase of products to customers in e-commerce; to recommend news articles or blog contents to online content readers; to recommend movies or music to users of streaming services, etc. The two classic entities considered in recommendation systems are users (those who choose) and items (what is chosen): the user-item matrix, also called preference or utility matrix, shows the users by row and the items by column, and each cell contains a number that represents the importance of that item for that user. This number can simply be 0/1 (the user has/hasn't chosen the item) or can be the rating expressed by the user for the item. The matrix is typically sparse because many or sometimes most of the entries are empty: recommender systems consist of filling in the empty cells with what similar users would choose. Additional information about users or items can be added to get better results. This article aims to use recommendation systems to predict the future skills that a person has to acquire to reach a particular profession or to develop himself to improve his chances of getting a job.

The data source is always huge, and the system must be able to produce timely responses by continuously updating the information set that is fed by users' feedback. Therefore, the problem is to combine traditional statistical methods used to develop professional skills with data mining and machine learning techniques, which are able to solve the computational complexity of the system and optimise its performance. Many different approaches can be used to solve this problem (Leskovec et al., 2019). We will refer to model based collaborative filtering methods (Chen et al., 2018), that have received great success in many fields of applications. In this case, no information on users or items is requested, and the user-item matrix is factorised by means of latent factor models to reduce its dimensionality. Different algorithms can be used to map each user and each

[1] https://ec.europa.eu/social/main.jsp?langId=en&catId=89&furtherNews=yes&newsId=10035

Adham Kahlawi, University of Florence, Italy, adham.kahlawi@unifi.it, 0000-0003-4040-5590
Francesca Giambona, University of Florence, Italy, francesca.giambona@unifi.it, 0000-0002-1760-2062
Lucia Buzzigoli, University of Florence, Italy, lucia.buzzigoli@unifi.it, 0000-0003-3297-1023
Laura Grassini, University of Florence, Italy, laura.grassini@unifi.it, 0000-0003-4678-6507
Cristina Martelli, University of Florence, Italy, cristina.martelli@unifi.it

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Adham Kahlawi, Francesca Giambona, Lucia Buzzigoli, Laura Grassini, Cristina Martelli, *A statistical information system in support of job policies orientation*, pp. 131-135, © 2021 Author(s), CC BY 4.0 International, DOI 10.36253/978-88-5518-461-8.25, in Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci (edited by), *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the on-site conference*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-461-8 (PDF), DOI 10.36253/978-88-5518-461-8

item to their corresponding factor vectors (Koren et al., 2009).

In our case, users and items are represented, respectively, by occupations and skills; the user-item matrix is built starting from a database produced by Burning Glass Technologies, which collects online job vacancy ads scanned from Italian online portals and company's websites in 2019 and 2020. The cell (i, j) of the matrix contains the number of ads that require skill j for occupation i. The use of recommender systems in this field of analysis is not new (Al-Otaibi and Ykhlef, 2012; Giabelli et al., 2021; Tavakoli et al., 2020; Valverde-Rebaza et al., 2018), but in this case the recommendation system is based on a dataset referred to Italy, in which occupations and skills follow the ESCO classification (European Skills, Competences, Qualifications and Occupations) (Kahlawi, 2020). In particular, the objective of the analysis is to help the vocational training systems and institutions to answer the question posed by every person looking for a new job or professional opportunities: which are the skills to have to enhance the professional profile? Finally, the matrix factorisation process is performed with the Alternating Least Squares (ALS) method and will be described in the next paragraph.

The results offered by the application of the proposed methodology will show which are the skills more requested in the framework of a specific occupation. Workers, job seekers, vocational training institutions, recruitment companies may take advantage of these results in different ways: Starting from the skills already owned by workers to suggest new skills for them, individuating the closest occupation that matches their skills based on the matrix and then comparing the actual profile with the most requested by the labour market. Alternatively, they may move from the concept of occupation to model updating skills politics.

2. Methodology

The methodology in this article is based on six basic actions, as shown in Figure 1.

- Action 1. The initial dataset contains different columns extracted from the job ads; for example, it has a column representing the occupation requested in the job ads after mapping it to the fourth level of the International Standard Classification of Occupations (ISCO-08). In addition, it has a column that represents the skills requested to be able to do this job. The user-item matrix is built using these two columns, and contains skills in the columns and occupations in the rows. Each matrix cell contains the number of times the skill has been requested for a particular occupation across all jobs ads.
- Action 2. We take the index of matrix cells that contain a value greater than zero, and then we randomly replace 20% of these values with zero. Afterwards, we replace each value greater than zero with the value of one.
- Action 3. For matrix factorisation, we use the ALS algorithm which is implemented in the Python implicit package[2], and built for large-scale collaborative filtering problems. ALS is doing a pretty good job at solving the scalability and sparseness of the compilation data, it is simple and scales well to enormous datasets. ALS has been used to solve different recommendation problems (Lakshmikanth et al., 2021).
- Action 4. First, we identify the occupations whose data has been hidden in preparing the test data. Second, we use the model that we built in the previous action to predict the values that have been hidden. Third, for each occupation, we calculate the Receiver Operating Characteristic Curve (ROC) to get the false positive rates and true positive rates which will be the input to calculate the Compute Area Under the Curve (AUC). Finally, we calculate the mean of AUC values of all occupations. The mean value represents the effectiveness of the model.

[2] https://implicit.readthedocs.io/en/latest/als.html

- Action 5. We use the model built in Action 3 to get the three best recommendations for a group of new job seekers.
- Action 6. We calculate the percentage of match between the current job seeker's skills and the skills required in the jobs ads (Match ratio). Then, we take the four job offers with the highest match ratio. Afterwards, we repeat these computations after adding the three skills recommended by the model to evaluate the improvement in job matching for the job seeker who has acquired these three skills.

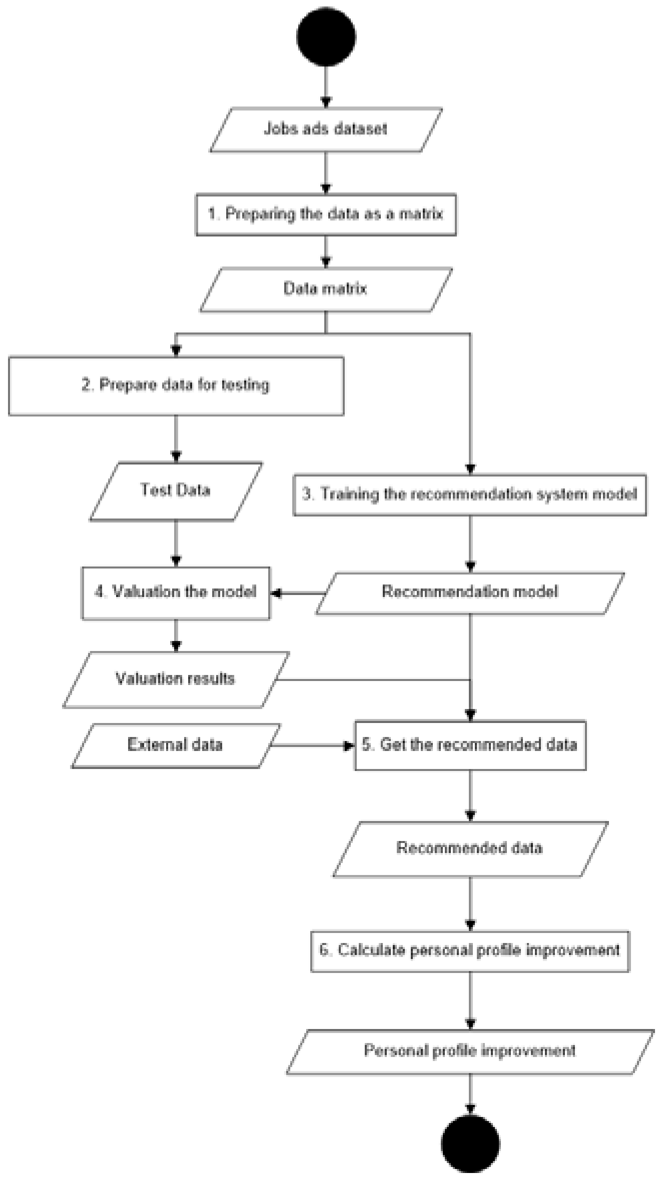

Figure 1 The methodology

3. Results and discussion

The effectiveness indicator of the recommendation system refers to the model's ability to recommended with an accuracy of up to 95.9 percent, as shown in Table One, which expresses the metadata of this model.

Table 1 Model metadata

Initial Database (job ads)	32,426,926
Matrix shape	(326,1213)
Matrix sparsity	95.5%
Occupations involved in the evaluation process	279
Model effectiveness	95.9%

Table 2 represents the profile of two job seekers and the best three recommendations of our model.

Table 2 Professional profile and the recommend skills for two hypothetical job seekers

Job seeker	Current skills	Recommended skills
First	Teamwork, database, communication, adapting to change, thinking proactively, assisting the client, English, providing information, identifying client wishes, managing time, finance methods, thinking creatively.	work independently, tolerate stress, show enthusiasm
Second	Database, marketing principles, PHP, adjust priorities, CSS, create the front-end design of a website, integrated development software environment, machinery functionality, event planning, communication, financing methods, Scala, prioritise homework, English, pandas	office software, administer ICT systems, SQL

Consequently, Table 3 represents the extent of development that the users will achieve after getting these three skills by showing the top 4 job ads they can apply for. Indeed, it appears clearly from the results how the recommendation system helped users improve their chances of getting jobs directly and based on market demands.

Table 3 Personal profile improvement of two hypothetical job seekers

Job seeker	Moment of progress	Job ads id	Match ratio (%)	Sector
First	Before recommendation	159492369	92	Professional, scientific and technical activities
		247268757	90	Wholesale and retail trade
		161885864	88	Construction
		180547711	88	Wholesale and retail trade
	After recommendation	159492369	92	Professional, scientific and technical activities
		166831331	92	Information and communication services
		180554644	91	Wholesale and retail trade
		180547711	90	Wholesale and retail trade
Second	Before recommendation	543695754	83	Administrative and support service activities
		724284081	80	Manufacturing activity
		175809178	75	Administrative and support service activities
		357363988	75	Accommodation and catering services
	After recommendation	543695754	83	Administrative and support service activities
		363981505	80	Administrative and support service activities
		615486253	80	Professional, scientific and technical activities
		605508081	80	Professional, scientific and technical activities

4. Conclusion

Choosing the skills that a person has to learn to get a job opportunity or develop his job position is an ongoing problem because the labour market is constantly changing, and the skills required to do the job are constantly changing. Thus, the solutions provided have to be able to be continuously updated based on market changes. Indeed, this article proposed a recommendation system based on a database collected from the labour market and capable of updating itself based on new data that can be obtained in the future from the labour market. Furthermore, the proposed recommendation system improved people's chances of getting new jobs through the skills that it recommended to these people. Finally, this work faced a set of limitations, the most important of which was the size of the matrix built from the initial data, which is why we used the same data for training and testing the model; for this reason, the proposed recommendation system is not considered a complete system and can not find all solutions for all users. Therefore, we will strive in future work to develop this system to become suitable for the largest possible segment of users.

References

Al-Otaibi S.T. and Ykhlef M. (2012), A survey of job recommender systems, International Journal of the Physical Sciences, **7**(29), pp. 5127-5142.

Chen R., Hua Q., Chang Y-S., Wang B., Zhang L., Kong X. (2018), A Survey of Collaborative Filtering-Based Recommender Systems: From Traditional Methods to Hybrid Methods Based on Social Networks, IEEE, **6**, pp.64301-64320.

Giabelli A., Malandri L., Mercorio F., Mezzanzanica M. and Seveso A., (2021), Skills2Job: A recommender system that encodes job offer embeddings on graph databases, Applied Soft Computing Journal **101,** pp.107049.

Jariha P. and Jain S.K. (2018), A state-of-the-art Recommender Systems: an overview on Concepts, Methodology and Challenges, Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 1769-1774.

Kahlawi A., (2020), An Ontology Driven ESCO LOD Quality Enhancement, International Journal of Advanced Computer Science and Applications (IJACSA), **11**(3).

Koren J., Bell R. and Volinsky C. (2009), Matrix Factorization Techniques For Recommender Systems, IEEE COMPUTER, **42**(8), pp. 30-37.

Lakshmikanth P, P. Radha K, J. V. R. M, (2021), Approaching the cold-start problem using community detection based alternating least square factorisation in recommendation systems, Evol. Intel. **14**, pp. 835–849.

Leskovec J., Rajaraman A., Ullman J. (2019), Mining of Massive Datasets, 3rd Edition, Cambridge University Press.

Martin P M., (2015), Activation and active labour market policies in OECD countries: stylised facts and evidence on their effectiveness, IZA Journal of Labor Policy **4**(4).

Mohla D., (2020), IEEE IAS Standards-A Pathway to Narrow the Technical Skills Gap [Standards News],in IEEE Industry Applications Magazine, **26**(5), pp. 94-95.

Ras E., Wild F., Stahl C., Baudet A., (2017), Bridging the skills gap of workers in Industry 4.0 by human performance augmentation tools: Challenges and roadmap, Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments, pp. 428-432.

Tavakoli M., Mol S.T. and Kismih G. (2020), Labour Market Information Driven, Personalized, OER Recommendation System for Lifelong Learners, Proceedings of the 12th International Conference on Computer Supported Education (CSEDU), **2**, pp. 96-104.

Valverde-Rebaza J., Puma R., Bustios P. and Silva N.C. (2018), Job Recommendation based on Job Seeker Skills: An Empirical Study, In: A. Jorge, R. Campos, A. Jatowt, S. Nunes (eds.): Proceedings of the Text2StoryIR’18 Workshop, Grenoble, France.

Varanasi B., (2021), Knowledge Obsolescence and the Future of Work: Relevance of Knowledge and Impact to Jobs, Computational Thinking for Problem Solving and Managerial Mindset Training. IGI Global, pp.64-181.