Recommender system in a non-stationary context: recommending job ads in pandemic times

Guillaume Bied, Solal Nathan, Elia Pérennes, Victor Alfonso Naya, Philippe Caillou, Bruno Crépon, Christophe Gaillac, Michèle Sebag

Recommender system in a non-stationary context: recommending job ads in pandemic times

Guillaume Bied[1,2], Solal Nathan[1], Elia Perennes[2], Victor Alfonso Naya[1], Philippe Caillou[1], Bruno Crépon[2], Christophe Gaillac[2,3], and Michele Sebag[1]

[1] Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), Orsay, France
[2] Centre de Recherche en Economie et Statistique (CREST), Palaiseau, France
[3] Oxford University, Oxford, United Kingdom

1 Introduction

This paper focuses on the recommendation of job ads to job seekers [1, 2], exploiting proprietary data from the French Public Employment Service (PES) and focusing more specifically on low or unskilled workers. Besides the usual challenges of data sparsity, the signal to noise ratio is high (few job seekers have diplomas), and the scalability requirements are high, hindering the use of joint job seekers-job ads features (e.g. geographical distance).

As a first contribution, a two-tiered approach, named VADORE, is designed to handle these requirements: a first system called VADORE.0 leverages heterogeneous information; a second system called VADORE.1 builds upon the assessment of job seeker/job ad pairs by VADORE.0 to narrow down the promising ads for a given job seeker, and support a more fine-grained recommendation. The comparison with XGBoost [17] shows significant computational gains with no performance loss.

A specific question examined in this paper concerns the non-stationarity of the job market, related with the Covid-19 pandemic. Two interpretations for the variation of the recall performance are discussed; a normalized recall indicator is proposed as second contribution of the paper.

Related work Job recommendation, surveyed in [15], has emerged as a major domain of "AI for good" [19, 17].

The winning algorithm of the RecSys 2017 challenge [17] relies on XGBoost [4], involving simple yet effective feature engineering. Quite a few other approaches have been developed by CareerBuilder [21], and LinkedIn [11, 7, 10, 14, 3, 20, 13]. A main difference lies in the fact that these approaches concern skilled job seekers, with significantly lesser sparsity of the interaction matrix. In CareerBuilder [22], embeddings related to different facets of recommendation are linearly aggregated. An originality of VADORE lies in the embedding related to the geographical distance of the job seeker and job ad. In LinkedIn, a filter [3] is used to narrow down the search, and generalized linear mixed models [20] are used to combine user and item features [13]. The main difference with VADORE.1 lies in the non-linear aggregation process.

2 Overview of VADORE

As said, VADORE is a two-tiered system: VADORE.0 narrows down the set of possible recommendations, enabling the use of the more expressive and complex VADORE.1 with good computational performance.

VADORE.0 adopts a "tower-shaped" architecture. As shown in Fig. 1, VADORE.0 learns separate embeddings (respectively noted ϕ_k and ψ_k) as neural nets, to model the different facets of job-seekers and job ads. The overall score associated to a pair (job seeker x_i, job ad y_j) is the aggregation of each facet through the learned A_k matrix, summed over all facets:

$$Vadore.0(x_i, y_j) = \sum_{k \in \{\text{geography, skills, other}\}} \phi_k(x_i)^t A_k \psi_k(y_j) \quad (1)$$

The facets modelled by these embeddings include: i) geographical aspects; ii) skills and occupation; iii) all "other" factors (ranging from age, required salary and type of contract for job seekers, to offered salary and company description for job ads). Let us detail the "Other" embedding (due to space limitations the geographical and skill modules are respectively detailed in Appendix B and C). The relevance of each embedding to VADORE.0 is assessed through ablation studies (Appendix E).

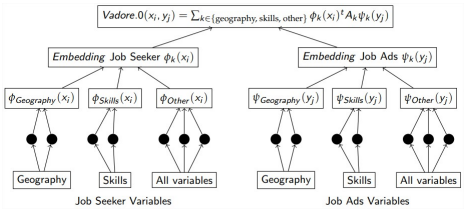


Fig. 1. VADORE.0 architecture overview

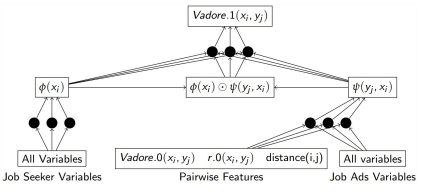


Fig. 2. VADORE.1 architecture overview

The "Other" module exploits all information related to the job seekers and job ads collected by the PES. The input representation size for job seekers (resp. job ads) is 498 (resp. 491), including their textual description and the list of skills associated with workers/ job ads after the French PES' ontology. In both cases, the description is reduced using singular value decomposition (SVD). Additional features include job seekers' search criteria, labor market profiles, administrative data and socio-demographic backgrounds, and job ad's requirements, characteristics and firm description.

The 'Other" module is trained using a triplet margin loss:

$$\mathcal{L}(M_{other}) = \sum_{x_i, y_j, y_k} \max\{M_{other}(x_i, y_k) - M_{other}(x_i, y_j) + \eta, 0\} \quad (2)$$

with $M_{other}(x_i, y_j) = \phi_{other}(x_i)^t A_k \psi_{other}(y_j)$ the output score of the module. The loss is meant to ensure that M_{other} scores pairs (x_i, y_j) (corresponding to an effective match) higher than (x_i, y_k), with y_k another job ad.

VADORE.0 is parameterized from the weights of embeddings ϕ_k and ψ_k and aggregation matrices A_k. Each embedding is first pre-trained independently using a triplet margin loss (Eq. 2). Thereafter, all embeddings are jointly trained end-to-end with a triplet margin loss. In all cases, the learning process relies on stochastic gradient descent and negative sampling. Negatives samples y_k are job ads sampled uniformly in the same week as the positive job ad y_j.

VADORE.1: filter-based recommendation model. VADORE.0 is used to filter out all job ads but the most promising 1,000 job ads for each job seeker. This filter enables VADORE.1 to consider more expressive and elaborate features while keeping computational costs in check. Notably, the negative sampling strategy used in VADORE.1 selects more relevant job ads.

Formally, VADORE.1 learns two embeddings ϕ and ψ describing the job seekers and the job ads in relation with the job seeker. Embedding $\phi(x_i)$ takes as input the job seeker x_i description provided to the module "Other". Embedding $\psi(y_j, x_i)$ takes as input the job ad y_j description likewise provided to the module "Other", augmented with the value $Vadore.0(x_i, y_j)$, the rank thereof noted $r.0(x_i, y_j)$, and the distance in kilometers between the locations of x_i and y_j.

VADORE.1 finally learns embeddings $\phi(x_i), \psi(y_j, x_i)$, and learns a neural net on the top of $\phi(x_i), \psi(y_j, x_i)$ and their elementwise multiplication $\phi(x_i) \odot \psi(y_j, x_i)$. VADORE.1 is learned end-to-end using stochastic gradient descent and a logistic loss,

$$\min \mathcal{L}(Vadore.1) = \sum_i \log(Vadore.1(x_i, y_j)) + \log(1 - Vadore.1(x_i, y_k)) \quad (3)$$

where y_j stands as usual for the match of x_i and y_k for another job ad uniformly selected after the filter selection operated by VADORE.0.

3 Empirical validation

This section reports on the comparative assessment of the VADORE recommendation models. After describing the experimental setting, the performances are reported and discussed. A decomposition of the performance of the modules composing VADORE.0 is also presented in Appendix E. Some preliminary analysis of biases in the data and in the performance of VADORE are presented in Appendix G.

Experimental settings We consider proprietary data of the French PES, recording all official job seekers and most available job ads. We restrict ourselves to the *Auvergne-Rhône-Alpes* region, starting from Jan. 2019 to Feb. 2022 (164 weeks), for a total number of 1.07 millions job seekers, 1.78 million job ads and 228k hires. More details about the dataset are presented in Appendix A.

The presented approach is comparatively assessed, using XGBoost, winner of the RecSys2017 challenge [17], as strong baseline. The baseline is trained from the same training weeks as VADORE, using the job ad and job seeker descriptions provided to the "Other" module (section 2). Following [17], XGBoost is provided with additional pairwise features. In our case, we compute the partial adequacy of the job ad and job seekers *re* the distance, skills, occupation, education, experience, contract type, spoken languages, driving licenses and wages.

VADORE.0 and VADORE.1, as described in section 2 are trained using the neural architecture hyperparameters described in Appendix D.

Three performance indicators are considered: the training time, the single recommendation computational time, and the mainstream recall indicator Recall@k (indicating the percentage of job seekers for whom the match is ranked among the top-k recommendations) for k ranging in $\{10, 50, 100, 1{,}000\}$.

Comparative results The performance indicators of the considered approaches are displayed in Table 1.

Table 1. Comparative Results of VADORE and XGBoost: Recalls@$\{10, 20, 50, 100, 1{,}000\}$. Overall training time and recommendation time (*per* job seeker, in seconds).

Recall@k	XGBoost	VADORE.0	VADORE.1
10	26.83	22.88	25.96
20	35.59	31.55	35.65
50	48.75	44.12	49.07
100	58.88	53.80	58.67
1000	86.47	82.13	-
Computational training time	10 hours	7 hours	+ 40 '
Computational single recommendation time	7"	0.002 "	+ 0.003"

The recall@1000 of VADORE.0 is greater than 80%, upper bounding per construction the recall@1000 of VADORE.1 as the filter removes the match for circa 20% of the job seekers (see appendix G). As expected, VADORE.1 significantly outperforms its first tier VADORE.0.

Overall, for a significant fraction of the job seekers, their match is ranked in the top 100.

The overall lesson learned from Table 1 is that VADORE.1 behaves on par with XGBoost, while improving the recommendation time by two orders of magnitudes. As known [17], a main source of efficiency of XGBoost is in the use of engineered pairwise features, finely confronting the job ad and the job seeker w.r.t. the various facets (e.g., distance, contract, wage, driving license).

4 Analysis of a non-stationary recommendation problem

This section focuses on the impacts of the Covid-19 pandemic on the job market (Fig. 3), and on the performance of VADORE. From Jan. 2019 to March

2022, three lock-downs have been associated with occasional shutdowns of "non-essential" firms' activity. The first period (Before-LDs: early 2019 to March 2020), is used as reference of the French labor market. The second period (During-LDs: May 2020 to Apr. 21) sees brutal adjustments and strong restrictions on the economy. The third period, (After-LDs: May 2021 up to Mars 2022) is characterized by lighter regulations. These unfortunate events, viewed as "natural experiment" [6, 9] enable to disentangle the respective impacts of the change in the *behavior* of the individuals, and in the *amount* of job ads, on the recommender system performance.

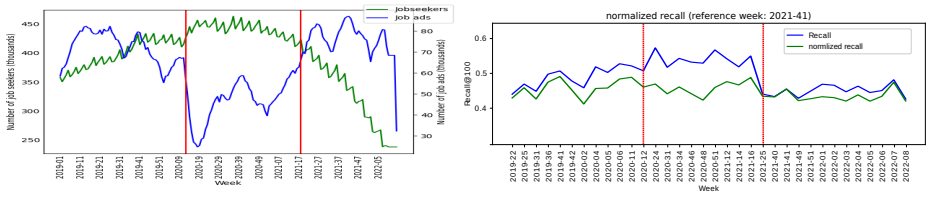


Fig. 3. Non-stationarity of the French job market from Jan. 2019 to March 2022 (partial data availability for the last two months). Left: distributions of job seekers (left scale) and job ads (right scale). Right: Evolution of the normalized recall (in green) vs the observed recall (in blue) of VADORE.1 during the Before-LDs, During-LDs and After-LDs periods.

Non-stationarity of the job market distribution. The non-stationarity of the distributions of respectively jobs seekers and job ads is evidenced through tackling a 3-class supervised learning problem, and predicting the class a given job ad/job seeker belongs to (i.e., is observed in the 1st, 2nd or 3rd period). The results suggest that the change of distribution is insignificant regarding the job seekers (the discrimination accuracy being close to chance) and very significant regarding the job ads (the discrimination accuracy is twice as high as chance). As could have been expected, the changes are related to the job sectors, with a surge of ads in the health sector, and a significant decrease in the Hotels-Restaurant-Leisure sector. The reader is referred to Appendix F for more details.

The recall of the VADORE.1 recommender system in all three periods is displayed in blue Fig 3 (right). Overall, the recall is significantly better in the During-LDs period, and recovers a performance on par or better than its reference level (in the Before-LDs period) in the After-LDs period.

A normalized recall indicator. A first conjecture is that the change of performance can be explained by a modification in the behavior of the individuals (job seekers applying to different job ads; recruiters taking advantage of the increased tension to uplift their expectation). A second conjecture is that the change is primarily and only due to the fact that part of the job ads have disappeared during the During-LDs period. Two further remarks are done. Firstly, the "missing"

job ads are not visible in the recall, that by design only considers the matches. Secondly, the task of the recommender system − recommending relevant job ads to people having found a match − is facilitated as the number of job ads decreases by circa 1/3 during the During-LDs period.

The alternative is investigated by proposing a normalized recall indicator, invariant w.r.t. the overall number of items (job ads) on the market. If the second conjecture holds, this normalized recall indicator is expected to be stable along the three periods.

Definition. Let us define a reference period with K^* offers. The normalized recall at k, noted $\overline{Recall@k}$, associated with a period with overall number of items K, is defined as the observed recall at k' such that $K/K^* = k'/k$:

$$\overline{Recall@k} = Recall@k', \text{ with } k' = \frac{K}{K^*}k \tag{4}$$

The evolution of the normalized recall, considering an average week of the Before-LDs period as reference, is displayed in green on Fig. 3 (right). The detail of its evolution per week suggests that the series of the normalized recall does not present changes after mid 2020.

5 Conclusion and Perspectives

The main contribution of the presented VADORE approach is to address the real-world complex problem of job recommendation for unqualified job seekers. The obtained results (circa 60% for recall@100) might contribute to easing the frictional information problem of retrieving interesting offers in the dozens of thousands available job ads. The proposed two-tier architecture supports a computational gain of circa two orders of magnitude compared to XGBoost, with similar performance.

A scientific question investigated in the paper concerns the impact of the Covid-19 pandemic on the job market. Based on a normalized recall indicator, it was possible to show that the changes of performance are mostly due to the change in the number of job ads (as opposed to, a change in the behavior of the job seekers or recruiters in terms of labor market choices).

A key perspective for further research is to investigate the biases of the data and of the recommender system itself, characterizing whether some categories of persons are less well taken into account than others.

References

1. Abel, F., Benczúr, A., Kohlsdorf, D., Larson, M., Pálovics, R.: Recsys challenge 2016: Job recommendations. In: Proceedings of the 10th ACM Conference on Recommender Systems. p. 425–426. RecSys '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2959100.2959207, https://doi.org/10.1145/2959100.2959207

2. Abel, F., Deldjoo, Y., Elahi, M., Kohlsdorf, D.: Recsys challenge 2017: Offline and online evaluation. In: Cremonesi, P., Ricci, F., Berkovsky, S., Tuzhilin, A. (eds.) Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017. pp. 372–373. ACM (2017). https://doi.org/10.1145/3109859.3109954, https://doi.org/10.1145/3109859.3109954
3. Borisyuk, F., Kenthapadi, K., Stein, D., Zhao, B.: Casmos: A framework for learning candidate selection models over structured queries and documents. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. p. 441–450. KDD '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2939672.2939718, https://doi.org/10.1145/2939672.2939718
4. Chen, T., Guestrin, C.: XGBoost. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM (aug 2016). https://doi.org/10.1145/2939672.2939785, https://doi.org/10.1145%2F2939672.2939785
5. Chzhen, E., Giraud, C., Stoltz, G.: A unified approach to fair online learning via blackwell approachability. In: Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W. (eds.) Advances in Neural Information Processing Systems. vol. 34, pp. 18280–18292. Curran Associates, Inc. (2021), https://proceedings.neurips.cc/paper/2021/file/97ea3cfb64eeaa1edba65501d0bb3c86-Paper.pdf
6. Dunning, T.: Natural Experiments in the Social Sciences: A Design-Based Approach. Cambridge University Press (2012)
7. Geyik, S.C., Guo, Q., Hu, B., Ozcaglar, C., Thakkar, K., Wu, X., Kenthapadi, K.: Talent search and recommendation systems at linkedin: Practical challenges and lessons learned. In: Collins-Thompson, K., Mei, Q., Davison, B.D., Liu, Y., Yilmaz, E. (eds.) The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval, SIGIR 2018, Ann Arbor, MI, USA, July 08-12, 2018. pp. 1353–1354. ACM (2018). https://doi.org/10.1145/3209978.3210205, https://doi.org/10.1145/3209978.3210205
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K. (eds.) Advances in Neural Information Processing Systems. vol. 27. Curran Associates, Inc. (2014), https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
9. Imbens, G.W., Rubin, D.B.: Causal inference in statistics, social, and biomedical sciences. Cambridge University Press (2015)
10. Kenthapadi, K., Le, B., Venkataraman, G.: Personalized job recommendation system at linkedin: Practical challenges and lessons learned. In: Cremonesi, P., Ricci, F., Berkovsky, S., Tuzhilin, A. (eds.) Proceedings of the Eleventh ACM Conference on Recommender Systems, RecSys 2017, Como, Italy, August 27-31, 2017. pp. 346–347. ACM (2017). https://doi.org/10.1145/3109859.3109921, https://doi.org/10.1145/3109859.3109921
11. Li, J., Arya, D., Ha-Thuc, V., Sinha, S.: How to get them a dream job?: Entity-aware features for personalized job search ranking. In: Krishnapuram, B., Shah, M., Smola, A.J., Aggarwal, C.C., Shen, D., Rastogi, R. (eds.) Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-

17, 2016. pp. 501–510. ACM (2016). https://doi.org/10.1145/2939672.2939721, https://doi.org/10.1145/2939672.2939721

12. Lian, D., Zhao, C., Xie, X., Sun, G., Chen, E., Rui, Y.: Geomf: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: Macskassy, S.A., Perlich, C., Leskovec, J., Wang, W., Ghani, R. (eds.) The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD. pp. 831–840. ACM (2014)
13. Ozcaglar, C., Geyik, S.C., Schmitz, B., Sharma, P., Shelkovnykov, A., Ma, Y., Buchanan, E.: Entity personalized talent search models with tree interaction features. In: Liu, L., White, R.W., Mantrach, A., Silvestri, F., McAuley, J.J., Baeza-Yates, R., Zia, L. (eds.) The World Wide Web Conference, WWW 2019, San Francisco, CA, USA, May 13-17, 2019. pp. 3116–3122. ACM (2019). https://doi.org/10.1145/3308558.3313672, https://doi.org/10.1145/3308558.3313672
14. Ramanath, R., Inan, H., Polatkan, G., Hu, B., Guo, Q., Ozcaglar, C., Wu, X., Kenthapadi, K., Geyik, S.C.: Towards deep and representation learning for talent search at linkedin. In: Cuzzocrea, A., Allan, J., Paton, N.W., Srivastava, D., Agrawal, R., Broder, A.Z., Zaki, M.J., Candan, K.S., Labrinidis, A., Schuster, A., Wang, H. (eds.) Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018. pp. 2253–2261. ACM (2018). https://doi.org/10.1145/3269206.3272030, https://doi.org/10.1145/3269206.3272030
15. de Ruijt, C., Bhulai, S.: Job recommender systems: A review (2021). https://doi.org/10.48550/ARXIV.2111.13576, https://arxiv.org/abs/2111.13576
16. Schmitt, T., Gonard, F., Caillou, P., Sebag, M.: Language modelling for collaborative filtering: Application to job applicant matching. In: 29th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2017, Boston, MA, USA, November 6-8, 2017. pp. 1226–1233. IEEE Computer Society (2017). https://doi.org/10.1109/ICTAI.2017.00186
17. Volkovs, M., Yu, G.W., Poutanen, T.: Content-based neighbor models for cold start in recommender systems. In: Proceedings of the Recommender Systems Challenge 2017 - RecSys Challenge 17. ACM Press (2017). https://doi.org/10.1145/3124791.3124792
18. Weinberger, K.Q., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**(2) (2009)
19. Xiao, W., Xu, X., Liang, K., Mao, J., Wang, J.: Job recommendation with hawkes process: an effective solution for recsys challenge 2016. In: Abel, F., Benczúr, A.A., Kohlsdorf, D., Larson, M.A., Pálovics, R. (eds.) Proceedings of the 2016 Recommender Systems Challenge, RecSys Challenge 2016, Boston, Massachusetts, USA, September 15, 2016. pp. 11:1–11:4. ACM (2016). https://doi.org/10.1145/2987538.2987543, https://doi.org/10.1145/2987538.2987543
20. Zhang, X., Zhou, Y., Ma, Y., Chen, B., Zhang, L., Agarwal, D.: Glmix: Generalized linear mixed models for large-scale response prediction. In: Krishnapuram, B., Shah, M., Smola, A.J., Aggarwal, C.C., Shen, D., Rastogi, R. (eds.) Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. pp. 363–372. ACM (2016). https://doi.org/10.1145/2939672.2939684, https://doi.org/10.1145/2939672.2939684

21. Zhao, J., Wang, J., Sigdel, M., Zhang, B., Hoang, P., Liu, M., Korayem, M.: Embedding-based recommender system for job to candidate matching on scale (2021). https://doi.org/10.48550/ARXIV.2107.00221, https://arxiv.org/abs/2107.00221
22. Zhao, J., Wang, J., Sigdel, M., Zhang, B., Hoang, P., Liu, M., Korayem, M.: Embedding-based recommender system for job to candidate matching on scale (2021). https://doi.org/10.48550/ARXIV.2107.00221, https://arxiv.org/abs/2107.00221

Supplementary material

Appendix A: The proprietary data

We consider proprietary data of the French Public Employment Service, recording all official job seekers and most available job ads. For convenience, only job seekers from the southeast-central region of France, referred to as *Auvergne-Rhône-Alpes*, are considered in the following, together with the job ads located in the region or in the neighboring departments. The considered time frame includes 164 weeks, starting from Jan. 2019 to February 2022.

The overall number of job seekers is 1.07 million, and the overall number of job ads is 1.78 million. The overall number of hires (matches) is 228k. Each hire is either directly documented through the PES (following a *Mise en relation*) or inferred from mandatory employment declarations by employers (DPAE). In the latter case, the job ad matched is deduced from the fact that the company posted a single job ad in the considered week.

The training set includes a random selection of 85% of the weeks from Jan. 2019 to Dec. 2021; the test set includes all remaining weeks from 2019 to 2021, plus the first 8 weeks of 2022. On average, circa 400k job seekers, 64k job ads and 1.4k matches are observed per week. The dimension of the job seekers and job ads description is reported in Table 2.

Appendix B: Geographical module

The geographical module uses a tiled representation of the locations, taking inspiration from kernel density estimation and matrix factorization [12].

The embeddings process the tile-based representation of the zip codes of x_i and y_j. A_{geo} (Eq. 1) is set to the identity matrix. These embeddings are pre-trained by minimizing a triplet margin loss [18]:

$$\mathcal{L}(M_{geo}) = \sum_{x_i, y_j, y_k} [M_{geo}(x_i, y_k) - M_{geo}(x_i, y_j) + \eta]_+ \tag{5}$$

where $[A]_+ = max(A, 0)$, and $\eta > 0$ a margin hyper-parameter (set to 1 in the experiments). The loss is meant to ensure that M_{geo} scores pair (x_i, y_j) (corresponding to an effective match) higher than (x_i, y_k), with y_k another job ad. M_{geo} is trained from triplets (x_i, y_j, y_k) where y_k is uniformly selected in the job ads contemporaneous of y_j and located farther away from x_i. This pre-training enables M_{geo} to capture the varying time or money cost of transport in different locations (tiles), notably depending on the public transport networks and/or traffic jams.

The geographical module eventually returns the scalar product of $\phi_{geo}(x_i)$ and $\psi_{geo}(y_j)$.

Appendix C: Skills and occupation module

A nomenclature of labor market skills is considered by the French Public Employment Service, including circa 14k terms (e.g., "welding techniques", "tax system knowledge"). An ontology of job sectors, called ROME, is also available, consisting of 11 general sectors ("agriculture", "healthcare"), composed of 110 coarse-grained sectors ("woodcutting and pruning", "medical practitioner") and 532 fine-grained types of job. Each ROME code is associated with a list of skills after expert knowledge. The description of each job ad includes the skills of its 3-level ROME code, and possibly extra required skills. Likewise, each job seeker describes their skills; and the skills associated with the desired occupation are added to their description. The ϕ_{skills} and ψ_{skills} embeddings respectively process the binary representation of the list of skills associated with x_i and y_j. This module is pre-trained using a margin loss (Eq. 5) like the geographical module, aimed to score a matched pair (x_i, y_j) higher than (x_i, y_k) with y_k uniformly selected among the contemporaneous job ads. It jointly learns embeddings ϕ_{skills} and ψ_{skills}, together with their aggregation matrix A_{skills}, expectedly capturing the occupational similarities among skills.[4]

Eventually, the skills module returns the scalar product of $\phi_{skills}(x_i)$ and $\psi_{skills}(y_j)$, mediated through matrix A_{skills}:

$$M_{skills}(x_i, y_j) = \phi_{skills}(x_i)^t \, A_{skills} \, \psi_{skills}(y_j)$$

Appendix D: Experimental settings

The main neural networks hyperparameters are described in Table 2.

In all cases, the models are trained and tested in the secured PES platform, using Xeon(R) Gold 5215 CPU @ 2.50GHz with 64 GB of RAM and a Tesla V-100S GPU.

Appendix E: Ablation studies

Ablation studies are conducted to investigate the impact of the different modules involved in VADORE.0, namely the geographical, skills and "Other" modules. Table 3 reports the recall@100 performance obtained by a single module (left part), and the cost of ablation, estimated from the performance of all modules but one. A first remark is that the overall performance of VADORE.0 (53.8) is close to the sum of the performances of its components ($15.43 + 34.79 + 4.80 = 55.02$), confirming the merits and the complementary nature of all three modules. The

[4] Note that a matrix of similarity among ROME codes is also available, defined from expert knowledge. The difficulty lies in its heterogeneous level of detail, as types of jobs more recently appeared on the job market, e.g. related to communication, are described in a less detailed way.

Table 2. VADORE hyper-parameters: Neural architecture of the VADORE.0 modules and of VADORE.1. In all cases, the learning rate is adapted using Adam.

	VADORE.0			VADORE.1
Parameter	Other	Geographical	Skills	
Input dimension	498 (Job seekers) 491 (Job ads)	573 (Job seekers) 571 (Job ads)	12.3k	483 (Job seekers) 472 (Job ads)
Hidden Layer size	500→100	573 → 571	200→100	ϕ, ψ: 200
Hidden Layer size				MLP: 200
Batch size	256	32	32	128
Learning rate	0.0001	0.0001	0.0001	0.00001

importance of the geographical module is witnessed by both its performance as standalone (around 15% of the matches are explained by the single geographical information) and by the fact that the performance is significantly decreased when *not* considering the geographical aspects. The importance of the skills module is significant though moderate, with a low associated standalone performance (4%), and removing this module decreases the overall performance by about the same amount. The most surprising result is related to the "Other" module: while its standalone performance is the best single one (34.79), removing it only entails a not too high decrease of the performance (and even outperforms the M_{other} alone!). A tentative interpretation is related to the redundancy between the information processed by M_{other} and by M_{skills}, both having access to the occupational profile of the job seekers.

Table 3. VADORE.0: Impact of the three geographical, skills and other modules. Left part: performance of modules as standalone. Right part: performance of the whole VADORE.0 when removing a single module.

	Single module			All modules but one		
	geo	other	skills	geo	other	skills
Recall@100	15.43	34.79	4.80	30.02	40.66	49.43

Appendix F: Impact of the pandemic on the job market

The changes in the job seeker and job ad distributions are measured along the celebrated adversarial principle [8], considering that two distributions differ iff datasets sampled from these distributions can be discriminated.

Change in the job seeker distribution

A 3 class classification problem is formulated, that of classifying a job seeker in the period they apply to the PES (noting that a job seeker can apply in

several periods). The question is whether some job seekers were mostly met during one of the three periods.[5] The classification problem, considering an equi-distributed sample of the job seekers in all three periods, yields a classification rate of 39%, that is close to chance (33%), as shown on Fig. 4. This suggests that the distribution of the job seekers is not significantly modified in nature due to the Covid-19: people arriving on, or leaving the job market are not significantly different.

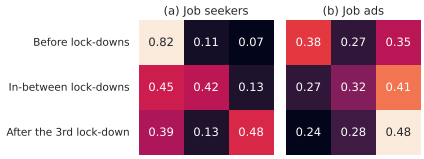


Fig. 4. Discriminating the 3 periods for the job seekers (left) and the job ads (right) distributions: Confusion matrices of a logistic regression, normalized by row. The discrimination accuracy is close to chance (39%) for the job seekers, and significantly higher (57%) for the job ads.

Change in the job ads distribution

The Covid-19 pandemic caused large shifts in terms of number of job ads (Fig. 3), with a weekly number of open job ads circa 70k in 2019, below 30k on the onset of the pandemic, reaching 50k in the middle of 2020, and returning to pre-pandemic levels after the end of the third lock-down.

The change of distribution between the job ads during the three periods is assessed as above, by tackling an equidistributed classification problem, involving 1.36 million job ads uniformly sampled in every period. In contrast with the above, however, the classifier yields a performance accuracy of 57%, thus significantly above the chance (Fig. 4, bottom).

This shift is essentially related to the job sectors: occupations linked to healthcare, as well as personal and community services, represent a larger part of the job ad population during the pandemic, while hotels, restaurants and other leisure-related occupations (closed throughout a large portion of the pandemic) are significantly reduced (Fig. 5).

Stability of the recall indicator

[5] The description includes: gender, number of children, maximum level of education reached and its type, geographic location as specified by latitude, longitude and department, desired occupation, willingness to work full or part time, accepted mobility, experience, qualification level, desired contract type, reason of inscription, administrative category of the job search, and zip-code level socio-demographic features. Some care was exercised to prevent information leakage between the job seekers description and the considered period (e.g. through the minimum wage).

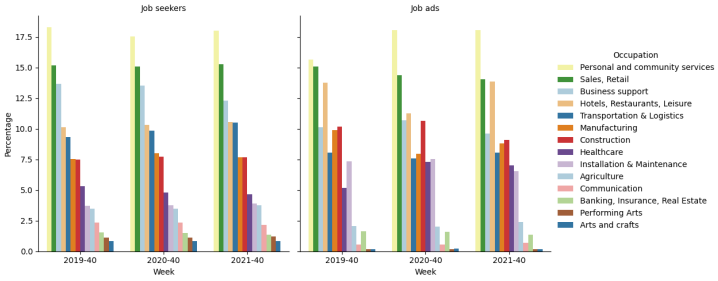


Fig. 5. Distributions of job seekers (left) and job ads (right) per sector, during the 40^{th} week of 2019, 2020 and 2021. Very moderate changes are visible concerning the job seekers (slight decrease of Business Support, slight increase of Transport&Logistics). Very significant changes are visible concerning the job ads, notably in the "Personal and community services", and in the "Hotels, restaurant and leisure" sectors.

Recall@k	Before-LDs	During-LDs	After-LDs
10	23.33	30.70	24.50
20	32.49	41.17	34.07
50	45.37	54.00	47.73
100	54.97	64.54	57.38

Fig. 6. Average recall of VADORE.1 during the Before-LDs, During-LDs and After-LDs periods.

Appendix G: Analysis of the biases

An important issue regarding the design and application of machine learning for social good [5] is related to the biases, usually present both in the data and in the behavior of the system. These biases are characterized using the same methodology: i) by defining a category of job seekers who are less "well served" by the PES (respectively by VADORE), ii) by tackling the discrimination between the less "well served" category and the well served category; iii) by analyzing and interpreting the feature importance provided by a random forest tackling the classification problem.

Biases in the data. The "less well served" category of job seekers is formed of the job seekers staying on the PES portal for eight consecutive weeks without finding a match (noting that the job seeker population includes circa 1 million individuals, with about 1.4k matches per week).

A random forest, considering an equidistributed dataset of size 278k, yields a predictive accuracy of 60.6%. The two considered categories of job seekers are therefore significantly statistically distinct, though the gap remains moderate.

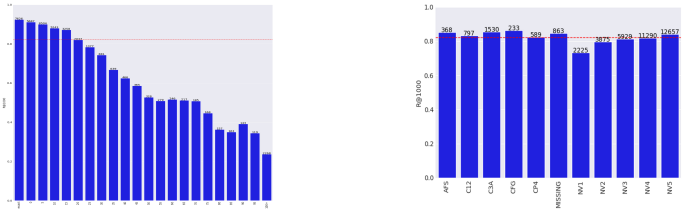
The most important features retained by the random forest to discriminate among both categories are: the reservation wage, the textual information ("business card" of the job seeker") and their spoken languages. A closer look at the

feature importance suggests that the random forest actually tackles another discrimination problem as the one intended: it discriminates the job seekers conducting an *active search on the PES portal*, from the job seekers conducting their search otherwise.

Biases in the VADORE *relevance.* Along the same line, the category of job seekers "less well served" by VADORE is set to the job seekers for whom the matched job ad has rank greater than 1,000 (circa 20% of the job seekers), missed by both VADORE.0 and VADORE.1.

A random forest, considering 34k job seekers in the test set, yields a predictive accuracy above mere chance, with an AUC .65. The two considered categories of job seekers are therefore significantly statistically distinct too.

The marginals of the "less well served" category show that these job seekers are those finding a job rather far apart from their location (Fig. 7, left), and (seemingly independently) with a comparatively high educational background (Fig. 7, right; NIV1 corresponds to 5 years after the end of high school). As noted by [16], the mobility of the job seekers tends to increase with their educational background. Overall, it is suggested that VADORE.1 serves less well the people



Recall *vs* distance to the job ad location Recall *vs* educational background

Fig. 7. VADORE.1: Who are the less "well served" job seekers ?

finding a job farther away, and those with a stronger educational background. This bias is interpreted as reflecting the distribution of job seekers present on the PES' portal.