Toward Job Recommendation for All

Guillaume Bied[1,2], Solal Nathan[1], Elia Pérennès[2,3], Morgane Hoffmann[2,3], Philippe Caillou[1], Bruno Crépon[2], Christophe Gaillac[4], Michèle Sebag[1]

[1]Laboratoire Interdisciplinaire des Sciences du Numérique (LISN), Université Paris-Saclay, France
[2]Centre de Recherche en Économie et Statistiques (CREST), France
[3]Pôle emploi, Paris, France
[4]Oxford University, United Kingdom

August 23, 2023

Summary

Summary

Context

- **Reducing unemployment**: key domain of AI for Social Good
  - Contributes to UN Sustainable Development Goals:
    - Goal 8: Decent work and Economics Growth
    - Goal 10: Reduced Inequalities

- Hope: **job recommendation** can reduce frictional unemployment and leave no one behind

Objectives

- Design a **Job Recommender System** based on proprietary data from *Pôle emploi*, the French Public Employment Service (PES)
- Challenges:
  - Handle **sparse** interaction matrix (hiring and application): 99.5% sparsity
  - Scale up: 400k job seekers ($\sim$ minimum wage)
  - Ensure **fairness** of the learned system
- Contributions:
  - MUSE: Multi-head Sparse e-Recruitment system
    - Comparatively assessed on public (RecSys 2017) and proprietary (*Pôle emploi*) data
  - Audit of MUSE's gender biases (compared to hiring and application data)

Summary

Related Work

- Expert systems, *e.g.* WCC ELISE
- Collaborative filtering — Bell et al., 2007 (Netflix prize)
- 2016 & 2017 RecSys challenges on job data

  Xiao et al., 2016; Volkovs et al., 2017
- e-recruitment systems based on proprietary data

  Kenthapadi et al., 2017 (LinkedIn)

  Zhao et al., 2021 (CareerBuilder)

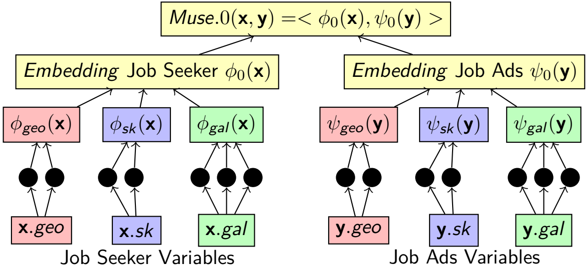
Summary

Available Data

- Job seeker / job ad descriptions:
  - Job seekers: zip code, skills, occupation, experience, education, accepted mobility, text, past experiences, search criteria... (dim. $\sim$ 500 after preprocessing)
  - Ads: zip code, skills, occupation, wage, contract, working hours, job & firm descriptions, requirements... (dim. $\sim$ 500 after preprocessing)
- Interactions: signed contracts and applications (99.5% sparsity)
- Where: Auvergne-Rhône-Alpes region
- When: 2019-2022
  - 400k job seekers, 70k job ads and 1.4k matches per week
- Specificities
  - Job seekers with low qualification
  - Few interactions per job seeker (vs. Xing dataset)

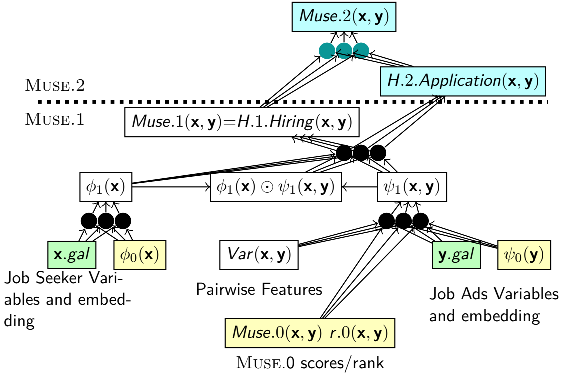
MUSE: First Tier

- Goal: rank job ads y for each job seeker x
- First tier serves as a filter, to support scalability
- Job seeker & ad embeddings are the concatenation of modules dedicated to:
  1. Geographic aspects;
  2. Declared / required skills
  3. General job seeker / job ad description
- Only top 1,000 job ads y selected for each job seeker x are considered in the following



MUSE: Second Tier

- Goal: re-rank the top 1,000 ads selected by the first tier, using more sophisticated attributes (*e.g.* geographic distance)
- One head to predict hiring
- One head to predict applications
- Last head: on top of both, to predict hirings



Results

- Baseline: XGBoost based on RecSys 2017 Challenge winner Volkovs et al., 2017

Recall@	XGBoost	Muse.0	Muse.2
10	26.83	22.88	**30.1***
20	35.59	31.55	**40.2***
100	58.88	53.80	**63.2***
1000	**86.47***	82.13	-
Training time (hours)	1.83	7.7	1.25
Recommendation time (seconds)	1.4	0.0004	0.02

Comparative results of Muse and XGBoost: recall, overall training time and recommendation time *per* job seeker.[1]

Muse: ✅ Scalability
✅ Recall

[1] The reported computational times are obtained on Intel® Xeon® Silver 4214Y CPU @ 2.20GHz, with 187 GB RAM and a Tesla T4 GPU.

Summary

Audit of gender biases: research questions

- A RS learned from real-world data may reproduce unwanted biases & discriminatory behavior — Ekstrand et al., 2021

Audit of gender biases: research questions

- A RS learned from real-world data may reproduce unwanted biases & discriminatory behavior Ekstrand et al., 2021

1. Is recommendation performance similar for men and women?
    - Recall@10 is 31% for women and 29% for men

Audit of gender biases: research questions

- A RS learned from real-world data may reproduce unwanted biases & discriminatory behavior Ekstrand et al., 2021

1. Is recommendation performance similar for men and women?
   - Recall@10 is 31% for women and 29% for men
2. Inspect differences in
   - jobs recommended to men / women w.r.t.
     - Wage
     - Distance (kilometers)
     - Executive status
     - Contract duration
     - Hours worked per week
     - Occupation (male-dominated or not)

   compared to:
   - jobs actually taken by men / women
   - jobs applied to by men / women

Audit of gender biases: research questions

- A RS learned from real-world data may reproduce unwanted biases & discriminatory behavior Ekstrand et al., 2021

1. Is recommendation performance similar for men and women?
   - Recall@10 is 31% for women and 29% for men
2. Inspect differences in
   - jobs recommended to men / women w.r.t.
     - Wage
     - Distance (kilometers)
     - Executive status
     - Contract duration
     - Hours worked per week
     - Occupation (male-dominated or not)

   compared to:
   - jobs actually taken by men / women
   - jobs applied to by men / women

   Question: does recommendation amplify the biases?

Gender biases in terms of ad characteristics: Methodology

- Y: characteristic of recommended job, *e.g.* wage
- Gender G (man / woman)
- Control for the specifics Z of job seekers (qualifications, declared preferences)
- Controlled model:

$$Y = \mu_0(Z) + \tau G + \varepsilon, \quad \mathbb{E}(\varepsilon | G, Z) = 0,$$

  - $\mu_0(z) := \mathbb{E}(Y | G = 0, Z = z)$: expected characteristics of jobs for men with preferences z
  - ϵ: noise variable
  - τ: gender gap in characteristics Y controlling for Z (parameter of interest)
- Estimation using Double Machine Learning

Chernozhukov et al., 2016

Gender biases in terms of ad characteristics: Results

	τ_{Hire}	τ_{App}	τ_{Muse}
LogWage	-0.005*	-0.006*	-0.006*
Distance	-2.628*	-1.942*	-0.182
Executive position	-0.006*	-0.004*	-0.004
Indefinite duration	-0.031*	-0.039*	-0.033*
Male-dominated occupation	-0.170*	-0.172*	-0.116*
Full time	-0.101*	-0.092*	-0.066*
Hours worked	-1.518*	-1.366*	-0.825*

Table: Gender-related gaps in recommendations, hires and applications

- Gender gaps (conditional on Z) exist in hiring and application behavior
  - For women: jobs paid less, closer to home, shorter working hours, more often in short-term duration contracts, less often in executive status and in male-dominated occupations
- MUSE-based recommendations do not amplify biases, and significantly reduce gaps w.r.t some features (more in the paper)

Summary

Conclusion and perspectives

- Contributions:
  - MUSE, a two-tier job RS tailored for the needs of a PES
  - On the *Pôle emploi* dataset, MUSE outperforms the XGBOOST baseline w.r.t. recall and scalability
  - In the paper: comparative assessment of MUSE w.r.t. Dropoutnet (Volkovs et al., 2017) on public data[2]
  - Audit the algorithm w.r.t. gender gaps
    - These gaps are shown to exist, but are similar or smaller than those observed in hiring and application data
  - Field experiment on 160,000 job seekers (June 2023)

[2] https://gitlab.com/solal.nathan/vadore_ijcai

Conclusion and perspectives

- Perspectives:
  - Proposing a subscription service at *Pôle emploi*
  - Impact on employment
  - Exploring the trade-off between recommendation quality and fairness objectives
  - Mitigating possible congestion effects?
  - Use of more sophisticated models (modeling job seekers' position on the labor market)

Thanks for your attention!

If you have any question or comment, please contact us by email:
`bied@lri.fr`