Job Recommandation on Real-World Interaction Data With Heterogeneous Graphs

TELECOM SudParis

IP PARIS

s@movar

Institut Mines-Télécom

Auteurs

Eric Behar

Julien Romero

Amel Bouzeghoub

Katarzyna Wegrzyn-Wolska

Partenaires

easy partner

efrei PARIS PANTHÉON-ASSAS UNIVERSITÉ

Libraries

PyTorch

PyTorch geometric

plotly

pandas

Problems

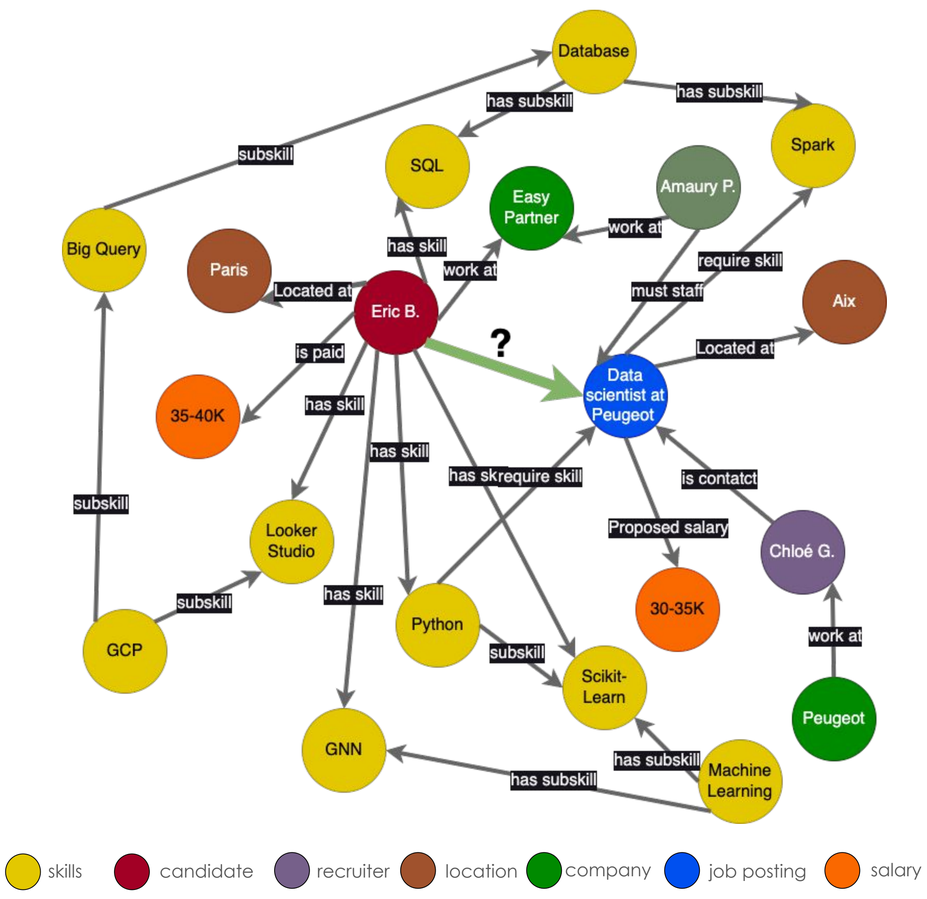
Goal: We want to recommend relevant open positions to candidates. However:

- People get few jobs during their life
  - Very few training point for each user
  - Cold start becomes an issue!
- How to represent candidates in a unified semantic graph?
- How to deal with the temporal dimension of recruiting?
- How to make the recommendations explainable?

Recruiting Steps at Easy Partner

Dataset Construction

- Existing datasets generated by candidates
  - We get noisy data
- We used data from Easy Partner, a company specialized in talent recruiting.
  - Data generated by expert recruiters
  - Recruiters annotate candidates with semantic information and notes.
  - We construct a heterogeneous graph
- We enriched the heterogeneous graph with other knowledge bases:
  - ESCO: The european job and skill taxonomy
  - Wikidata: A general knowledge base
- In the end, we get a graph with 52,728 nodes, 26,000 candidates, and 4,000 jobs.



Sample of the Heterogeneous Job-Candidate Graph

Future Works

- We are implementing a graph neural network on top of our heterogeneous job-candidate graph
  - Integration into a general evaluation framework
- We want to integrate the temporal dimension:
  - We cannot use future recommendations
  - We have to represent the most recent recommendations
- Given the recommendations from our engine, how can we explain it to a recruiter?
- Can we estimate the bias of our algorithm?

Comparison of Existing Datasets

Dataset	#Users	#Items	Items/User	Users/Item
MovieLens	6k	9k	137	85.5
Yelp	2M	150k	3.50	46.5
CareerBuilder	321k	365k	4.99	4.38
Our Dataset	26k	4k	2.58	16.7

Comparison of Existing Algorithms

- Collaborative Filtering: item2Vec, DeepWalk, NGCF, ALS
- Content-based We compare these datasets according to several metrics
- (usage of additional features): Wide & Deep, FM, DeepFM, YoutubeRanking, AutoInt, DIN
- Graph-based : LightGCN, GraphSage, PinSage, NGCF

=> LightGCN and GraphSage provide the best results

Algorithms	Log Loss	ROC	P@10	R@10	MAP@10	NDCG@10	Features?
Item2Vec	0,691	0,7468	0,0018	0,0146	0,0096	0,0112	No
DeepWalk	1,5621	0,9023	0,0091	0,0763	0,0328	0,0449	No
NGCF	**0,4967**	0,8876	0,005	0,0424	0,0229	0,0288	No
LightGCN	0,5601	**0,9081**	**0,0104**	**0,0862**	**0,0423**	**0,0557**	No
GraphSage	0,5093	0,8993	0,0053	0,0422	0,0193	0,0268	Yes
PinSage	1,9526	0,8899	0,0048	0,0393	0,0205	0,0265	Yes
ALS	0,6367	0,8554	0,0045	0,0382	0,0178	0,0238	No
BPR	0,684	0,7687	0,001	0,0081	0,0022	0,0039	No
NCF	0,6626	0,8361	0,001	0,0076	0,0028	0,0044	No
SVD	0,6472	0,7947	0,0012	0,0095	0,0019	0,0041	No
SVDpp	0,5437	0,8828	0,0083	0,0703	0,0376	0,0476	No
Caser	0,6917	0,7066	0,0008	0,0063	0,0033	0,0044	No
RNN4Rec	0,585	0,7719	0,001	0,0082	0,0019	0,0037	No
WaveNet	0,6918	0,7623	0,0012	0,0091	0,0029	0,005	No
Din	0,8335	0,565	0,0005	0,0035	0,0013	0,0021	Yes
FM	1,6901	0,5171	0,0001	0,0009	0,0003	0,0005	Yes
AutoInt	7,3647	0,542	0,0001	0,001	0,0003	0,0006	Yes
DeepFM	0,6923	0,5898	0,0002	0,0013	0,0003	0,0006	Yes
WideDeep	0,6838	0,5979	0,0002	0,0019	0,0011	0,0014	Yes
YouTubeRanking	0,6838	0,5979	0,0002	0,0019	0,0011	0,0014	Yes
LogReg	0,6914	0,7446	0,0015	0,0124	0,0094	0,0104	Yes

Journée des doctorants

mai 2023

Contact eric.behar@telecom-sudparis.eu