

JobBERT: Understanding Job Titles through Skills

Jens-Joris Decorte, Jeroen Van Haute, Thomas Demeester, Chris Develder
jensjoris@techwolf.ai jeroen@techwolf.ai thomas.demeester@ugent.be chris.develder@ugent.be
jensjoris.decorate@ugent.be

TECHWOLF


Context & Goal

Task: classifying free-form job titles with respect to a standardized job title list (**Job Title Normalization**)

Goal: design and build a solution that

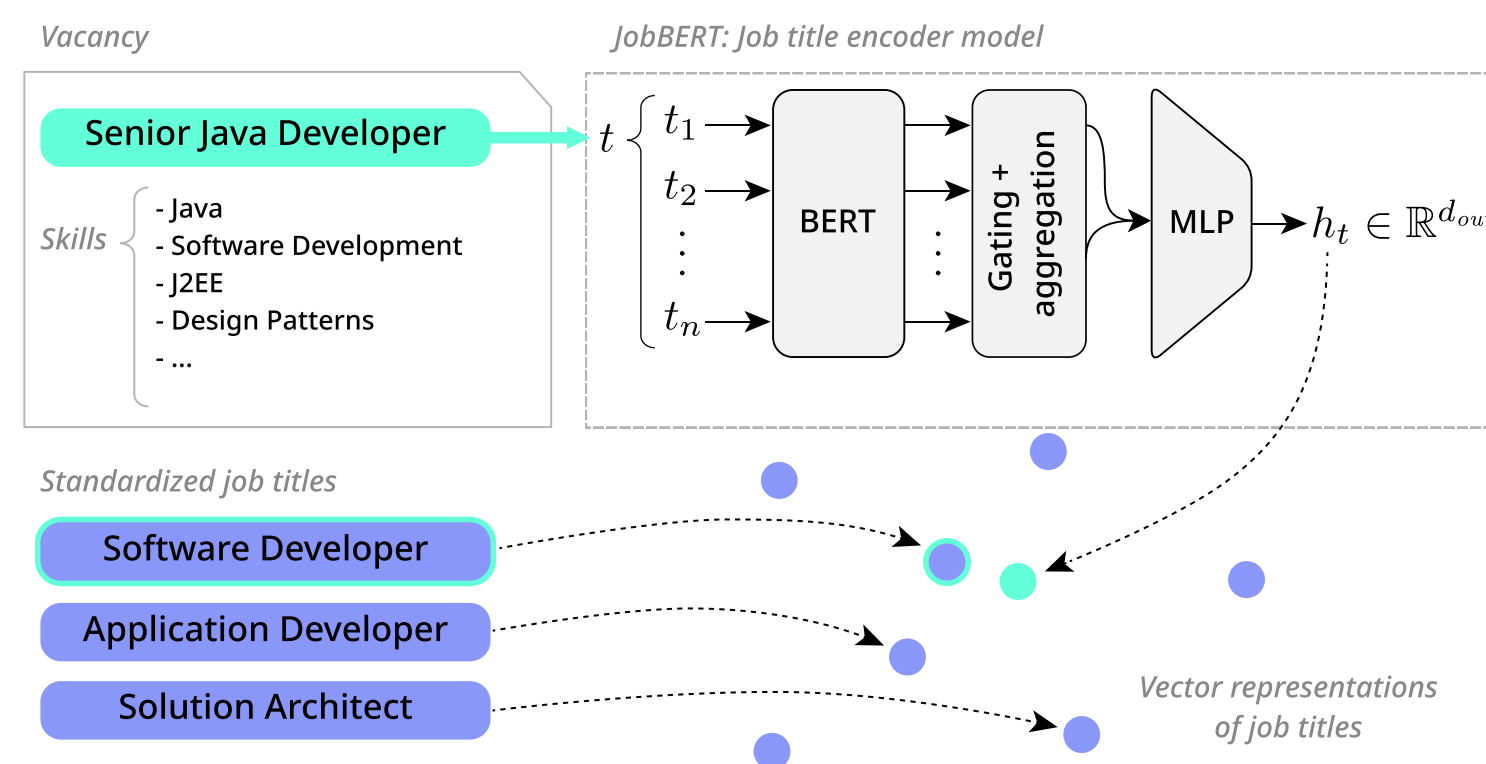
- requires **no manual labeling efforts**
- can be **applied to any taxonomy without retraining**

Contributions

- **Semantic text similarity** (STS) approach for job title normalization
- Learning the meaning of job titles through their entailed skills (**weak supervision**)
- **Token gating mechanism** for model inspection
- **Evaluation dataset:**
 <https://github.com/jensjorisdecorate/JobBERT-Understanding-Job-Titles-through-Skills>

JobBERT Architecture & Training

- **BERT-based encoder** with token gating mechanism
- Trained on corpus of vacancies to **predict mentioned skills given a vacancy title**
- **Skip-Gram objective** with negative sampling
- Similar job titles in terms of skills result in similar vector representations



Gating Mechanism

- Token embeddings are dynamically weighted before aggregation
- Learns to assign a lower weight to words that do not contribute to the meaning of a job title

Experimental Results

Experimental setup

- Training data: scraped 10.5M online vacancies
- Skill extraction: literal string match with 35K vocab
- Evaluation data: 30,926 vacancy titles tagged with ESCO occupation labels
 - dev/test: 50/50%
- BERT weights fixed (JobBERT) or fine-tuned (JobBERT_{FT})

Results

Method	MRR		Recall@1	
	Macro	Micro	Macro	Micro
LASER	0.219	0.255	0.160	0.172
fastText	0.253	0.321	0.171	0.160
BERT _{avg}	0.200	0.206	0.136	0.149
Sentence-BERT _{avg}	0.270	0.265	0.186	0.193
JobBERT	0.326	0.269	0.230	0.192
JobBERT _{FT}	0.364	0.309	0.267	0.225

Conclusions

- Effective fine-tuning of BERT for job title normalization based on job titles and their skills
- Outperforms state-of-the-art STS models, even without fine-tuning BERT weights
- Gating mechanism: inspect what the model has learned

ECML PKDD 2021

VIRTUAL
13-17 September