IPOD: An Industrial and Professional Occupations Dataset and its Applications to Occupational Data Mining and Analysis

Junhua Liu[1], Yung Chuen Ng[2], Kristin L. Wood[1], and Kwan Hui Lim[1]

[1] Singapore University of Technology and Design
junhua_liu@mymail.sutd.edu.sg, kristinwood@sutd.edu.sg,
kwanhui_lim@sutd.edu.sg
[2] National University of Singapore
e0201912@u.nus.edu

Abstract. Occupational data mining and analysis is an important task in understanding today's industry and job market. Various machine learning techniques are proposed and gradually deployed to improve companies' operations for upstream tasks, such as employee churn prediction, career trajectory modelling and automated interview. Job titles analysis and embedding, as the fundamental building blocks, are crucial upstream tasks to address these occupational data mining and analysis problems. In this work, we present the Industrial and Professional Occupations Dataset (IPOD), which consists of over 190,000 job titles crawled from over 56,000 profiles from Linkedin. We also illustrate the usefulness of IPOD by addressing two challenging upstream tasks, including: (i) proposing *Title2vec*, a contextual job title vector representation using a bidirectional Language Model (biLM) approach; and (ii) addressing the important occupational Named Entity Recognition problem using Conditional Random Fields (CRF) and bidirectional Long Short-Term Memory with CRF (LSTM-CRF). Both CRF and LSTM-CRF outperform human and baselines in both exact-match accuracy and F1 scores. The dataset and pre-trained embeddings are available at https://www.github.com/junhua/ipod.

Keywords: Occupational Data Mining · Named Entity Recognition · Natural Language Processing · Title2Vec · dataset

1 Introduction

There is a growing interest in occupational data mining and analysis tasks in recent years, especially with the rapid digitization of today's economy and jobs. Furthermore, the advancement of AI and robotics are changing every industry and every sector, challenging the employability of work force especially those with high level of repetition.

Occupational data mining and analysis is also an important topic in academia research. In the literature, various downstream tasks of occupational analysis showed promising results using machine learning techniques, such as employee churn prediction

Literature	Source	Size	Avail.
IPOD	**Linkedin**	**190K**	**Yes**
Mimno et al., 2008	Resumes	54K	No
Lou et al., 2010	Linkedin	67K	No
Paparrizos et al., 2011	Web	5M	No
Zhang et al., 2014	Job site	7K	No
Liu et al., 2016	Social network	30K	No
Li et al., 2017	Linkedin	-	No
Li et al., 2017	High tech co.	-	No
Yang et al., 2017	Resumes	823K	No
zhu et al., 2018	Job portals	2M	No
James et al., 2018	APS	60K	Yes
Yang et al., 2018	Var. channels	-	No
Xu et al., 2018	Pro. networks	20M	No
Qin et al., 2018	High tech co.	1M	No
Lim et al., 2018	Linkedin	10K	No
Shen et al., 2018	High tech co.	14K	No

Table 1: A survey of datasets used for related works. No available datasets can be found publicly except a dataset of publications and authors from American Physics Society (APS) [13] that only describes the names and affiliations of physics scientists without titles.

[13,44,46], professional career trajectory modelling [21,25] and predicting employee behaviors with various factors [8,7], among others.

These earlier works on occupational data mining and analysis fulfil industrial demands and create substantial value for both the companies and professionals. However, these tasks remain challenging due to a lack of publicly available datasets. For many years, the relevant data resides with a small number of enterprises, which utilize such data privately for maintaining their competitive edge in the industry. Thus, such data is not publicly available for smaller companies or individuals to better understand the industry and job market or for career planning purposes. Table 1 shows a survey of 15 related works that utilizes similar types of dataset, of which only one is publicly available (apart from our proposed dataset).

1.1 Main Contributions

In this paper, we make the following contributions:

- To address the needs of career analysis for industry, we present and make publicly available the Industrial and Professional Occupation Dataset (IPOD). This dataset consists of 192,295 occupation entries, drafted by working professionals for their Linkedin profiles, with the motivations of displaying their career achievement, attracting recruiters or expanding professional networks. As shown in Table 1, IPOD is the largest publicly available dataset out of a total of 16 used in various recent works.

- To improve the usability of IPOD, we propose *Title2vec*, a contextual job title vector representation with a bidirectional Language Model (biLM) [30] approach. This upstream embedding task map the raw job titles into a high-dimensional vector space that allows and boosts the performance of the downstream occupational NER task.
- To further demonstrate the usefulness of IPOD, we propose two models for a challenging Named Entity Recognition (NER) task, alongside with 2 baselines and human performance. The two models include a probabilistic machine learning model, namely Conditional Random Field (CRF), and a state-of-the-art recurrent neural network model, namely bidirectional LSTM-CRF [20]. Both of two models outperform human and baselines in terms of Exact Match (EM) accuracy and F1 scores for both overall and tag-specific results.

Existing corpora for Named Entity Recognition (NER) tasks [11,38,42,6] typically use general tags such as **LOC**ation, **PER**son, **ORG**anization, **MISC**ellaneous, etc.. On the contrary, IPOD provides domain-specific NE tags to denote the properties of occupations, such as ***RES****ponsibility*, ***FUN****ction* and ***LOC****ation*. All named entities are tagged using a comprehensive gazetteer created by three experts, which reports high inter-rater reliability, achieving 0.853 on Percentage Agreement [41] and 0.778 on Cohen's Kappa [3], with no instances where all three annotators disagree. The labels are further processed by adding prefix using BIOES tagging scheme [36], i.e., **B**egin, **I**nside, **E**nding, **S**ingle, and **O** indicating that a token belongs to no chunk, indicating the positional features of each token in a title.

2 Description of the Industrial and Professional Occupations Dataset (IPOD)

In this section, we describe our data collection process, characteristics of the IPOD dataset and results from an exploratory data analysis.

2.1 Data Collection

We obtained over 192K job titles based on Linkedin profiles from Asia and the United States, as representatives of the world's most competitive economies [2]. Subsequently, the raw data underwent a series of processing, including converting to lowercase, substituting meaningful punctuation to words (i.e. changing $\&$ to and) and removing special symbols. We decided not to lemmatize or stem the words because the original forms suggest its most accurate named entity, i.e., strategist is labeled as RES while strategy is labeled as FUN.

2.2 Dataset Analysis

This section discusses the exploratory data analysis conducted to better understand the properties of IPOD. The statistics and histogram of the length of job titles can be found in Table 2 and Fig. 1 respectively. The corpus comprises of 192,295 English

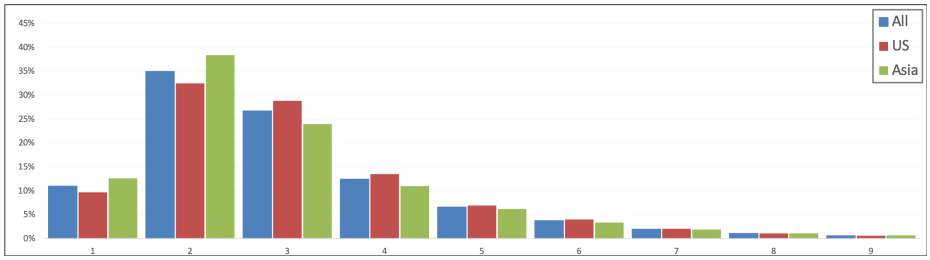

Fig. 1: Histogram of occupation entries. The x-axis shows the number of words in the occupation title and the y-axis shows the usage frequency in percentage.

occupation entries from 56,648 unique profiles. These profiles are mainly from United States (56.7%) and Asia (43.3%). Most of the titles fall within five words, contributing to 91.7% of the entries, as shown in both Table 2 and Fig. 1). The median statistics and the histogram also suggest that job titles written by Asian professionals tend to be shorter, i.e., within two words, than that by US professionals.

	All	US	Asia
min	1	1	1
max	21	17	21
avg	3.0	3.1	2.9
med	3	3	2

Table 2: Statistics of entries

NE	Count
RES	310570
FUN	255974
LOC	9998
O	66948

Table 3: NE counts

Figure 3 shows the distribution of top 20 Unigrams and Bigrams [9] of IPOD. In the Unigram case, the most popular token, *manager*, appears in 34,065 entries, about twice as much as the next few popular ones, i.e., *and (18,466)*, *senior (16,475)*, *engineer (15,593)* and *director (14,182)*. On the contrary, the Bigram case shows a gentler curve (i.e. lower slope), with *project manager (3,536)* and *vice president (3,458)* being the top two choices.

2.3 Domain-specific Sequence Tagging

Job titles serve as a concise indicator for one's level of responsibility, seniority and scope of work, described with a combination of *responsibility*, *function* and *location*. Table 4 shows examples of the occupational NE tags.

Responsibility, as its name suggests, describes the role and duty of a working professional. As shown in figure 2, responsibility may include indicators of managerial levels, such as *director, manager* and *lead*, seniority levels, such as *vice, assistant* and *associate*, and operational role, such as *engineer, accountant* and *technician*. A combination of the three sub-categories draws the full picture of one's responsibility.

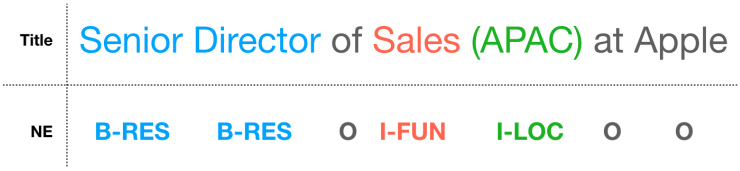

Fig. 2: An example of occupational title and its domain-specific NE tags. Tokens in a title indicate the person's responsibility (**RES**), function (**FUN**), and location (**LOC**). Furthermore, the NE tags are also added with positional prefixes using BIOES scheme, i.e., **B**egin, **I**nside, **O**thers, **E**nding and **S**ingle.

Function describes business functions in various dimensions. Specifically, *Departments* describes the company's departments the staffers are in, such as *sales, marketing* and *operations*; *Scope* indicates one's scope of work, such as *enterprise, project* and *national*; lastly, *Content* indicates one's content of work, such as *data, r&d* and *security*.

Finally, **Location** indicates the geographic scope that the title owner is responsible of. Examples of this NE tag include geographic regions such as *APAC, Asia, European*, and counties/states/cities such as *China, America* and *Colorado*.

Formally, we define the occupational domain-specific NE tags as *RES, FUN, LOC* and *O*, indicating the *responsibility, function, location* and *others* respectively. For instance, a job title of *chief financial officer asia pacific* is tagged as *S-RES S-FUN S-RES B-LOC E-LOC* with the BIOES scheme [36]. The distribution of the four labels are shown in Table 3. We adopt a knowledge-based NE tagging strategy by creating a gazetteer of word tokens. This is achieved by first running a Unigram analysis of the job titles, sorted in descending order. Subsequently, the top 1,500 tokens are tagged by three annotators, who are a HR personnel, a senior recruiter and a seasoned business professional. Among 1,500 tokens tagged, every tag is agreed with at least two annotators, where 1,169 (77.9%) are commonly agreed among all three annotators, and 331 (22.1%) are agreed with two annotators. We further assess the Inter-Rater Reliability with two inter-coder agreements, achieving 0.853 on Percentage Agreement [41] and 0.778 on Cohen's Kappa [3], a *Strong* level of agreement. Finally, the job titles are labelled with NE tags using BIOES scheme and formatted for NER tasks.

Responsibility	Managerial level: *lead, supervisor, manager, director, president*
	Operational role: *engineer, designer, accountant, technician*
	Seniority: *junior, vice, associate, assistant, senior*
Function	Departments: *sales, marketing, finance, operations, strategy*
	Scope: *enterprise, project, customer, national, site*
	Content: *data, r&d, security, training, integration, education*
Location	Regions: *APAC, SEA, Asia, European, north, central*
	Countries/States/Cities: *China, America, Singapore, Colorado*

Table 4: Examples of occupational NE tags.

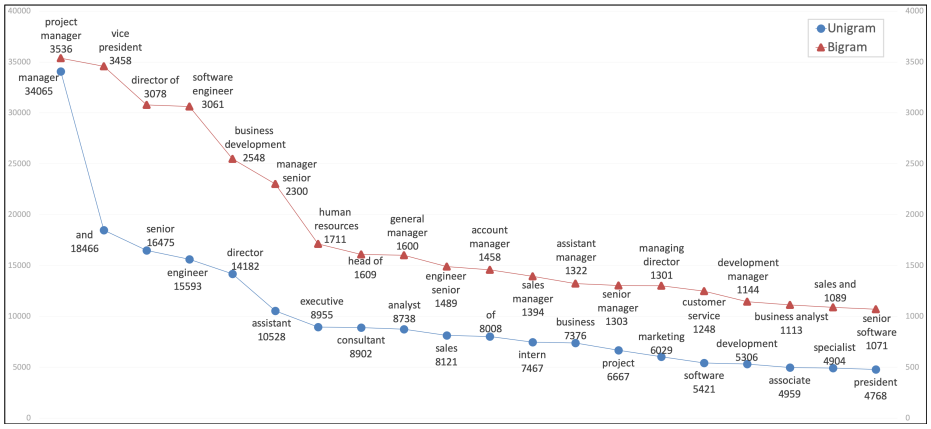

Fig. 3: N-grams analysis of the occupation title entries, where the y-axis shows the usage frequency.

3 Applications of IPOD: Named Entity Recognition and Job Embedding

To demonstrate the usefulness of IPOD, we use this dataset for an occupational NER task and propose a job title vector representation called Title2vec, which we describe next.

3.1 NER Models

We propose two encoders to address the occupational NER task. One of the models is Conditional Random Fields (CRF), which is a probabilistic machine learning model that produce joint probability of the co-occurrence of output sequence [16]. The other model falls in the recurrent neural networks family, namely a bidirectional Long Short-Term Memory model with a CRF layer adding to the output layer (LSTM-CRF). Variations of the LSTM-CRF model showed state-of-the-art results in some recent works for different downstream NLP tasks [17,20]. Both the CRF and LSTM-CRF are decoded using a first-order Viterbi algorithm [12] that finds the sequence of NE tags with highest scores. We also construct two baseline encoders, namely a Logistic Regression (LogReg) classifier and a standard LSTM, both of which are decoded using a softmax layer.

3.2 Job Embedding

We propose a contextual job title vector representation model, *Title2vec*, using a bidirectional Language Model (biLM) approach [30] where each token is represented with a contextual vector with 3072 dimensions, before passing through a bi-directional LSTM network. The forward LSTM predicts the probability of each token given its history, and the backward LSTM takes the same approach with reverse order.

Instead of training from scratch, we construct *Title2vec* by fine-tuning from a pre-trained model, namely the Embedding from Language Models (ELMo) [30]. The choice

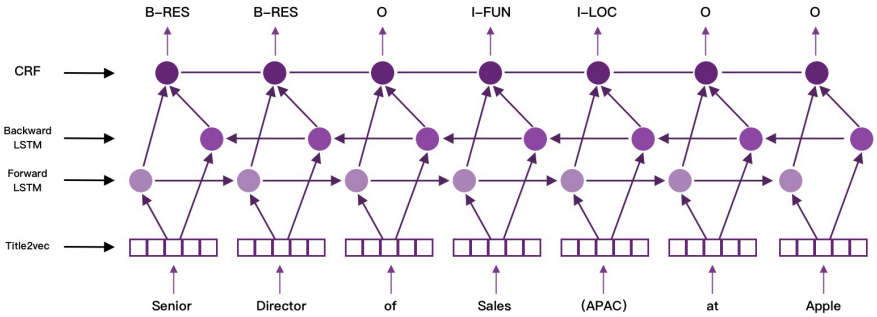

Fig. 4: Bidirectional LSTM-CRF model for occupational NER task

is because ELMo provides a language-level contextual meaning for word tokens that is highly similar, if not identical, to that in job titles. For instance, the word *director* appearing in a job title is the same as that appears in a Wikipedia article.

3.3 Hyper-parameter Optimization

We conduct grid search of hyper-parameters for both CRF and LSTM-CRF models to fine-tune the performance of proposed methods. The search space includes varying learning rates (i.e., 0.1 or 0.01), number of LSTM hidden layers (1 or 2), number of hidden states (128 or 256), mini-batch size (32 or 128) and type of optimizer(Adam or SGD). Word Dropout and Variational Dropout [15] are used to prevent over-fitting, with probability of 0.05 and 0.5 respectively. In total, we evaluate over 100 hyper-parameters sets, where each set of hyper-parameters is run with 10 epochs.

Hyper Parameter	CRF		LSTM-CRF	
	Final	Range	Final	Range
Learning Rate	0.1	{0.01,0.1}	0.1	{0.01,0.1}
Mini-batch Size	32	{32,128}	128	{32,128}
Word Dropout	0.05	-	0.05	-
Variational Dropout	0.5	-	0.5	-
Type of Optimizer	SGD	{Adam, SGD}	SGD	{Adam, SGD}
LSTM Layers	-	-	1	{1,2}
LSTM State Size	-	-	256	{128,256}

Table 5: Hyper-parameter search space and final values used

Table 5 shows the breakdown of the search space and the final hyper-parameters used for both models. We deploy a Cross Entropy loss function and a SGD optimizer with an initial learning rate of 0.1 and a mini-batch size of 32 for both proposed CRF model and baselines. For the two LSTM-based models, we use a single hidden layer with an initial learning rate of 0.1, LSTM state size of 256 and a mini-batch size of 128.

4 Experiments and Results

In this section, we discuss our experimental evaluation and results.

4.1 Metrics

Our work uses two metrics to assess performance of various machine models and human performance, namely *Exact Match (EM)* and the *F1* score, formally defined as $F1 = 2 * Precision * Recall * (Precision + Recall)^{-1}$. The *EM* metric measures the percentage agreement between the ground truth and predicted labels with exact matches, while the *F1* score metric is designed to measure the average overlaps between the ground truth and prediction. Furthermore, the overall *Precision* and *Recall* metrics of all models are also reported.

4.2 Human Performance

We construct the human performance baseline for IPOD using the NE tags annotated by the three domain experts. We choose the set of labels tagged by annotator 1 as the ground truth labels and compute against the other two annotation sets. We then take the average *EM* and *F1* to indicate human performance. We record an *EM* accuracy of 91.3%, and an *F1* of 95.4%. This shows a strong human performance as compared to those of other datasets, such as 91% *EM* for the CHEMDNER corpus[23], 86.8% EM and 89.5% F1 for SQuAD2.0 [33], and 77.0% *EM* and 86.8 *F1* for SQuAD1.0 [34].

4.3 Model Performance

Table 6 shows the overall performance of our models and human performance on IPOD, in terms of precision(P), recall(R), exact match(EM) and F1. While the performance of LSTM, CRF and LSTM-CRF models are very close to each others (± 0.2 difference), all three models outperform human in precision, recall, exact match and F1.

Models	P	R	EM	F1
LogReg	90.80	93.20	85.10	92.00
LSTM	99.71	99.90	99.61	99.80
CRF	**99.90**	99.81	99.71	99.85
LSTM-CRF	99.86	**99.97**	**99.83**	**99.91**
Human	91.60	99.60	91.30	95.40

Table 6: Overall results of Job Title NER

Table 7 shows the per-tag breakdown of NER results, in terms of EM and F1. CRF and LSTM-CRF perform similarly to each others, and outperform LogReg and LSTM for all three categories, though LSTM model also has a good performance. CRF also shows a significant advantage in classifying RES tags (99.99 EM and 99.99 F1).

	FUN		LOC		RES	
	EM	F1	EM	F1	EM	F1
LogReg	78.30	87.80	93.70	96.80	90.10	94.80
LSTM	99.49	99.74	97.68	98.83	99.77	99.88
CRF	99.35	99.67	**98.96**	**99.48**	**99.99**	**99.99**
LSTM-CRF	**99.88**	**99.94**	98.70	99.35	99.82	99.91

Table 7: Performance stratified by NE tags (EM, F1)

5 Related Work

In this section, we review related works, in the area of occupational data mining and analysis, contextual embedding and Named Entity Recognition (NER).

5.1 Occupational Data Mining and Analysis

Prior works on occupational data mining and analysis aim to accomplish a wide range of tasks, such as Career Modeling and Job Recommendation. In the area of Career Modeling, prior works address downstream tasks including career path modeling [21,25], career movement prediction [13,44], job title ranking [43], and employability [24]. In Job Recommendation, past works focus on analysing Person-Job Fit [47,39,31] which commonly aims to suggest employment suitability for companies, and Job Recommendation [22,45] which on the other hand provides decision analysis for the job seekers. These works commonly leverage real-world data from different sources, including Linkedin [21,19], resumes [25,44], job portals [45,47] and tech companies [39].

The proposed solutions to these problems are based on different approaches. Most works utilize various machine learning approaches, such as linear classification models [21,18,13,44], generative models [25,43] and Neural Networks [25,47,31]. Some take algorithmic approaches, such as statistical inference [25,13,39], Graph-theoretic models [24,28] and recommmender systems with content-based and collaborative filtering [45,22]. Some works report their time complexity to be polynomial [21,18].

5.2 Natural Language Processing

Word Embedding. Classic word embedding methods construct word-level vector representations to capture the contextual meaning of words. A Neural Network Language Model (NNLM) was proposed with Continuous Bag of Words (CBoW) model and skip-gram model [4], which lead to a series of NNLM-based embedding works [40]. Pennington et al., 2014 proposed GloVe [29], which uses a much simpler approach, i.e., constructing global vectors to represent contextual knowledge of the vocabulary, that achieves good results. More recently, a series of high quality contextual models are proposed, such as ELMo [30], FastText [5] and Flair [1]. Both word-level contextualization and character-level features are commonly used for these works.

Document Embedding. While Word Embedding constructs static continuous vectors on wordlevel, recent works also propose methods to represent document-level embeddings. A transformer-based approach receives high popularity in recent literature. It uses pre-trained transformer-based models with very large datasets to construct the document-level embeddings, such as Bert [10] and GPT [32], among others. This approach enables contextual embedding in both word level and document level. Lample et al. (2016) proposes a remarkable Stacked Embedding approach that constructs a hierarchical embedding architecture of document-level embedding by stacking word-level embedding with character-level features and concatenating with an RNN, which performs well in NER tasks [17].

Named Entity Recognition. Named entity recognition is a challenging task that traditionally involves a high degree of manually crafted features for different domains. The good news is that numerous large-scale corpora, such as CoNLL-2003 [38] and Ontonotes [42], are made available for training with deep neural architectures. State-of-the-art NER models are LSTM-based [17,20], where feeding the sentence embeddings into uni- or bi-directional LSTM encoder. Instead of decoding directly, some works also add a Conditional Random Field (CRF) layer at the end while training the classifier, and use Viterbi [12] to decode the probabilistic output into NE labels. The recently popular transformer-based models [10,32] are also capable of producing good results.

While manually tagging a large dataset requires tremendous amount of efforts, prior works leverage knowledge-based gazetteers developed by various unsupervised or semi-supervised learning approaches [14,37], or rely on generative models [27,26]. Tags can be further formatted with tagging schemes such as IOB [35] or BIOES [36], to indicate the position of tags in a chunk.

6 Conclusion

In this work, we present the IPOD corpus that comprises a large number of job titles, with a knowledge-based gazetteer that includes manual NE tags from three domain exports annotators. We also address two challenging upstream tasks of occupational data mining and analysis, namely job title embeddings and occupational NER. Despite strong human performance records of 91.3% *EM* and 95.4% *F1*, our proposed models, namely CRF and bidirectional LSTM-CRF, outperform human and baselines in EM and F1 for overall results and per-tag breakdown. Finally, we release a pre-trained *Title2vec* job title vector representation that can serve as basic building blocks and improve the performance for a wide spectrum of downstream tasks. To the best of our knowledge, our work is the first attempt to address the challenging occupational NER task and both the dataset and pre-trained embeddings are first made available in the literature of occupational analysis.

References

1. Akbik, A., Blythe, D., Vollgraf, R.: Contextual string embeddings for sequence labeling. In: Proc. of COLINGs. pp. 1638–1649 (2018)
2. Akhtar, A.: Singapore and hong kong have overtaken the us as the most competitive economies. here's how 25 countries rank. (2019), https://www.businessinsider.com/most-competitive-economies-in-the-world-2019-5

3. Artstein, R., Poesio, M.: Inter-coder agreement for computational linguistics. Computational Linguistics **34**(4), 555–596 (2008)
4. Bengio, Y., Ducharme, R., Vincent, P., Jauvin, C.: A neural probabilistic language model. JMLR **3**(Feb), 1137–1155 (2003)
5. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. Trans. ACL (2017)
6. Borchmann, Ł., Gretkowski, A., Gralinski, F.: Approaching nested named entity recognition with parallel lstm-crfs. Proceedings of the PolEval2018Workshop p. 63 (2018)
7. Cetintas, S., Rogati, M., Si, L., Fang, Y.: Identifying similar people in professional social networks with discriminative probabilistic models. In: Proc. of SIGIR. pp. 1209–1210 (2011)
8. Chen, Z.: Mining individual behavior pattern based on significant locations and spatial trajectories. In: Proc. of PerCom Workshops. pp. 540–541 (2012)
9. Damashek, M.: Gauging similarity with n-grams: Language-independent categorization of text. Science **267**(5199), 843–848 (1995)
10. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv:1810.04805 (2018)
11. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by gibbs sampling. In: Proc. of ACL. pp. 363–370 (2005)
12. Forney, G.D.: The viterbi algorithm. Proceedings of the IEEE **61**(3), 268–278 (1973)
13. James, C., Pappalardo, L., Sîrbu, A., Simini, F.: Prediction of next career moves from scientific profiles. arXiv:1802.04830 (2018)
14. Kazama, J., Torisawa, K.: Inducing gazetteers for named entity recognition by large-scale clustering of dependency relations. In: Proc. of ACL-HLT. pp. 407–415 (2008)
15. Kingma, D.P., Salimans, T., Welling, M.: Variational dropout and the local reparameterization trick. In: Proc. of NIPS. pp. 2575–2583 (2015)
16. Lafferty, J., et al.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data (2001)
17. Lample, G., et al.: Neural architectures for named entity recognition. arXiv:1603.01360 (2016)
18. Li, H., Ge, Y., Zhu, H., Xiong, H., Zhao, H.: Prospecting the career development of talents: A survival analysis perspective. In: Proc. of KDD. pp. 917–925 (2017)
19. Li, L., Jing, H., Tong, H., Yang, J., He, Q., Chen, B.C.: Nemo: Next career move prediction with contextual embedding. In: Proc. of WWW Companion. pp. 505–513 (2017)
20. Liu, L., Shang, J., Ren, X., Xu, F.F., Gui, H., Peng, J., Han, J.: Empower sequence labeling with task-aware neural language model. In: Proc. of AAAI (2018)
21. Liu, Y., Zhang, L., Nie, L., Yan, Y., Rosenblum, D.S.: Fortune teller: predicting your career path. In: Proc. of AAAI (2016)
22. Malinowski, J., Keim, T., Wendt, O., Weitzel, T.: Matching people and jobs: A bilateral recommendation approach. In: Proc. of HICSS. vol. 6, pp. 137c–137c (2006)
23. Martin, K., et al: The chemdner corpus of chemicals and drugs and its annotation principles. Journal of Cheminformatics **7**(S1), S2 (2015)
24. Massoni, S., Olteanu, M., Rousset, P.: Career-path analysis using optimal matching and self-organizing maps. In: Proc. of WSOM. pp. 154–162 (2009)
25. Mimno, D., McCallum, A.: Modeling career path trajectories (2008)
26. Mukund, S., Srihari, R.K.: Ne tagging for urdu based on bootstrap pos learning. In: Proc. of CLIAWS. pp. 61–69 (2009)
27. Nallapati, R., Surdeanu, M., Manning, C.: Blind domain transfer for named entity recognition using generative latent topic models. In: Proc. of NIPS Workshop. pp. 281–289 (2010)
28. Paparrizos, I., Cambazoglu, B.B., Gionis, A.: Machine learned job recommendation. In: Proc. of RecSys. pp. 325–328 (2011)
29. Pennington, J., Socher, R., Manning, C.: Glove: Global vectors for word representation. In: Proc. of EMNLP (2014)

30. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv:1802.05365 (2018)
31. Qin, C., et al.: Enhancing person-job fit for talent recruitment: An ability-aware neural network approach. In: Proc. of SIGIR (2018)
32. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I.: Language models are unsupervised multitask learners. OpenAI Blog **1**(8) (2019)
33. Rajpurkar, P., Jia, R., Liang, P.: Know what you don't know: Unanswerable questions for squad. Proc. of ACL (2018)
34. Rajpurkar, P., Zhang, J., Lopyrev, K., Liang, P.: Squad: 100,000+ questions for machine comprehension of text. Proc. of EMNLP (2016)
35. Ramshaw, L.A., Marcus, M.P.: Text chunking using transformation-based learning. In: Natural language processing using very large corpora, pp. 157–176 (1999)
36. Ratinov, L., Roth, D.: Design challenges and misconceptions in named entity recognition. In: Proc. of CoNLL. pp. 147–155 (2009)
37. Saha, S.K., Sarkar, S., Mitra, P.: Gazetteer preparation for named entity recognition in indian languages. In: Proceedings of the 6th Workshop on Asian Language Resources (2008)
38. Sang, E.F., De Meulder, F.: Introduction to the conll-2003 shared task: Language-independent named entity recognition. arXiv preprint cs/0306050 (2003)
39. Shen, D., Zhu, H., Zhu, C., Xu, T., Ma, C., Xiong, H.: A joint learning approach to intelligent job interview assessment. In: IJCAI. pp. 3542–3548 (2018)
40. Turian, J., Ratinov, L., Bengio, Y.: Word representations: a simple and general method for semi-supervised learning. In: Proc. of ACL (2010)
41. Viera, A.J., Garrett, J.M., et al.: Understanding interobserver agreement: the kappa statistic. Fam med **37**(5), 360–363 (2005)
42. Weischedel, R., et al.: Ontonotes release 5.0 ldc2013t19. Linguistic Data Consortium, Philadelphia, PA **23** (2013)
43. Xu, H., Yu, Z., Guo, B., Teng, M., Xiong, H.: Extracting job title hierarchy from career trajectories: A bayesian perspective. In: IJCAI. pp. 3599–3605 (2018)
44. Yang, Y., Zhan, D.C., Jiang, Y.: Which one will be next? an analysis of talent demission (2018)
45. Zhang, Y., Yang, C., Niu, Z.: A research of job recommendation system based on collaborative filtering. In: Proc. of ISCID. vol. 1, pp. 533–538 (2014)
46. Zhao, Y., Hryniewicki, M.K., Cheng, F., Fu, B., Zhu, X.: Employee turnover prediction with machine learning: A reliable approach. In: Proc. of IntelliSys. pp. 737–758 (2018)
47. Zhu, C., Zhu, H., Xiong, H., Ma, C., Xie, F., Ding, P., Li, P.: Person-job fit: Adapting the right talent for the right job with joint representation learning. ACM TMIS **9**(3), 12 (2018)