

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/324223036>

Matching Resumes to Jobs via Deep Siamese Network

Conference Paper · April 2018

DOI: 10.1145/3184558.3186942

CITATIONS

6

READS

1,803

2 authors:



Saket Maheshwary

International Institute of Information Technology, Hyderabad

5 PUBLICATIONS 31 CITATIONS

[SEE PROFILE](#)



Hemant Misra

Prolong Pharmaceuticals

42 PUBLICATIONS 406 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Sanguinate [View project](#)

Matching Resumes to Jobs via Deep Siamese Network

Saket Maheshwary

IIIT Hyderabad, India

saket.maheshwary@research.iiit.ac.in

Hemant Misra

IBM Research & Global Decision Management, Citi Group

hemant.misra@citi.com

ABSTRACT

In this paper we investigate the important and challenging task of recommending appropriate jobs for job seeking candidates by matching semi structured resumes of candidates to job descriptions. To perform this task, we propose to use a siamese adaptation of convolutional neural network. The proposed approach effectively captures the underlying semantics thus enabling to project similar resumes and job descriptions closer to each other, and make dissimilar resumes and job descriptions distant from each other in the semantic space. Our experimental results on a set of 1314 resumes and a set of 3809 job descriptions (5,005,026 resume-job description pairs) demonstrate that our approach is better than the current state-of-the-art approaches.

1 INTRODUCTION

Job search via online matching engines is very important and extremely beneficial for both job seekers and employers. We are interested in finding job descriptions (JDs) that are an appropriate match to a candidate's resume, where an appropriate match means that a prospective candidate would be interested in reading and applying for the retrieved job descriptions. Given a collection of semi-structured resumes R , a collection of semi-structured job descriptions J , and some known matched and unmatched resume/job description pairs $\langle r, j \rangle$, the task is to retrieve matching job descriptions for any existing or new resume r .

For a trained human, reading a resume or a job description and understanding its relevance is an easy task. In contrast, a computer system that parses such documents needs to be continuously trained and adapted to deal with the endless expressivity of human language. The existing traditional engines fail to understand the underlying semantic meanings of different resumes and have not kept pace with the recent progress in machine learning and natural language processing (NLP) techniques. These solutions are generally driven by manually engineered features and set of rules with predefined weights to keywords which lead to an inefficient and ineffective search experience for job seeking candidates. Besides, these solutions are not scalable as well.

Text based tasks rely heavily on representations that could be effectively learned. Such representations must capture the underlying semantics among textual elements, where textual elements can be a sequences of sentences, words or characters. Such effective learning model should not be affected by the variation of words in text used for expressing the same idea. This is a hard and complex

problem due to the lack of availability of annotated data, complex structure and variable length of text. Though state-of-the-art models like bag-of-words (BOW) and TF-IDF are effective in many NLP tasks, in the context of understanding the underlying semantics they prove to be ineffective because of their inherent term specificity [5]. Recently proposed models such as Word2vec and GloVe are less effective at document level [3].

In this paper, we used an intuitive approach to successfully handle the problem of matching appropriate jobs for resumes by learning a similarity metric between the JDs and the resumes. Siamese networks [1, 2] possess the capability of learning a similarity metric from the available training records. Our proposed network contains a siamese adaptation of convolutional neural networks (CNN) with a contrastive loss energy function joining the twin network at the top. These twin networks share the weights with each other (parameter sharing). The main contributions of our paper are as follows: (1) We investigate the application of deep siamese network for the task of matching resumes with appropriate JDs. (2) Our approach is universal and robust as it effectively handles JDs from variety of domains. (3) Our model beats state-of-the-art methods by a large margin.

2 PROPOSED APPROACH

Our deep siamese network consists of a pair of identical CNN that contains repeating convolution, max pooling and leaky rectified linear unit layers with a fully connected layer on the top. As expected, any part of text in resumes and job descriptions can influence the semantics of a word. In order to effectively capture such an influence, we want our network to see the entire input at once – here CNN comes to our rescue. Our use of CNN can be contrasted with the possible use of Long Short-Term Memory (LSTM) networks for the same task. LSTM networks read their input from left to right (or right to left) but could be ineffective when the context at the end of a sentence influence the thoughts about its beginning. Though we can overcome this drawback by running two LSTMs, one left to right and the other right to left, and concatenating their outputs. However this approach is inefficient as it doubles our computational load. CNN on the other hand, grows a larger receptive field as we stack more and more layers. It does give us the desired effect in a controllable fashion and with a low computational cost. Besides, CNN helps to give a non linear projection to the resumes and JD in the semantic space. The semantic vectors yielded from CNN are connected to a layer measuring similarity between resumes and JD. The contrastive loss function combines the measured distance and the label. The gradient of the loss function with respect to the weights and biases shared by the sub-networks is computed using back-propagation.

Parameter sharing between twin networks is a fundamental property of siamese network. Training the network with a shared set of parameters not only reduces the number of parameters (thus,

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3186942>

saving a lot of computations) but also ensures consistency of the representation of resumes and job descriptions in semantic space. The shared parameters of the network are learnt with the aim to minimize the semantic distance between the resumes and the relevant JDs and maximize the semantic distance between the resumes and irrelevant JDs. The document embedding of resumes and JDs (each of 200 dimension), generated using Doc2Vec [3] are given as inputs to our network.

3 EXPERIMENTS

Baseline Methods: As baselines, we experiment with six commonly used representations. (1) Word n-grams and (2) TF-IDF: Baseline methods to compare with due to their effectiveness in variety of NLP tasks. (3) BOW: The word frequency from the training set (resumes and job descriptions) is selected and the count of each word is used as features. (4) Bag-of-Means: The average word2vec embedding of the training data is used as a feature set. (5) Doc2Vec: An unsupervised algorithm [3] that learns fixed-length feature representations from variable-length pieces of texts. (6) CNN: Recently used method [4, 6], showing significant performance improvement over that of other traditional methods for resume job matching task. We used cosine similarity as final step with all these baseline models in order to find the similarity of $\langle r, j \rangle$ pair. The value of threshold for cosine similarity is determined by grid search on values $\{0.1, 0.15, 0.2, \dots, 0.80\}$.

Dataset and Experimental Settings: The dataset consists of 1314 resumes which came in as a part of summer research intern application at IBM Research Labs and a set of 3809 JDs from various domains were crawled from a popular India based job portal, generating a corpus of 5,005,026 distinct $\langle r, j \rangle$ pairs with a label as either 0 (Match) or 1 (Not Match). We annotated the corpus in semi supervised fashion. At first, few resume-JD pairs were manually annotated by interns. An exercise was carried out where 40 interns who were part of the research internship program were asked to annotate 15 distinct JDs each. Then we (the authors) manually annotated 100 JDs, 50 out of these were a match (class label 1) and the other 50 were a non-match (class label 0) to all these 40 interns. Thus in total we now had 115 JDs annotated for each intern. These 40 interns were from engineering background pursuing either bachelors, masters or doctorate degrees. Besides, their expertise in diverse technology domains makes our model robust and effective. Rest of the pairs were labelled via label propagation. We split 40 interns into two groups of 30 and 10 respectively. For all the experiments the 30 interns set was taken as training set. Testing was carried out on the group of 10 interns (test set A) and also on the remaining corpus that was annotated using label propagation (test set B). Note that the 30 interns used for training were never present in the test set. Each resume and JD was pre-processed by lower-casing, stemming and special character removal. We performed training in batches of size 128. The length of semantic vector is 256, depth of CNN is 4, kernel width of max pooling is 100 and that of convolution is 10. The learning rate was set to 0.01. Adam optimization method is used to update the parameters of the sub-networks.

Results and Analysis: Table 1-3 shows the results of various methods on the task of matching resumes to jobs. We compare our approach with 6 baseline methods. It is evident from tables 1-3 that the performance of our proposed siamese adaptation of CNN

(parameters are sharing is done) is significantly better than that of the simple CNN model (no parameter sharing) is used. Among baselines, TF-IDF is better than BOW and Word n-gram.

Table 1: Accuracy comparison with baseline models

Method	Accuracy on test set A	Accuracy on test set B
BOW	56.92	57.06
Word n-gram	58.24	59.01
TF-IDF	60.46	61.23
Bag of Means	62.34	63.92
Doc2Vec	65.24	67.24
CNN	73.31	74.68
Our Approach	81.34	83.19

Table 2: Comparisons with baseline models on test set A

Method	Class Label	Precision	Recall	f1-score
BOW	0	0.62	0.63	0.62
	1	0.54	0.54	0.54
Word n-gram	0	0.65	0.66	0.65
	1	0.54	0.56	0.56
TF-IDF	0	0.70	0.74	0.72
	1	0.54	0.58	0.56
Bag of Means	0	0.70	0.74	0.73
	1	0.60	0.62	0.62
Doc2Vec	0	0.78	0.78	0.78
	1	0.60	0.64	0.64
CNN	0	0.80	0.82	0.82
	1	0.70	0.72	0.70
Our Approach	0	0.92	0.92	0.92
	1	0.75	0.78	0.78

Table 3: Comparisons with baseline models on test set B

Method	Class Label	Precision	Recall	f1-score
BOW	0	0.62	0.63	0.62
	1	0.55	0.56	0.56
Word n-gram	0	0.65	0.66	0.65
	1	0.54	0.56	0.56
TF-IDF	0	0.70	0.74	0.72
	1	0.59	0.60	0.60
Bag of Means	0	0.70	0.74	0.73
	1	0.62	0.64	0.63
Doc2Vec	0	0.78	0.78	0.78
	1	0.62	0.66	0.64
CNN	0	0.82	0.82	0.82
	1	0.70	0.72	0.72
Our Approach	0	0.92	0.94	0.92
	1	0.79	0.80	0.80

4 CONCLUSION

In this paper, we propose a siamese adaptation of CNN to tackle the task of recommending appropriate jobs to job seeking candidates. To evaluate the approach, we curated a dataset of 5,005,026 resume-job description pairs and annotated it in a semi-supervised manner. The proposed approach effectively captures the underlying semantics and significantly outperforms the existing methods.

REFERENCES

- [1] Jane Bromley, James W. Bentz, Léon Bottou, Isabelle Guyon, Yann LeCun, Cliff Moore, Eduard Säckinger, and Roopak Shah. 1993. Signature Verification Using A "Siamese" Time Delay Neural Network. *IJPRAI* 7 (1993), 669–688.
- [2] Sumit Chopra, Raia Hadsell, and Yann LeCun. 2005. Learning a similarity metric discriminatively, with application to face verification. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* 1 (2005), 539–546 vol. 1.
- [3] Quoc V. Le and Tomas Mikolov. 2014. Distributed Representations of Sentences and Documents. In *ICML*.
- [4] Yiu Lin, Hang Lei, Prince Clement Addo, and Xiaoyu Li. 2016. Machine Learned Resume-Job Matching Solution. *CoRR* abs/1607.07657 (2016).
- [5] Rada Mihalcea, Courtney Corley, and Carlo Strapparava. 2006. Corpus-based and Knowledge-based Measures of Text Semantic Similarity. In *AAAI*.
- [6] Luiza Sayfullina, Eric Malmi, Yiping Liao, and Alex Jung. 2017. Domain Adaptation for Resume Classification Using Convolutional Neural Networks. *CoRR* abs/1707.05576 (2017).