Job Classification Based on LinkedIn Summaries CS 224D

Eric Boucher, Clement Renault
Stanford University

Abstract

Using data provided by HiQ Lab - a company specialized in data analysis for Human Resources - we want to predict someone's job category based on his job summary. Job summaries are created by users to describe their skills and tasks. Our goal is to use NLP to extract information from these free-form text fields and predict the occupation of the user.

1 Introduction

Despite the existence of some classifications, there is no clear normalization of job-titles and people are often using their own words to shortly describe what they are doing. Summaries provide an interesting platform for testing text understanding and classification techniques.

2 Task Definition

As suggested earlier, job-titles appear in many variations. Thus, in our project, our first task has been to classify a large number of job titles (around 20,000) into the 133 level-3 categories from the ISCO 2008 Standard Classification. From there, we trained a one-hidden-layer neural network able to classify the associated summaries. Ideally, we would like our neural network is able to understand subtleties of language such as power relationships in order to understand hierarchical positions as well.

3 Background and Related Work

We could not find previous paper which treated exactly of the same subject and tried to classify summaries. However, numerous paper exist on text understanding and text classification using various machine-learning techniques. Based on what we were trying to achieve, we focused on papers regarding text understanding with the help of neural networks. We thought that using a Recurrent Neural Network would make sense. Other interesting techniques include extended RNN by Garen Arevian and Recurrent Convolutional Neural Networks who have shown interesting results. This last model developed by Siwei Lai, Liheng Xu, Kang Liu and Jun Zhao captures contextual information with the recurrent structure and constructs the representation of text using a convolutional neural network. On the downside, text understanding and text classification is most commonly used on texts composed of complete and grammatically correct sentences such as news articles or short stories. Summaries are structurally different as people often use bullet point and abbreviations to make their points more effectively.

4 Approach

4.1 Data

We will use a dataset provided by the company HiQ Labs (`http://www.hiqlabs.com`). This company specializes in HR analytics and has compiled a large dataset of professional profiles from various public online sources. This dataset contains LinkedIn summaries of approximately 25k people as well as the job titles gathered by the company about these profiles. An example of entry is:

```
Senior Vice President, Financial Advisor, Portfolio Management
Director
"As a licensed investment professional for more than 16 years,
David leads a team that implements risk controlled wealth
management strategies for high net worth clients.  Specialities:
Wealth Management, Portfolio Management, Certified Financial
Planner, Retirement Planning, Asset Allocation, Investment
Strategies, Estate Planning".
```

4.2 Label Acquisition

As suggested above, one of the big challenges we encountered was the fact that job title are also filled by users. This means that many titles will be slightly different although the jobs are very similar. Our first task was to regroup job titles in a smaller number of categories. To do so we considered many approaches. First we thought about clustering those job titles using: vectors indicating which words are in each title (binary vectors) and then using K-mean clustering on these vectors. Although this last approach gave interesting results, we thought that this classification was very important for the next steps of our project and decided to go a little further. In the end we used vectors obtained as mean of word vectors of the words in each title (Google Trained wordToVec - dim 300) K-mean clustering Closest occupation based on ISCO 2008 Standard (level 3 i.e. 133 categories) also taken as average word vectors.

4.2.1 Results

Table with interesting results showing off the power of word vectors

Input Job Title	Output Category
Senior Analyst Pricing Profitability and Margin Analysis	Finance Professionals
Commercial Loan Officer	Financial and Mathematical Associate Professionals
Youth and Emerging Adult Project Coordinator	Legal Social Cultural and Related Associate Professionals
Architect and Engineering Lead Data Center and Cloud Networking	Science and Engineering Associate Professionals

Table 1: Sample Clustering Results

The distribution of the different classes is given in Figure 1.

4.3 Recurrent Neural Network

We have tested a good variety of models and methods on the data: at first, we tried to use a nave Recurrent Neural Network on the summaries and read word by word to predict the outcome category. At first, we decided to train word vectors on the go as we feared that the structural differences

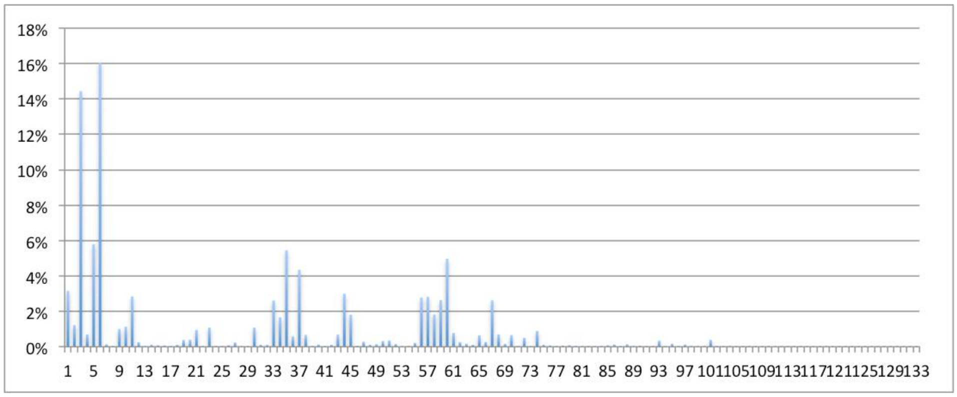


Figure 1: Distribution of the different classes.

between the type of texts used to train commonly available word vectors and summaries would deter the results. In front of poor results, we then tried to use pre-trained word-vectors (Google). Unfortunately, results were still doubtful.

$$
\begin{aligned}
h^{(t)} &= \operatorname{sigmoid}\left(Hh^{(t-1)} + Lx^{(t)}\right) \\
\hat{y}^{(t)} &= \operatorname{softmax}\left(Uh^{(t)}\right)
\end{aligned}
$$

Figure 2: Equations of the Recurrent Neural Network

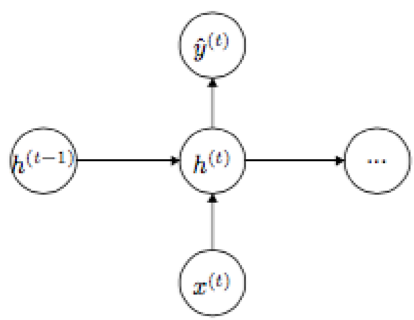


Figure 3: Sketch of the RNN at a Single Timestep

4.4 One Hidden Layer Neural Network

Eventually, we tried a very simple approach and settled on a basic one-hidden neural network taking as entry the average of the word vectors for the summary. Surprisingly (or not), we got much improved results.

5 Experiment and Results

5.1 Data and Job Classification

As mentioned earlier, classifying the labels before training the model was a much longer task than previously expected. However, we are quite happy with the results and believe that this part would

$$h = \tanh(Wx^{(t)} + b_1)$$
$$\hat{y} = \text{softmax}(Uh + b_2)$$

Figure 4: Equations of the One-Hidden-Layer Neural Network

be worth spending more time on as it could be very useful in itself to help people classify job-titles into less categories. The closest occupation method is very interesting in itself. As an exemple, one could get a global overview of occupations in a country using public LinkedIn data for instance. This would be very useful to get a live overview of social strata.

5.1.1 Improvements

Looking at the histogram of the 133 classes, we thought that it might be interesting to reduce the number of classes and regroup some categories to have more homogeneously distributed labels.

5.2 Recurrent Neural Network

The results of the RNN were inconclusive. We believe that the average length of the summaries created too much noise for the neural network to interpret. Implementing GRU or SLTM model could improve this model significantly. Unfortunately, we did not have enough time to implement these techniques.

5.3 One Hidden Layer Neural Network

We found that a middleDim of around 140 yielded good results. Our best model was able to achieve 31.7 percent on the training set and 29.4 percent on the dev set for 133 classes. Our baseline was to classify everything in the highest category, which represented 16 percent of our dataset.

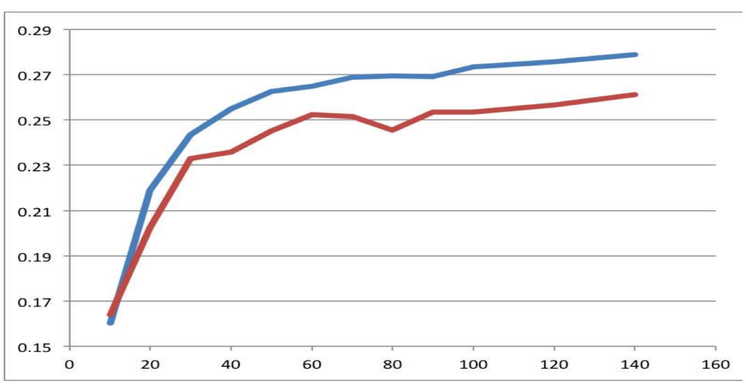
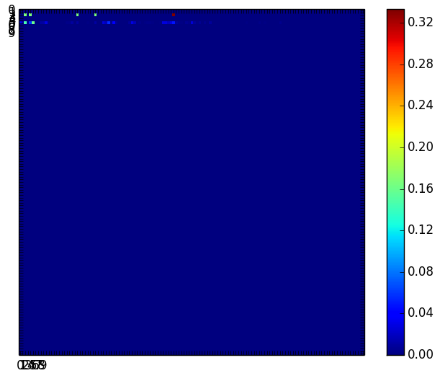


Figure 5: Training Accuracy and Test Accuracy vs middleDim

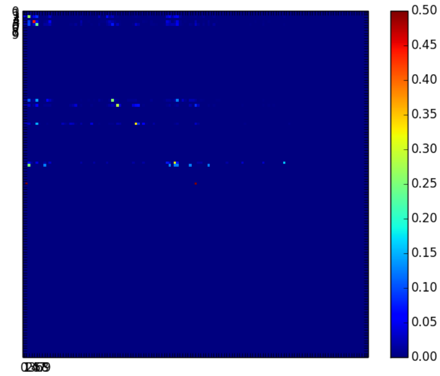
Confusion matrices for different models are plotted in Figure 6. One can clearly see that the model has learnt some pattern as a diagonal starts to appear when middleDim increases. We can also see that the model does not learn very well the last classes which is explained by the very small number of samples that classes above 77 regroup (see Figure 1).

5.4 Final Results and Discussion

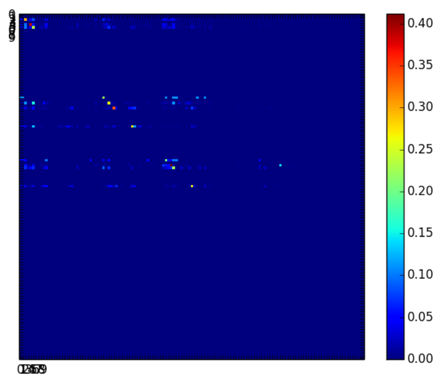
In the end, we were able to achieve 31.7 percent accuracy on 133 classes with a baseline at 16 percent. However, accuracy is calculated binarily (in or out) and it would be interesting to see how the model is performing based on a more flexible calculation. For example, we could take the cosine distance between the predicted and target categories to take into that some job categories might actually be quite close to one-another.



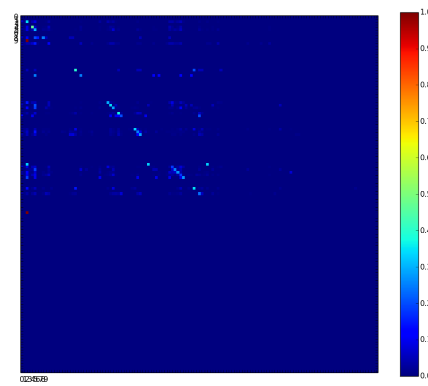
(a) Training set confusion matrix for middleDim = 10



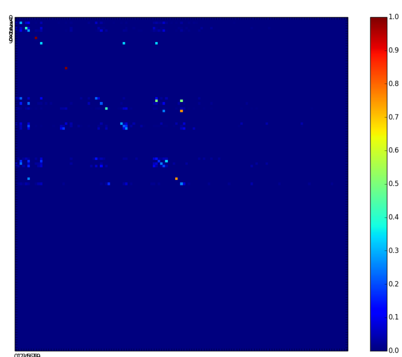
(b) Training set confusion matrix for middleDim = 60



(c) Training set confusion matrix for middleDim = 120



(d) Training set confusion matrix for best model



(e) Dev set confusion matrix for best model

Figure 6: Confusion Matrices Evolution

6 Conclusion

The different algorithms implemented gave mixed results. In particular, We believe that our implementation does not give RNN a fair chance and that changes such as implementation of GRU or LSTM would improve this model significantly. This project was the occasion to witness by ourselves the power of well trained word vectors. We were happily surprised to obtain such interesting results on the K-means of word vectors of the title, which proves the power of well trained word vectors. Adding the tf-idf as a weight could help us improve the classification process.

This model could be interesting for two main applications:

- Associate a description and a job-title
- Describe what you like and find a job. Indeed, students are often able to list activities that they would like to do in their job, but are rarely informed enough to know what job title would suit their needs. Similarly, job seekers might know what skills and activities from their current job they would like to keep using but not what other position they could use them for.

References

[1] Garen Arevian (2007) *Recurrent Neural Networks for Robust Real-World Text Classification.* IEEE/WIC/ACM International Conference on Web Intelligence , Issue Date: 2-5 Nov. 2007

[2] Siwei Lai, Liheng Xu, Kang Liu, Jun Zhao (2013) *Recurrent Convolutional Neural Networks for Text Classification.* Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

[3] Ilya Sutskever (2013) *Training Recurrent Neural Networks.* PHD Thesis for the University of Toronto

[4] Richard Socher (2015) *CS224D - Deep Learning for Natural Language Processing - Stanford University*