Understanding Job-Skill Relationships using Big Data and Neural Networks

Abhinav Maurya
Carnegie Mellon University
Pittsburgh, PA – 15213
ahmaurya@cmu.edu

ABSTRACT

Nonlinear vector space embeddings have shown great promise in many applications such as similarity search, analogy mappings, dataset visualization, etc. In this paper, we propose a simple distribution-to-distribution regression model based on neural networks which provides (i) *job2vec*: interpretable embeddings of jobs in terms of associated skills, (ii) *skill2vec*: converse embeddings of skills in terms of the jobs that require them, and (iii) *SkillRank*: a mechanism for ranking skills associated with each job that can be recommended to members aspiring for that job. Due to the simplicity of our model, it has no hyperparameters that need tuning except for the number of stochastic gradient descent iterations which is easily determined using the early stopping criterion. We dicsuss preliminary results and insights from our model using data from a major professional social network.

Keywords

Machine Learning, Neural Networks, Labor Market, Job-Skill Ranking, Vector Space Embeddings

1. INTRODUCTION

The exchange of skilled services forms the driving force of modern economies. We spend the formative years of our lives acquiring skills that prepare us for our desired careers, and continue to acquire new skills throughout our professional careers. However, the decision of skill acquisition for a desired career is often fraught with uncertainty due to lack of information. It is therefore important to study the association between jobs and the skills most commonly associated with these jobs.

The study of job-skill association has been challenging for two reasons: (i) lack of large-scale data on jobs held and skills acquired by members of the workforce, and (ii) the multiple attribution problem which prevents us from associating the acquisition of a particular skill with a specific job held by a person. With the advent of online professional social networks, large-scale data is now available to study the relationship between jobs and skills. In this paper, we tackle the second problem of multiple attribution in job-skill association by exploiting naturally occurring variation in the sets of jobs and skills across large-scale labor market data.

Our proposed model is inspired by topic modeling where latent topics occur throughout the documents of a corpus in varying proportions, which enables their identification to characterize the text corpus. Similarly, in our model, variation in the co-occurrence of jobs and skills can be used to disentangle which skills are associated with which jobs. However, unlike unsupervised topic models such as LDA [1], we cast our problem of identifying job-skill associations as a supervised distribution-to-distribution regression, where the input is the empirical distribution over job titles and the output is the corresponding empirical distribution over skills for that person. Moreover, unlike unsupervised topic models such as LDA and supervised variants such as [4], our neural network model allows greater flexibility in job-skill association formulation. Our results show that most job-skill associations are negative, which is intuitive since only a small subset of all job-skill pairs are useful in the real world. Topic models often allow only a non-negative modeling of association which would be restrictive in this application.

In the same vein as [5], our method provides embeddings for jobs and skills, known as *job2vec* and *skill2vec* respectively. However, the model used in [5] is quite different from ours, since it is geared towards identifying word synonyms and analogies. A more complete version of this work will seek to distinguish between the predictive and semantic power of embeddings obtained from the method proposed here and an adaptation of the *word2vec* model investiagted by [5].

[3] matches members and jobs to make skill recommendations based on the demand of skills in jobs. However, it does not consider a user's choice of career in making skill recommendations. In this paper, we identify specific skills for each job that a user can acquire if she aspires to get the job.

Our work bears similarity to Canonical Correlation Analysis [2] in that it would be possible to map job and skill vectors of a user such that the distance between the mappings is minimized in the projected space. However, CCA would lose the benefit of using an asymmetric loss such as KL Divergence which allows us to model only the present skills.

2. DATA

Our dataset consists of online profiles crawled from a major worldwide professional social network. Each profile is annotated with a city and an industry. We focus on a subset of the entire data from a diverse set of cities and industries

Bloomberg Data for Good Exchange Conference.
16-Sep-2018, New York City, NY, USA.

Industry
Computer Software
Hospital and Health Care
Financial Services
Real Estate
Research
Marketing and Advertising

Cities
San Francisco Bay Area
New York Metropolitan Area
Seattle
Pittsburgh
Mumbai
New Delhi

Table 1: Industries and cities included in the evaluation.

listed in table 1 for the purpose of our preliminary analysis. From each profile, we extract and use the set of job titles held by the person as well as the set of skills the person has listed on the profile. Summary statistics about the used data are provided in table 2. Datapoints are divided into 90% training, 5% validation, and 5% test data. We identified 12,652 unique job titles and 15,185 unique skills from our data. Most frequent job titles and skills are listed in table 3.

Statistic	Value
Number of datapoints	157252
Number of train datapoints	125801
Number of validation datapoints	15725
Number of test datapoints	15726
Number of unique skills	15185
Number of unique titles	12652

Table 2: Dataset Summary Statistics

3. METHODOLOGY

Since we have over 12 thousand unique job titles and over 15 thousand unique skills in our data, the challenge lies in effectively formulating a neural network model that can learn to perform a high-dimensional distribution-to-distribution regression by ingesting vast amount of examples available in our data. Our simple neural network model is designed to perform this high-dimensional map from a distribution over positions to a distribution over skills.

The architecture of our *SkillRank* neural network model is shown in figure 1. The model maps an empirical distribution over job titles to an empirical distribution over skills: $f : \mathcal{J} \to \mathcal{S}$. Given an empirical distribution over jobs $\mathbf{j} \in \mathcal{J}$, the model first performs a linear transformation $\phi \cdot \mathbf{j}$, where $\phi \in \mathbb{R}^{S\text{x}J}$, S is the number of skills and J is the number of job titles. The result of the linear transformation is a vector of unnormalized real-valued skill scores. The model then applies a softmax transformation to normalize the skill scores and output a distribution over skills. The model is trained by minimizing the KL Divergence loss between the groundtruth empirical distribution over skills and the predicted distribution over skills.

Since the KL Divergence is asymmetric, it does not penalize any absent skills in the groundtruth empirical distribution of skills. This is a major reason why we choose the KL Divergence loss in our model optimization. In the case of a single categorical output, the KL Divergence loss simplifies to the popular cross-entropy loss widely used in neural network classifiers.

Most Frequent Positions
Software Engineer
Intern
Research Assistant
Director
Manager
Vice President
Consultant
Account Executive
Project Manager
Owner

Most Frequent Skills
Microsoft Office
Management
Microsoft Excel
Marketing
Customer Service
Research
Leadership
Social Media
PowerPoint
Sales

Table 3: Most frequent job titles and skills in the data.

The model was implemented in PyTorch[1]. It is notable that the number of training epochs is the only parameter we need to tune in our model, and it needs a single training run to find the right number of iterations to avoid overfitting. There are no other model-specific parameters. This is in contrast to hyperparameter-heavy models that may need expensive grid search, random search, or Bayesian optimization to tune the hyperparameters.

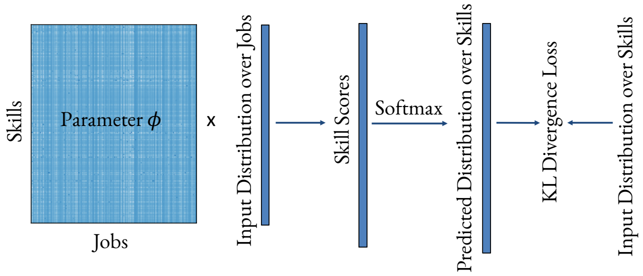

Skills
Parameter ϕ
Jobs
x
Input Distribution over Jobs
Skill Scores
Softmax
Predicted Distribution over Skills
KL Divergence Loss
Input Distribution over Skills

Figure 1: Our Neural Distribution-to-Distribution Regression Model.

3.1 Embeddings

We recall that $\phi \in \mathbb{R}^{S\text{x}J}$, S is the number of skills and J is the number of job titles. Thus, the rows of ϕ correspond to skills and columns to jobs. The matrix ϕ encodes the associations between jobs and skills. A positive value indicates that a skill is desired by a job, and a negative value indicates that it is highly unlikely for the corresponding position-skill combination to occur in the data.

[1] http://pytorch.org/

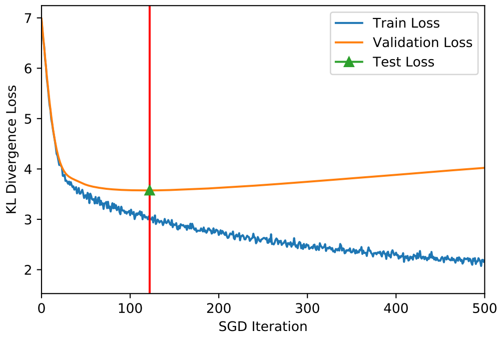


Figure 2: Training, Validation, and Test Losses.

job2vec: To calculate the embedding of a job, we pass its corresponding column through a softmax transformation. The softmax transformation of a vector exponentiates its entries and normalizes the exponentiated entries so that they sum to 1. Thus, each job's embedding indicates its distribution over skills. The exponentiation is also useful to get meaningful job embedding vectors, since it diminishes the negative values drastically and therefore represents a job using the skills its positively associated with.

skill2vec: Similar to job2vec, skills can also be embedded to get a useful representation. Skills correspond to rows of the ϕ matrix. Therefore, their embeddings are obtained by applying the softmax transformation to each row of ϕ.

4. PRELIMINARY RESULTS

In our implementation, we used minibatch gradient descent with a minibatch size of 1000. The training and validation loss curves during the optimization procedure and the final test loss are shown in figure 2. Since the validation loss starts rising after epoch 122 (shown with a vertical red line in the figure), we choose the model obtained at the end of 122 iterations as our final model as per validation-based early stopping criterion to prevent overfitting.

Figure 3 shows the histogram of values from the parameter matrix ϕ. Most values in the matrix are negative. This makes intuitive sense because most job-skill pairs are not naturally associated with each other. Only 1.13% values in the matrix are positive indicating only a small subset of job-skills are positively associated with each other.

In table 4, we show the job-skill pairs which have the highest corresponding ϕ values. It is interesting to note the job-skill associations are non-trivial i.e. it's not the most common skills such as "Microsoft Office" that are repeatedly recommended for every position. Instead, each skill recommended for the job is precisely relevant for the job.

Figure 6 shows a 2-dimensional t-SNE visualization of job embeddings. Job titles that are tightly clustered in the visualization are often related to each other. For example, "postdoctoral scholar" and "postdoctoral fellow" are situated very close to each. A number of medical industry jobs such as "medical director," "consultant physiotherapist," "clinical trial manager," "pharmacy manager," etc. are also located closely in the visualization. Figure 7 shows a corresponding 2-dimensional t-SNE visualization for skill embeddings. Not surprisingly, related skills are often located close to each other in the embedding vector space. For example, "commercial real estate," "investment properties," "tenant retention," etc. are located close to each other.

From figure 4, we see the parameter matrix ϕ visualized before and after spectral coclustering. The existence of bands in figure 4b indicates that there is the possibility of coclustering or mixed-membership stochastic blockmodel structure on parameter ϕ. This can help in identification of closely-related job and skill bundles.

Figure 5 shows two histograms. The first one shows the number of skills on the X-axis and the number of positions that were positively associated with a certain number of skills on the Y-axis. Here, positive association means a positive value for the corresponding ϕ entry. Similarly, the second figure shows the number of jobs versus the number of skills positively associated with a particular number of jobs. The figure illustrates that there is considerable diversity and specificity in job-skill associations, since most jobs are associated with a moderate number of skills and most skills are associated with a moderate number of jobs, as indicated by the rapid tapering of the histograms to the right of the X-axis.

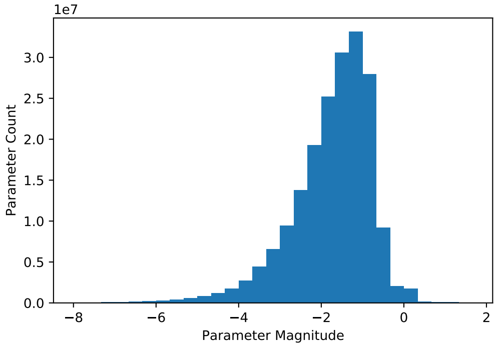
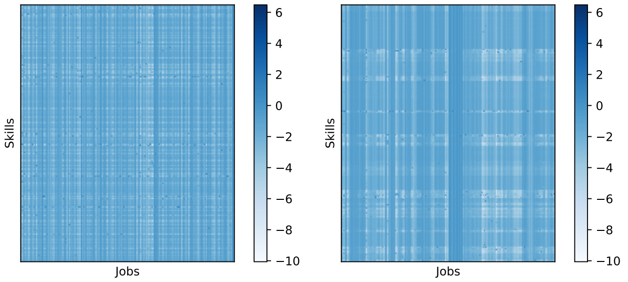


Figure 3: Histogram of values in ϕ.



(a) Parameter ϕ

(b) After Spectral Coclustering

Figure 4: Fig (a): Parameter matrix ϕ visualized. Fig (b): Parameter matrix ϕ visualized after spectral coclustering with 100 coclusters.

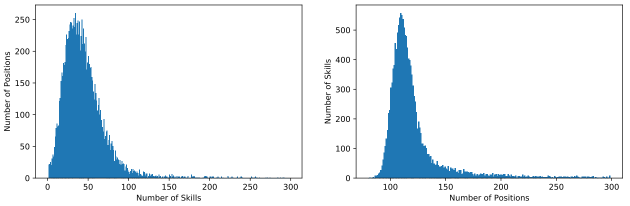
5. CONCLUSIONS

In this paper, we presented a simple neural approach to identifying job-skill associations from large-scale labor market data. The model helps us rank skills by the order of

Job	Skill
Associate Media Director	Mobile Advertising
Account Director	Integrated Marketing
Management Supervisor	Interactive Advertising
Clinical Dietitian	Medical Nutrition Therapy
Product Management	Go-to market Strategy
Senior Art Director	Corporate Identity
Chief Compliance Officer	Securities Regulation
Speech Language Pathologist	Articulation
Senior Art Director	Typography
Account Director	Creative Direction
Sr. Product Manager	Go-to market Strategy
Art Director	Logo Design
Freelance Art Director	Typography
Associate Creative Director	Concept Development
Speech Language Pathologist	Speech Therapy
iOS Developer	Swift
Copywriter	Creative Writing
Management Supervisor	Relationship Marketing
Law Clerk	Legal Research
Enterprise Account Manager	Solution Selling
iOS Developer	Xcode
Sr. Product Manager	Product Launch
Social Work Intern	Crisis Intervention
Group Creative Director	Art Direction
Financial Advisor	Retirement Planning
Media Supervisor	Mobile Marketing
Law Clerk	Legal Writing
Software Development Manager	Scalability
Associate Creative Director	Interactive Advertising
Mortgage Consultant	FHA
Account Director	Brand Development
Graphic Designer	Corel Draw
Consultant Physiotherapist	Physical Therapy
Editorial Intern	Journalism
Creative Director	Logo Design
Account Director	Customer Insight
Graphic Designer	InDesign
Editorial Intern	Proofreading
QA Engineer	Selenium

Table 4: Examples of most positively associated job-skill pairs.

their importance to each position, and allows us to embed jobs and skills using *job2vec* and *skill2vec* to enable tasks such as similarity searches and ontology visualization.



(a) Histogram of number of skills per position.

(b) Histogram of number of positions per skill.

Figure 5: Fig (a): Number of skills versus number of positions positively associated with a certain number of skills. Fig (b): Number of positions versus number of skills positively associated in ϕ with a certain number of positions.

6. REFERENCES

[1] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.

[2] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

[3] A. Maurya and R. Telang. Bayesian multi-view models for member-job matching and personalized skill recommendations. In *IEEE International Conference on Big Data (Big Data), 2017*, 2017.

[4] J. D. Mcauliffe and D. M. Blei. Supervised topic models. In *Advances in neural information processing systems*, pages 121–128, 2008.

[5] T. Mikolov, K. Chen, G. Corrado, and J. Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

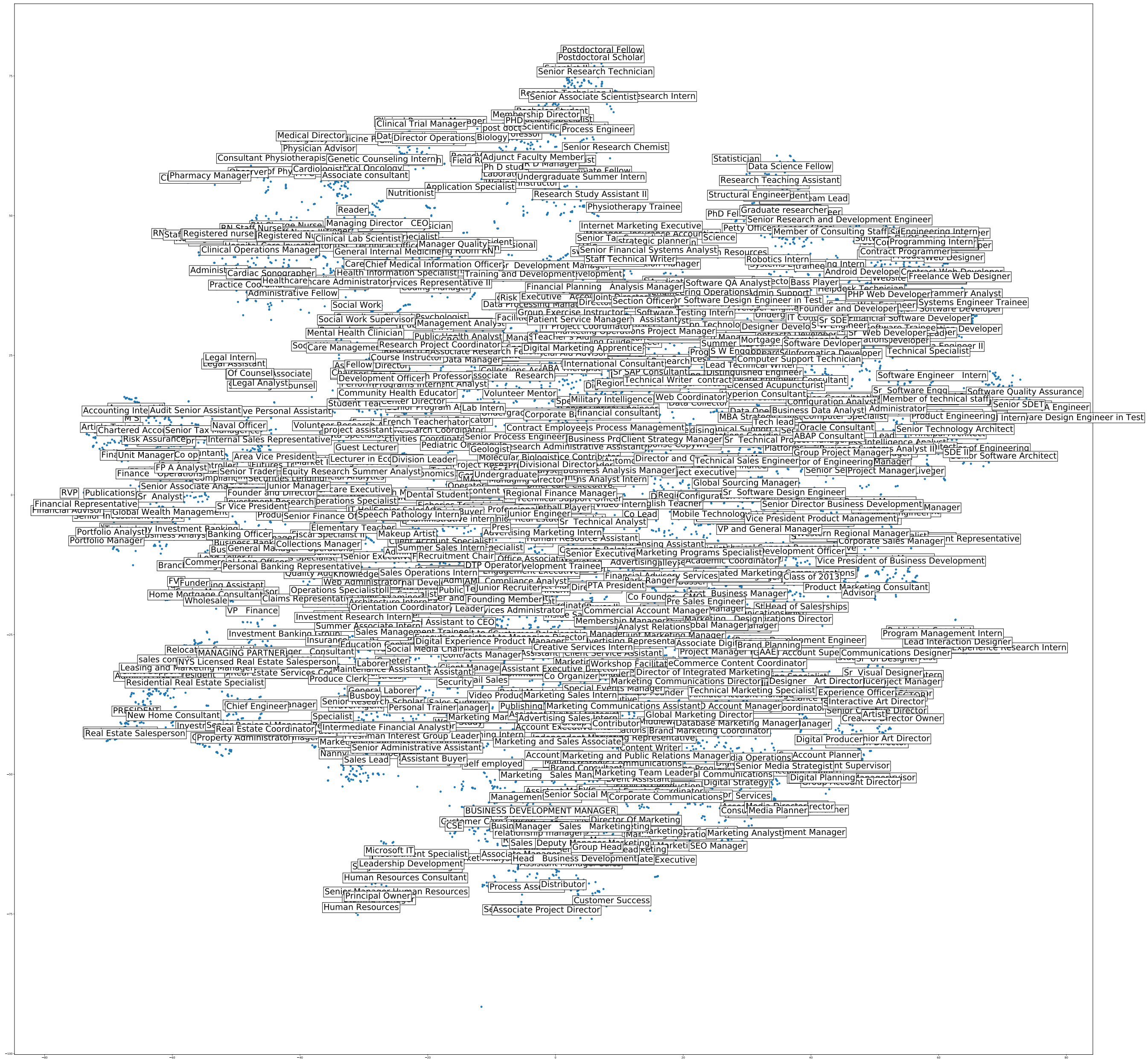


Figure 6: Job Embeddings. Please zoom in for an easier read.

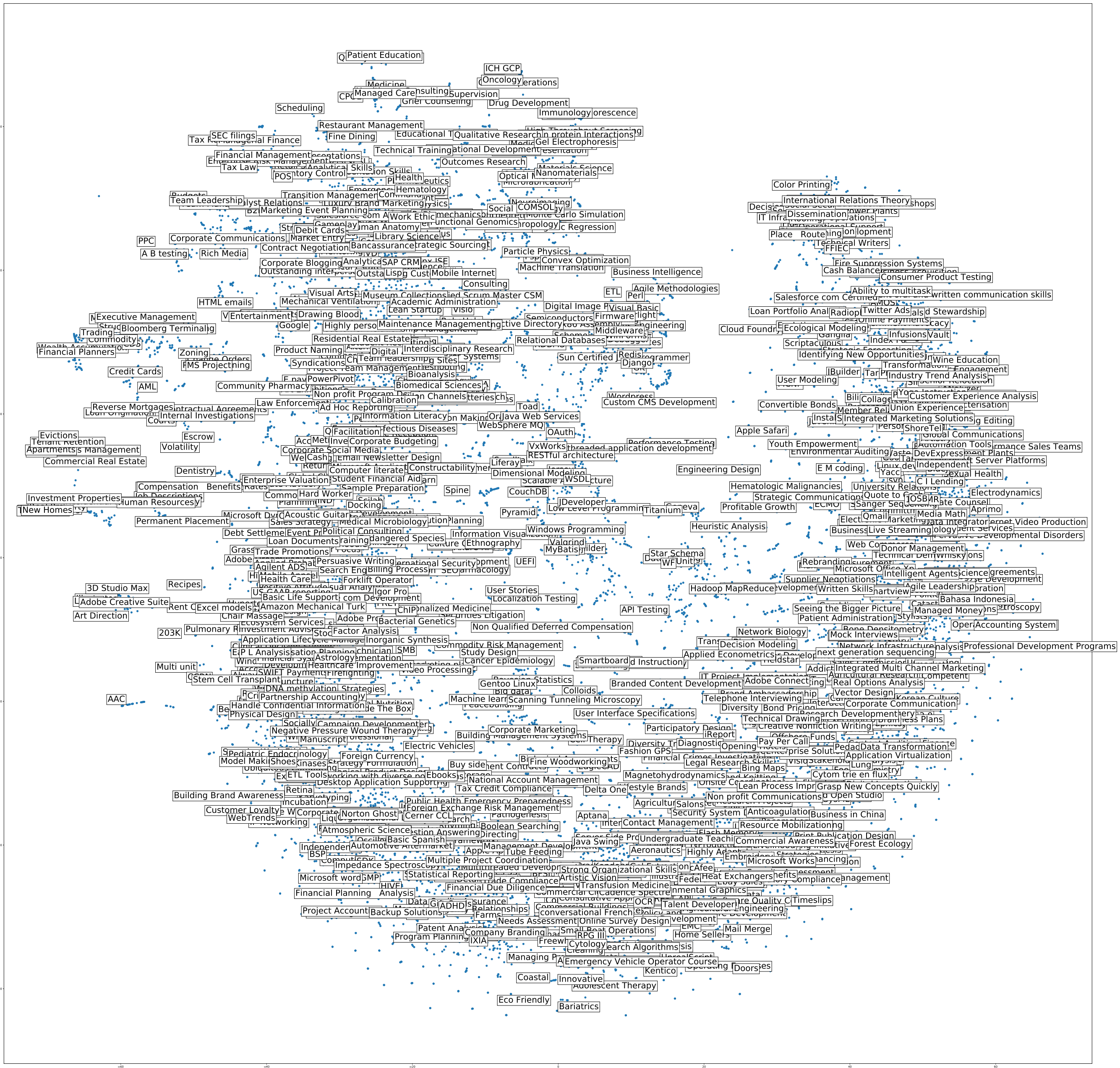


Figure 7: Skill Embeddings. Please zoom in for an easier read.