work2vec: Learning the Latent Structure of the Labor Market

Sarah Bana (Chapman University & Stanford Digital Economy Lab)

Erik Brynjolfsson

Daniel Rock

Sebastian Steffen

ASSA 2023

January 6, 2022

Ample Evidence of Occupations Changing Over Time

And Occupations Differ Across Space

- Through new job titles introduced (Lin, 2011)
- Through changes in tasks within occupation (Atalay et al., 2020)

Yet, no systematic approach to measuring occupational dynamism

Ample Evidence of Occupations Changing Over Time

And Occupations Differ Across Space

- Through new job titles introduced (Lin, 2011)
- Through changes in tasks within occupation (Atalay et al., 2020)

Yet, no systematic approach to measuring occupational dynamism

Traditional data sources in the United States have limited scope for this question

- O*NET compresses any occupation level variation, and updates only periodically
- Admin data, if it does have occupational information, omits the composition of tasks

Ample Evidence of Occupations Changing Over Time

And Occupations Differ Across Space

- Through new job titles introduced (Lin, 2011)
- Through changes in tasks within occupation (Atalay et al., 2020)

Yet, no systematic approach to measuring occupational dynamism

Traditional data sources in the United States have limited scope for this question

- O*NET compresses any occupation level variation, and updates only periodically
- Admin data, if it does have occupational information, omits the composition of tasks
  - Teller at a bank branch with an ATM might have different responsibilities than a teller at a bank branch without ATMs (Bessen, 2015)

Information on Skills and Direction of Change is Hiding in Plain Sight

In the Form of Online Job Postings

Job postings provide

- Rich text data about workers and firms
- In close to real-time
- Preserving variation

Information on Skills and Direction of Change is Hiding in Plain Sight

In the Form of Online Job Postings

Job postings provide

- Rich text data about workers and firms
- In close to real-time
- Preserving variation

We propose an approach to turn this inherently unstructured data into spatial representations

- Turning work into vectors $\rightarrow work2vec$

Information on Skills and Direction of Change is Hiding in Plain Sight

In the Form of Online Job Postings

Job postings provide

- Rich text data about workers and firms
- In close to real-time
- Preserving variation

We propose an approach to turn this inherently unstructured data into spatial representations

- Turning work into vectors $\rightarrow work2vec$
- Primarily descriptive work

How is the Labor Market Changing?

Labor market is a moving target – can we identify the direction and magnitude of change?

Outline:

- Methods: Variational autoencoder (VAE) trained on job postings from 2010-2019
- Results: Factors used to measure how occupations have changed over time
- Results: Factors used to measure how the overall job space has changed over time
- Application: A data-driven alternative to occupation classification

Methods

Many different attributes of jobs are discussed in postings $\rightarrow$ postings have many dimensions!

Methods

Many different attributes of jobs are discussed in postings $\rightarrow$ postings have many dimensions!

First Step: Turn words into vectors using a Large Language Model (LLM)

- BERT embeddings
- Produces a 512 x 768 dimensional matrix for each posting

Methods

Many different attributes of jobs are discussed in postings $\rightarrow$ postings have many dimensions!

First Step: Turn words into vectors using a Large Language Model (LLM)

- BERT embeddings
- Produces a 512 x 768 dimensional matrix for each posting

Inherently an unsupervised learning problem

Methods

Common Solution: Dimensionality reduction

- Linear factors (e.g. Principal Components Analysis (PCA))

Methods

Common Solution: Dimensionality reduction

- Linear factors (e.g. Principal Components Analysis (PCA))
- Nonlinear factors (e.g. t-distributed stochastic neighbor (t-SNE) embeddings)

However, factors derived from these types of models lack regularity

Methods

Common Solution: Dimensionality reduction

- Linear factors (e.g. Principal Components Analysis (PCA))
- Nonlinear factors (e.g. t-distributed stochastic neighbor (t-SNE) embeddings)

However, factors derived from these types of models lack regularity

A **Variational Autoencoder (VAE)** provides factors that are meaningful

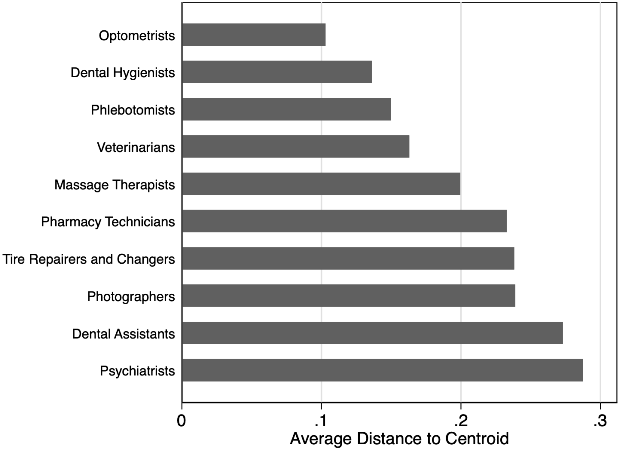
- "Training the data itself as the label"
  - Encoder: encodes the data into low-dimensional space
  - Decoder: reconstructs the original data from the latent representation
- Training data expressed in a compact way, grouping similar data together in latent space

Methods

1. Train a VAE with 30 factors
2. Micro-level: Compare movements in these factors to understand which occupations are changing most over time
   - Use Euclidean distance to look at differences between 2010 and 2019
3. Macro-level: Measure the volume of the “jobspace” over time
   - Bootstrap samples of postings to measure the volume of the convex hull

Occupation Level Results

Figure: Lowest Diversity Occupations in 2019



Occupation Level Results

Figure: Lowest Diversity Occupations in 2019

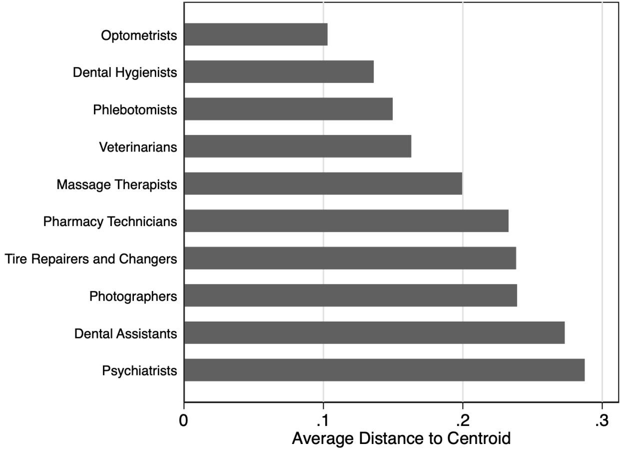
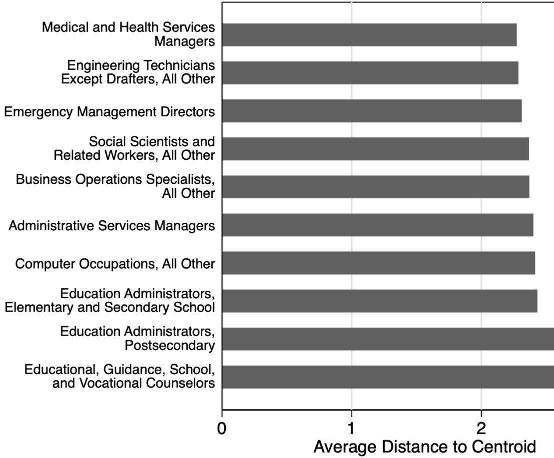
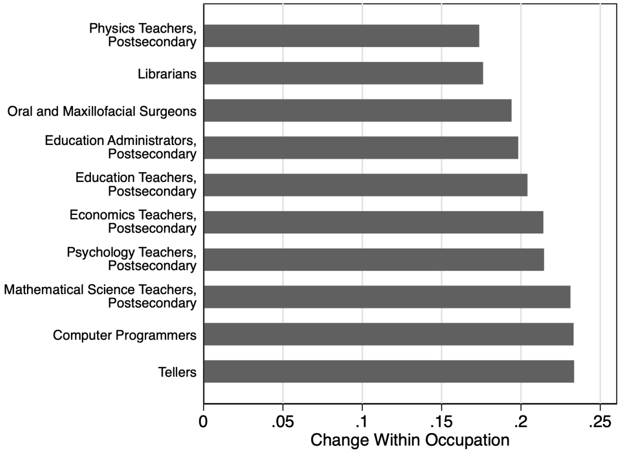


Figure: Highest Diversity Occupations in 2019



Occupation Level Results

Figure: Occupations Changing the Least



Occupation Level Results

Figure: Occupations Changing the Least

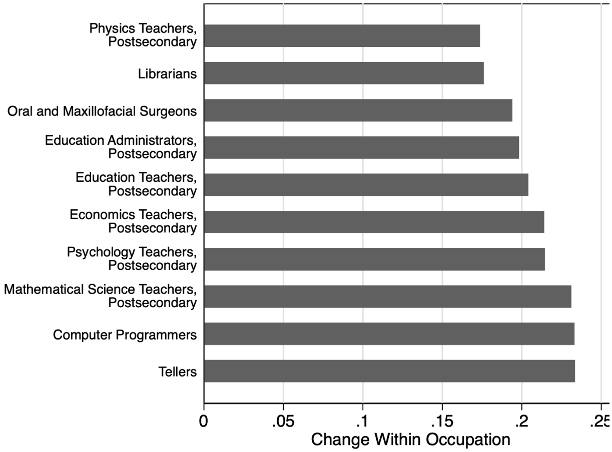
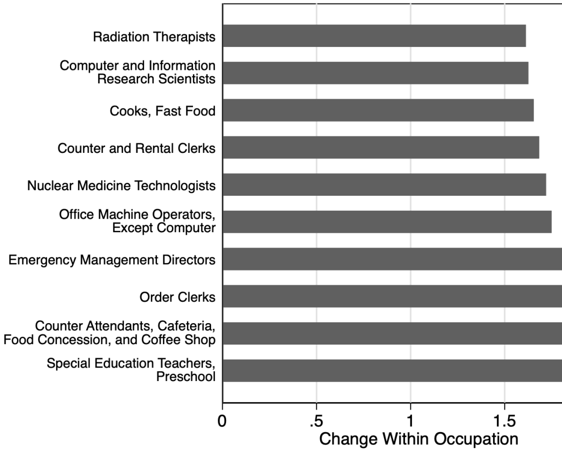


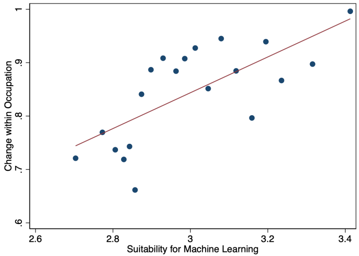
Figure: Occupations Changing the Most



How Do Changes Over Time Relate to Other Measures?

Change in Factors Between 2010 and 2019 as the Outcome

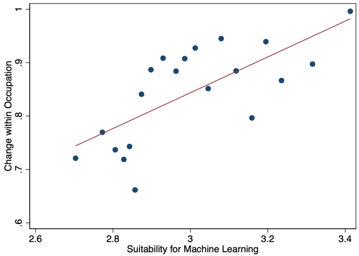
(a) SML



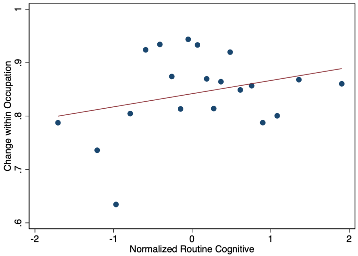
How Do Changes Over Time Relate to Other Measures?

Change in Factors Between 2010 and 2019 as the Outcome

(a) SML



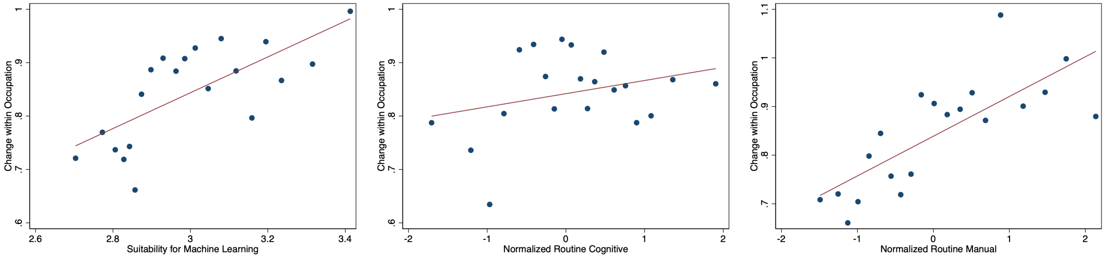
(b) Routine Cognitive



How Do Changes Over Time Relate to Other Measures?

Change in Factors Between 2010 and 2019 as the Outcome

(a) SML *(b)* Routine Cognitive *(c)* Routine Manual



Occupations changing the most were ones measured as most suitable for machine learning and having larger shares of routine manual tasks

Aggregate Labor Market

- Each posting has 30 factors
- **Convex hull**: smallest convex set that contains it
- **Volume of the convex hull**: spatial representation of the “jobspace”
  - **Challenge**: Computationally very challenging
  - **Solution**: Bootstrapping samples of postings

Aggregate Labor Market Results

Recombination of Existing Roles:

points moving around *within* the existing "jobspace"

Expansion of the "Jobspace":

points that are *outside* of the existing "jobspace"

Analogous classifications for patents in Cheng et al. (2022) – also conceptually motivated by Autor (2019)

Aggregate Labor Market Results

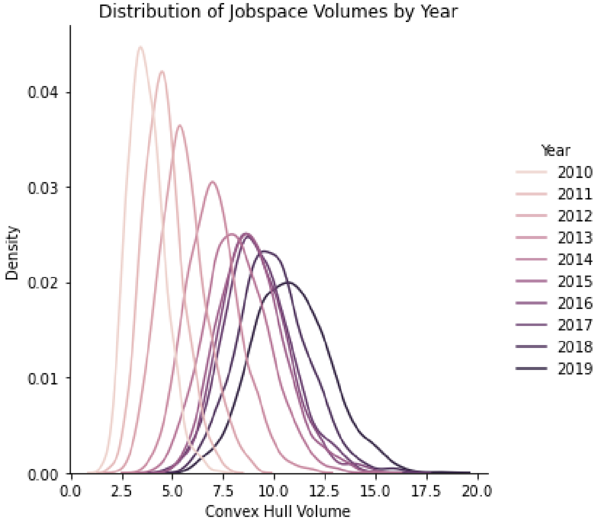
Recombination of Existing Roles:

points moving around *within* the existing "jobspace"

Expansion of the "Jobspace":

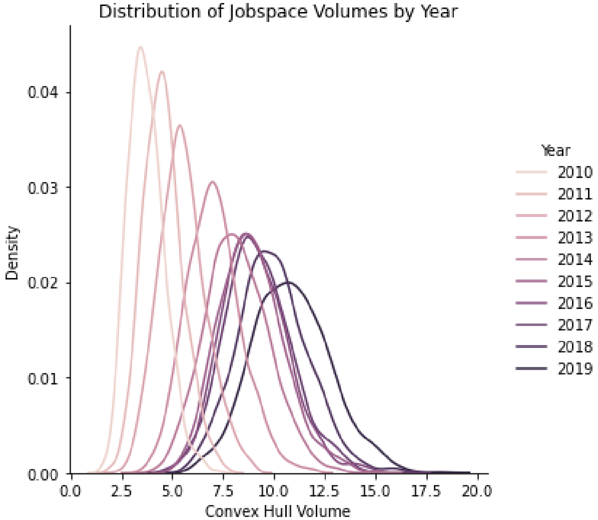
points that are *outside* of the existing "jobspace"

Analogous classifications for patents in Cheng et al. (2022) – also conceptually motivated by Autor (2019)



Aggregate Labor Market Results

- The volumes of the jobspace are monotonically increasing each year
- The samples are less uniform over time $\rightarrow$ not only the frontier expanding, but also filling in!
- Since 2015, average growth of 4-5% per year
- From 2010-2019, 2.5x increase



Application: Hierarchical Clustering/Aggregation

The Standard Occupation Classification (SOC) system in the U.S. is used to segment the labor market

But an alternative segmentation might be more consistent with work performed on the job (Turrell et al., 2022)

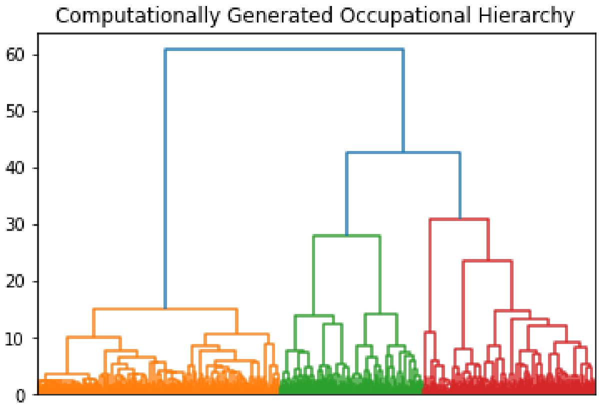
Can we use the VAE factors to generate a parallel SOC system?

Application: Hierarchical Clustering/Aggregation

The Standard Occupation Classification (SOC) system in the U.S. is used to segment the labor market

But an alternative segmentation might be more consistent with work performed on the job (Turrell et al., 2022)

Can we use the VAE factors to generate a parallel SOC system?



Conclusion

- New methods to characterize within occupation change and aggregate labor market change over time using the text of job postings
- Routine manual occupations and those suitable for machine learning are changing most
  - Important to take into account within-occupation change when measuring the effect of technology on labor demand
- We are seeing a recombination of work, along with an expansion of the "jobspace"
  - From 2015 to 2019, we saw the "jobspace" expanding at a rate of 4-5% per year

Thank You!

Any feedback is appreciated

sarah.bana@gmail.com

erik.brynjolfsson@gmail.com

rockdi@wharton.upenn.edu

sebastian.steffen@bc.edu