

New Methods for Studying Old Work

José Ignacio Velarde Morales

February 25, 2020

Contents

1	Introduction	2
1.1	Motivation for studying new work	2
1.2	ML for econ	2
2	Related Literature	2
2.1	Literature about new work	2
2.2	NLP literature	2
2.3	Econ and ML literature	2
3	Data	2
3.1	Dictionary of Occupational Titles	2
3.2	1940 Census Complete Count Files	3
3.3	CAI	4
4	Methods	4
4.1	Methods used	4
4.1.1	TF-IDF	4
4.1.2	RoBERTa	4
4.2	Matching procedure	4
4.3	Prediction	4
5	Results	4
5.1	Matching Rate/Validation	4
5.2	Attribute Prediction	4
6	Conclusion	4

1 Introduction

1.1 Motivation for studying new work

1.2 ML for econ

2 Related Literature

2.1 Literature about new work

2.2 NLP literature

2.3 Econ and ML literature

3 Data

3.1 Dictionary of Occupational Titles

The Dictionary of Occupational Titles (DOT) is a volume that was created by the United States Department of Labor to help match job openings and job seekers. It provides detailed textual descriptions of thousands of jobs. The job descriptions generally describe what work is done, who it is done by, and why it is done. The descriptions also include information regarding the industry and occupation group the job is in.

Figure 1: Example definition from the 1939 DOT

ACCOUNTING CLERK; accountant clerk (clerical)
1-01.31. Performs the more routine calculating, typing, and posting duties necessary in accounting: checks items on various reports, summarizing and posting the data to designated books; performs a variety of other clerical duties, such as making up invoices or monthly statements to be sent to customers, preparing pay rolls, verifying the company bank account, keeping files of various records, making up a periodic report of the company's business activities, and listing and checking detail as instructed. May operate a calculating machine, *ref.* CALCULATING-MACHINE OPERATOR. May operate an adding machine, *ref.* ADDING-MACHINE OPERATOR.

There are 5 editions of the DOT, published in 1939, 1949, 1965, 1977, and 1991. Each edition has about 20,000 job definitions. The 1965, 1977, and 1991 editions of the DOT include structured information about the jobs in addition to the textual description. The structured information is in the form of numerical codes that quantify job complexity, training requirements, and skill requirements. These codes are further described below.

Data, People, Things Codes

Each job is assigned a 3-digit code that describes the job's complexity in relation to data, people, and things. Each digit in the code corresponds the job's complexity in one of those three categories. For example, the job of Aeronautical Test Engineer has a code of 061. The code tells us the job has high data complexity, medium/low people complexity, and high things complexity. The figure below shows the meanings of the different codes.

Figure 2: Explanation of Data, People, Things Code

DATA (4th digit)	PEOPLE (5th digit)	THINGS (6th digit)
0 Synthesizing	0 Mentoring	0 Setting-Up
1 Coordinating	1 Negotiating	1 Precision Working
2 Analyzing	2 Instructing	2 Operating-Controlling
3 Compiling	3 Supervising	3 Driving-Operating
4 Computing	4 Diverting	4 Manipulating
5 Copying	5 Persuading	5 Tending
6 Comparing	6 Speaking-Signaling	6 Feeding-Offbearing
7 } No significant relationship	7 Serving	7 Handling
8 }	8 No significant relationship	8 No significant relationship

Job Attributes

Job descriptions in the DOT include descriptions of the mathematical education, aptitudes, and temperaments required to perform each job. Mathematical education refers to formal and informal education that develops basic reasoning skills in math. Aptitudes are specific abilities that an individual should have in order to perform a specific job. Examples of aptitudes are finger dexterity and hand-eye coordination. Temperaments are adaptability requirements made on the worker by specific types of job situations. These can include directing activities or performing repetitive work. The full list of job attributes we predict can be found in the table below.

Attribute	Description	Scale
GED Math	Formal and informal education developing general math skills	1-6
Finger Dexterity	Ability to use fingers to manipulate small objects	1-5
Eye-hand-foot Coordination	Motor responsiveness to visual stimuli	1-5
Direct, control, plan	Situations involving the direction, control and planning of activities	0/1
Attain limits	Situations involving the precise attainment of set limits or standards	0/1

3.2 1940 Census Complete Count Files

The United States conducts a census every ten years to collect data about its population. The Census collects a wide range of information about respondents including occupation, income, education, and place of residence. The data is confidential, but becomes publicly available 72 years after Census Day. The data from the 1940 Census became publicly available in April of 2012.

3.3 CAI

4 Methods

4.1 Methods used

4.1.1 TF-IDF

4.1.2 RoBERTa

4.2 Matching procedure

4.3 Prediction

5 Results

5.1 Matching Rate/Validation

5.2 Attribute Prediction

6 Conclusion