Reproducing SML scores

Based on the article by Brynjolfsson et. al (2018)

Honorata Bogusz, Szymon Winnicki

Faculty of Economic Sciences, University of Warsaw

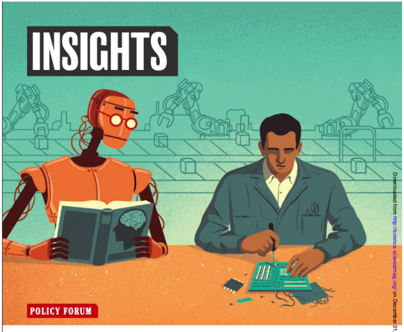
09.05.2021

Outline

Purpose

- We want to recreate the SML scores from Brynjolfsson et al. (2018) using the python2 script available at: https://www.aeaweb.org/articles?id=10.1257/pandp.20181019
- Then, we need to recode them from O*NET (American occupation classification) to ISCO (international occupation classification)
- This will facilitate Honorata's work during the PhD studies

Suitability for Machine Learning scores



Downloaded from http://science.sciencemag.org/ on December 21, 2017

TECHNOLOGY AND THE ECONOMY

What can machine learning do? Workforce implications

Profound change is coming, but roles for humans remain

By **Erik Brynjolfsson**[1,2] *and* **Tom Mitchell**[3]

Digital computers have transformed work in almost every sector of the economy over the past several decades (*1*). We are now at the beginning of an even larger and more rapid transformation due to recent advances in machine learning (ML), which is capable of accelerating the pace of automation itself. However, although it is clear that ML is a "general purpose technology," like the steam engine and electricity, which spawns a plethora of additional innovations and capabilities (*2*), there is no widely shared agreement on the tasks where ML systems excel, and thus little agreement on the specific expected impacts on the workforce and on the economy more broadly. We discuss what we see to be key implications for the workforce, drawing on our rubric of what the current generation of ML systems can and cannot do [see the supplementary materials (SM)]. Although parts of many jobs may be "suitable for ML" (SML), other tasks within these same jobs do not fit the criteria for ML well; hence, effects on employment are more complex than the simple replacement and substitution story emphasized by some. Although economic effects of ML are relatively limited today, and we are not facing the imminent "end of work" as is sometimes proclaimed, the implications for the economy and the workforce going forward are profound.

Any discussion of what ML can and cannot do, and how this might affect the economy, should first recognize two broad, underlying considerations. We remain very far from artificial general intelligence (*3*). Machines cannot do the full range of tasks that humans can do (*4*). In addition, although innovations

[1]Sloan School of Management, Massachussetts Institute of Technology, Cambridge, MA 02139, USA. [2]National Bureau of Economic Research, Cambridge, MA 02138, USA. [3]Carnegie Mellon University, Pittsburgh, PA 15213, USA. Email: erikb@ mit.edu

ILLUSTRATION: DAVIDE BONAZZI/@SALZMANART

Published by AAAS

AEA Papers and Proceedings 2018, 108: 43–47
https://doi.org/10.1257/pandp.20181019

ECONOMIC CONSEQUENCES OF ARTIFICIAL INTELLIGENCE AND ROBOTICS[‡]

What Can Machines Learn and What Does It Mean for Occupations and the Economy?[†]

By ERIK BRYNJOLFSSON, TOM MITCHELL, AND DANIEL ROCK*

Rapid advances in machine learning (ML) are poised to generate significant economic value and transform numerous occupations and industries. Machine learning, as described in Brynjolfsson and Mitchell (2017), is a subfield of artificial intelligence (AI) that studies the question "How can we build computer programs that automatically improve their performance at some task through experience?" We believe it is also a "general purpose technology" (GPT), a technology that becomes pervasive, improves over time, and generates complementary innovation (Bresnahan and Trajtenberg 1995).

Recent rapid progress in ML has been driven largely by an approach called deep learning, and has made it possible for machines to match or surpass humans in certain types of tasks, especially those involving image and speech recognition, natural language processing, and predictive analytics. So far, the realized economic effects are small relative to the potential offered by this new GPT (Brynjolfsson, Rock, and Syverson 2017). This reflects the time lags of years or even decades before GPTs generate substantial economic value. Entrepreneurs and innovators take time to adopt new technologies, reconfigure existing work, discover new business processes, and co-invent complementary technologies (Bresnahan and Greenstein 1996). Reorganization of economic activity is an important determinant of the returns to innovation.

Concern about the coming wave of automation's impact on employment is growing. For instance, Acemoglu and Restrepo (2017) connect the adoption of robots to reduced employment and wages in local labor markets. A study by the McKinsey Global Institute suggested that about half of the work activities people perform could be automated with current technology (Manyika et al. 2017). While advances in ML are impressive, and automation is already having significant effects on many parts of the workforce, we are far from artificial general intelligence (AGI) which would match humans in all cognitive areas. This raises the question of which tasks will be most affected by ML and which will be relatively unaffected.

In particular, a key insight of Autor, Levy, and Murnane (2003) is that an occupation can be viewed as a bundle of tasks, some of which offer better applications for technology than others. As with studies of routine task automation, the impact of machine learning on employment is a function of the suitability of machine learning for specific work activities.

[‡]*Discussants:* Jason Furman, Peterson Institute for International Economics; Ariel Burstein, University of California-Los Angeles; Susan Helper, Case Western Reserve University; Hal Varian, Google.

*Brynjolfsson: MIT Sloan School of Management, 100 Main Street, Cambridge, MA 02142, and NBER (email: erikb@mit.edu); Mitchell: Machine Learning Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15222 (email: Tom.Mitchell@cmu.edu); Rock: MIT Sloan School of Management, 100 Main Street, Cambridge, MA 02142 (email: drock@mit.edu). We thank the MIT Initiative on the Digital Economy for financial support, and Daron Acemoglu, David Autor, Seth Benzell, Rodney Brooks, Shane Greenstein, Yann LeCun, Frank Levy, James Manyika, Andrew Ng, Alex Peysakhovich, Daniela Rus, Guillaume Saint-Jacques, Yo Shavit, Chad Syverson, and Sebastian Thrun for helpful comments on earlier drafts of this research. We are grateful to Eric Bradford, Alenta Demissew, Francisco Proskauer, Tim Schoen, and Kathleen Zhu for excellent research assistance. We are, of course, responsible for any remaining errors.

[†]Go to https://doi.org/10.1257/pandp.20181019 to visit the article page for additional materials and author disclosure statement(s).

Outline

Steps of reproducing SML scores

1. Collecting the script from the AEA website
2. Collecting necessary data files from the O*NET website
3. Modifying the python2 script
4. Running the python2 script

Modifications to the python2 script

- Older version of 'xlrd' package (version 1.2.0) needed to be installed (the newest one does not support reading xlsx files)
- A mistake was made in the original code with respect to reading the "Task Statements.xlsx" file from O*NET. The authors add comment do load the xlsx file but use the command for reading a csv file.
- The file "Task Statements.xlsx" has a non-unique column name "Title" which the script confuses with the "Title" column from another file. The names had, thus, to be changed.
- We also deleted the part with American wages from BLS which does not concern us

Outline

Crosswalk

- We used a O*NET-ISCO crosswalk by Wojciech Hardy available at: https://ibs.org.pl/en/resources/occupation-classifications-crosswalks-from-onet-soc-to-isco/
- The crosswalk is in Stata
- This was very straight forward and we didn't encounter any problems
- However, the ISCO classification is less thorough than the ONET classification, so there is some aggregation in the recoded scores.

Min and Max recoded score

Min: The ISCO88 profession with min. SML is 2211: "Biologists, botanists, zoologists and related professionals", who are a subgroup of "PROFESSIONALS/Life science and health professionals/Life science professionals".

Max: The ISCO88 profession with max. SML is 3118: "Draughtspersons" who are a subgroup of "TECHNICIANS AND ASSOCIATE PROFESSIONALS/Physical and engineering science associate professionals/Physical and engineering science technicians".

Thank you for your attention!