Skilled Technical Workforce Classification

Guy Leonel Siwe
Social & Decision Analytics Division, Biocomplexity Institute & Initiative

FCSM October 2022

UNIVERSITY of VIRGINIA

UVA STW Team

University of Virginia, Biocomplexity Institute, Social and Decision Analytics

- Vicki Lancaster, Principal Scientist
- Cesar Montalvo, Postdoc Researcher
- Guy Leonel Siwe, Postdoc Researcher
- Haleigh Tomlin, Student Intern
- Stephanie Shipp, Interim Deputy Division Director & Professor

This presentation provides results of exploratory research sponsored by the National Center for Science and Engineering Statistics (NCSES) within the National Science Foundation (NSF). This information is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any views expressed are those of the author and not necessarily those of NCSES or NSF.

What is the Skilled Technical Workforce?

The **Skilled Technical Workforce** (STW) comprises individuals

- without a bachelor's degree but
- with a post-secondary nondegree credential or training that provides them with STEM knowledge and skills.

There are an estimated 16 million skilled technical workers.

In 2017 the National Science Board (NSB) raised concerns the United States is not adequately developing and sustaining a STW with the skills needed to compete in the 21st century.

Who is in the STW?

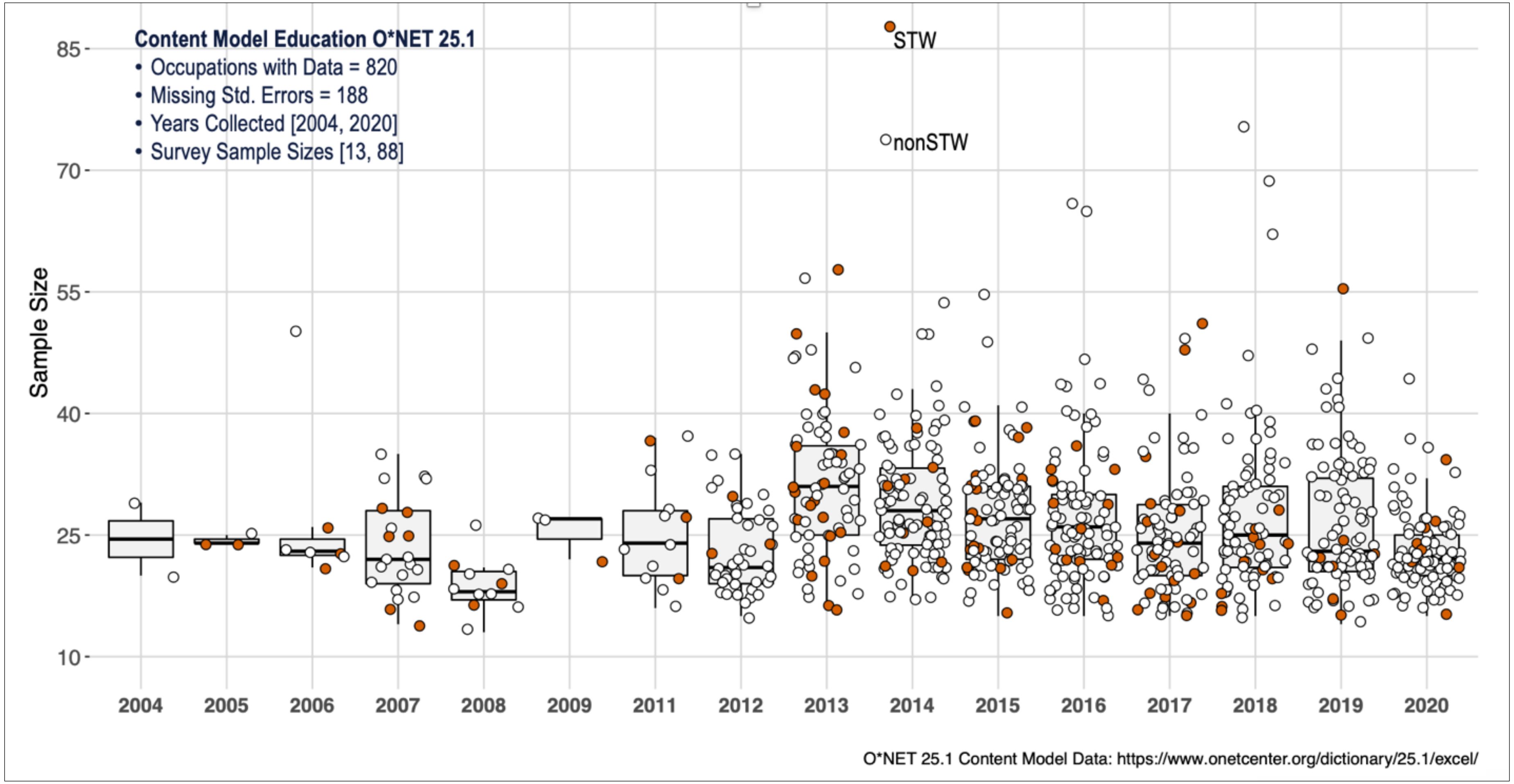
The STW occupations are classified using the definition operationalized by Rothwell (2015). He proposed using the education and knowledge survey data from the **Occupational Information Network (O*NET) Content Model**. His criteria are:

1. **education -** ≥ **50%** of the survey respondents do not have a bachelors degree or higher; and
2. **knowledge -** ≥ **4.5** level of knowledge in any of the 14 knowledge domains in the table below.

Fourteen Knowledge Domains of the Skilled Technical Workforce

Biology	Economics and Accounting	Medicine and Dentistry
Building and Construction	Engineering and Technology	Physics
Chemistry	Food Production	Production and Processing
Computers and Electronics	Mathematics	Telecommunications
Design	Mechanical	

Fitness-for-Use of the Content Model Data



- Small sample sizes
- **Untimely data (update every 10 years)**
- Of the 1,016 occupations, only 820 have both education & knowledge Content Model Data

Is there a timelier data source that can be used to designate STW occupations?

Real-time Job postings have virtually no lag time and there is now an aggregator that scrapes tens of thousands job posting websites a day and provides the data free to researchers. Job-postings capture the rapidly changing skills demanded by employers making skills the currency used to measure technology adoption in occupations.

Can we use online Job postings where we only observe skills required to designate STW occupations?

Data Sources

O*NET Content Model Knowledge and Skill Survey Data

- 873 occupations, 35 skills and 33 knowledge domains

Non-technical → Reading_Comprehension

Technical → Mathematics

OnetSoc	SkillName	Value
11-1011.00	Reading_Comprehension	4.75
11-1011.00	Active_Listening	4.88
11-1011.00	Writing	4.38
11-1011.00	Speaking	4.88
11-1011.00	Mathematics	3.62
11-1011.00	Science	1.12
11-1011.00	Critical_Thinking	4.75

Level

OnetSoc	DomainName	Value
11-1011.00	Administration_and_Management	6.23
11-1011.00	Clerical	3.50
11-1011.00	Economics_and_Accounting	4.36
11-1011.00	Sales_and_Marketing	3.90
11-1011.00	Customer_and_Personal_Service	5.55
11-1011.00	Personnel_and_Human_Resources	5.02
11-1011.00	Production_and_Processing	2.92

Burning Glass Technology Job-ads (Virginia in 2019)

- We focus on 4 Major Occupation Groups (MOG): Construction & Extraction (47), Installation, Maintenance, & Repair (49), Production (51), Transportation & Material Moving (53)
- 91,377 job postings distributed in 271 occupations in the 4 MOGs
- For each job posting we have the following BGT skill taxonomy:

Skills (5,212) → Skill Clusters (575) → Skill Cluster Families (29)

Methodology for a New STW Classification

1 Can we substitute technical skills for knowledge domains?

Evaluate the relationship between O*NET technical skills and 14 knowledge domains

2 Can we identify technical skills in job postings?

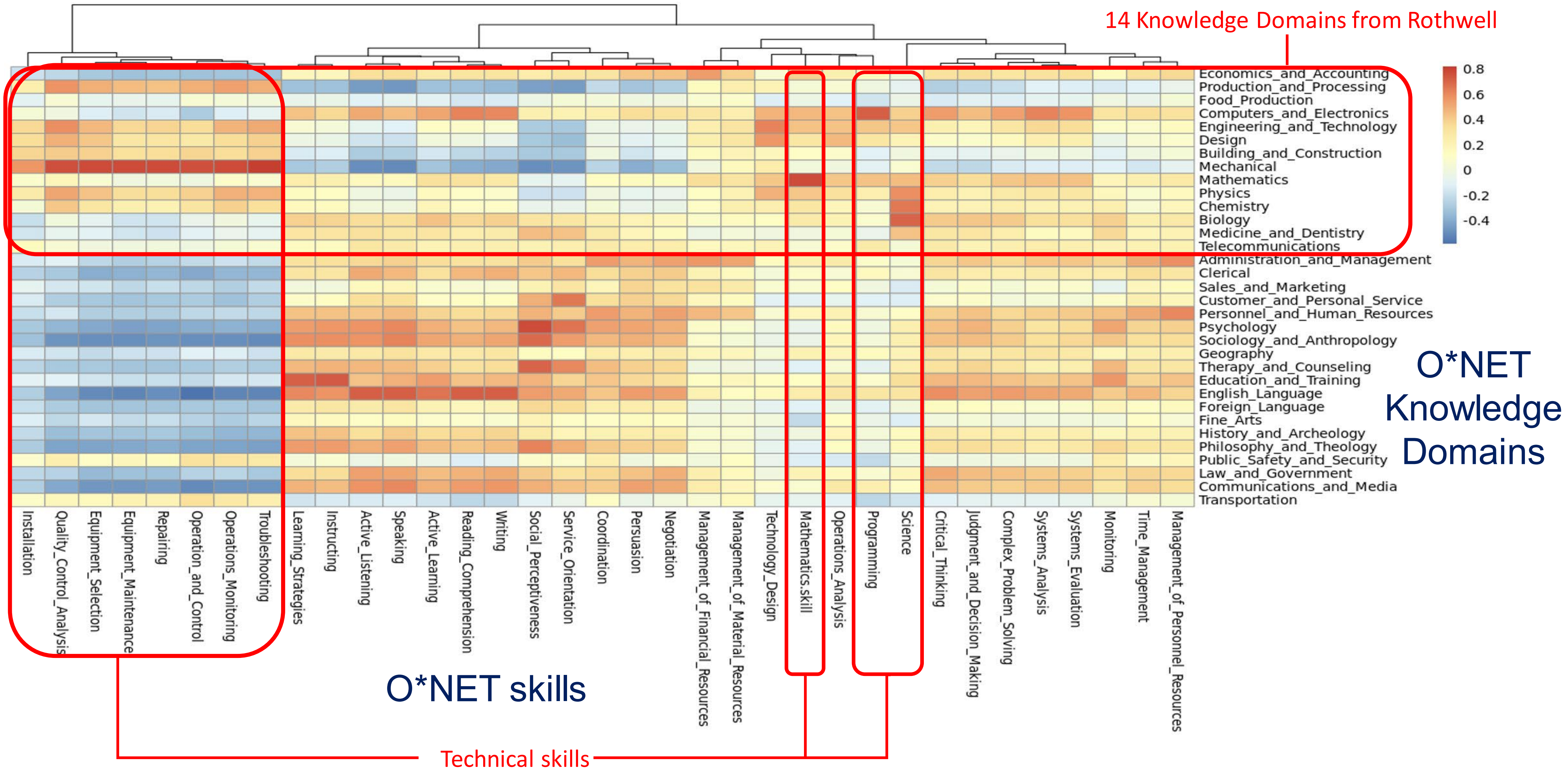
Use **Natural Language Processing** to connect O*NET technical skills to BGT technical skill clusters

3 Define STW using technical skill intensity.

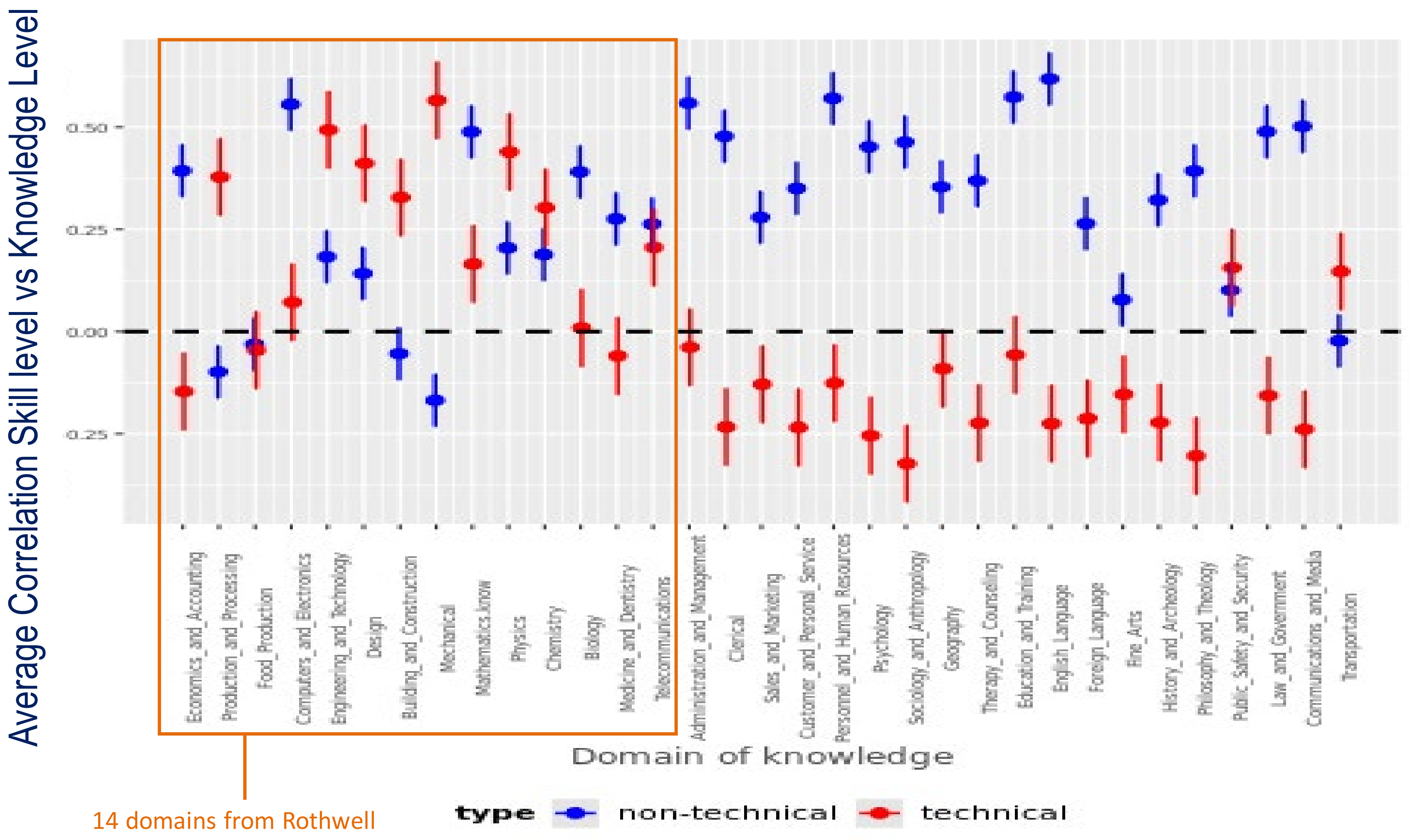
Use the **Projection Method** to compute the technical skill intensity for each occupations

Compare the projection method with Rothwell

1 Correlation between Skills & Knowledge Domains



1 Correlation between Technical Skills & Knowledge Domains



Conclude: Technical skill level can replace Rothwell's 14 knowledge domains.

- O*NET technical skills are positively correlated with the 14 knowledge domains used by Rothwell.
- O*NET technical skills are negatively correlated with the other 19 knowledge domains.

2 Identifying Technical Skills in Job Postings: Matching Model

Challenges:

- O*NET and BGT skill clusters (more granular) use different dictionaries. The solution is to:

BGT skill cluster ⟷ (Use NLP) ⟷ O*NET skill competencies ⟷ (O*NET classification) ⟷ O*NET competency groups (technical and non-technical)

Matching model:

1. Build the vector representation of each O*NET skill and BGT skill cluster using BERT.

Predict the masked word
(langauge modeling)

man →	0.6	−0.2	0.8	0.9	−0.1	−0.9	−0.7
woman →	0.7	0.3	0.9	−0.7	0.1	−0.5	−0.4
king →	0.5	−0.4	0.7	0.8	0.9	−0.7	−0.6
queen →	0.8	−0.1	0.8	−0.9	0.8	−0.5	−0.9

2. Compute similarity-based between a O*NET skill and BGT skill cluster using the cosine.

$$\mathrm{F}(\textit{skill cluster}) = \arg\max_{\textit{Onet skill}} \left(\mathrm{cosine}(\textit{vector skill cluster}, \textit{vector } O{*}\textit{NET skill}) \right)$$

2 Matching Model Results

BGT Skill Cluster distribution by O*NET Skill/Competency Group

O*NET Skill/Competency Group	Mean cosine similarity score	Number of BGT-Skill Clusters
Non-technical skills	0.4350	319
Technical skills	0.4332	256

Matching model performance

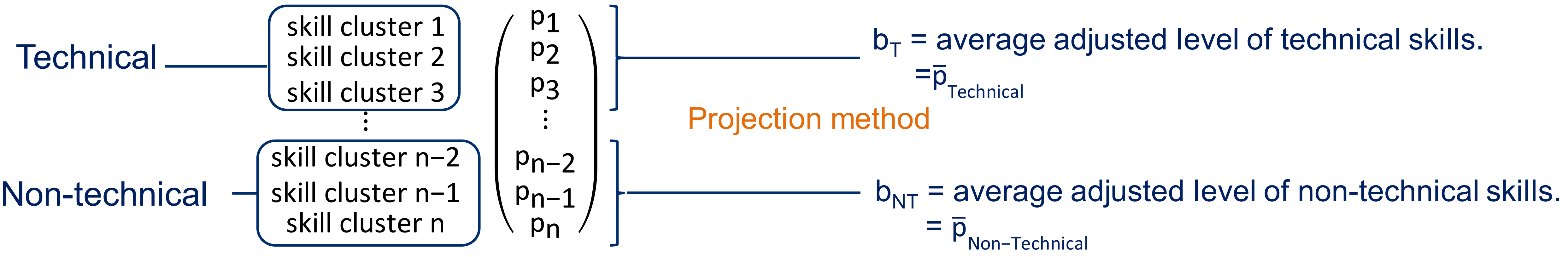
- Use the classification of skills into skill clusters from job postings as **ground truth.**
- Reclassify skills into skill clusters using the matching model
- Use resampling of skill clusters
- Precision = 70% +/- 5 %.

3 Identify STW Occupations using BGT: Projection Method

- An occupation is defined as a distribution of skill clusters with 2 measures:
  - Skill cluster level = Number of skills listed in the skill cluster.
  - Skill cluster Intensity = Proportion of job postings that list the skill cluster.
  - Adjusted skill cluster level = p = skill cluster level x skill cluster intensity.

- Projection method: Compute the average level (weight) of technical skill and non-technical skill.

Occupation profile



3 Projection Method Results

- We compute the Likelihood that an occupation is intensively technical

$$\text{Likelihood} = \frac{b_T}{b_T + b_{NT}}$$

- Define STW using projection methods as an occupation that:
  1. Doesn't require a bachelor degree
  2. Intensively use technical skill (meaning a Likelihood > 0.5)

Classification comparison from BGT

MOG	Rothwell		Projection	
	STW	Non-STW	STW	Non-STW
Construction & Extraction (47)	30	31	47	14
Install., Maintenance & Repair (49)	40	11	40	11
Production (51)	22	83	56	49
Transport & Material Moving (53)	4	47	21	30

The projection method classifies more occupations as STW than the Rothwell (96.

3 Comparative Study with Rothwell

Confusion table by MOG

MOG	Rothwell	Projection	
		Non-STW	STW
Construction and Extraction (47)	Non-STW	**0.226**	0.774
	STW	0.233	0.767
Installation, Maintenance, and Repair (49)	Non-STW	**0.455**	0.545
	STW	0.150	0.850
Production (51)	Non-STW	**0.494**	0.506
	STW	0.364	0.636
Transportation and Material Moving (53)	Non-STW	**0.563**	0.437
	STW	0.250	0.750
Overall	Non-STW	**0.465**	0.535
	STW	0.250	**0.750**

If the Rothwell approach is considered as ground true, how perform the projection method?

- The projection method identify 75% of occupations listed by Rothwell as effectively STW.
- However, the projection method shows that many non-STW from Rothwell are misclassified; 46% of those listed by Rothwell are non-STW.
- The result varies across MOG

3 Skills Profile and STW Classification

P-value from Fisher's Exact Test

MOG	Fisher's exact test using Rothwell	Fisher's exact test using Projection
Construction and Extraction (47)	0.000	0.227
Installation, Maintenance, and Repair (49)	0.504	0.188
Production (51)	0.311	0.006
Transportation and Material Moving (53)	1.000	0.008

- The projection method shows a lower p-value
- We likely have an association between skill profile and the STW classification using the projection method than with the Rothwell definition.

Why the **Projection** perform better?

- Occupations with a similar skill cluster profiles would have the same classification as STW.

Strategy:

- By MOG, cluster occupations with the same skill profile using KNN.
- Test the relation between the cluster and STW classification for each method using Fisher exact test.
- Compare the p-value associate with the test (**probability to likely have no relation**).

Conclusion

- We have shown that skills from Labor Market Information can be an alternative data source to the O*NET Content Model for classifying occupations as being the in the Skilled Technical Workforce.
- We focus on the skill profile in each occupation to identify STW occupations and develop a a novel projection method.
  - We estimated a precision of 75% for identifying the same STW occupations as Rothwell.
  - However, the projection method shows that Rothwell misclassifies occupations that have the same skill profiles.
- This framework allows us to track the STW using real-time data that can reveal the rapidly changing nature of technical occupations and as a consequence employer demands.

References

- Rothwell, J. T. (2015). “Defining skilled technical work.” Available at SSRN 2709141.
- Lancaster, V. et al. (2021). “Designating the Skilled Technical Workforce Using O*NET-SOC (2019)”.
- Occupational Information Network content model (1995). “Development of Prototype Occupational Information System: Content Model”.
- Cunningham, Wendy V. and Villaseñor, P. (2016). Employer Voices, Employer Demands, and Implications for Public Skills Development Policy Connecting the Labor and Education Sectors. Published by Oxford University Press on behalf of the World Bank. © World Bank. https://openknowledge.worldbank.org/handle/10986/27700 License: CC BY-NC-ND 3.0 IGO.