The complementarity between ICT and cognitive skills in the third industrial revolution: an empirical assessment

ODISSEI Conference for Social Science in the Netherlands 2022

> Marie Labussière & Thijs Bol University of Amsterdam



European Research Council



Introduction 0 Theo

Data

Method 000 Result

Mapping changing labour markets

"Third industrial revolution" (Liu and Grusky, 2013) : digitalisation, automation, robotisation, Al...

Introduction

Theo 0 Dat

Methoo 000 Re

Mapping changing labour markets

"Third industrial revolution" (Liu and Grusky, 2013) : digitalisation, automation, robotisation, Al...

How should we train and educate workers?

Introduction

Theor

Da

Metho 000 Resu 0000

Mapping changing labour markets

"Third industrial revolution" (Liu and Grusky, 2013) : digitalisation, automation, robotisation, Al...

How should we train and educate workers?

Main hypothesis: technological change is skill-biased

 $\hookrightarrow$  complementarity between cognitive tasks and technological devices

Introduction Theory

Data o Method 000

sults

Mapping changing labour markets

"Third industrial revolution" (Liu and Grusky, 2013) : digitalisation, automation, robotisation, Al...

How should we train and educate workers?

Main hypothesis: technological change is skill-biased

 $\hookrightarrow$  complementarity between cognitive tasks and technological devices

Are employers looking for hybrid skill profiles combining ICT and cognitive skills?

Analyse skill profiles over time and within occupations

Introduction

Mapping changing labour markets

"Third industrial revolution" (Liu and Grusky, 2013) : digitalisation, automation, robotisation, Al...

How should we train and educate workers?

Main hypothesis: technological change is skill-biased

 $\hookrightarrow$  complementarity between cognitive tasks and technological devices

Are employers looking for hybrid skill profiles combining ICT and cognitive skills?

Analyse skill profiles over time and within occupations

Case: the UK

Introduction

Data 0 Method 000 Resu 000



Skills are often poorly operationalised:

• Broad categories of skills, e.g. "ICT skills" (Liu and Grusky, 2013; Buchmann et al., 2020)

ntroduction Theory

d

Results



Skills are often poorly operationalised:

- Broad categories of skills, e.g. "ICT skills" (Liu and Grusky, 2013; Buchmann et al., 2020)
- Skills are analysed in isolation (e.g., Deming, 2017; Deming and Kahn, 2018)

 $\neq$  skill *profiles* 

 $\rightarrow\,$  mixes of (different types of) hard and soft skills



Skills are often poorly operationalised:

- Broad categories of skills, e.g. "ICT skills" (Liu and Grusky, 2013; Buchmann et al., 2020)
- Skills are analysed in isolation (e.g., Deming, 2017; Deming and Kahn, 2018)

 $\neq$  skill *profiles* 

 $\rightarrow\,$  mixes of (different types of) hard and soft skills

 $\rightarrow\,$  Where does complementarity take place?

Results

#### Job ads as data



- ✓ A direct measure of the actual demand for skills
- Employers have an interest in making them accurate
- ✓ Raw skills  $\neq$  pre-existing categories

Introduction 0 Theory

Data

Method 000

Results 00000

#### Job ads as data



- ✓ A direct measure of the actual demand for skills
- Employers have an interest in making them accurate
- ✓ Raw skills  $\neq$  pre-existing categories

#### Data

Online job ads collected by Burning Glass Technologies (BGT)

- $\rightarrow$  Period 2012-2019
- ightarrow Random sample of 128,000 ads

Introduction 0 Theo

Data

Method 000 Resu 000

#### Operationalisation

How to measure complementarities?

Job ads are already pre-processed by the BGT: extracts skills requirements and standardises their labels

- 13,435 distinct skills
- How to measure their association within the ads?

topic modeling with LDA

Aim to identify **latent skill sets** and analyse them over **time** and **within occupations** 

Result

Job ads are short texts: data sparsity and lack of context

Introduction 0 Theo 0 Da

Method ○●○

Results

Job ads are short texts: data sparsity and lack of context

The biterm topic model (Yan et al., 2013)

• BTM uses biterms instead of words as semantic units

```
"planning sales Excel"=

{"planning sales"

"sales Excel"

"Excel planning"
```

Theor 0 Data

Method

Results

Job ads are short texts: data sparsity and lack of context

The biterm topic model (Yan et al., 2013)

• BTM uses biterms instead of words as semantic units

```
"planning sales Excel"=

{"planning sales"

"sales Excel"

"Excel planning"
```

• It directly models the generation of words co-occurrence patterns in the whole corpus ( $\neq$  in each single document)

Introduction 0 Theo

Data

Method

Resu 0000

Job ads are short texts: data sparsity and lack of context

The biterm topic model (Yan et al., 2013)

• BTM uses biterms instead of words as semantic units

```
"planning sales Excel"=

{"planning sales"

"sales Excel"

"Excel planning"
```

- It directly models the generation of words co-occurrence patterns in the whole corpus (≠ in each single document)
- The document generative process can be estimated

Intr	od		~+	n
III UI		u	- L	
0				

Data o Method

Results

#### Setting for LDA Biterm model with *k* topics

maximtrp/bitermplus (cythonized)

Priors  $\alpha = \beta = \frac{1}{k}$  i.e., the ads are specialised Iterations 2,000 Choice of k Perplexity  $U_{mass}$  Coherence (Röder et al., 2015) Visual inspection

The criteria converge towards k = 8 topics

Introduction 0 Theo 0 Data o Method

Res

	Label	Top-10 most relevant words ( $\lambda=0.8$ )
1	Pure IT	Javascript, Microsoft C, SQL, net, java, software development, active server pages ASP, ASPNET, Jquery, sql server
2	IT support	Microsoft active directory, VmWare, trouble shoot- ing, Cisco, Windows server, Microsoft Windows, Linux, technical support, ITIL, communication skills
3	Business & data analytics	SQL, business analysis, project management, com- munication skills, Oracle, stakeholder management, business process, data analysis, problem solving, business intelligence
4	Engineering	project management, mechanical engineering, com- munication skills, AutoCAD, commissioning, plan- ning, quality assurance control, quality management, problem solving, budgeting
5	Digital marketing	creativity, social media, marketing, Adobe Pho- toshop, communication skills, Adobe InDesign, writ- ing, copy writing, editing, research

Introduction 0 Theo 0 Dat

Method 000 Results

6	Administrative management	communication skills, Microsoft excel, organisational skills, <b>budgeting</b> , detail-orientated, planning, <b>ac-</b> <b>counting</b> , Microsoft office, teamwork collaboration, <b>customer service</b>
7	Sales	sales, communication skills, customer service, busi- ness development, sales management, building ef- fective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with pa- tients: mental health, surgery, research, care plan- ning, patient care, cleaning, staff management, child care

L	-	÷		~		~	4	-	-
		L	Γ.			C			

Theory 0 Dat

Method 000 Results

6	Administrative management	communication skills, Microsoft excel, organisational skills, budgeting, detail-orientated, planning, ac- counting, Microsoft office, teamwork collaboration, customer service
7	Sales	sales, communication skills, customer service, busi- ness development, sales management, building ef- fective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with pa- tients: mental health, surgery, research, care plan- ning, patient care, cleaning, staff management, child care

- $\rightarrow$  omnipresence of (different types of) ICT skills
- $\rightarrow$  different mixes of ICT and soft skills
- ightarrow each job ad defined by its distribution over the 8 topics

Results

6	Administrative management	communication skills, Microsoft excel, organisational skills, <b>budgeting</b> , detail-orientated, planning, <b>ac</b> - <b>counting</b> , Microsoft office, teamwork collaboration, <b>customer service</b>
7	Sales	sales, communication skills, customer service, busi- ness development, sales management, building ef- fective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with pa- tients: mental health, surgery, research, care plan- ning, patient care, cleaning, staff management, child care

- $\rightarrow$  omnipresence of (different types of) ICT skills
- $\rightarrow$  different mixes of ICT and soft skills
- ightarrow each job ad defined by its distribution over the 8 topics
  - $\hookrightarrow$  skill profiles, over time and within occupations

# Prevalence of skill sets over time Dichotomous skill profile

```
Job ad j contains topic k if Prob_j(k) > c
```



0

#### Skill sets within occupations

#### Do ads in the same occupation have similar skill profiles?

► Clustering to analyse the diversity of skill profiles within the ads of a given occupation

 $\hookrightarrow$  *k***-means** for each occupation

with k determined based on the silhouette score and SSE

Introduction Theory o o

D

Method 000 Results

Conclusior 00

#### Skill sets within occupations

Example with two occupations (UK SOC 2010 major groups 4 and 8)



Topic distribution of the occupation-specific cluster medoids

ntrodu	uction.

Theory 0 Dat o Method 000 Results

#### Conclusion Work in progress!

Preliminary but encouraging results!

Introduction 0 Theo 0 Dat

Methoo 000

Results

#### Conclusion Work in progress!

Preliminary but encouraging results!

Limitations/what to do next:

- skill sets vs. skill profiles
- More explicit link with ICT/cognitive skills

   → analyse share of cognitive/ICT skills within each topic?
- The composition of topics may change over time
  - $\hookrightarrow \mathsf{Dynamic \ biterm \ LDA?}$
- $\square$  Combine occupation x time
  - $\hookrightarrow \ \mathsf{Dynamic} \ \mathsf{clustering}?$

#### Discussion Want to know more?



projectcareer.eu

@CAREER\_erc

Introduction 0 Theory

Data

Method

F

Results

#### References I

- Buchmann, M., Buchs, H., and Gnehm, A.-S. (2020). Occupational Inequality in Wage Returns to Employer Demand for Types of Information and Communications Technology (ICT) Skills: 1991-2017. KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie, 72(S1):455-482.
- Deming, D. and Kahn, L. B. (2018). Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, 36(S1):S337–S369.
- Deming, D. J. (2017). The Growing Importance of Social Skills in the Labor Market\*. The Quarterly Journal of Economics, 132(4):1593–1640.
- Liu, Y. and Grusky, D. B. (2013). The Payoff to Skill in the Third Industrial Revolution. American Journal of Sociology, 118(5):1330–1374.
- Röder, M., Both, A., and Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. In Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, pages 399–408, Shanghai China. ACM.
- Yan, X., Guo, J., Lan, Y., and Cheng, X. (2013). A biterm topic model for short texts. In Proceedings of the 22nd international conference on World Wide Web - WWW '13, pages 1445–1456, Rio de Janeiro, Brazil. ACM Press.

References

Appendix

Biterm topic 000 Biterm Model statistics 00000

Data Burning Glass Technologies (BGT)

Online job ads gathered by the BGT in the UK:

Observation period January 2012-December 2019 Full sample 59,9 million ( $\approx$ 5.7M/year) Random Samples 8\*16,000 ads released the months of January only



References

pendix Data

Biterm topic model 000 Biterm Model statistics 00000

#### Biterm topic model From Yan et al. (2013)

Biterm: unordered word-pair co-occurrence in a short context

#### Generative process

- The corpus consists of a mixture of topics
- Each biterm is drawn from a specific topic
- **1** For each topic *z*, draw a topic-specific word distribution  $\phi_z \sim Dir(\beta)$
- **2** Draw a topic distribution  $\theta \backsim Dir(\alpha)$  for the whole collection
- **3** For each biterm b in the biterm set B
  - draw a topic assignment  $z \backsim Multi(\theta)$
  - draw two words:  $w_i, w_j \sim Multi(\phi_z)$

#### Biterm topic model From Yan et al. (2013)

The joint probability of a biterm  $b = (w_i, w_j)$ :

$$P(b) = \sum_{z} P(z)P(w_i|z)P(w_j|z) = \sum_{z} \theta_z \phi_{i|z} \phi_{j|z}$$

The likelihood of the whole corpus:

$$P(B) = \prod_{(i,j)} \sum_{z} \theta_{z} \phi_{i|z} \phi_{j|z}$$

The topic proportion of a document can be estimated via Bayes and the empirical distribution of the generated biterms:

$$P(z|d) = \sum_{b} P(z|b)P(b|d) = \frac{P(z)P(w_i|z)P(w_j|z)}{\sum_{z} P(z)P(w_i|z)P(w_j|z)} \frac{n_d b}{\sum_{b} n_d(b)}$$

References

Appendix D

Biterm topic model

Biterm Model statistics 00000

#### Biterm topic model Comparison with LDA (from Yan et al. (2013))



Plate notation LDA

Plate notation Biterm

Appendix

Biterm topic model 000

#### Choice of the number of topics Coherence (Röder et al., 2015)



#### Choice of the number of topics Perplexity



#### Distribution of job ads over the topics



▲ Back

References

Appendix D

Biterm topic mode

Biterm Model statistics

#### Distribution of job ads over the topics



🕻 🖣 Back

References

Appendix D

Biterm topic mode

Biterm Model statistics 00000

#### Distribution of job ads over the topics









Rank







c ......

▲ Back

es Appendi>

< Data o Biterm topic model

Biterm Model statistics  $0000 \bullet$ 

8

#### Occupation classification

# Occupations: **nine major groups** of the UK Standard Occupational Classification (SOC 2010)

- 1. Managers, directors and senior officials
- 2. Professional occupations
- 3. Associate professionals and technical occupations
- 4. Administrative and secretarial occupations
- 5. Skilled trades occupations
- 6. Caring, leisure and other service occupations
- 7. Sales and customer service occupations
- 8. Process, plant and machine operatives
- 9. Elementary occupations

▲ Back

References

pendix

Biterm topic mo 000 Biterm Model statistics 00000

#### Clustering by occupations Choice of *k* – Silhouette width



References

Appendix

Biterm topic mode

Biterm Model statistics

#### Clustering by occupations Choice of *k* – Sum of squared errors (SSE)



References

Appendix

Biterm topic m

Biterm Model statistics 00000

## Skill sets within occupations (1/2)

Topic distribution for the SOC cluster medoids (k-means)



References

Appendix D

Biterm topic model 000 Biterm Model statistics 00000

## Skill sets within occupations (2/2)

#### Topic distribution for the SOC cluster medoids (k-means)



# ICT support

Skilled trades occ.

Sales and customer service occ.



References

Appendix D

Biterm topic model

Biterm Model statistics 00000

## Clustering by occupations (all)



#### Topic distribution for the SOC cluster medoids (k-means)



Sales and customer service occ. Process, plant and machine operatives Elementary occ. ICT support ICT support ICT support 0.8 0.8 Pure IT Digital mark Pure II Digital mark Pure IT Digital mark. 0.6 0.4 0.4 0.4 0.2 0.2 0.2 Admin me Care Admin m Care Admin Care Band D analytics Engineering Engineering 8 and D analytics Engineering Band D analytics Sales Sales Sales

References

Appendix Da

Biterm topic model

Biterm Model statistics