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

career

Introduction

Mapping changing labour markets

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

Introduction

Mapping changing labour markets

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

- å How should we train and educate workers?

Introduction

Mapping changing labour markets

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

å How should we train and educate workers?

Main hypothesis: technological change is **skill-biased**

, ! **complementarity** between cognitive tasks and technological devices

Introduction

Mapping changing labour markets

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

å How should we train and educate workers?

Main hypothesis: technological change is **skill-biased**

,! **complementarity** between cognitive tasks and technological devices

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

å Analyse skill profiles over time and within occupations

Introduction

Mapping changing labour markets

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

å How should we train and educate workers?

Main hypothesis: technological change is **skill-biased**

,! **complementarity** between cognitive tasks and technological devices

Are employers looking for hybrid skill profiles combining ICT and cognitive skills?
å Analyse skill profiles over time and within occupations

Case: the UK

Theory

From skills to skill profiles

Skills are often poorly operationalised:

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

Theory

From skills to skill profiles

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*
  - ! mixes of (different types of) hard and soft skills

Theory

From skills to skill profiles

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*
  - ! mixes of (different types of) hard and soft skills

! Where does complementarity take place?

Job ads as data

- 3 A direct measure of the actual demand for skills
- 3 Employers have an interest in making them accurate
- 3 Raw skills ≠ pre-existing categories

Job ads as data

3 A direct measure of the actual demand for skills

3 Employers have an interest in making them accurate

3 Raw skills 6= pre-existing categories

Data

Online job ads collected by Burning Glass Technologies (BGT)

! Period 2012-2019

! Random sample of 128,000 ads

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?

å **topic modeling with LDA**

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

Biterm Topic model (BTM)

A variant of Latent Dirichlet Allocation (LDA)

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

Biterm Topic model (BTM)

A variant of Latent Dirichlet Allocation (LDA)

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

$$\text{“planning sales Excel”} = \begin{cases} \text{“planning sales”} \\ \text{“sales Excel”} \\ \text{“Excel planning”} \end{cases}$$

Biterm Topic model (BTM)

A variant of Latent Dirichlet Allocation (LDA)

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

$$\text{“planning sales Excel”} = \left\{ \begin{array}{l} \text{“planning sales”} \\ \text{“sales Excel”} \\ \text{“Excel planning”} \end{array} \right.$$

It directly models the generation of words co-occurrence patterns in the whole corpus (≠ in each single document)

Biterm Topic model (BTM)

A variant of Latent Dirichlet Allocation (LDA)

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)

The document generative process can be estimated

Setting for LDA

Biterm model with k topics

`maximtrp/bitermplus` (cythonized)

Priors $=$ $= \frac{1}{k}$ 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

Results Biterm: 8 skill sets

	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 shooting, Cisco, Windows server, Microsoft Windows, Linux, technical support, ITIL, communication skills
3	Business & data analytics	SQL, business analysis, project management, communication skills, Oracle, stakeholder management, business process, data analysis, problem solving, business intelligence
4	Engineering	project management, mechanical engineering, communication skills, AutoCAD, commissioning, planning, quality assurance control, quality management, problem solving, budgeting
5	Digital marketing	creativity, social media, marketing, Adobe Photoshop, communication skills, Adobe InDesign, writing, copy writing, editing, research

Results Biterm: 8 skill sets

6	Administrative management	communication skills, Microsoft excel, organisational skills, budgeting, detail-orientated, planning, accounting, Microsoft o□ce, teamwork collaboration, customer service
7	Sales	sales, communication skills, customer service, business development, sales management, building effective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with patients: mental health, surgery, research, care planning, patient care, cleaning, sta□ management, child care

Results Biterm: 8 skill sets

6	Administrative management	communication skills, Microsoft excel, organisational skills, budgeting, detail-orientated, planning, accounting, Microsoft o□ce, teamwork collaboration, customer service
7	Sales	sales, communication skills, customer service, business development, sales management, building effective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with patients: mental health, surgery, research, care planning, patient care, cleaning, sta□ management, child care

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

Results Biterm: 8 skill sets

6	Administrative management	communication skills, Microsoft excel, organisational skills, budgeting, detail-orientated, planning, accounting, Microsoft o□ce, teamwork collaboration, customer service
7	Sales	sales, communication skills, customer service, business development, sales management, building effective relationships, sales goals, organisational skills, teamwork collaboration, account management
8	Care	teaching, communication skills, working with patients: mental health, surgery, research, care planning, patient care, cleaning, sta□ management, child care

! omnipresence of (different types of) ICT skills

! different mixes of ICT and soft skills

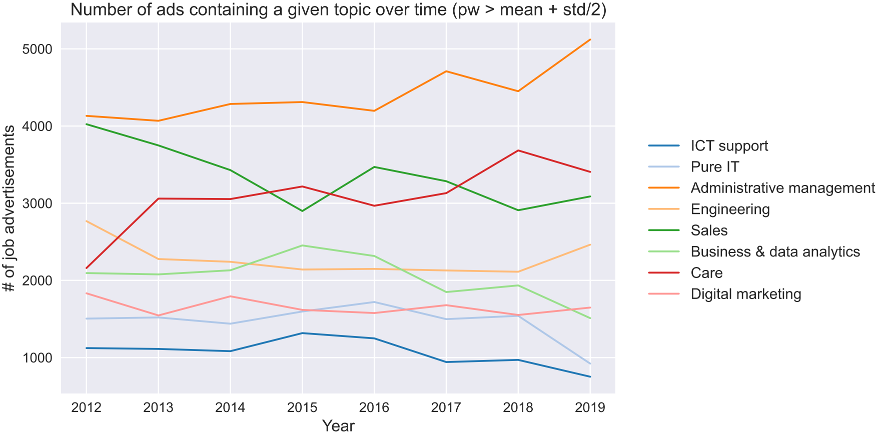
! each job ad defined by its distribution over the 8 topics

,! skill profiles, over time and within occupations

Prevalence of skill sets over time

Dichotomous skill profile

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



Skill sets within occupations

Do ads in the same occupation have similar skill profiles?

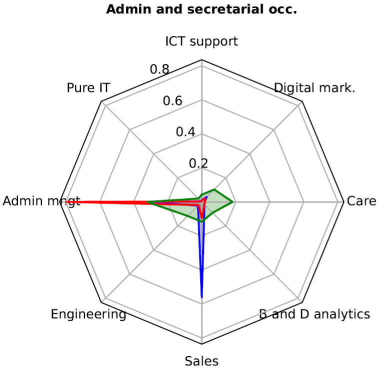
ä **Clustering** to analyse the diversity of skill profiles within the ads of a given occupation

,! ***k*-means** for each occupation

with k determined based on the silhouette score and SSE

Skill sets within occupations

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



Topic distribution of the occupation-specific cluster medoids

Conclusion

Work in progress!

3 Preliminary but encouraging results!

Conclusion

Work in progress!

3 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
  - ,! Dynamic biterm LDA?
+ Combine occupation *x* time
  - ,! Dynamic clustering?

Discussion

Want to know more?

projectcareer.eu

@CAREER_erc

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.

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 (5.7M/year)
Random Samples	8*16,000 ads released the months of January only

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 $_z \backsim Dir(\quad)$
2. Draw a topic distribution $\quad \backsim Dir(\quad)$ for the whole collection
3. For each biterm b in the biterm set B
   - draw a topic assignment $z \backsim Multi(\quad)$
   - draw two words: $w_i; w_j \backsim Multi(\quad_z)$

Biterm topic model

From Yan et al. (2013)

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

$$P(b) = \sum_z P(z)P(w_i jz)P(w_j jz) = \sum_z \quad {}_z \quad {}_{i|z} \quad {}_{j|z}$$

The likelihood of the whole corpus:

$$P(B) = \prod_{(i,j)} \sum_z \quad {}_z \quad {}_{i|z} \quad {}_{j|z}$$

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

$$P(zjd) = \sum_b P(zjb)P(bjd) = \frac{P(z)P(w_i jz)P(w_j jz)}{\sum_z P(z)P(w_i jz)P(w_j jz)} \frac{n_d b}{\sum_b n_d(b)}$$

Biterm topic model

Comparison with LDA (from Yan et al. (2013))

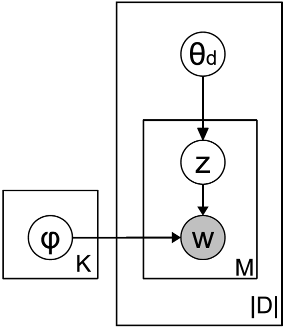


Plate notation LDA

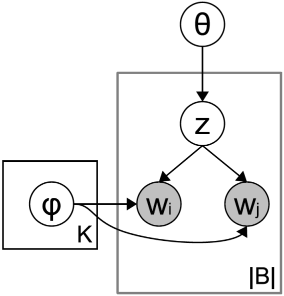
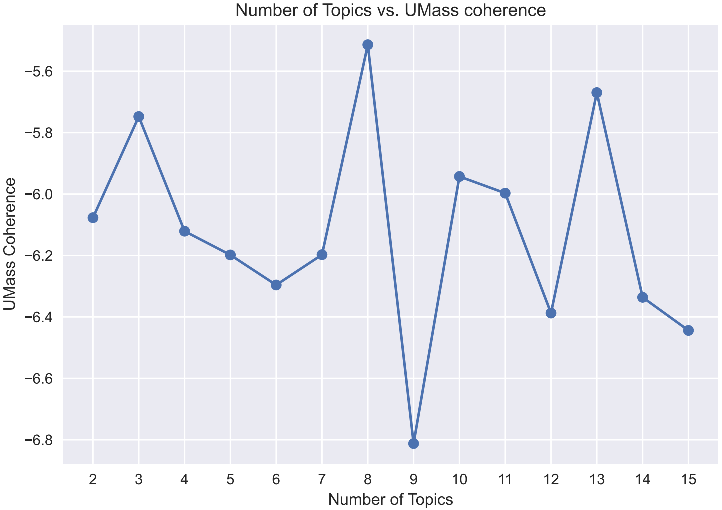


Plate notation Biterm

Choice of the number of topics

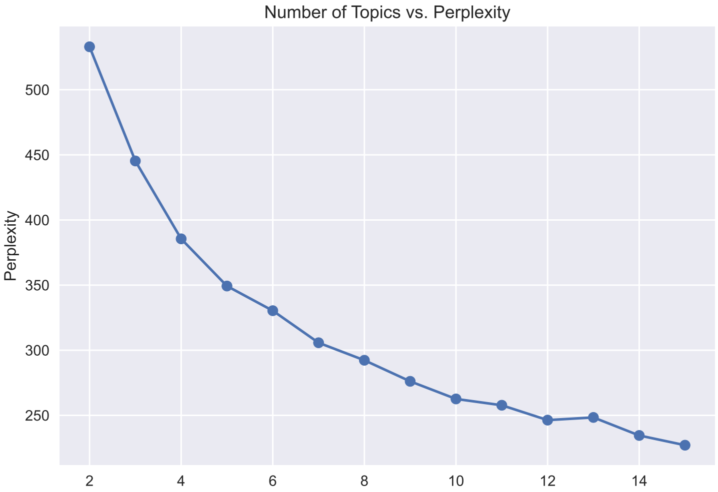
Coherence (Röder et al., 2015)



◂ Back

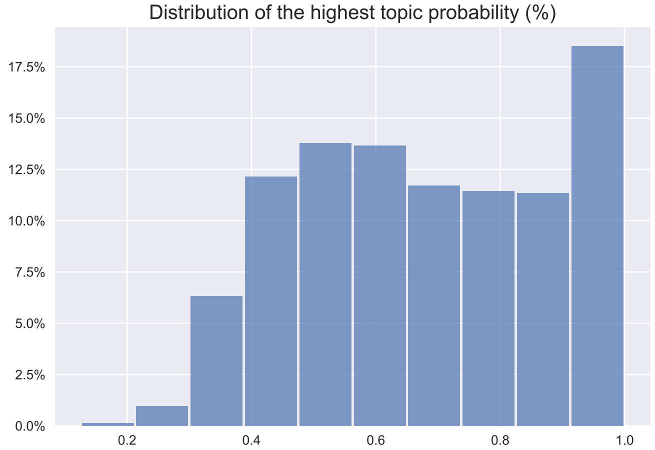
Choice of the number of topics

Perplexity



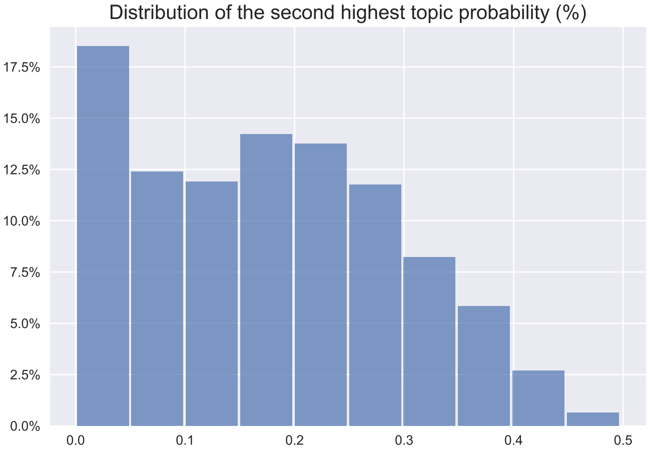
Back

Distribution of job ads over the topics



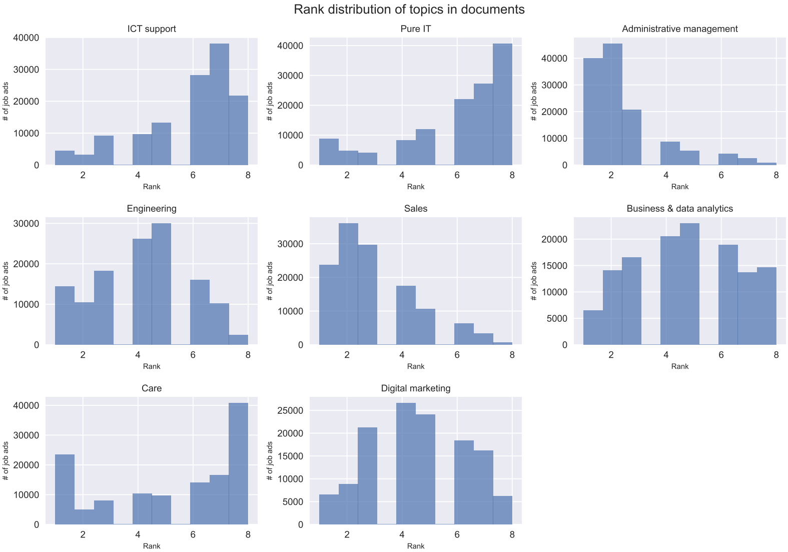
Back

Distribution of job ads over the topics



Back

Distribution of job ads over the topics



Back

Occupation classification

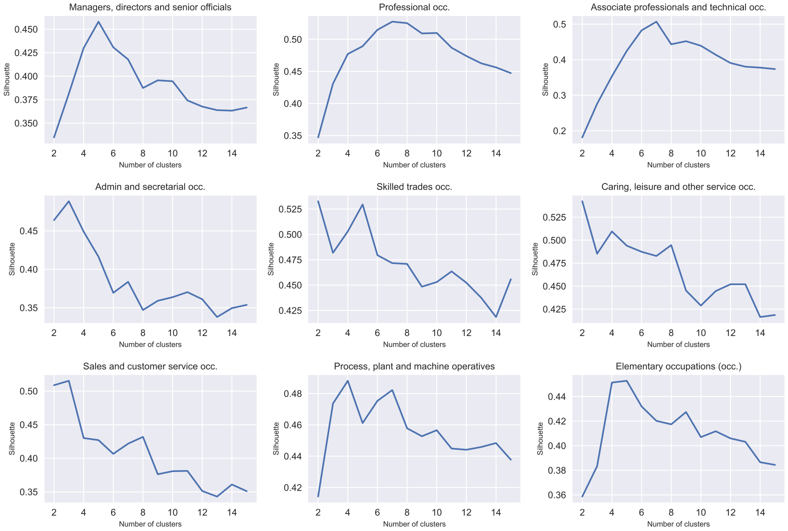
Occupations: **nine major groups** of the UK Standard Occupational Classification (SOC 2010)

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

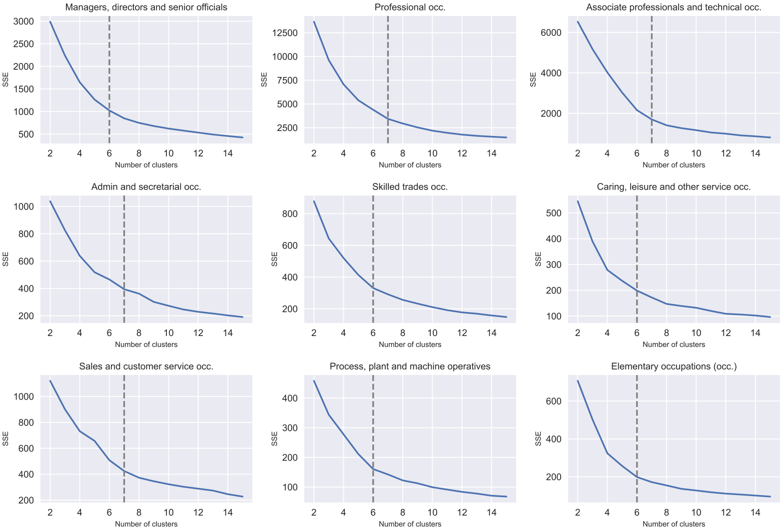
Clustering by occupations

Choice of k – Silhouette width

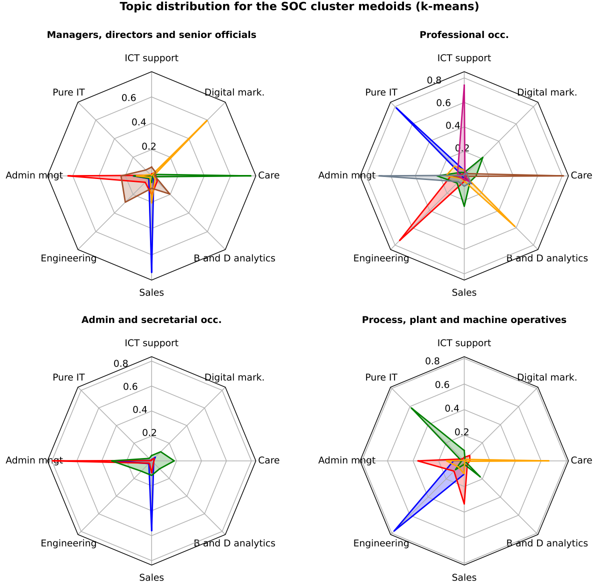


Clustering by occupations

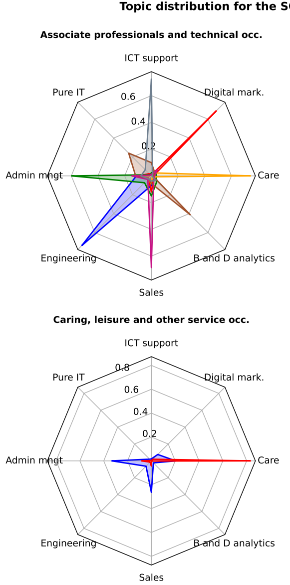
Choice of k – Sum of squared errors (SSE)



Skill sets within occupations (1/2)



Skill sets within occupations (2/2)



Clustering by occupations (all)