Karlsruhe Institute of Technology

A Patent Search Strategy based on Machine Learning for the Emerging Field of Service Robotics

September 16th 2015

Institute for Economics, Chair in Economic Policy

Vladimir Korzinov, Ingrid Ott, Florian Kreuchauff

Outline

- Motivation
- Service Robotics
- Methodology
  - Data Source
  - Patent Extraction
  - Structuring Text Data
- Machine Learning
- Results
- Conclusion
- Literature

Source: EFFIROB Fraunhofer IPA, page 33

Motivation and Research Gap

Problem: The technology field of Service Robotics so far is…

- neither part of any existing official industry, patent or trademark classication system
- nor of any concordances – for example ISI-OST-INPI and NACE-ISIC, see Hinze et al. (1997) and Schmoch et al. (2003)
- not to mention national account systems

… that allow to frame the technology field and to estimate the corresponding economic implications – similar to identification problem of nanotech some years ago, see Porter et al. (2008), Mogoutov & Kahane (2007), Arora et al. (2013).

At the same time, especially high technologies and expected future industries are in the core focus of (supra-)national **innovation policies**.

For being effective and efficient, these **policies strongly rely on credible data bases** that include entire value creation chains, starting from research and **development** over production and sales.

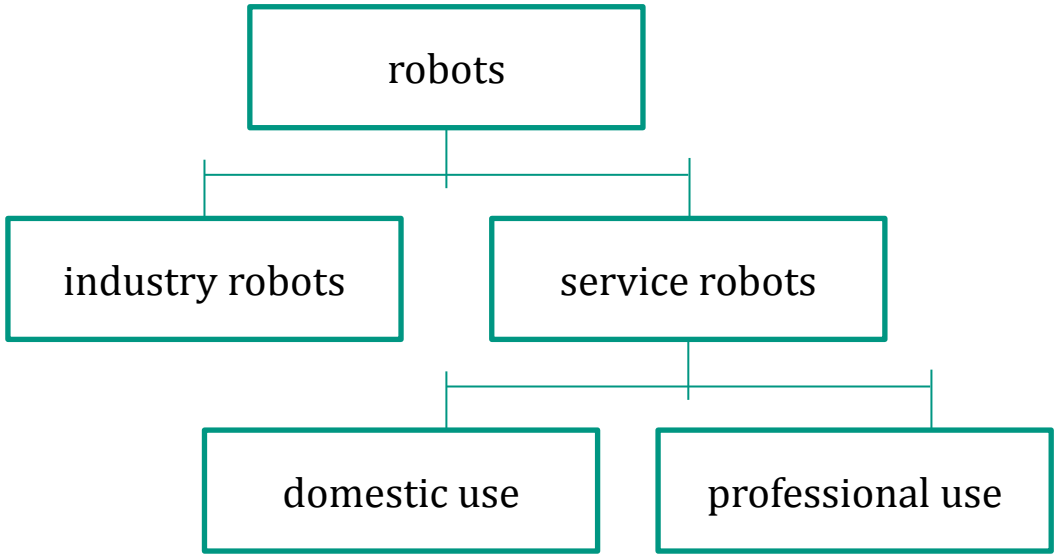
Service Robotics (SR)

- **Preliminary definition – *according to the International Federation of Robotics (IFR)***

A service robot…

- is a robot that **performs useful tasks** for humans or equipment
- **excluding industrial automation** applications.
- Industrial automation applications include, but are not limited to, manufacturing, inspection, packaging, and assembly

(compare **www.ifr.org** and ISO 8373:2012).



Source: Ott (2012)

Methodology 1/5

- **Data sources**
  - Patents: PATSTAT database October 2013 (European Patent Office)

- **Patent Extraction**
  - Step1: Retrieval of suspect set of potential SR patents
    a) Retrieval of patents classified as *robotics* according to EPO (IPC class **B25J** and '**robot**') from various leading countries (Japan, USA, Germany, Korea, etc…)
    b) Confinement via lexical query regarding hundreds of keywords as well as various IPC subclasses describing 11 SR application fields (according to IFR)

```
…(TTL.appln_title LIKE '%clean%' AND NOT (TTL.appln_title LIKE '%house%' OR TTL.appln_title LIKE '%domestic%' OR
TTL.appln_title LIKE '%pool%')) OR
(TTL.appln_title LIKE '%climb%' AND NOT TTL.appln_title LIKE '%wheelchair%')
AND NOT (TTL.appln_title LIKE '%vacuum%' OR AST.appln_abstract LIKE '%vacuum%' OR TTL.appln_title LIKE '%wafer%' OR
AST.appln_abstract LIKE '%wafer%' OR TTL.appln_title LIKE '%semiconductor%' OR AST.appln_abstract LIKE '%semiconductor%'
AND AST.appln_abstract LIKE '%industr%' AND AST.appln_abstract LIKE '%milk%' OR TTL.appln_title LIKE '%milk%' OR
AST.appln_abstract LIKE '%paint%' OR TTL.appln_title LIKE '%paint%' OR AST.appln_abstract LIKE '%weld%' OR
TTL.appln_title LIKE '%weld%' OR …
…
(SUBSTRING(IPC.ipc_class_symbol,1,5)='"A61B' AND NOT IPC.ipc_class_symbol LIKE '%A61B 5%') OR
SUBSTRING(IPC.ipc_class_symbol,1,5)='"A61C' OR SUBSTRING(IPC.ipc_class_symbol,1,5)='"A61D' OR
SUBSTRING(IPC.ipc_class_symbol,1,5)='"A61F' …
```


Institute for Economics
Chair in Economic Policy

Methodology 2/5

c) Screening of resulting patent set by an **independent** group of robotic **experts** – inter alia

- *High Performance Humanoid Technologies* (H2T) from the *Institute for Anthropomatics and Robotics* at KIT, and the
- *Delft Center for Systems and Control / Robotics Institute at TU Delft, NL*

Running robot Phides. Source: TU Delft

ARMA. Source: KIT

d) Identification of **core set of SR patents** (at the moment **98**) and complementing industrial robotics set (at the moment **130**)

- Step 2: Transforming unstructured document text into structured data
  a) Removing Stop-words
  b) Extraction of apt **n-grams** of words in patent abstracts and titles and, if possible, in patent claims (language and thus patent authority dependent)
  c) Stemming: reducing inflected words to their **stem**
  d) Deriving **frequencies** for each document

Structured Data

patent	Attribute vectors **x**									binary
	$word_{w1}$	$word_{w2}$	...	$bigram_{b1}$	$bigram_{b2}$	...	$trigram_{t1}$	$trigram_{t2}$	...	decision y
1	$freq._{1\|w1}$	$freq._{1\|w2}$	...							1
2	$freq._{2\|w1}$	...								-1
...	...									...
205	$freq._{205\|w1}$	...			$freq._{205\|b2}$	...				-1
206	$freq._{206\|w1}$	...			...					1
...	...									...
228	$freq._{228\|w1}$	...						$freq._{228\|t2}$	...	-1
xxx	$freq._{xxx\|w1}$	...						...		?
xxx	$freq._{xxx\|w1}$	...								?
...	...									...

Institute for Economics
Chair in Economic Policy

Methodology 4/5

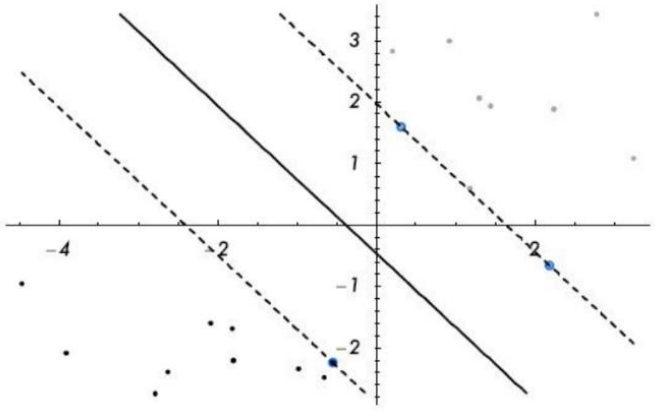
- **Classification of patents**
  - Step1: Identifying appropriate machine learning / classification algorithms
    1. **K-Nearest neighbour**
       - Sensitive to irrelevant or redundunt features
    2. **Neural Networks**
       - Gets stuck in local maximums
       - Requres very large training sets
    3. **Genetic algorithms**
       - No assurance of optimum solution
       - Hard to find a best fit function
    4. **Support Vector Machine**
       - No local maxima
       - Problem of overfitting

       + Fast classification
       + High precision
       + Clear approach and solution
       + Effective for significant number of attributes

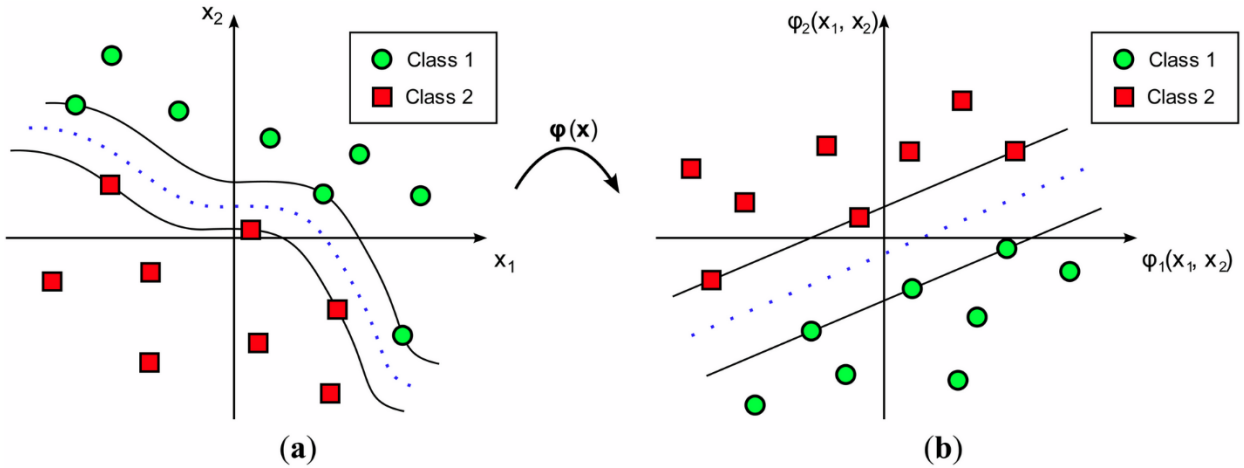
Institute for Economics
Chair in Economic Policy

Methodology 5/5

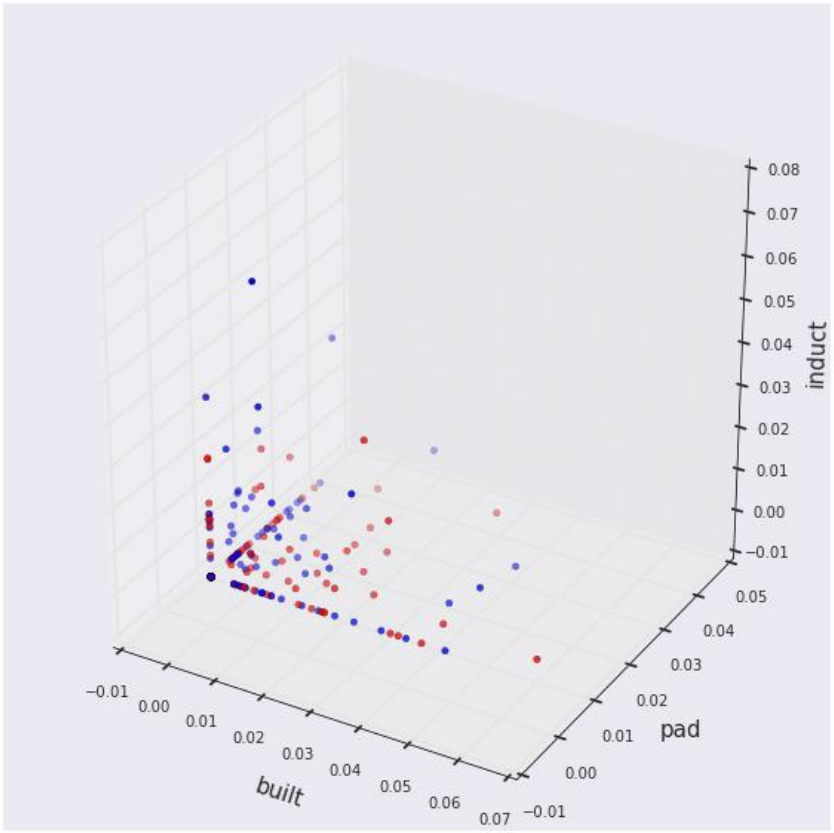
- Step2: Applying Support Vector Machine
  - Formally defined as a minimization problem with a constrain
  - Constructs a **hyperplane in a high-dimensional space** for separation between to classes
  - Maximization of distance to the nearest training data (here: patent vector) point of any class to achieve best discrimination

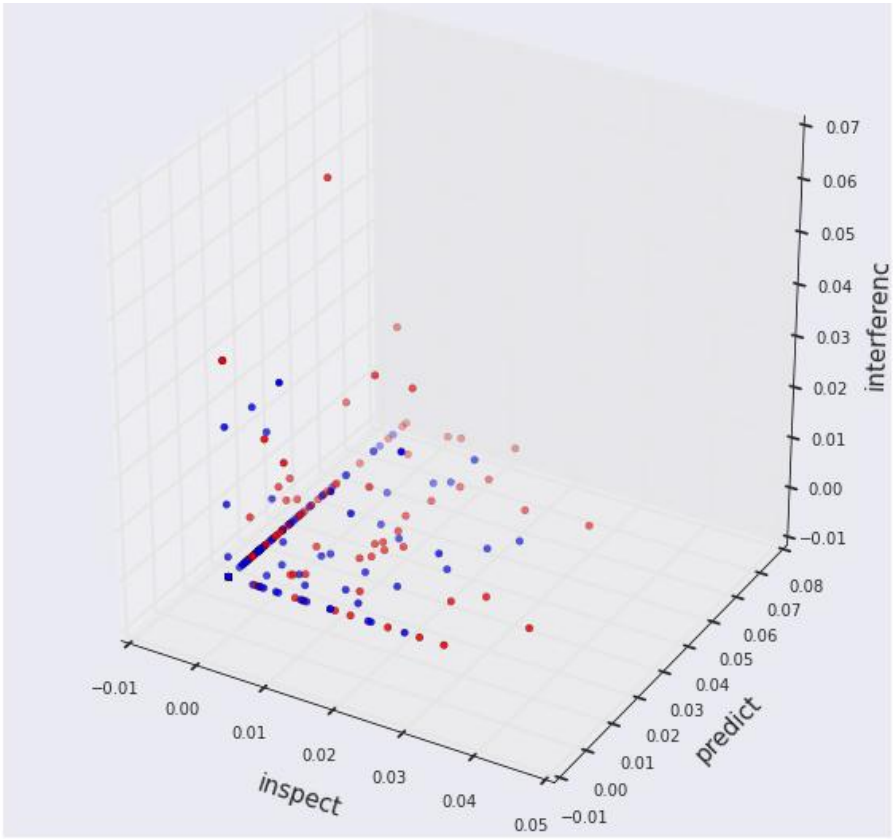


Source: Nilsson et al. (2006)

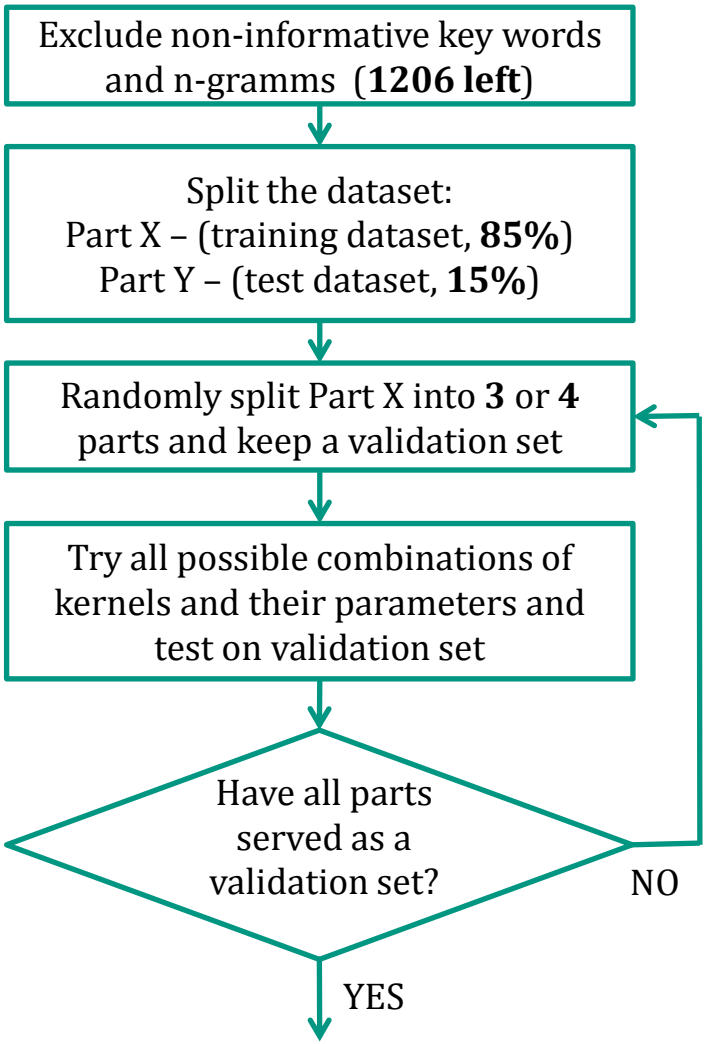


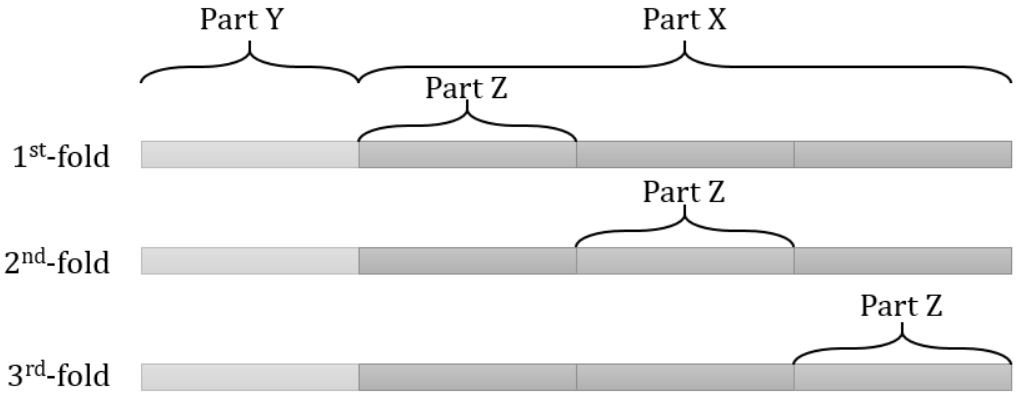
SVM Input





Machine Learning Algorithm 1/2

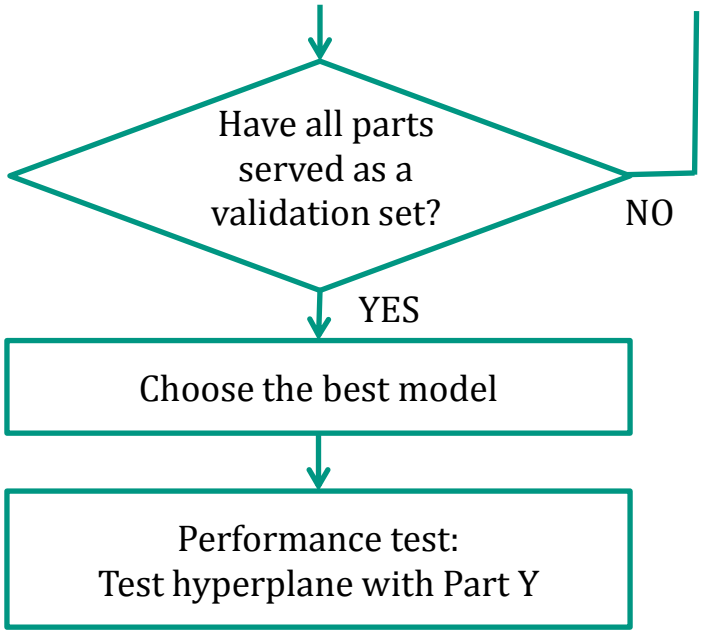




Parameter	Values
CV (cross-validation)	3,4
C (complexity)	10,… , 1000
γ of rbf kernel	1e-6, … , 1e-2
γ of polynomial kernel	1e-6, … , 1e-2
d of polynomial kernel	1,2,3
r of polynomial kernel	1,2,3
γ of sigmoid kernel	1e-6, … , 1e-2
r of sigmoid kernel	1,2,3

Institute for Economics
Chair in Economic Policy

Machine Learning Algorithm 2/2



The best hyperplane:

Parameter	Values
Kernel function	rbf
C (complexity)	10 (smooth surface)
γ	0.005

	Precision	Recall	f1-score	Number of Patents in Test Set
SR	75 %	94 %	83 %	16
IR	93 %	74 %	82 %	19
Avg. / total	**85 %**	**83 %**	**83 %**	35

Precision = tp/ (tp+fp)
Recall = tp/ (tp+fn)
f1-score = 2 * (Precision*Recall) / (Precision + Recall)

Institute for Economics
Chair in Economic Policy

Results
Examples of new classified patents

Title	Application ID	Filing Date
'Autonomous surface cleaning robot for wet and dry cleaning.'	273225364	2006-02-21
'A mobile robot … arranged to operate in an environment is described as well as a method for building a map …'	16204492	2005-12-28
'The present invention relates to apparatus for milking an animal and provides a milking parlour…'	16410754	2007-11-16
'The present invention provides a technique that makes a robot continue a stabilized walk…'	56207682	2007-07-17
'Medical tele-robotic system. A remote controlled robotic system that is coupled to a broadband network…'	315565124	2003-07-25
'Surgical robotic tools; data architecture; and use…'	267647377	1999-12-06

Conclusions

- Classifier between industrial and service robotic patents
- Advantages
  - Experts do not choose the terms
  - Not limited to Service Robotics
  - Not limited to patents
- Current work
  - Analysis of the new classified patents
- Future work
  - Increase the initial training sample
  - Extend to publications

Thank you for attention.

QUESTIONS?

Literature 1/2

- Arora, S. K., Porter, A. L., Youtie, J. and Shapira, P.: 2013, Capturing new Developments in an Emerging Technology: An Updated Search Strategy for Identifying Nanotechnology Research Outputs, *Scientometrics* 95, 351-370.
- Autor, D. H., Levy, F. and Murnane, R. J.: 2003, The Skill Content of Recent Technological Change: An Empirical Exploration, *The Quarterly Journal of Economics* 118(4), 1279-1333.
- Bresnahan, T. and Trajtenberg, M.: 1995, General Purpose Technologies: 'Engines of Growth'?, *Journal of Econometrics* 65, 83–108.
- Casella, G. and Berger, R. L.: 2002 , *Statistical Inference*, 2nd ed., Belmont, CA: Duxbury Press.
- Chih-Wie Hsu, Chih-Chung Chang, and Chih-Jen Lin 2003, A Practical Guide to Support Vector Classification*, Bioinformatics*
- David, P.: 1991, Computer and dynamo: The modern productivity paradox in a not-toodistant mirror, *Technology and Productivity: The Challenge for Economic Policy* pp. 315–347.
- David, P. and Wright, G.: 1999, General purpose technologies and surges in productivity: Historical reflections on the future ot the ICT revolution, Paper presented to the International Symposium on 'Economic Challenges of the 21st Century in Historical Perspective', Oxford, UK.
- Erdi, P., Makovi, K., Smomogyvari, Z., Strandburg, K., Tobochnik, J., Volf, P. and Zalangi, L.:2013, Prediction of Emerging Technologies Based on Analysis of the US Patent Citation Network, *Scientometrics* 95, 225-242.
- Griliches, Z.: 1990, Patent statistics as economic indicators: A survey, *Journal of Economic Literature* 28, 1661–1707.
- Haykin, S.: 1999, *Neural Networks: A Comprehensive Foundation*, 2nd ed., Englewood Cliffs, NJ: Prentice Hall.

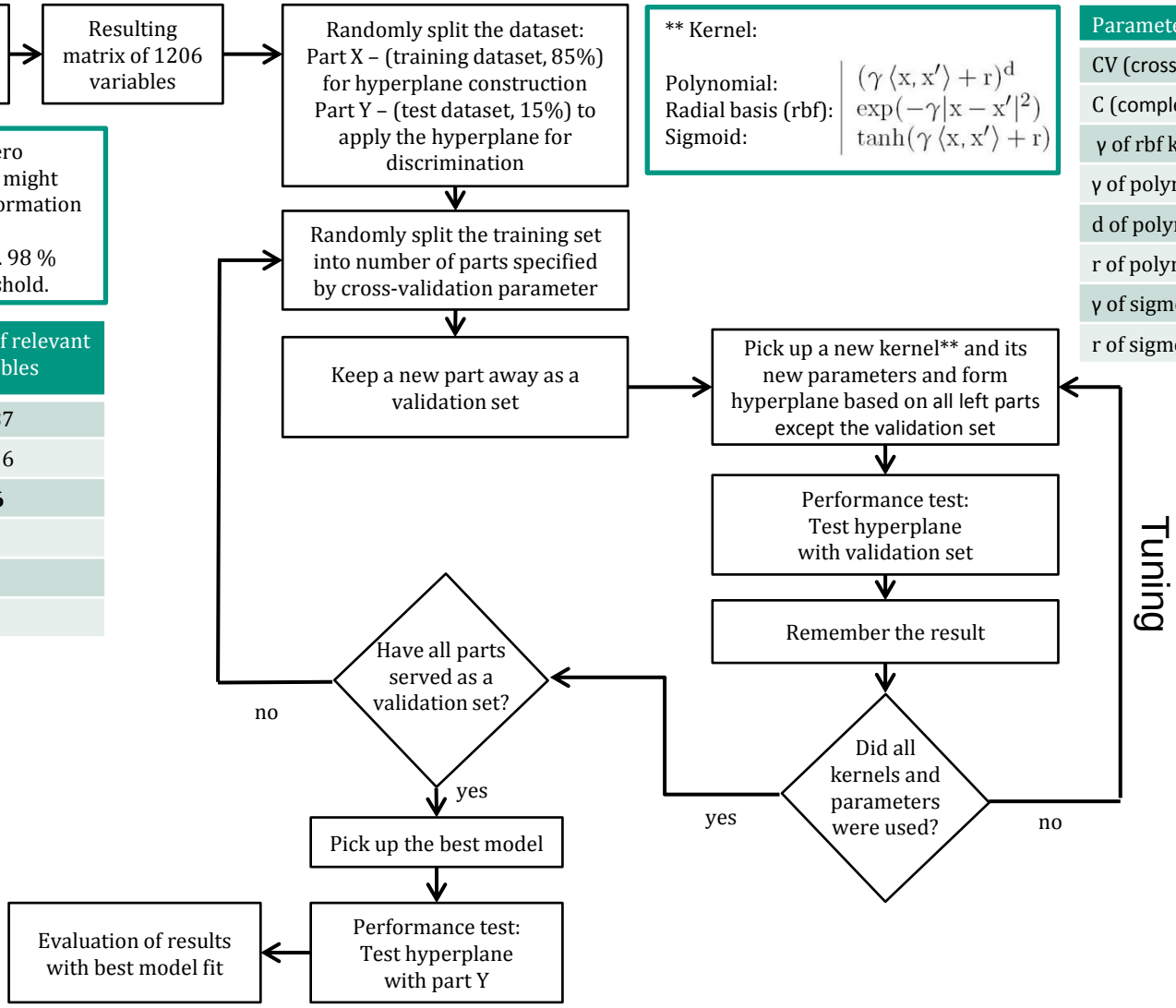
Literature 2/2

- Hinze, S., Reiss, T. und Schmoch, U.: 1997, Statistical Analysis on the Distance between Fields of Technology, Report for European Commission TSER Project - Technological Entry: Diversication vs. New Innovators.
- Helpman, E. and Trajtenberg, M.: 1998, A time to sow and a time to reap, *General Purpose Technologies and economic growth*, Helpman, Elanahan, Cambridge, pp. 55–83.
- Huang, C., Notten, A. and Rasters, N.: 2010, Nanoscience and technology publications and patents: a review of social science studies and search strategies, *Journal of Technology Transfer* .
- Jovanovic, B. and Rousseau, P.: 2005, General Purpose Technologies, *in* P. Aghion and S. Durlauf (eds), *Handbook of Economic Growth*, Vol. 1B, Elsevier, Amsterdam.
- Mogoutov, A. and Kahane, B.: 2007, Data Search Strategy for Science and Technology Emergence: A Scalable and Evolutionary Query for Nanotechnology Tracking, Research Policy 36, 893-903.
- Nilsson, R., Björkegren, J. and Tegnér, J.: 2006, A Flexible Implementation for Support Vector Machines, *The Mathematica Journal* 10(1) Wolfram Media, Inc.
- Ott, I.: 2012, Service Robotics: An Emergent Technology Field at the Interface between Industry and Services, Poiesis and Praxis 9(3-4), 219-229.
- Porter, A., Youtie, J., Shapira, P. and Schoeneck, D.: 2008, Refining Search Terms for Nanotechnology, *Journal of Nanoparticle Research* 10, 712–728.
- Schmoch, U., Laville, F., Patel, P. and Frietsch, R.: 2003, Linking Technology Areas to Industrial Sectors, Final Report to the European Commission, DG Research.

Backup slides
Additional Definitions (ISO 8373:2012)

- **Robot:**
  - actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks
    - Note 1 to entry: A robot includes the control system and interface of the control system.
    - Note 2 to entry: The classification of robot into industrial robot or service robot is done according to its intended application.
- **Autonomy:**
  - ability to perform intended tasks based on current state and sensing, without human intervention
- **Control System:**
  - set of logic control and power functions which allows monitoring and control of the mechanical structure of the robot and communication with the environment (equipment and users)
- **Robotic Device**
  - actuated mechanism fulfilling the characteristics of an industrial robot or a service robot, but lacking either the number of programmable axes or the degree of autonomy

Backup slides
Machine Learning Algorithm



** Kernel:

Polynomial:	$(\gamma \langle x, x' \rangle + r)^d$
Radial basis (rbf):	$\exp(-\gamma \lvert x - x' \rvert^2)$
Sigmoid:	$\tanh(\gamma \langle x, x' \rangle + r)$

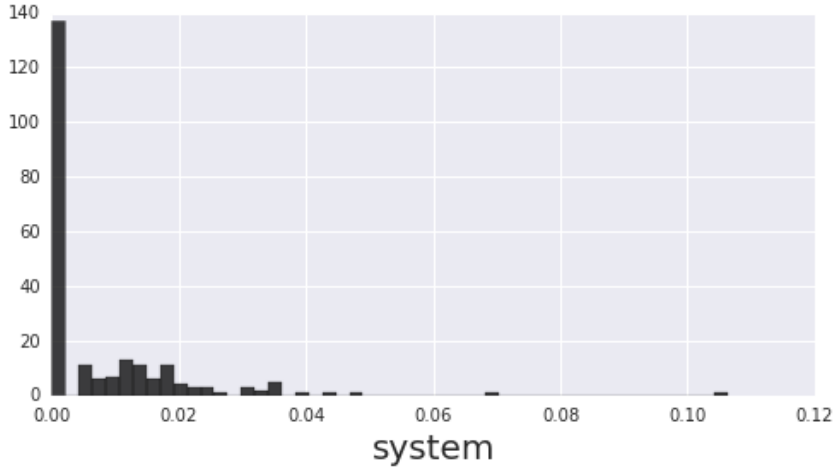
Parameter	Values
CV (cross-validation)	3,4
C (complexity)	10,… , 1000
γ of rbf kernel	1e-6, … , 1e-2
γ of polynomial kernel	1e-6, … , 1e-2
d of polynomial kernel	1,2,3
r of polynomial kernel	1,2,3
γ of sigmoid kernel	1e-6, … , 1e-2
r of sigmoid kernel	1,2,3

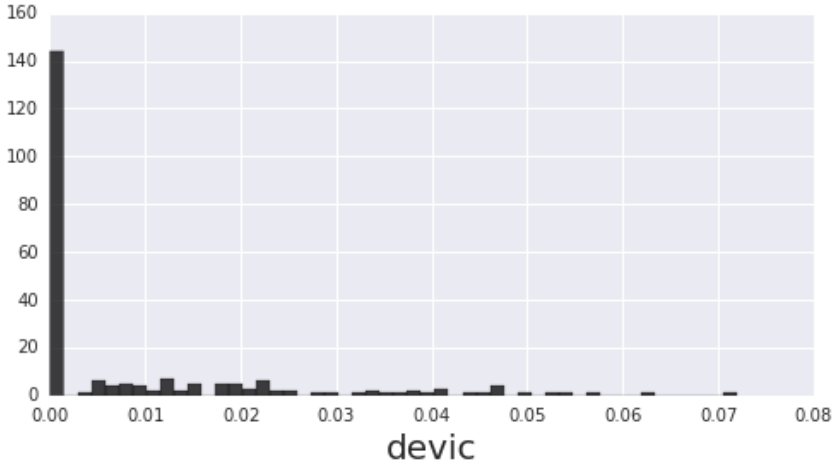
* If there are mostly zero entries in a variable, it might not contain a lot of information for classfication and is considered redundant. 98 % were chosen as a threshold.

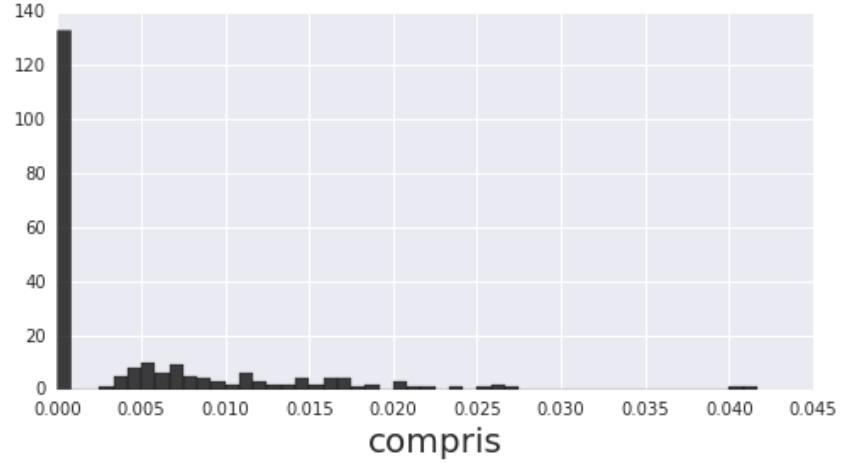
Percentage	No. of relevant variables
100	30987
99	14116
98	**1206**
95	252
90	99
80	28

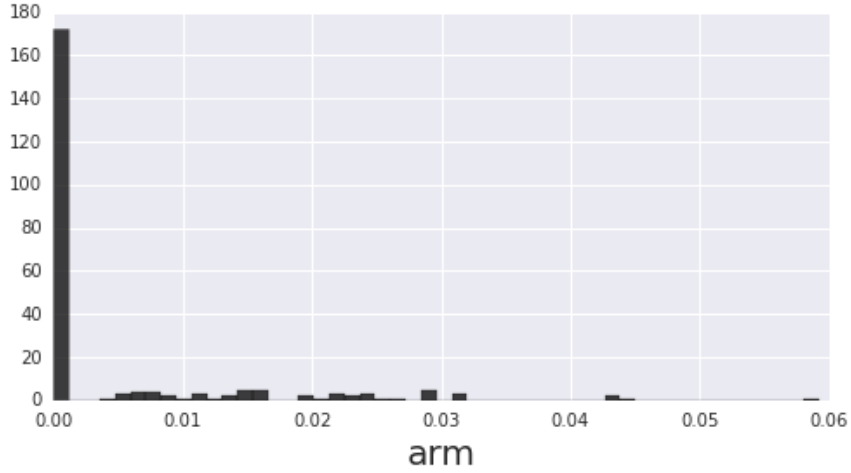
Institute for Economics
Chair in Economic Policy

Backup slides
Key words statistic

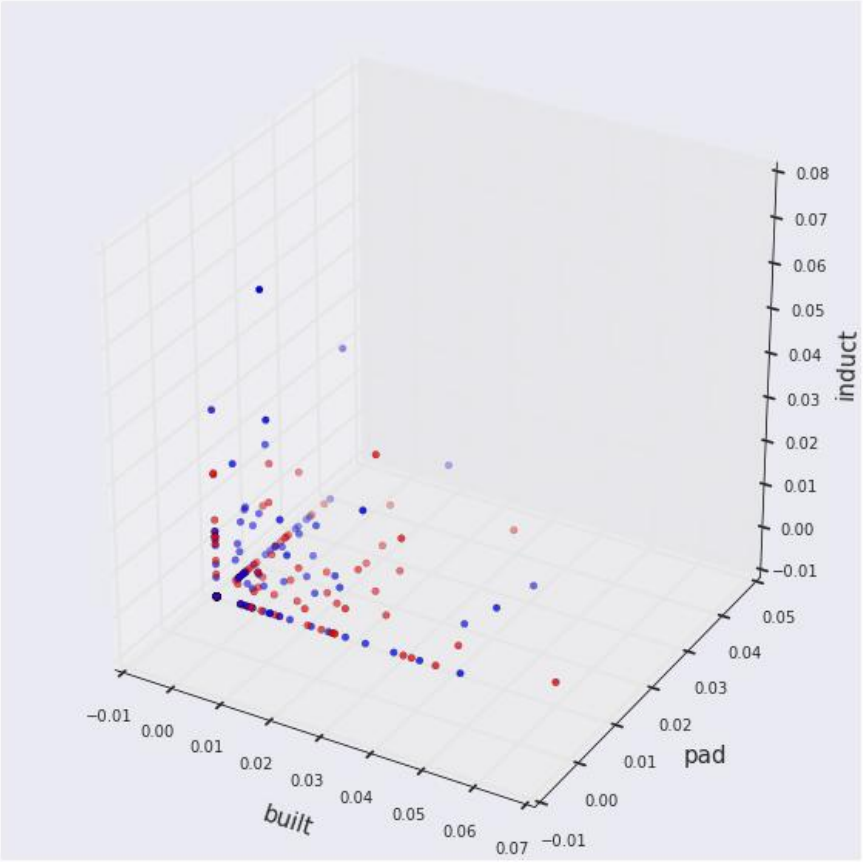


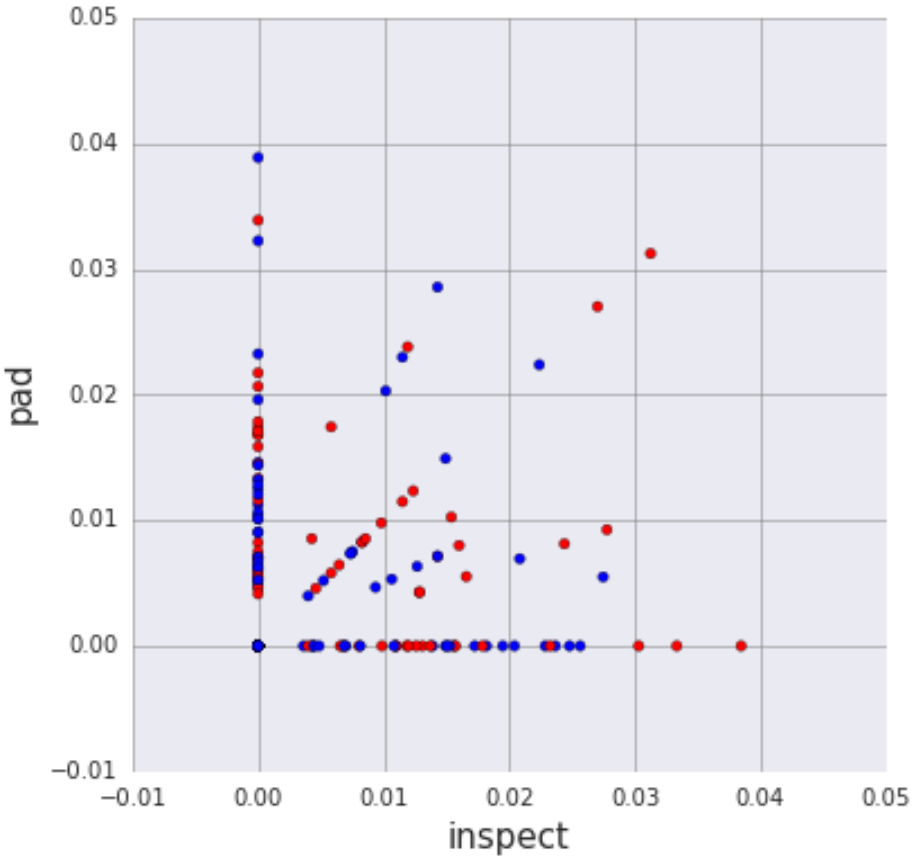






Backup slides
Key words statistic





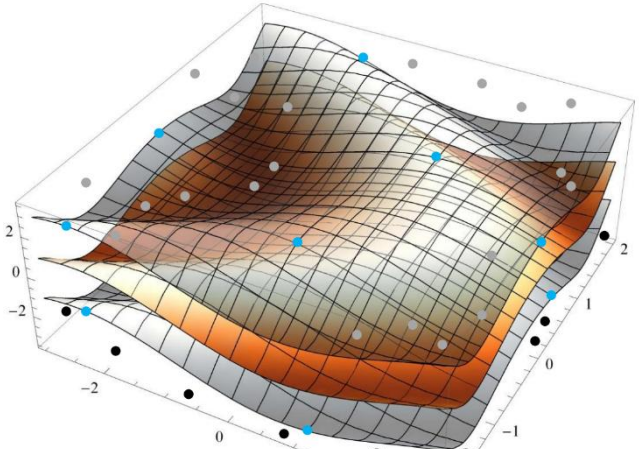
Institute for Economics
Chair in Economic Policy

Backup slides
Support Vector Machine

Formally defined, given a training set of pairs $(x_i, y_i), i = 1, ..., l$ where $x_i \in R^n$ and $y \in \{-1, 1\}^l$, SVM requires a solution to the following optimization problem:

$$\underset{w,b,\xi}{\text{minimize}} \quad \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i$$
$$\text{subject to} \quad y_i(w^T \phi/(x_i) + b) \geq 1 - \xi_i$$
$$\xi_i \geq 0.$$

Source: C Hsu et. al (2003)




```
SVC(C=10, cache_size=200, class_weight='auto', coef0=0.0, degree=3,
  gamma=0.005, kernel='rbf', max_iter=-1, probability=False,
  random_state=None, shrinking=True, tol=0.001, verbose=False)
```

Number of support vectors - 192

Which patents where significant for classification problem?

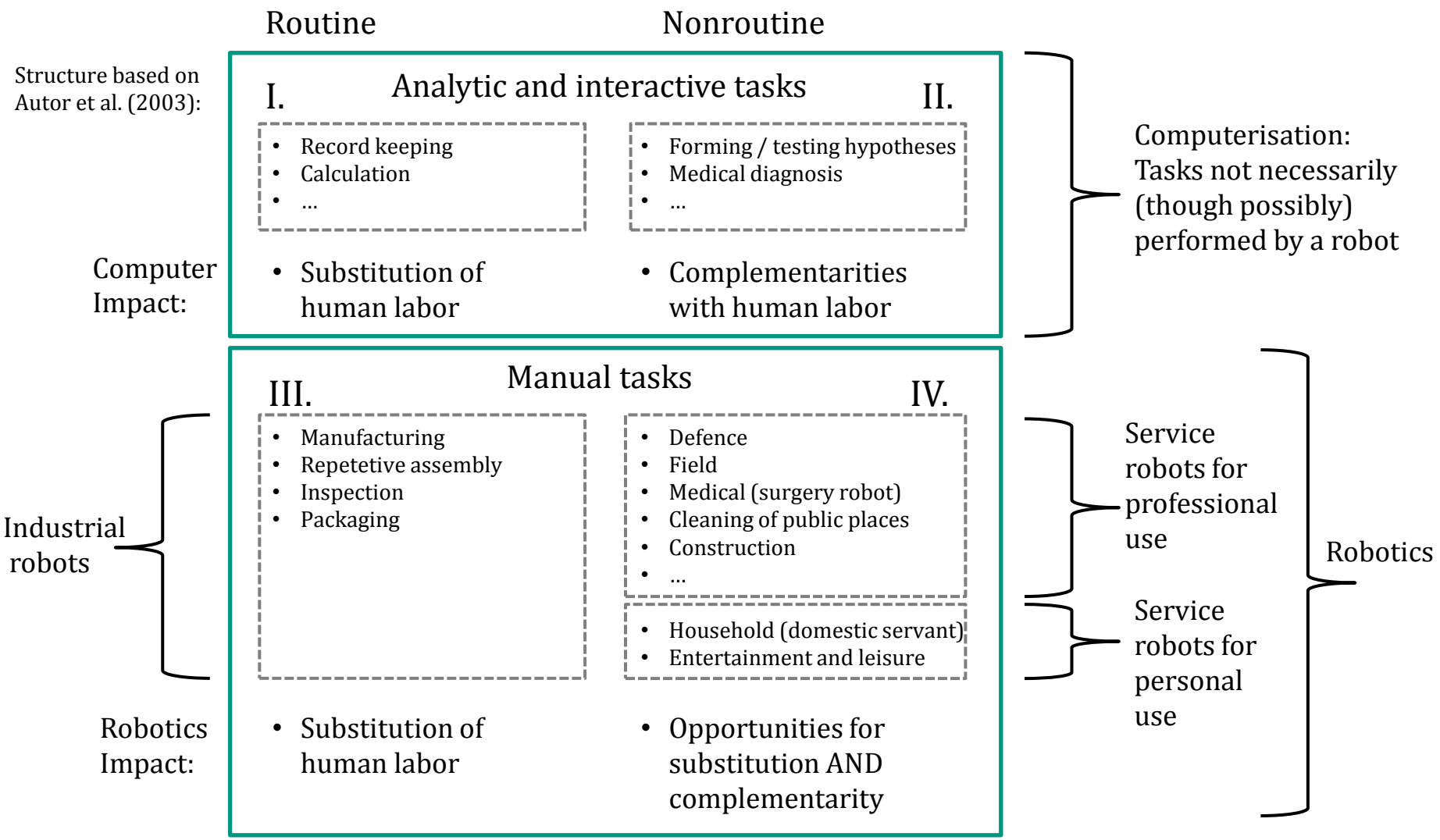
Indexes of support vectors -

```
[  2   4   6   7   8   9  10  12  15  16  19  20  23  26  27  28  29  30
  31  32  34  35  36  38  40  41  42  43  44  45  46  47  48  49  50  52
  54  56  59  60  63  64  65  67  69  71  73  76  78  80  85  87  88  89
  93  96  98 100 101 102 104 107 110 112 113 114 115 119 122 124 125 127
 129 133 134 137 138 139 142 143 146 148 150 152 154 155 156 158 159 162
 163 164 169 171 172 175 176 178 179 181 182 184 189 190 192 193 194 195
 196 197 199 200 203   0   1   3   5  13  14  17  18  21  22  24  25  33
  37  51  53  55  57  58  61  62  66  68  70  72  74  75  77  79  81  82
  83  84  86  91  92  94  95  97 106 108 109 116 118 120 121 123 126 128
 131 135 140 141 144 145 147 149 151 153 157 160 161 165 166 167 168 170
 173 174 177 180 183 186 187 188 191 198 201 202]
```

Backup slides
Examples of used key-words and n-gramms

abnormal	characteris	fasten	accord,invention	devic,robot,arm
accelerat	characteristic	featur	actual,position	effector,control,input
access	characteriz	feedback	actuat,control	effector,correspond,movem
accommodat	charg	field	addition,equipment	ent
accord	chassi	fig	adjust,position	effector,handl,move
accordanc	check	figur	adjustabl,surgeon	effector,manipulat,hold
accurat	circuit	fill	allow,surgeon	effector,move,button
achiev	claim	filter	angl,adjust	effector,movement,handl
acquir	clamp	finger	apparatus,compris	effector,movement,moveme
act	clean	fit	apparatus,method	nt
action	cleaner	fix	apparatus,perform	factor,adjustabl,surgeon
activ	climb	flang	arm,coupl	front,robot,arm
actual	clip	flat	arm,includ	hand,surgeon,scale
actuat	close	flexibl	arm,instrument	handl,controller,handl
adapt	coat	floor	arm,join	handl,move,comfortabl
adapter	code	flow	assembl,method	handl,move,effector
addition		fluid	automatic,clean	handl,move,surgeon
			thereof,invention	instrument,coupl,pair

Backup slides
Paradigm Changes in Employment Structures

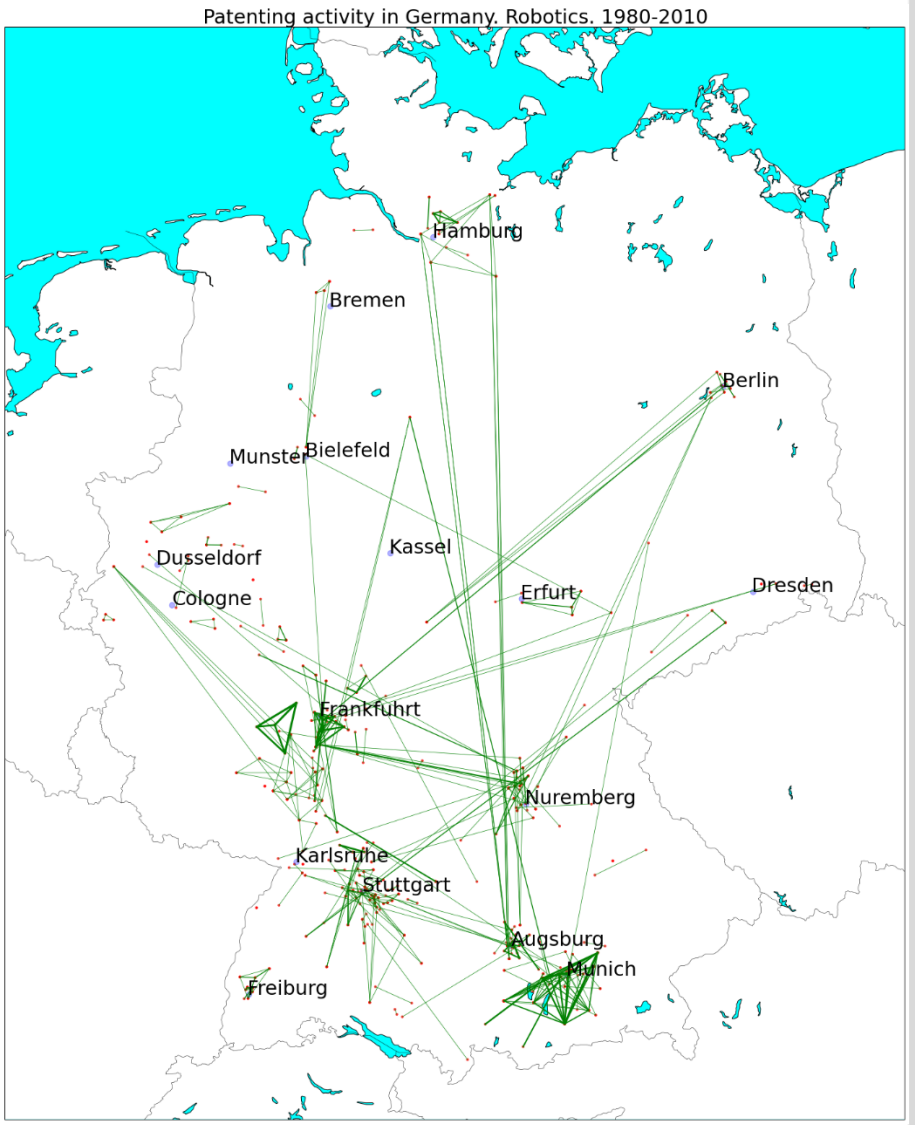


Backup slides
Employment Structures

- How the emergence and diffusion of innovations into the production process challenge the established division of tasks between human and machines and creates societal opportunities and challenges?
- Does innovations of a certain kind, namely in robotics and service robotics, influence a shift in skills of workers employed by firms?
- Data:
  - Database from German Institute for Employment Research (IAB)
  - PATSTAT Database from European Patent Office

Backup slides
Innovation Networks in Robotics

- Example:
  - Co-patenting network of inventors in Germany
  - 1980-2010
- Who are the key players in the field and how do they interact?
- Where in the world is the activity clustered and how does knowledge spill over?
- What kind of patterns can be found concerning developments over time and which application fields prevail?

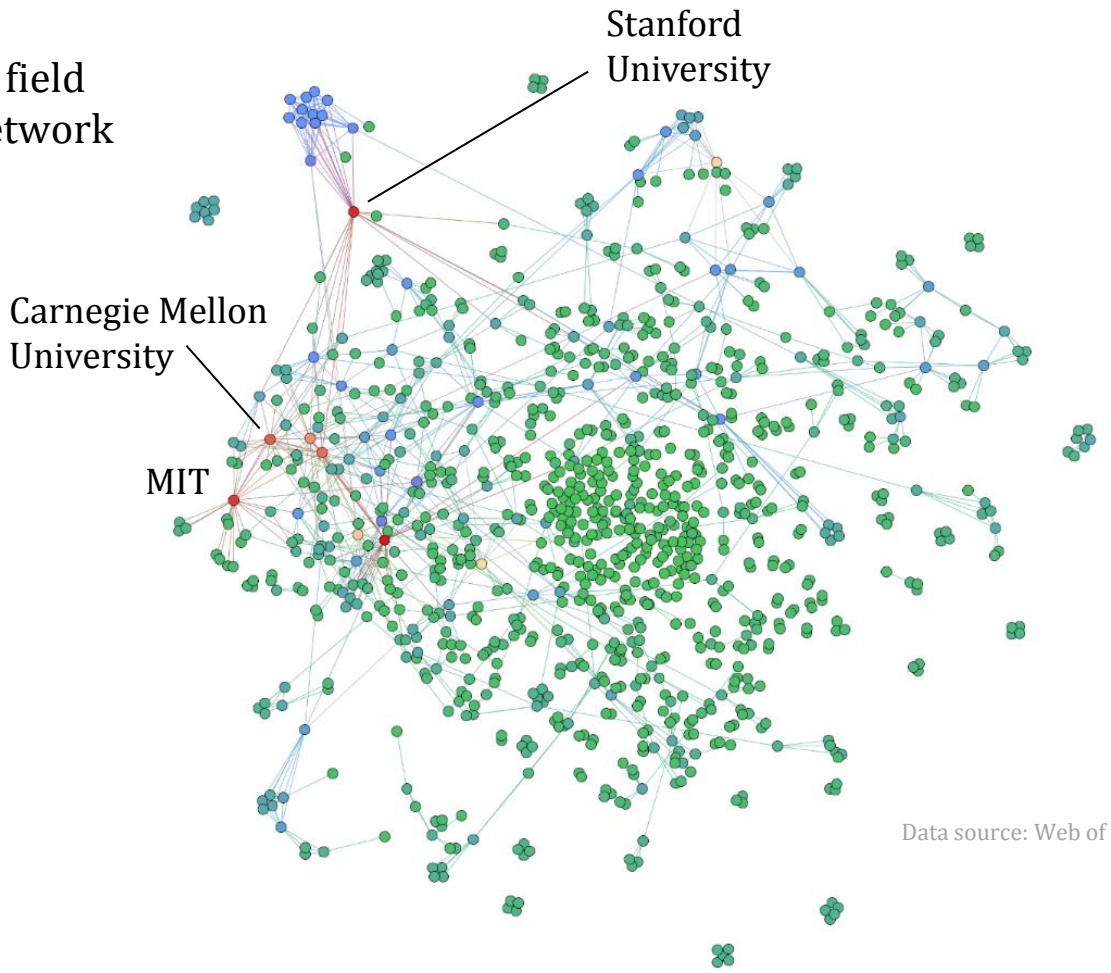


Backup slides
Analyses – Innovation networks

- Identifying key players in the field utilizing methods of Social Network Analyses (SNA)

	Name	Number of publications
1	Vukobratovic, M	93
2	Fukuda, T	68
3	Dawson, D	62
4	Arimoto, S	52
5	Woo, S	45
6	Ohnishi, K	44
7	Dario, P	40
8	Korayem, M	39
9	Ishiguro, H	38
10	Kumar, V	37

Individuals

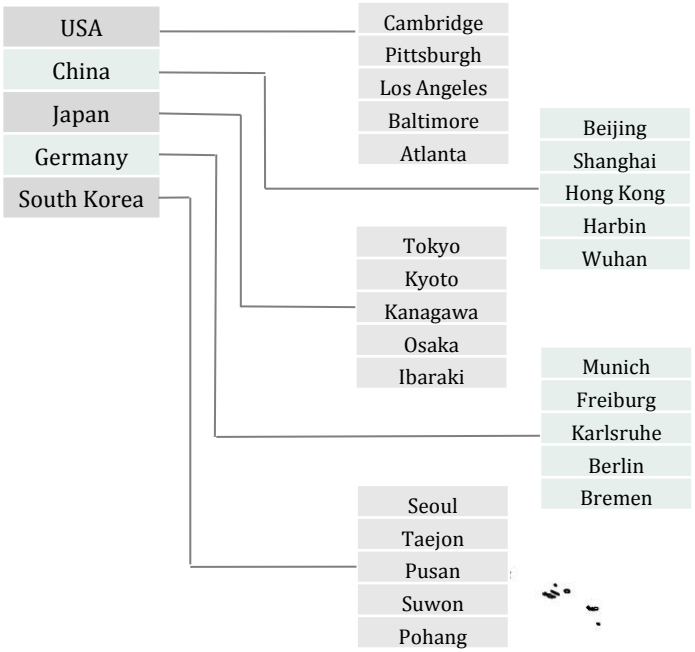


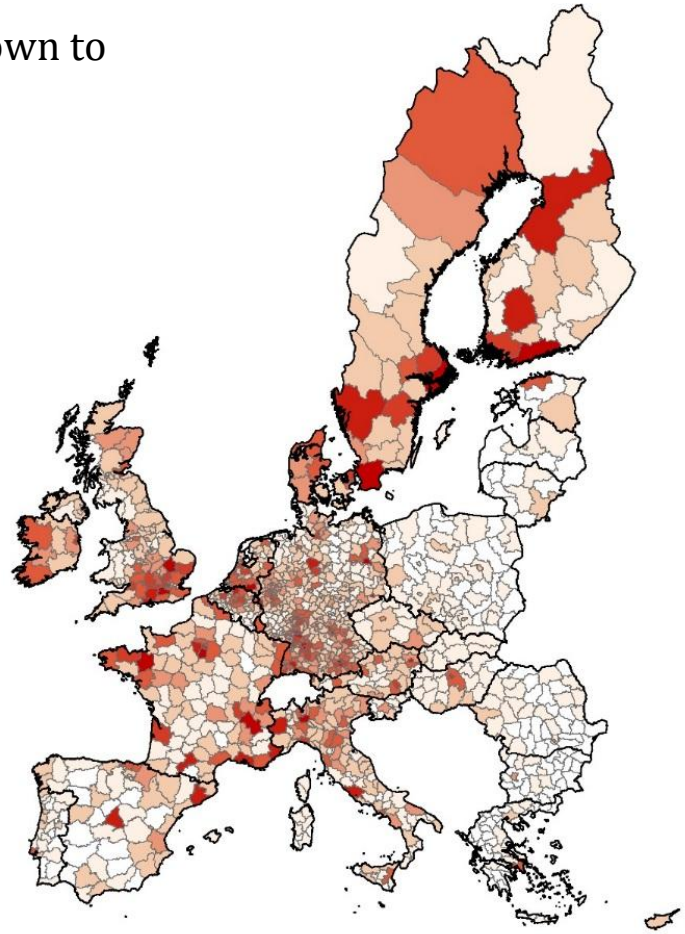
Data source: Web of Science

Organizations

Backup slides
Analyses – Geographical perspective

- Measuring regional allocation of activity (down to NUTS-3 level regarding regional policies)

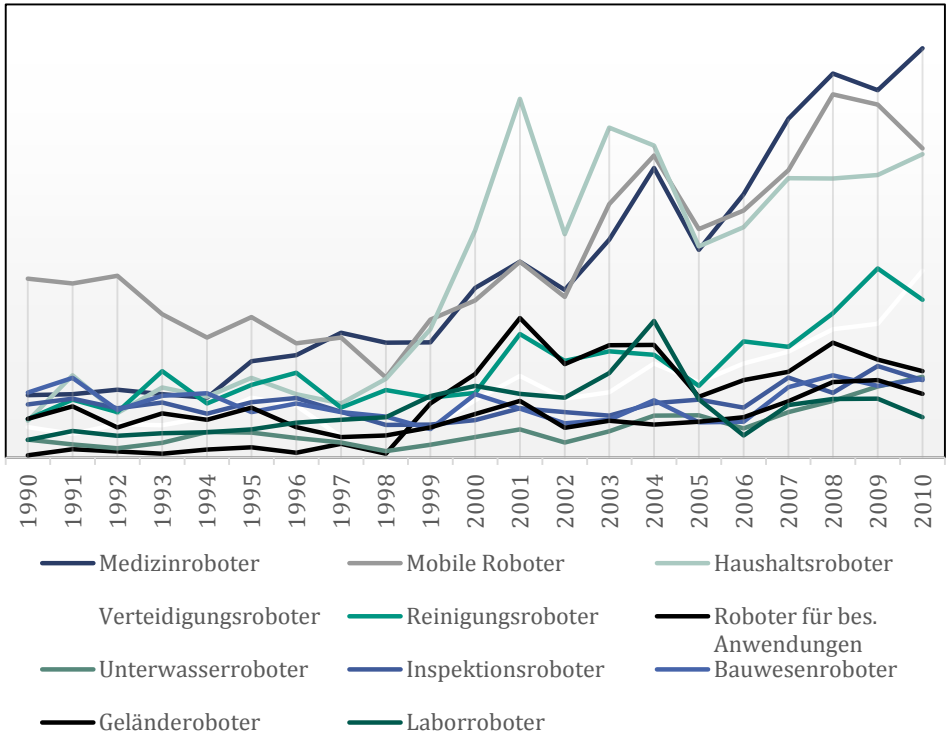


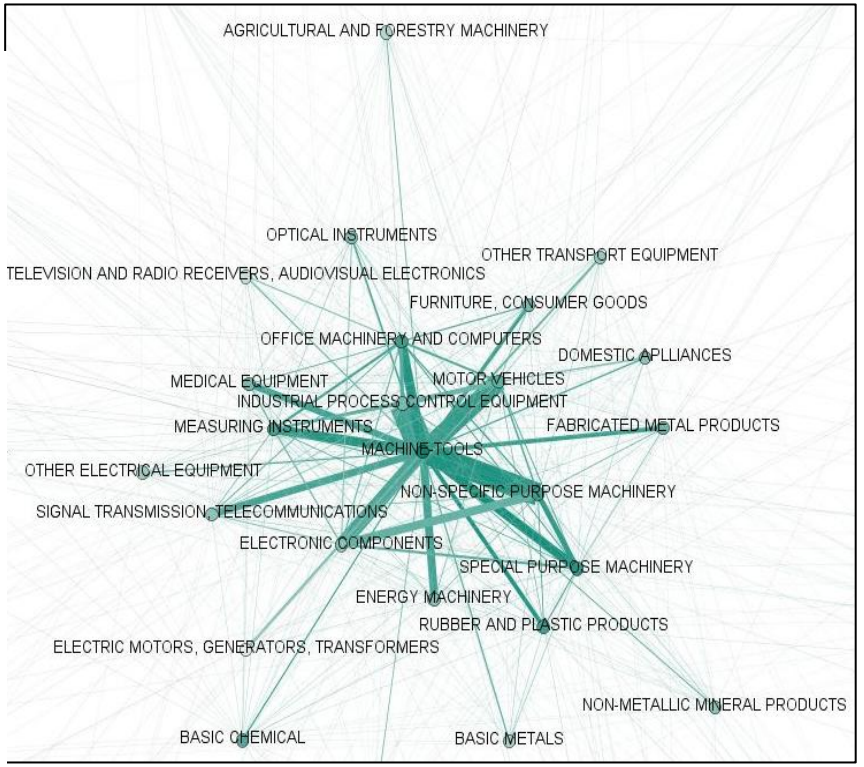


Data sources: Web of Science, Patstat

Backup slides
Analyses – Knowledge structures

- Identifying patterns of developments over time and in application fields, scientific areas and functional categories





Data source: Web of Science, Patstat

Institute for Economics
Chair in Economic Policy