INTERUNIVERSITY PH.D. PROGRAM BETWEEN POLITECNICO DI BARI AND UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO IN INDUSTRY 4.0

Cycle XXXVI

Assessment of Italian Intellectual Property Data: A Machine Learning Approach

Ph.D candidate
Annamaria Demarinis Loiotile

Tutors
Roberto Bellotti
Nicola Amoroso

Research seminar
02/03/2022

Politecnico di Bari

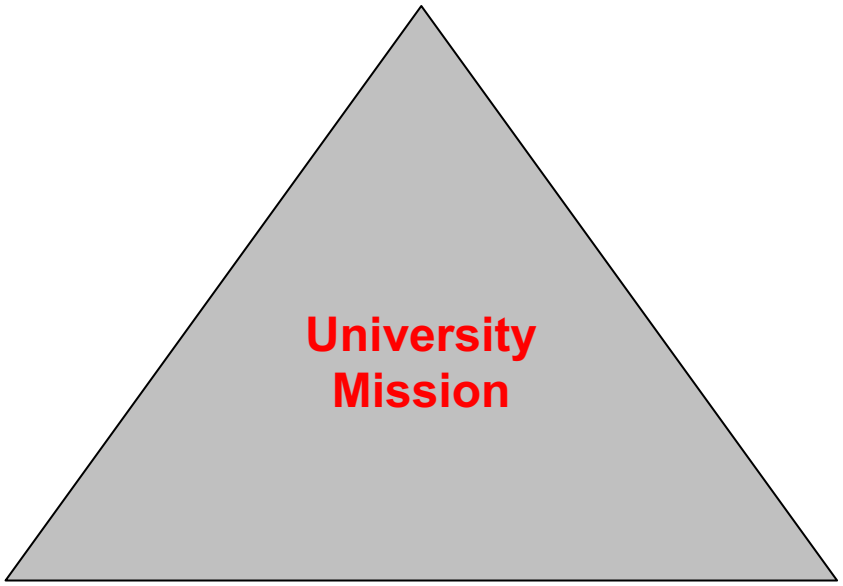
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Abstract

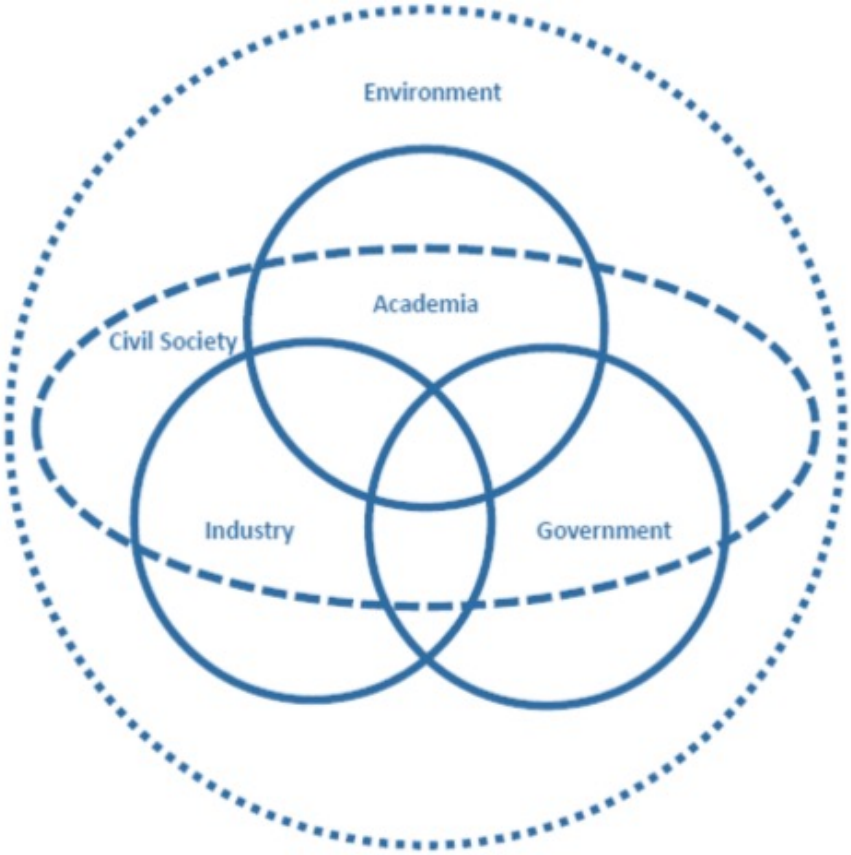
In the multidisciplinary and complex context of Industry 4.0 the technology transfer plays a key role for the adsorption and dissemination of technologies, resources and knowledge in order to transform the invention into tangible and useful innovation. In the framework of technology transfer, the patent analytics represents an important tool for more effective exploitation of the largest repository of technological information.

NLP techniques and machine learning approach have been applyed to the italian intellectual property data to realize a technology and knowledge assessment and to identify some relevant trends and emerging technologies.

Technology Transfer



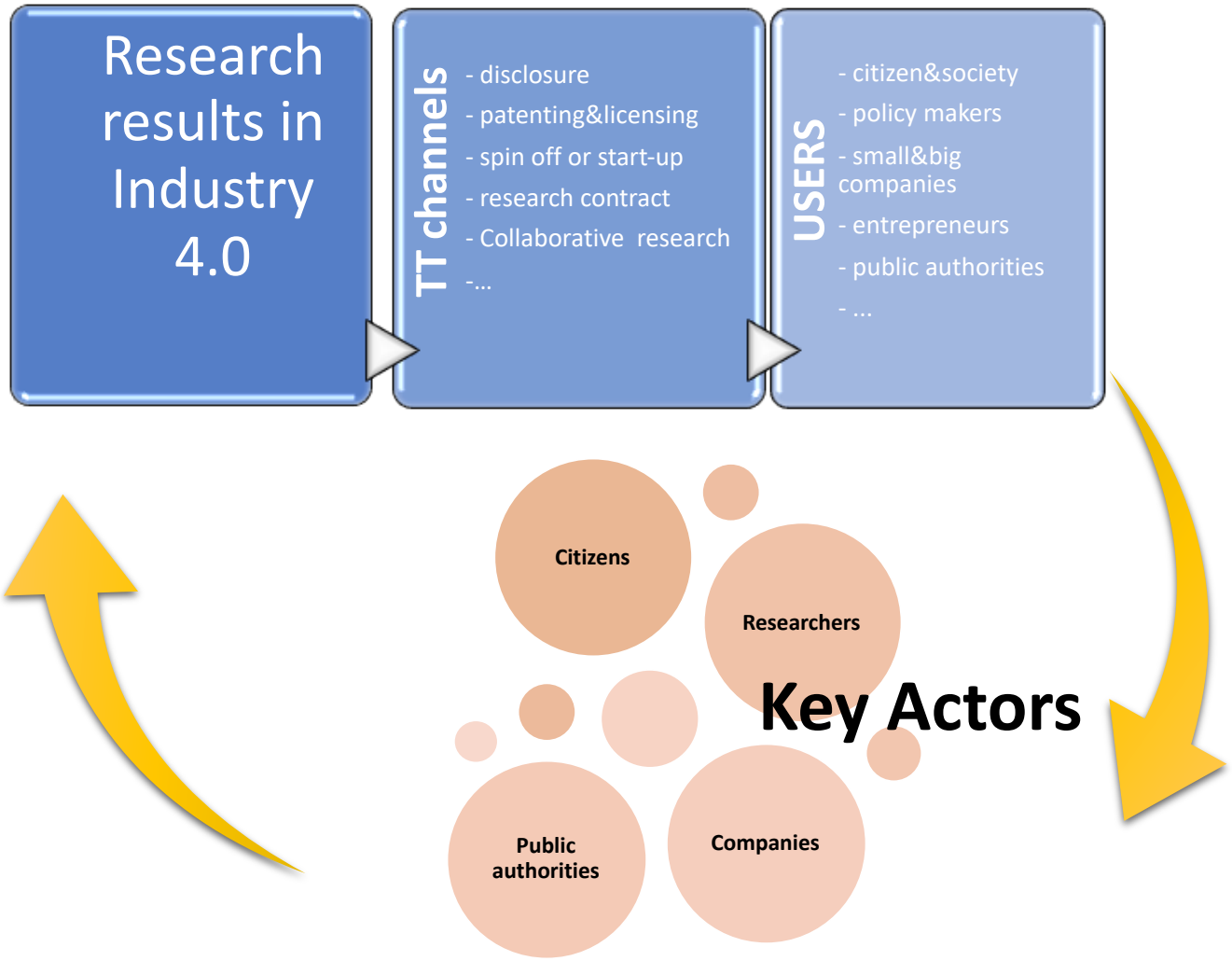
Quadruple/Quintuple Helix Model and Industry 4.0



The integration of Industry 4.0 technologies within society is pivotal for resolving many challenges that the world and its population are facing presently. However, academics and practitioners still struggle to fully understand I4.0 outcomes outside of the manufacturing domain, thereby unravelling their potential for society at large.

In this scenario, Society 5.0 is arising as a new paradigm that places humans at the centre of innovation.

The flow of knowledge and technologies towards the creation of impact



«Current Forth Industrial Revolution is identificated with three realities:

(1) Exponential evolutions (velocity),

(2) the integration of technologies (the breadth and depth of what digitisation offers),

(3) holistic system impact across society, industry and countries»

From «Schwab, K., 2017, The fourth industrial revolution»

The value of Data in Industry 4.0

Data presents value for enabling a competitive data-driven economy, which is at the heart of the Internet of things and Industry 4.0.

Increased data availability presents an opportunity for better decision making and strategy development, to introduce the next generation of innovative and disruptive technologies.

With the rise of artificial intelligence (AI), and the increase in the usage of methods such as machine learning (ML) and deep learning (DL), a number of these have been applied to **analyse Intellectual Property data**.

Intellectual Property Analytics

My research belongs to the domain «Intellectual Property Analytics (IPA)» - **the data science of analysing large amount of intellectual property information**, to discover relationships, trends and patterns in the data for decisiong making.

It is a multidisciplinary approach that makes use of mathematics, statistics, computer programming, and operations research to gain **valuable knowledge from data, to support decision making** rooted in the business context.

Patent data has long been considered the world's largest repository of technological information.

A summary of the intellectual property analytics methods

Intellectual Property Analytics Methods (i.e. Artificial intelligence, machine learning and deep learning techniques analysing patent data), arranged in alphabetical order.

Approach	Method	Authors
Artificial Neural Networks (ANN)	Back Propagation learning (BP)	[30,31,36,39–51]
	Evolutionary sigmoidal unit, Evolutionalry product unit	[52]
	Extension theory	[53,54]
	Extreme learning machine (ELM)	[43,47,55]
	Growing cell structure, paired with Girvan-Newman clustering algorithm	[56]
	Restricted Boltzmann machines	[57]
Clustering	K-means (and derivations)	[33,35,52,58,59]
	Self organising maps (SOM)	[36,39,40,60]
Deep Learning (DL)	Deep Belief Networks (DBN)	[57]
	Reinforcement Learning (RL)	[61]
Ensemble	Bootstrapping	[29]
	Random Forest	[62]
	Stacking	[63]
Decision tree	Classification and Regression Tree (CART)	[64,65];
	C4.5	[62]
Dimensionality Reduction	Linear Discriminant Analysis (LDA)	[50,66]
	Multi-dimensional scaling (MDS)	[67]
	Principal Component Analysis (PCA)	[31,33,54]
	Quadratic Discriminant Analysis (QDA)	[50]
	Singular Value Decomposition (SVD)	[33]
Regression	Linear	[33,35,37,54]
	Logistic	[62,68,69]
Statistical and probabilistic modelling	Conditional random fields (CRF)	[29,34,58,70]
	Latent Dirichlet Allocation (LDA)	[56,71]
	Naive Bayes	[62,65]
	Hidden Markov Model (HMM)	[72]
Support Vector Networks (SVN)	Support Vector Clustering (SVC)	[33]
	Support Vector Machine (SVM)	[34,38,45,60,73–76]
	Semantic Support Vector Machine (SVM)	[70]
Text mining approaches	Dictionary-based approach	[34,58]
	Natural Language Processing (NLP)	[34,68]
	Rule-based approach	[34,62]
	Semantic based ontology	[49,70,77]

- The majority of articles are concentrated around artificial neural networks (ANN) and the use of the back propagation learning method, followed by support vector machine (SVM), as well as conditional random fields (CRF).

- The majority of articles focus on classification methods, with some combining both clustering and classification methods.

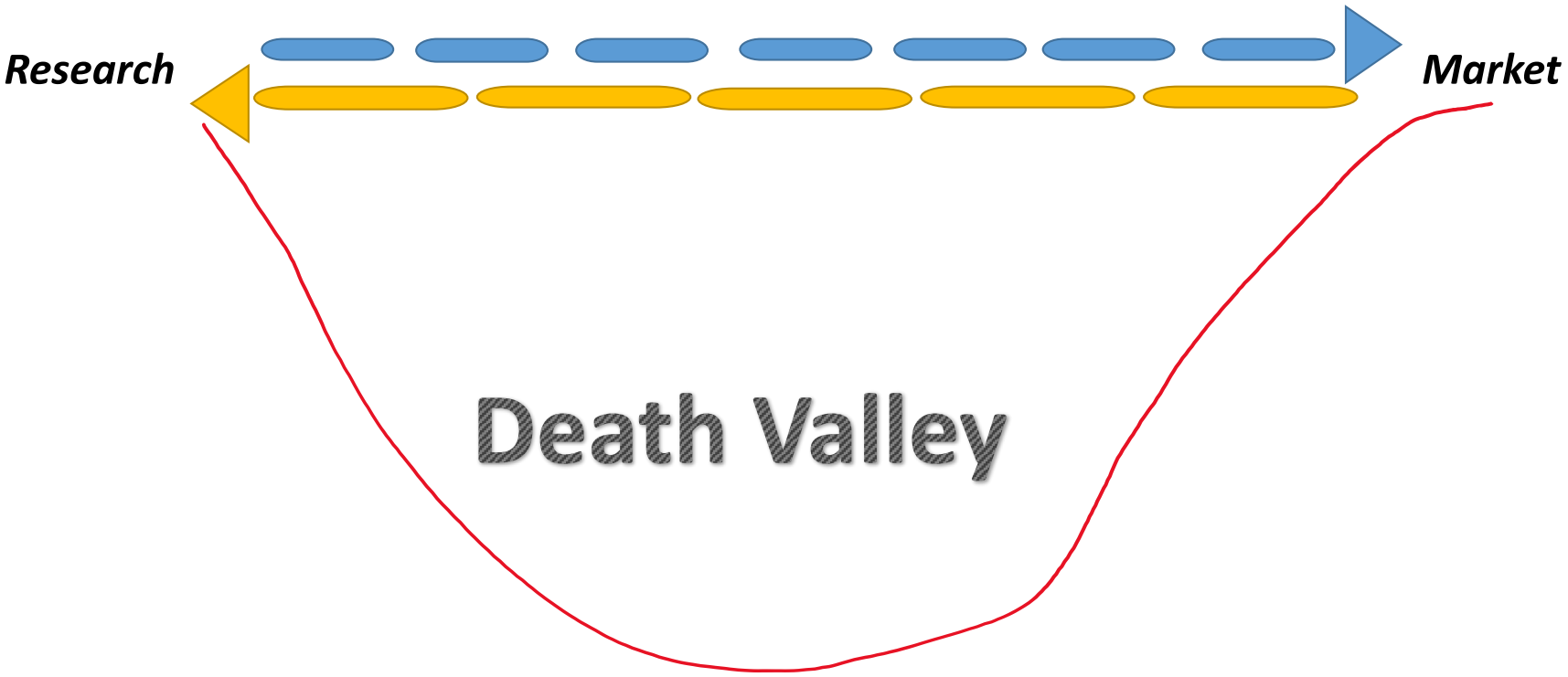
From «The state-of-the-art on Intellectual Property Analytics (IPA): A literature review on artificial intelligence, machine learning and deep learning methods for analysing intellectual property (IP) data» - Leonidas Aristodemou, Frank Tietze (University of Cambridge)

The use of artificial intelligence methods with intellectual property data

Four categories emerge in which the use of artificial intelligence methods with intellectual property data is implemented:

- The first category is **knowledge management**: patent evaluation and patent quality classification
- The second category is **technology management**: technological patentability, R&D planning within organisations, technology intelligence including monitoring technological changes, identification and forecasting of emerging technologies
- The third category is the **economic value of intellectual property** (in this case patents), and its impact in other areas
- The fourth category is a **hybrid category**: extraction of information and effective management of the information.

The valorization of Patents



The Death Valley represents “the gap between where publicly available research funding stops and where private investment or commercial funding starts”

Mutual non-disclosure Agreement signed with Netval – Network per la Valorizzazione della ricerca universitaria for the elaboration of the "Knowledge Share" database

MUTUAL NON-DISCLOSURE AGREEMENT

This Confidentiality Agreement is made by and between **Netval – Network per la Valorizzazione della ricerca universitaria**, via Gaetano Previati 1/C - 23900 LECCO, Partita Iva/Codice fiscale: 03092240138 - 92053760135

and

Dipartimento Interateneo di Fisica, Università degli Studi di Bari Aldo Moro, via Orabona, 4, Bari (Italy), P.I.01086760723 | C.F.80002170720

In connection with the Potential Project (as defined below), we expect to give each other information concerning us and/or our respective Companies and/or Group Companies. This agreement governs the terms of that exchange of information. As a condition to each of us providing to the other Confidential Information, we respectively agree and undertake to be bound by the terms of this Confidentiality Agreement.

1. Definitions
In this Confidentiality Agreement, the following words and phrases shall have the following meanings:
"**Associates**" means, with respect to each of us, our agents, advisers, employees, sub-contractors, PhD, officers, directors
"**Confidential Information**" means (i) all information disclosed by one Party and/or its respective Group Companies and Associates to the other Party and/or respective Group Companies and Associates, in written, oral or other tangible forms of any scope or subject matter whatsoever relating to the meetings held by theParties for the evaluation of the feasibility of the Project, and (ii) all information,disclosed by one Party and/or its respective Group Companies and Associates to theother, in written, oral or other tangible forms of any type, scope or subject matter whatsoever relating to the Potential Project, including without limitation all plans and strategies, market studies, samples, inventions whether or not patented, know-how, data, computer programs, disks, diskettes, tapes, reports, letters, drawings and models, manuals, specifications, designs, documents, records, notebooks and similar repositories of information and all other technical, financial or business information.
"**Confidentiality Agreement**" means this Mutual Confidentiality Agreement and the relevant terms and conditions.
"**Disclosing Party**" means the Party disclosing Confidential Information.
"**Group Companies**" means our respective subsidiaries; our holding or parent companies, including our ultimate holding or parent companies; and any subsidiary of any of our holding or parent companies, including our ultimate holding or parent companies.
"**Party**" means, depending on the context, either Travel Appeal or the other signing entity to this Confidentiality Agreement as set out in the first paragraph of this agreement.
"**Potential Project**" means:

The object of the collaboration is the application of AI - machine learning - complex networks techniques to the data contained in the "Knowledgeshare" platform that today represents the main source of information on the patent production of 89 Italian research institutions.

Objectives of the collaboration

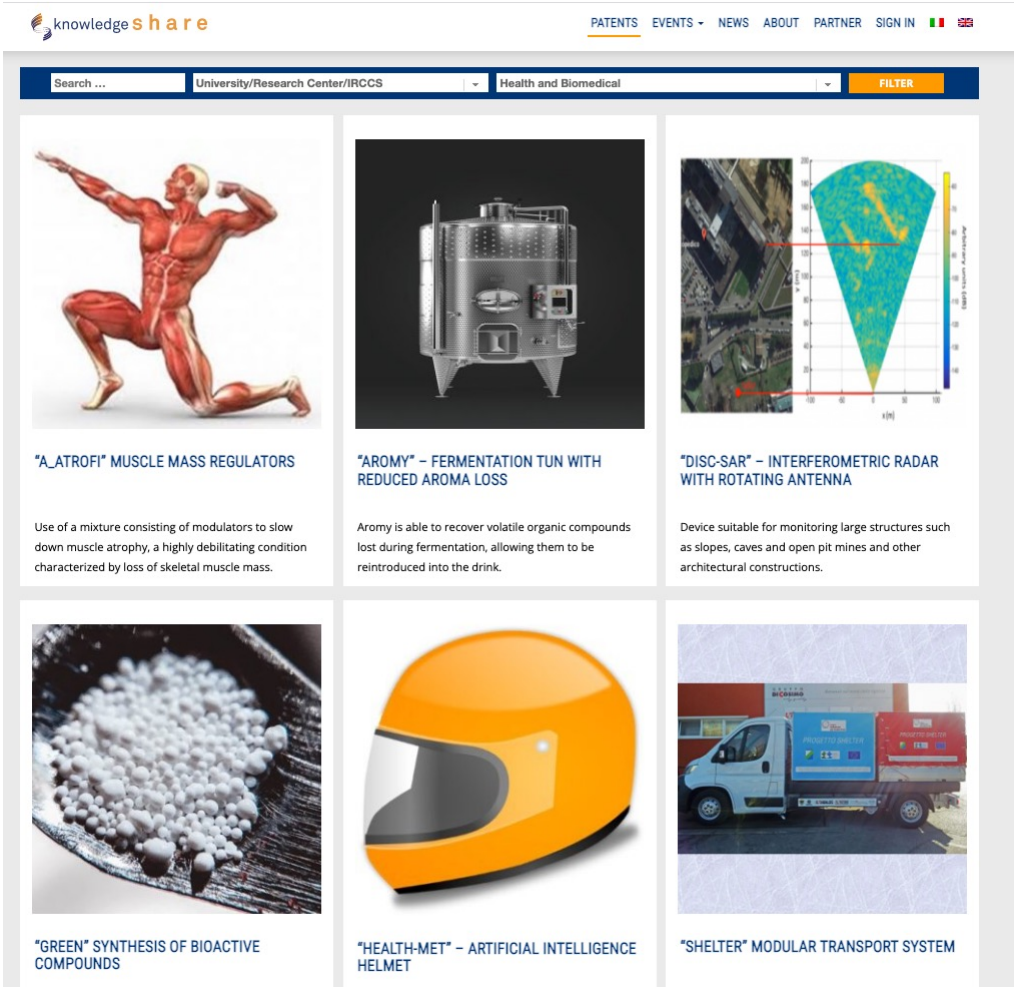
- Realize a pre-matching system between demand and supply of innovation
- Use keywords to improve and refine the search for technologies
- Evaluate a territorial characterization by sector/technological area and Top Technology domains in Italy
- Identify clusters of technologies to meet specific research needs
- Aggregate similar and complementary technologies - and therefore research institutes - able to converge in a common industrial development
- Hypothesize a connection with patent databases (UIBM, EPO, WIPO, etc.) to integrate even more information
- Identify technological trends and forecast emerging technologies
- Landascape by technology clusters
- Creation of patents pool to be selled with a higher value

KNOWLEDGE SHARE

- **Knowledge Share** is a portal as the meeting point for Italian companies with the expertise developed by the Italian universities and Research Centers, which can become practical applications.
- Is a joint project involving Politecnico di Torino, Italian Patent and Trademark Office (UIBM) at MISE (Ministry for Economic Development) and Netval. (https://www.knowledge-share.eu/).
- It is a portal created to make available, in a clear and understandable way, information related to patents and technologies that represent the excellence of the scientific know-how of the Italian Universities and Research Centers.
- Access is open to Italian and foreign SMEs/investors that have access to the patents and technologies' information.

KNOWLEDGE SHARE Structure

- 89 Research Institutions (Universities, Research Centers, IRCCS…)
- 1694 patents
- 10 technological areas:
  - Aerospace and aviation,
  - Agrifood,
  - Architecture and design,
  - Chemistry, Physics, New materials and Workflows,
  - Energy and Renewables,
  - Environment and Constructions,
  - Health and Biomedical,
  - Informatics, Electronics and Communication System,
  - Manifacturing and Packaging,
  - Transports



Marketing form

ACTIVE BIOMIMETIC ANKLE-FOOT PROSTHESIS

Ankle-foot | biomechatronics | biomimetic | gait analysis | Prosthesis | Robotics

INTRODUCTION

The very few existing prostheses that can accurately reproduce the ankle-foot system are still experimental. The structure of this patented active ankle-foot prosthesis is directly inspired by the actual anatomic configuration of the human ankle-foot system and can therefore mimic its different functions.

TECHNICAL FEATURES

This patented prosthesis contains many specific innovative functions that make it stand out both among energy storing and returning feet (ESR) as well as among bionic feet. Its structure includes six sensorized articulations, two arches of the foot connected to the forefoot, an ankle joint composed of a talocrural and subtalar articulation, along with series elastic elements in the motor. Together these elements offer those who study biomechanics a versatile and functional instrument for in depth analysis of the walking gait and for the development of control algorithms based on multiple parameters. Once certified, the prosthesis will offer lower limb amputees a new perspective in terms of comfort and natural walking gait. It can be used in either the active (motorised) or passive mode, without any changes to the design. The prototype is currently undergoing in vivo testing inside the compatible patented test bench (link).

Related Patent: Walking Simulator for Lower Limb Prothesis Testing

POSSIBLE APPLICATIONS

- Prosthesis for lower-limb amputees;
- Lower limb biomechanics, especially ankle-foot joint research for prosthetics;
- Highly accurate footwear tests (with the test bench);
- Highly sensorized diagnostic device for walking gait analysis of prosthesis users.

ADVANTAGES

- Accurate reproduction of the biological ankle-foot joint (6 articulations);
- Highly sensorized;
- Suitable for different walking surfaces;
- Maximum energy efficiency, thanks to the elastic elements;
- Promotes greater symmetry in the gait of one-legged amputees.

PATENT INFO

REGISTER TO CONTACT US

DOWNLOADS

Login to download the documents

PATENT OWNER

Università degli Studi di Padova

INVENTORS

Petrone Nicola | Mistretta Paolo | Faggian Riccardo

PATENT STATUS

Pending

PRIORITY NUMBER

IT102020000013141

PRIORITY DATE

03/06/2020

LICENSE

International

COMMERCIAL RIGHTS

Exclusive

AVAILABILITY

Available

TECHNOLOGICAL AREA

Manufacturing and Packaging - Health and Biomedical

Focus on patents for Industry 4.0

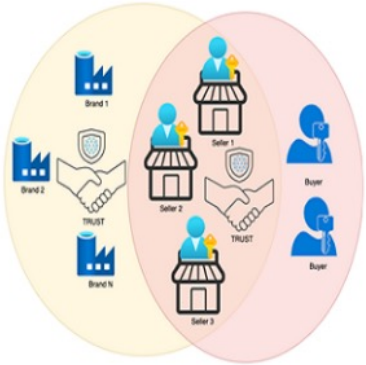
knowledge share

PATENTS EVENTS NEWS ABOUT PARTNER SIGN IN

PATENTS

"HEALTH-MET" – ARTIFICIAL INTELLIGENCE HELMET

Helmet able to call for help in the event of an accident, and able to estimate the damage suffered to the head through an AI module.



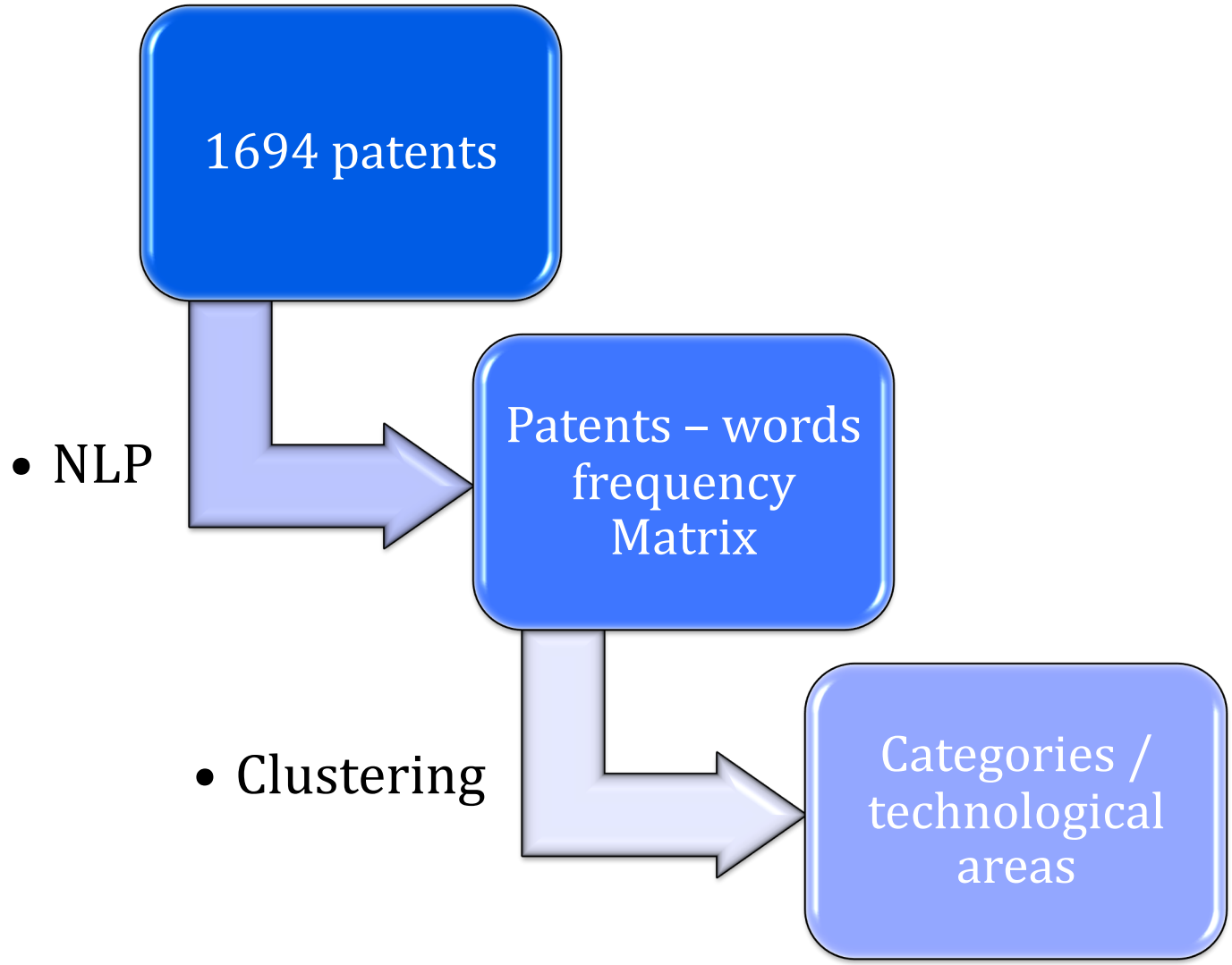
IDENTIFICATION AND TRACKING OF AN ASSET USING BLOCKCHAIN TECHNOLOGY

ICT system with software and hardware solutions, for the identification and tracking of an asset based on blockchain technology.

AI BASED QUANTUM COMPILER

Novel computer-implemented method for real-time quantum compilation based on reinforcement learning techniques.

A machine learning approach



First step

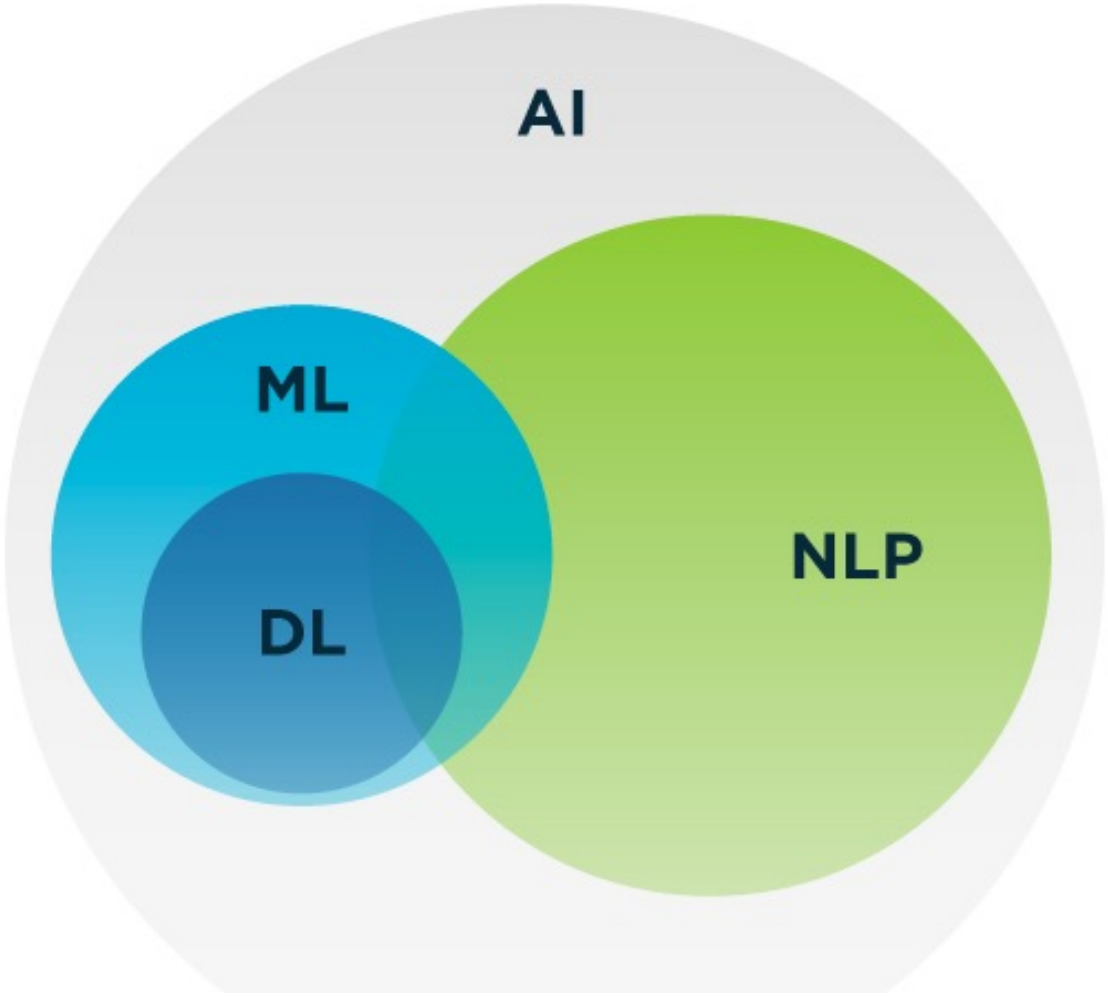
The Marketing form is a mix of text, date, names, pictures and so on.

Documents have to be converted into structured data that are suitable for analytical models.

We transform document data to document-term matrix using text mining techniques.

The first step is NLP - Natural Language Processing (natural language understanding and processing)

NLP



Artificial Intelligence

Machine Learning

Language Processing

Deep Learning

NLP Application and results

Application of NLP on «introduction» and «technical features» of 1694 patents

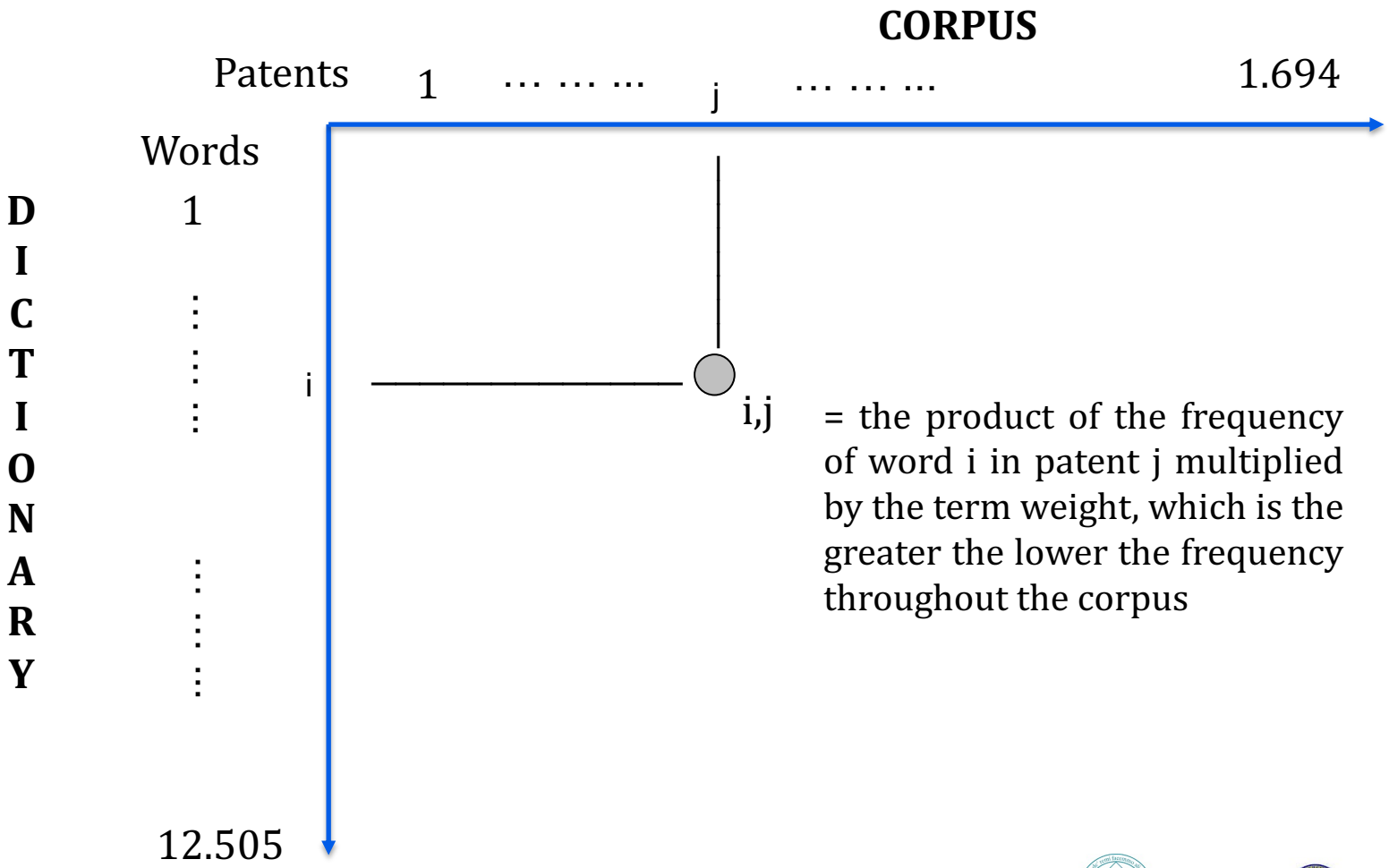
Steps:

- 1. Tokenization - a way of separating a piece of text into smaller units called tokens
- 2. Lower casing - Converting a word to lower case
- 3. Stop words removal – removal of very commonly used words (a, an, the, etc.) in the documents which do not really signify any importance in distinguishing two documents
- 4. Stemming - process of transforming a word to its root form.

Result: a list of 1694 «cleaned and treated» texts

Matrix TF_IDF Construction

Matrix TF_IDF returns a Term Frequency-Inverse Document Frequency (tf-idf) matrix constructed in this way:



= the product of the frequency of word i in patent j multiplied by the term weight, which is the greater the lower the frequency throughout the corpus

Sparseness problem and its resolution to clustering

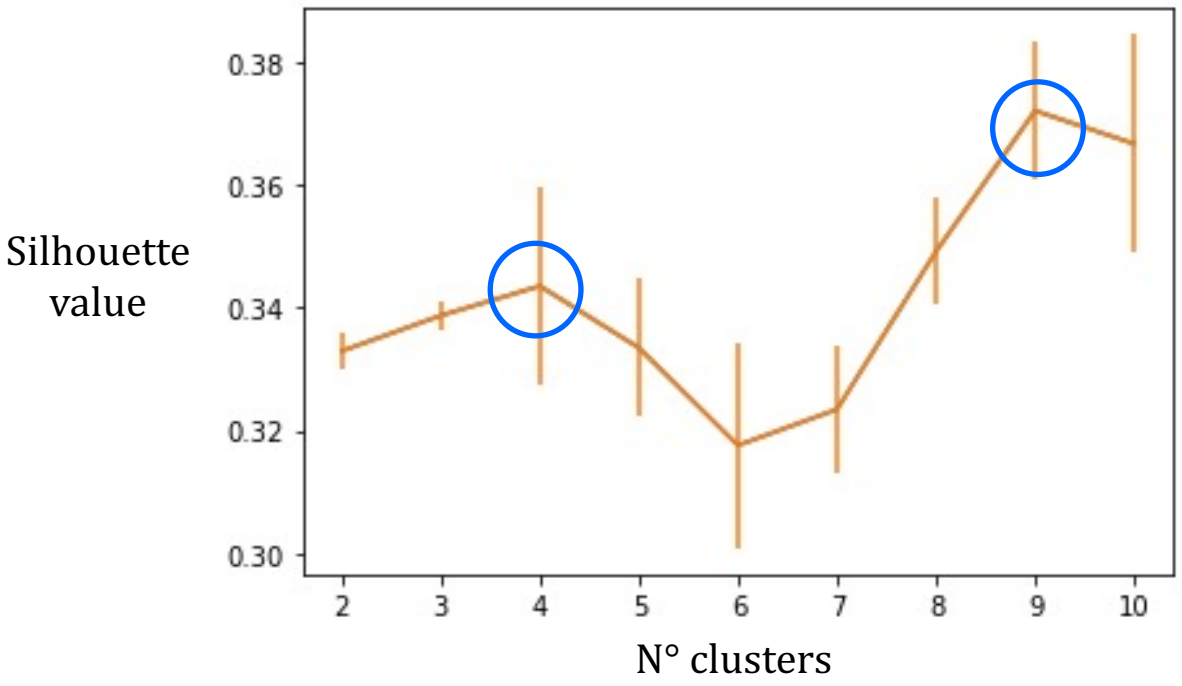
- The size of the rows (terms) is larger than that of the columns (patents) – we have too many features!
- In addition, many of the elements in the matrix have a value of zero (Sparseness).
- Although the matrix is suitable for statistics and machine learning, it is difficult to analyze it because it has a very sparse data structure.
- To overcome the sparseness in document data clustering, we perform:
  1. **Dimension reduction using singular value decomposition (SVD)**
  2. **Clustering using K-MEANS clustering algorithm**
  3. **Clustering quality measurement with Silhouette measure**

Choice of free parameters

- SVD, K-MEANS and SILHOUETTE have free parameters that it is necessary to fix:
  - SVD - the number of word combinations
  - K-means - the number of clusters
  - Silhouette - the metric used to measure the distance between points
- To choose these parameters, I have analyzed the parameter space and fixed:
  - For SVD - word combinations interval between **10 to 1694** (step of 10)
  - K-means - the number of clusters from **2 to 10**
  - Silhouette - the most common distance metric are the **Euclidean distance, the Manhattan distance and cosine distance.**

The best combination of the parametres

- By reiterating n times the program in Phyton, the best combination of the three free parameters is found:
  - For SVD - word combinations interval **10**
  - K-means - the number of **clusters 9** (absolute maximum) **and 4** (local maximum)
  - Silhouette - **cosine distance**



Future perspectives and conclusion

- A Focus on healthcare patents that represent a very populated cluster and, in particular, identification of patents on Healthcare 4.0
- Increase the number of clusters (K = 20, 50 etc) to understand if you can identify patterns of complementary technologies and maybe sellable/transferable together
- Realization of a recommendation system to classify correctly the patent in a technological area
- Analysis of valorized patents among the 1694 to understand the impact of licensing, selling and so on
- Discover the Technological emerging trends and associations of technologies

In conclusion:

- this research aims to emphasize the importance of technology transfer in the fourth industrial revolution, from the perspective of the quadruple and quintuple helix.
- Companies and investors could take advantage of innovations produced by research institutions, we could matchmaking between supply and demand for innovation, predicting emerging technologies.
- The potential application of a quantitative framework for innovation are countless, ranging from scientific policy, to R&D strategies for firms, regions and even nations, it can be connected to socioeconomic data, to products and can be embedded in frameworks for industrial development.

Thank you for your attention!