A patent review on machine learning techniques and applications: depicting main players, relations and technology landscapes

Zambetti M.*, Sala R.*, Russo D.*, Pezzotta G.*, Pinto R.*

* *Department of Management, Information and Production Engineering, University of Bergamo, Viale Marconi, 5, Dalmine (BG), 24044, Italy (michela.zambetti@unibg.it, roberto.sala@unibg.it, davide.russo@unibg.it, giuditta.pezzotta@unibg.it, roberto.pinto@unibg.it)*

Abstract: The increasing availability of data, promised by the 4th industrial revolution wave, is challenging companies and organizations in diverse industry sectors to extract useful and actionable information. To this end, a vast array of data management strategies and new analytical methods is becoming available to the large audience of researchers and practitioners. Although traditional statistical approaches are still applicable for different purposes, artificial intelligence techniques, particularly machine learning algorithms, are increasingly being explored and adopted to approach data analysis. Artificial intelligence becomes a necessary ingredient for technology progress. The machine learning domain, in particular, has been extensively investigated by academics, who mainly focused on algorithms and suitable applications, and it is also permeating business reality at an unprecedented rate. Against this background, instead of eliciting knowledge from academics, the proposed research adopts a patent review and analysis approach, with the specific purpose of understanding the ongoing industrial effort on the subject, and new as well as expected trends on machine learning technologies and applications. The paper analyses technological development in various industries by defining patents trend over the years and investigating the different areas of applications according to the Cooperative Patent Classification (CPC), a patent classification system jointly developed by the European and US patent authorities. Patent applicants are also investigated in order to highlight active and competitive players in the domain, as well as collaboration between different companies. Furthermore, the paper includes a patent citation network analysis, which is useful to show critical technologies developed, and to understand applicants' behaviours, such as influences or infringement trials. Overall, the paper provides an original and "literature-complementary" outlook on the machine learning landscape, giving an understanding on industrial R&D effort in this context, delineating trends related to technology diffusion and innovation from an industrial perspective.

Keywords: machine learning; data analytics; patents review; cooperative patent classification (CPC); technology landscaping; Industry 4.0

1. Introduction

The development and the adoption of new technologies, belonging to the so-called 4th industrial revolution, is creating an impressive availability of data, which is reaching an unprecedented scale. Information is gathered from various sources, both outside and inside the companies: structured and unstructured data came from processes, logistic operators, social networks, machine sensors, images, voice file, GPS, products usage, and other sources (Maklan, Peppard and Klaus, 2015).

However, data alone does not create competitive advantage: extracting useful and actionable information from this data deluge represents an opportunity and a big challenge at the same time. Indeed, extracting valuable insights from data requires structured approaches and techniques able to support knowledge extraction. Moreover, to leverage the value of information, companies need to be supported in the interpretation of the information and, ultimately, in their decision-making processes.

As a result, the Business Intelligence and Analytics (BI&A) discipline is growing (Chen *et al.*, 2012). The term refers to all the techniques, technologies, practices, methodologies, and applications that analyze critical business data to help an enterprise to understand its business and market better and make timely business decisions. Although BI&A is not a recent discovery, it is now becoming increasingly essential to every business, and performances of such tools may be enhanced with the possibilities to rely on a dataset of sufficient size to provide meaningful learning as results.

Particularly, one of the faster growing techniques in this field is machine learning (ML) (Smola and Vishwanathan, 2008), which is also a well-known and discussed topic in statistic and computer science. Recently, ML can explore its potential based on rich datasets that provide an encouraging environment for information discovery through iteration. Moreover, this data availability is creating the right background to embrace ML at different levels and from different users: technologies that previously were limited to academic research and big players, are now becoming available and also used from small organization and even individuals (e.g. Amazon ML, TensorFlow Serving). Companies are recognizing the potential of ML for both enhancing internal performance and offering better solutions to customers by building profiles of buying habits, prices, and also anticipating shoppers' needs ahead of time.

ML is being increasingly adopted at large scale in both the academic and the industrial communities. To investigate the adoption trend, the authors propose a study based on the analysis of companies' patents investment and portfolio to highlight the main direction along which they are moving. Indeed, patents provides evidences on the inventive activities and research covering multiple fields, geographical location and actors (Hullmann & Meyer 2003). They represent a good guidance with respect to the level of detail and information contained on inventions, and the possibility of disaggregating data for different technologies (Griliches 1990; Johnstone et al. 2010; Popp 2005). Moreover, the systematic filing and records provide accessibility to those documents over decades (Debackere et al. 2002). The paper is structured as follows: Section 2 introduces the background on the Industry 4.0 and ML domain, as well as the potential of patents analysis. Section 3 outlines the specific questions addressed in this study. Section 4 presents the methodology used to answer those questions, while results and discussion are presented in Section 5. Section 6 concludes the paper, underlying the main findings and limitations of the presented work.

2. Background

2.1 Machine Learning

With the 4th industrial revolution, an increasing amount of software and embedded intelligence are integrated into both industrial systems and end-products, enabling the collection of data about the internal performance as well as about the product usage. This enables real-time control and optimization across all the value chain and along all the product lifecycle (Kagermann, Wahlster and Helbig, 2013), creating the potential for product and service enhancement and customization. Big data analytics offers benefits to organizations (J.Q. Li et al. 2015; J. Li et al. 2015), enabling them to analyze data collected from various sources and in different formats such as text, video, and social media data (Maklan, Peppard and Klaus, 2015).

Nonetheless, the development of algorithms for dealing with data has been recognize as one of the main challenge in Industry 4.0 (Wang *et al.*, 2017)

From the beginning of the 1960s, when the term Machine Learning (ML) has been defined for the first time (Samuel, 1963), different approaches have been proposed, attempting to extract useful information from data. One of the most interesting and intensively discusses is the ML domain. ML represents "computational methods which rely on experience to improve performance to make accurate prediction" (Mohri, Rostamizadeh and Talwalkar, 2012). Such methods embrace different technologies and capabilities, including the ones defined in the artificial intelligence (AI) domain, which refers to the specific capacity of a system to learn.

The success of learning algorithms depends on the initial set of data used, and the recent possibility created by such availability of data, have led to the rebirth of the machine learning research and real adoption from companies. Indeed, they can be used to autonomously predict product performance degradation, manage and optimize product service needs and enhance customization. Studies on the application of those techniques are widespread over different fields and just to name a few, they can range from maintenance (Susto *et al.*, 2015) to electricity forecast (Naimur Rahman, Esmailpour and Zhao, 2016) and brand mining (Pournarakis, Sotiropoulos and Giaglis, 2017).

The increasing interest in ML and AI domain, from the academic side, can be recognised from the number of publications on these topics during years. Figure 1 represents the trend of scientific publications over the past 20 years mentioning "MACHINE LEARNING" and "ARITIFICIAL INTELLIGENCE". It is possible to notice the massive increment in the last few years.

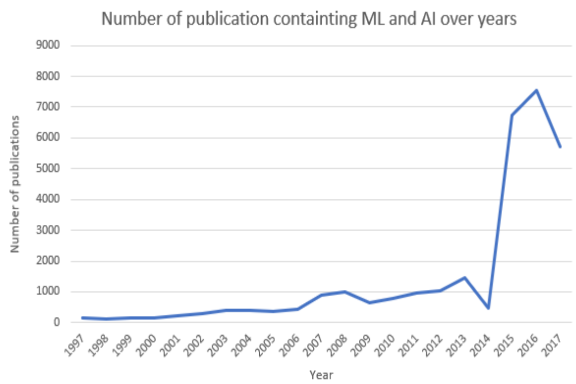


Figure 1 – Number of publications on Machine Learning and Artificial Intelligence over year (source: Scopus)

Some corporate initiatives in the ML learning domain are also well known, as the IBM Watson, Amazon Machine Learning and Microsoft Cortana Analytics, but the real commitment of the industrial context in this domain is still under investigated.

2.2 Patents analysis

Understanding where innovation and research are going and if there is particular interest from the corporate side in a particular area is an important, yet not easy question to answer. Certainly, companies do not freely share their core strategies and knowledge, but often protect their critical innovations through patent protection, and intrinsically they are forced to share information related to them.

Indeed, trying to define technological landscapes, patents provide an appealing dataset to use for several reasons. The first reason lies in the main proposition of patents: it can be defined as a deal between the owner of an invention and the society. Thus, from one side, the owner benefits for the exclusive monopoly over the exploitation of the patented device for a defined period of time, and on the other side, the society benefits from the knowledge of the patent content and the possibility to use the innovation after the patent is expired. (Connolly &

Hirschey 1988; Debackere et al. 2002; Debackere et al. 2002; Gupta 1999). This means that the general public can benefit from the accessibility to those documents (Huang *et al.*, 2003), which retained, according to the statistics of the World Intellectual Property Organization (WIPO), 90–95% of economically valuable human innovation results (Brockhoff, Chakrabarti and Hauschildt, 1999). Secondly, patents must be submitted in standard format and content, providing a set of documents homogeneous respect to the information. Even if differences among the various patent systems exist (e.g., United States Patent and Trademark Office, European Patent Office, etc.), because of variations in legal, economic, geographic, and cultural factors (Debackere *et al.*, 2002), the World Intellectual Property Organization (WIPO) is also attempting to standardize them worldwide, and the document for the patent application must contain specific content. Particularly, a patent must be submitted including, apart from the description of the technological innovation, detailed information, e.g. the inventor(s), the assignee(s), the applicant(s), the geographical scope, classification codes, dates (priority, publication) and also citations to previous patent and innovations which are supporting the new one.

As a result, it is possible to state that patents represent the inventive activity and output from applied research over different fields, countries, and time (Trajtenberg 1990; Hullmann & Meyer 2003) and at the same time an interesting pool in terms of quality of data, accessibility, and level of detail (Debackere et al. 2002; Griliches 1990). Indeed, during years, patent analysis has been widely used both from academics and from companies, for different purposes, for example:

- Define the economic value of an invention (Albert et al. 1991; Gay & Le Bas 2005);
- investigate the success of a R&D program or policy, for technology, department, institution, region, or country (Connolly & Hirschey 1988; Narin et al. 1992);
- evaluate inventive activity at the corporate, industry, or national levels simple statistics, such as counting, clustering, and citation analysis have been used (Wallin, 2005);
- analyze the commercialization of academic knowledge (Baldini et al. 2007; Leydesdorff & Meyer 2007);
- asses the corporate technological level and plan the firm's technological innovation and competitiveness strategies (Ernst, 2003);
- trace the flow of knowledge in technology sectors and discover critical patents troughs the patent citations analysis (Nemet, 2009).

This list is not intended to be exhaustive: in terms of previous knowledge, there are also other applications, and in terms of future knowledge, new insights in analyzing patents might be explored.

3. Definition of the research scope

Based on the previously mentioned considerations, the main scope of this research is to define and depict a clear image of how industries are moving in the context of machine learning after the so-called 4th industrial revolution, exploring patents applications. To this end, the following specific research questions have been defined in Table 1.

Table 1: Research questions

Main questions	Sub-questions
Q1: How does the industrial context embrace ML technologies?	**Q1.1:** When they started investigating ML potential? **Q1.2:** How much effort do companies put in the R&D? **Q1.3:** Where are located the most active R&D teams and which is the patent geographical domain? **Q1.4:** How the 4th industrial revolution influenced the innovation activity in the field? **Q1.5:** Which are the most critical area of application of ML?
Q2: Are companies undertaking new strategies, and are they developing new competences?	**Q2.1:** Who are the companies most active? **Q2.2:** In which sectors do they operate? **Q2.3:** Which is the behavior among applicants?

RQ1 addresses the issue of how much companies are investing in ML, while RQ2 wants to investigate whether companies which are collecting data are approaching directly data analysis or are they relying on other companies. Sub-questions have been formulated in accordance with the information provided in the patents. Indeed, for each (sub-)question, a patent-based approach has been undertaken, considering, for example, the patents evolution of application over time, the location of filed among the various national offices, areas of applications and technology domain, for what concern RQ1. RQ2 has been addressed with an analysis of applicants, since it reveals which companies are working in the field and if they are collaborating with others; it also suggests if universities are involved in the development of their knowledge.

With this regard, some of the terms most frequently adopted in the analysis need to be clarified. All the following explanations are in accordance with the European Patent Office (EPO) definitions.

The *priority date* is the date used to establish the novelty of a particular invention relative to prior art, and is the filing date of the very first patent application for a specific invention. The *publication date* is the one on which the patent application is available to public. This normally occurs 18 months after the priority date. The *patent family* identify a collection of several applications or publications for an individual invention (in different countries) claiming the same priority. All the "family members" have priority dates in common, but the publication date may differ. The *inventor(s)* is usually an individual who has played a role in conceiving the invention, while the *assignee* is the person or "juristic entity," with whom the inventor has transferred the entire ownership interest or a percentage of it and it is the person who owns all the rights that are inherent with the patent. *The applicant* usually coincides with the assignee. In few cases, the inventor can be identified as the applicant, if the assignee has opted not to file the application. When more than one actor own the right for the patent, we will refer to *co-assignee*. *Cited documents* are documents, both patent and non-patent literature, cited during any of the procedures before the patent acceptance, e.g. examination, review, opposition or by the applicant. *Citing documents* reports all the documents that have cited the patent. The *International Patent Classification (IPC)* is a hierarchical system for the classification of patents with respect to the different areas of technology to which they belong and it was established in 1971. *Cooperative Patent Classification* (CPC) is the classification system introduced in January 2013 in order to standardize the classification systems of all major patent offices. This taxonomy covers all EPO and US documents and contains 250,000 classes, the highest number of subdivisions, resulting in the most granular and precise classification.

4. Methodology

For the patent search, the FAMPAT patent database was used. It contains a complete collection of full text patent documents, recording more than 90 million among patent applications and utility models (used to protect minor improvements of existing products, which does not fulfill the patentability requirements, for a limited period of time). This database covers the most important patent offices and contains bibliographic data from more than 100 authorities, in English language, and for the ones not filled in English, the database provide the possibility to make a translation leveraging automatic machine translates. The first step of the research was to build a thesaurus about ML constituted by synonyms of concept, algorithms, technologies and techniques supporting ML (i.e. deep learning, support vector machine, neural network, LDA, artificial intelligence, predictive analysis, clustering, pattern recognition, etc.). The thesaurus was built according to most used terminologies in the literature. Then, a set of queries have been built by crossing these terms together. It was thus possible to estimate the error that would have been committed using a search based on the keyword "machine learning", and also the individual contributions of each term of the thesaurus. Particularly, matching machine learning and artificial intelligence with: (a) deep learning; (b) support vector machine, (c) neural network, (d) Lda; (e) predictive analysis; (f) clustering; (g) pattern recognition. The initial hypothesis is that, if machine learning and artificial intelligence would have been different queries, the patent pool found with queries with of the two main terms and one reported in the previously identified list, should have been disjoint. Patents which overlap ML and AI cover for the 50% to the 99% of the total. Exception made for the Neural Network, for all the other terms it is possible to state that ML and Artificial intelligence are used as synonymous and thus, the final query should include both terms to result in an exhaustive and comprehensive set.

However, to limit the total number of patents within 100,000 units and to have a statistically and significant sample representative of the sector, it was decided to use only "machine learning" and "artificial intelligence". These terms were combined in OR, searched in full text considering only patents which have a priority date (that is, the first date of filing of an application, anywhere in the world, to protect the invention) in and after 2011. The selection of the time coverage has been made according to the scope of the research. Since the word Industry 4.0 came about starting from 2011, that year was used as the beginning of the considered time horizon. Regarding the geographical coverage, all countries and patent collections represented in the patent database have been considered.

Thus, the following query has been performed, using Orbit software:

((MACHINE LEARNING) OR (ARTIFICIAL INTELLIGENCE)) /KEYW/TI/AB/IW/TX AND PRD >= 2011

The total number of patents resulted was 85000.

A second query has been also defined to extrapolate patents on the same topic before 2011, and to be able to compare patent activity in the two different time window. Specifically:

((MACHINE LEARNING) OR (ARTIFICIAL INTELLIGENCE)) /KEYW/TI/AB/IW/TX AND PRD <= 2011

Two different queries were used since all the analysis not explicitly related to the investigation on the trends over year aims to investigate only the invention after 2011.

5. Dataset analysis and discussion

In this section, we present the analysis of the patent dataset extracted via the first query illustrated in section 4. Only in Figure 2 and Figure 3, the research included patent prior 2011 to investigate if and when specific interest emerged.

Figure 2 illustrates the evolution of patent applications over time, making it possible to evaluate the dynamics of inventiveness in the field of ML. The priority date is the most reliable data to outline the profile the number of patents over year. Because of an 18-month delay between the filing of an application and its publication, the year 2017 has been discarded from the analysis. The graph depicts a consistently increasing effort over time, and particularly after 2011, since when the yearly incremental growth of patent filling became exponential. This is indicative massive and rapid investment from companies in the ML related technologies.
The two different phases, before and after 2011, can be easily identified in the time line, and verified by computing the Compound Annual Growth Rate (CAGR), from 2000 until 2010, it is around 9%, while between 2011 and 2016 it has more than doubled, reaching 28%.

Additionally, Figure 3 illustrates the evolution of published applications over time, indicating the dynamics of the inventiveness of the ML techniques, and is a proxy of the demand for patent rights in one or more countries. The graph supported the already defined positive trend and the significant growth from 2011, underlying the beliefs in the potential of ML approaches, since companies seem to construct massive portfolios and undertake ML as a strategic decision for their development and competitiveness.

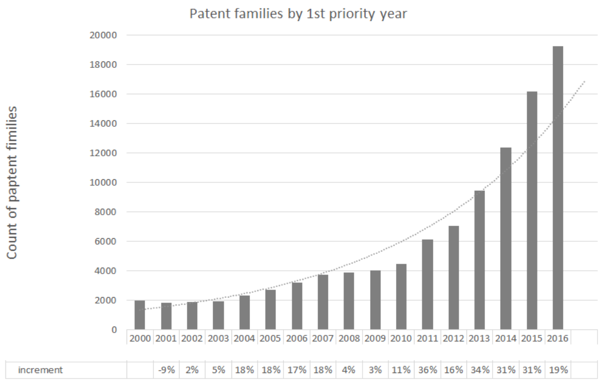


Figure 2 - Evolution of patent applications over time (priority date)

For what concern the geographical location of patents filling, Figure 3 reports the number of priority applications in the various national office. This representation missed the World Wide (WO) and European (EP) requests, but since most players file the first application locally, it is a good starting point in identifying the location of R&D investments.

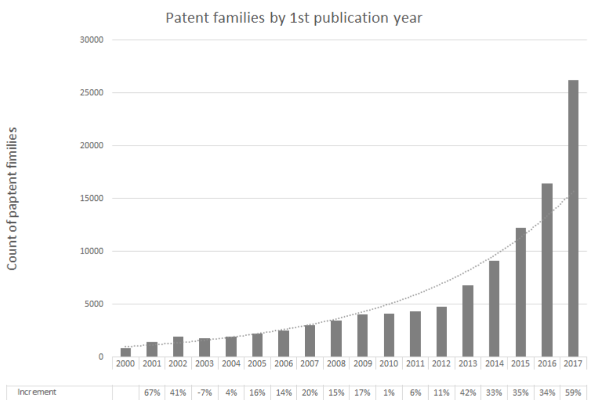


Figure 3 - Evolution of patent applications over time (application date)

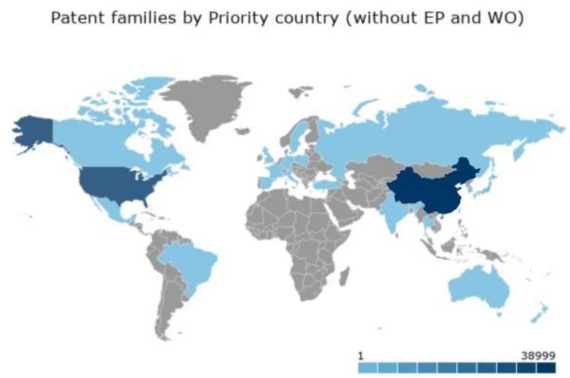


Figure 4 - Number of priority applications filed in the various national offices (2011-2017)

As it is possible to notice, China and US are the most active countries, fact that is also supported by the applicant analysis (Figure 5 and Appendix A). In fact, if we look into the size of the applicants' portfolios, we can observe that most of the main player are located in the US and China.

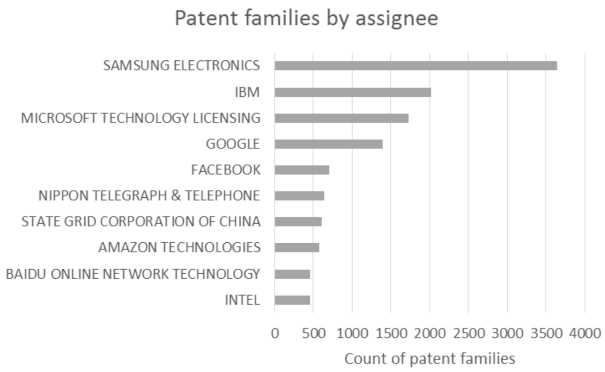


Figure 5 - Top 10 most active assignee (2011-2017)

In this analysis two main categories emerged: firstly, core technology companies like Samsung, IBM, Microsoft and Intel, whose business is mainly related to the computer science and some of their artificial intelligence initiatives are well known, like Samsung's A.I. research center, IBM's Watson project, and Microsoft Azure platform. Second, it is possible to figure out the high commitment of players like Google, Alibaba, Amazon, Baidu, and Facebook, who own a platform base asset.

Those platforms represent high scalable environments to gain users and customers insight, based on hundreds of thousands of data. ML adoption represents for those companies a critical opportunity to define customer preferences and customize offers. On the other side, it is possible to notice that they are also creating ML learning services, since they are also offering ML accessible for any developer, such as Google Cloud Machine Learning platform and Amazon Machine Learning platform.

State Grid Corporation of China needs to be treated differently since recently this company has pervaded patent landscape in all fields, with a tremendous amount of patents filled, without providing real advancement in state of the art.

The presence of some university indicates that many researchers are working on innovations. It is also shown in the co-assignee graphs (Appendix B) that the collaboration between private players and universities are not rare. We can notice that preferred collaborations are always done with universities: for example, Huawei and Baidu are cooperating with Chinese universities, as well as Samsung and IBM who owns co-application with the Massachusetts Institute of Technology (MIT).

The relationship among companies is rare, instead. It is possible to see just few cases, but it seems that for now companies are leveraging internal resources to innovate and there is no propensity in collaboration. Even though collaborations are rare between companies, through the analysis of the patent citations among the different actors, reported in Appendix C, it is possible to notice that Microsoft, IBM, Samsung and Google are focal center point for other patents. This analysis, in fact, discovered portfolios that have strong interactions with each other, and whenever a portfolio is strongly cited by most players, it is likely that it is a pioneering or a reference patents collection. Finally, the technology domain and principal field of application have been studied through the analysis of International Patent Classification (IPC) and Cooperative Patent Classification (CPC). Each patent contains many CPC/IPC codes, which associate patents to the different areas of technology to which they relate; for this reason, not all codes are useful for the scope of this analysis, making results not reliable in absolute terms but representative to identify core business in the field and potential applications and uses of patents.
As it is possible to notice in Figure 6 the primary domain of ML patents are computer technologies, which covers the 36,21%. Digital communication and telecommunication coves together the 14,84 and an interesting percentage is related to IT methods for management. Control and measurements field are also covered massively, followed by medical technology.

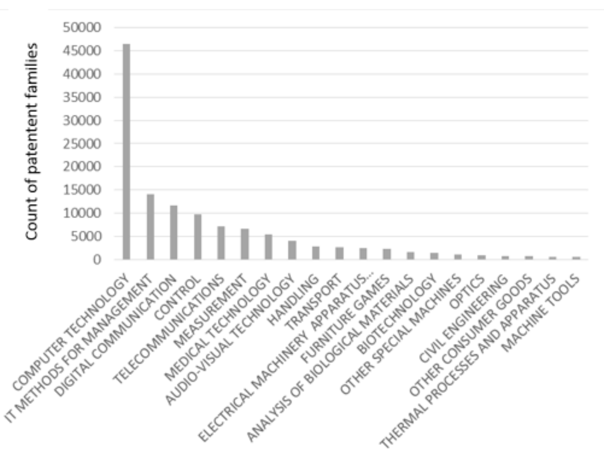


Figure 6 – Patent families in the top 20 technology domain (IPC) (2011-2017)

When investigating the CPC classification (Figure 7), most of the patents belong to the "computing; calculating; counting" class (G6), and particularly, interest has been devoted to the electrical digital data processing subclass, followed by data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes as well as data recognition and processing. Electric communication technique (H04) results in another important field of application.

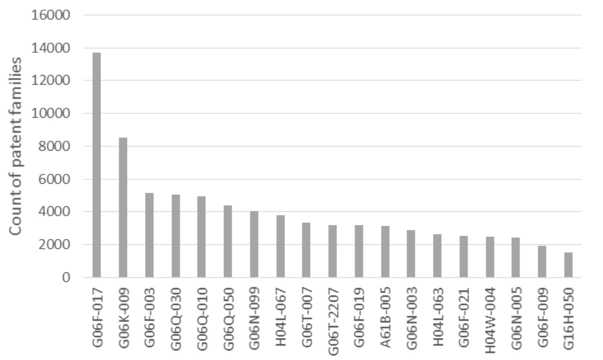


Figure 7 – Patent families in the top 20 CPC subclass (CPC classes description is reported in Appendix E)(2011–2017)

Graphs that are more complete, both for the IPC and CPC classification are shown, respectively in Appendix D and Appendix E.

6. Conclusion

The presented paper illustrates how companies are embracing the digital era leveraging the potential of machine learning. The authors investigated a consistent pool of ML and AI related patents, which represent the most comprehensive mean to discover the companies' activity in innovation and knowledge creation. The analysis has been pursued wondering at first (RQ1) how the industrial context is embracing ML technologies, and results showed that there is an increasing interest in the field, which started spreading around 2011 with the advent of the 4th industrial revolution. This information supports the intuition that firms attempt to drive business value from their data and analytics capabilities. For what concerns geographical considerations, US and China emerged as the most competitive counties, where the main players are also located. A second important focus (RQ2) was devoted to the assignee examination, understanding if companies are moving toward new strategies and if they are developing new competencies. High tech, computer and software providers are the most active, as well as platform-based companies. Samsung, IBM, Microsoft, Google and Amazon results as the leading player in the field and the one that owns the most critical technologies. Interesting collaborations have been found between companies and universities, particularly in China, while cooperation among different companies seems to be scant.

As a final remark, it is important to notice that patents by themselves do not offer insight into the adoption of a technology, but the study represents the R&D companies' investment. Further elaboration could include an analysis of offers and technologies available on the market and the evaluation of the patents in terms of their real application.

Since the nature of the research, which is based on the patent review, some important limitations have to be considered. Indeed, the range of patentable innovations constitutes just a sub-set of all research outcomes. Patenting is a strategic decision, and hence not all patentable innovations are patented. Nonetheless, the research includes an analysis of a pool of more than 85000 patents, emerging from a structured defined query. Finally, companies and applicants often use difficult languages in their patents or turn of phrase instead of specific terminology, and in some cases, they even create new term to describe the invention. This research addresses this shortcoming in the structured definition of the query, but future works may include different queries to verify the relative differences in results.

References

Albert, M. B. *et al.* (1991) 'Direct validation of citation counts as indicators of industrially important patents', *Research Policy*. North-Holland, 20(3), pp. 251–259.

Baldini, N., Grimaldi, R. and Sobrero, M. (2007) 'To patent or not to patent? A survey of Italian inventors on motivations, incentives, and obstacles to university patenting', *Scientometrics*. Kluwer Academic Publishers, 70(2), pp. 333–354.

Brockhoff, K., Chakrabarti, A. K. and Hauschildt, J. (1999) *The Dynamics of Innovation : Strategic and Managerial Implications*. Springer Berlin Heidelberg.

Chen, H. *et al.* (2012) 'Business intelligence and analytics: from big data to big impact', *MIS Quarterly*, 36(4), pp. 1168–1188.

Connolly, R. A. and Hirschey, M. (1988) 'Market value and patents: A Bayesian approach', *Economics Letters*. North-Holland, 27(1), pp. 83–87.

Debackere, K. *et al.* (2002) 'Measuring progress and evolution in science and technology - II: The multiple uses of technometric indicators', *International Journal of Management Reviews*. Wiley/Blackwell (10.1111), 4(3), pp. 213–231.

Ernst, H. (2003) 'Patent information for strategic technology management', *World Patent Information*. Pergamon, 25(3), pp. 233–242.

Gay, C. and Le Bas, C. (2005) 'Uses without too many abuses of patent citations or the simple economics of patent citations as a measure of value and flows of knowledge*', *Economics of Innovation and New Technology*. Routledge, 14(5), pp. 333–338.

Griliches, Z. (1990) 'Patent Statistics as Economic Indicators: A Survey', . *Journal of Economic Literature*. National Bureau of Economic Research, Inc, 28, pp. 1661–1707.

Gupta, V. K. (1999) 'Technological trends in the area of fullerenes using bibliometric analysis of patents', *Scientometrics*, 44(1), pp. 17–31.

Huang, Z. *et al.* (2003) 'Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field', *Journal of Nanoparticle Research*. Kluwer Academic Publishers, 5(3/4), pp. 333–363.

Hullmann, A. and Meyer, M. (2003) 'Publications and patents in nanotechnology', *Scientometrics*. Kluwer Academic Publishers, 58(3), pp. 507–527.

Johnstone, N., Haščič, I. and Popp, D. (2010) 'Renewable Energy Policies and Technological Innovation: Evidence Based on Patent Counts', *Environmental and Resource Economics*. Springer Netherlands, 45(1), pp. 133–155.

Kagermann, H., Wahlster, W. and Helbig, J. (2013) *Securing the future of German manufacturing industry: recommendations for implementing the strategic initiative INDUSTRIE 4.0. Final report of the Industrie 4.0 working group.*

Leydesdorff, L. and Meyer, M. (2007) 'The scientometrics of a Triple Helix of university-industry-government relations (Introduction to the topical issue)', *Scientometrics*. Kluwer Academic Publishers, 70(2), pp. 207–222.

Li, J. *et al.* (2015) 'Big Data in product lifecycle management', *The International Journal of Advanced Manufacturing Technology*. Springer London, 81(1–4), pp. 667–684.

Li, J. Q. *et al.* (2015) 'The Value of Field Experiments', *Management Science*. INFORMS, 61(7), pp. 1722–1740.

Maklan, S., Peppard, J. and Klaus, P. (2015) 'Show me the money', *European Journal of Marketing*. Emerald Group Publishing Limited, 49(3/4), pp. 561–595.

Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2012) *Foundations of machine learning*. Cambridge MA: MIT Press.

Naimur Rahman, M., Esmailpour, A. and Zhao, J. (2016) 'Machine Learning with Big Data An Efficient Electricity Generation Forecasting System', *Big Data Research*. Elsevier, 5, pp. 9–15.

Narin, F., Albert, M. B. and Smith, V. M. (1992) 'Technology indicators in strategic planning', *Science and Public Policy*. Oxford University Press, 19(6), pp. 369–381.

Nemet, G. F. (2009) 'Demand-pull, technology-push, and government-led incentives for non-incremental technical change', *Research Policy*. Elsevier Science Publishers B.V. (North-Holland), 38(5), pp. 700–709.

Popp, D. (2005) *Using the Triadic Patent Family Database to study environmental innovation*. Paris, France: OECD.

Pournarakis, D. E., Sotiropoulos, D. N. and Giaglis, G. M. (2017) 'A computational model for mining consumer perceptions in social media', *Decision Support Systems*. North-Holland, 93, pp. 98–110.

Samuel, A. (1963) 'Some studies in machine learning using the game of checkers', *Computers and Thought*, pp. 14–36.

Smola, A. and Vishwanathan, S. V. N. (2008) *Introduction to Machine Learning*.

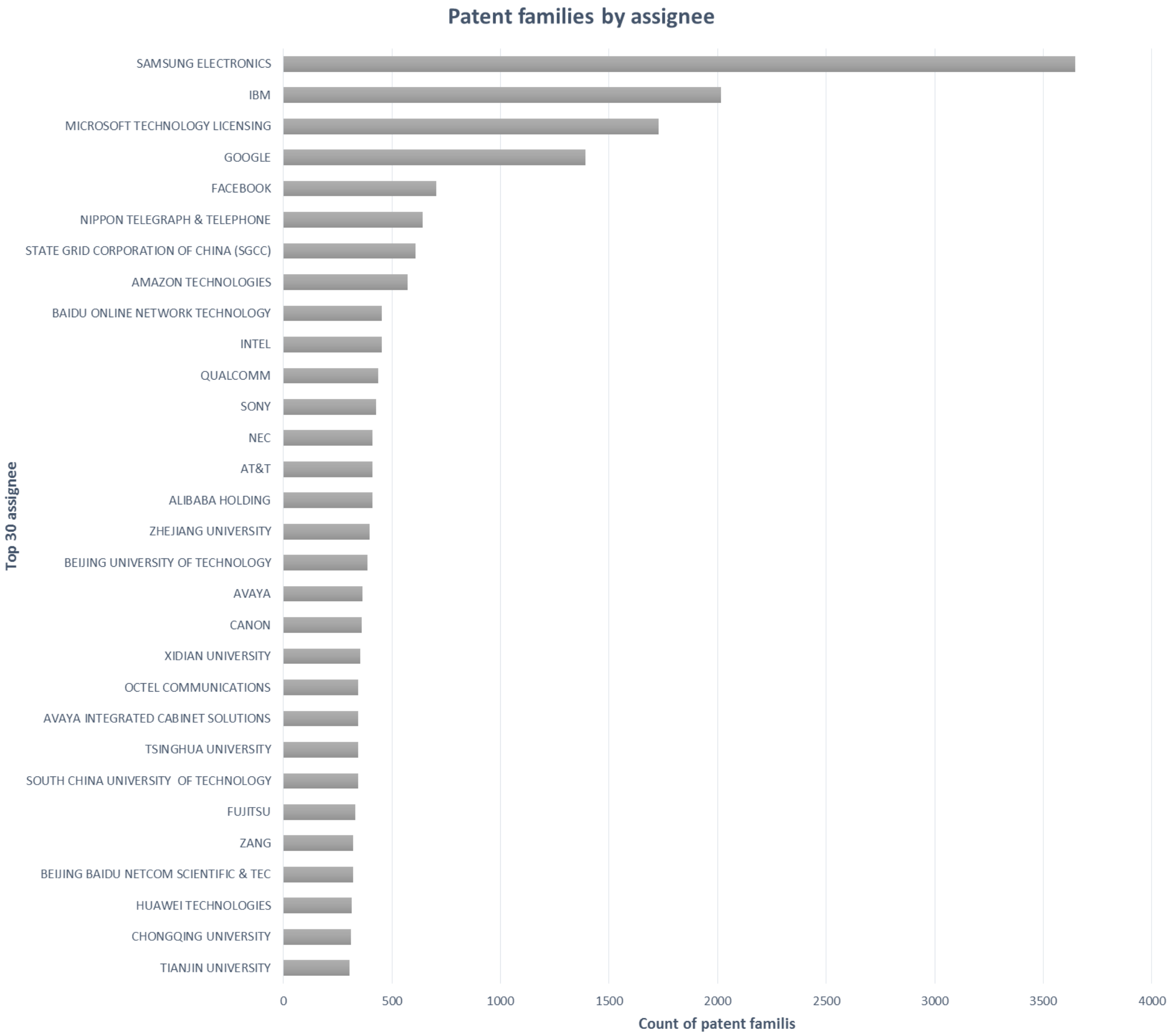
Susto, G. A. *et al.* (2015) 'Machine Learning for Predictive Maintenance: A Multiple Classifier Approach', *IEEE Transactions on Industrial Informatics*, 11(3), pp. 812–820.

Trajtenberg, M. (1990) 'A Penny for Your Quotes: Patent Citations and the Value of Innovations', *The RAND Journal of Economics*. WileyRAND Corporation, 21(1), p. 172.

Wallin, J. A. (2005) 'Bibliometric Methods: Pitfalls and Possibilities', *Basic & Clinical Pharmacology & Toxicology*, 97(5), pp. 261–275.

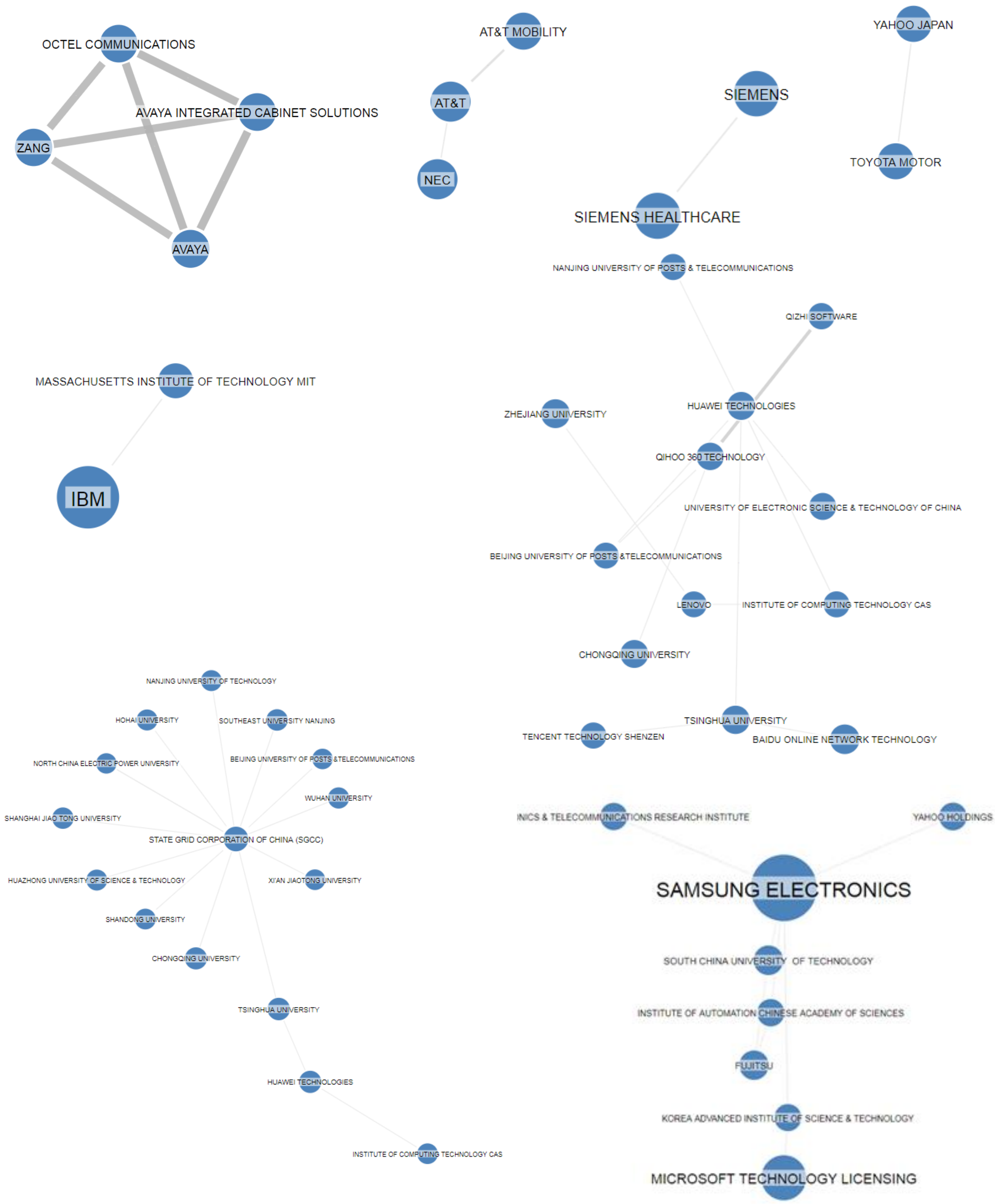
Wang, Y. *et al.* (2017) 'Industry 4.0 : a way from mass customization to mass personalization production', *Advances in Manufacturing*. Shanghai University.

Appendix A. NUMBER OF PATENS FAMILIES BY ASIGNEE (TOP 30)



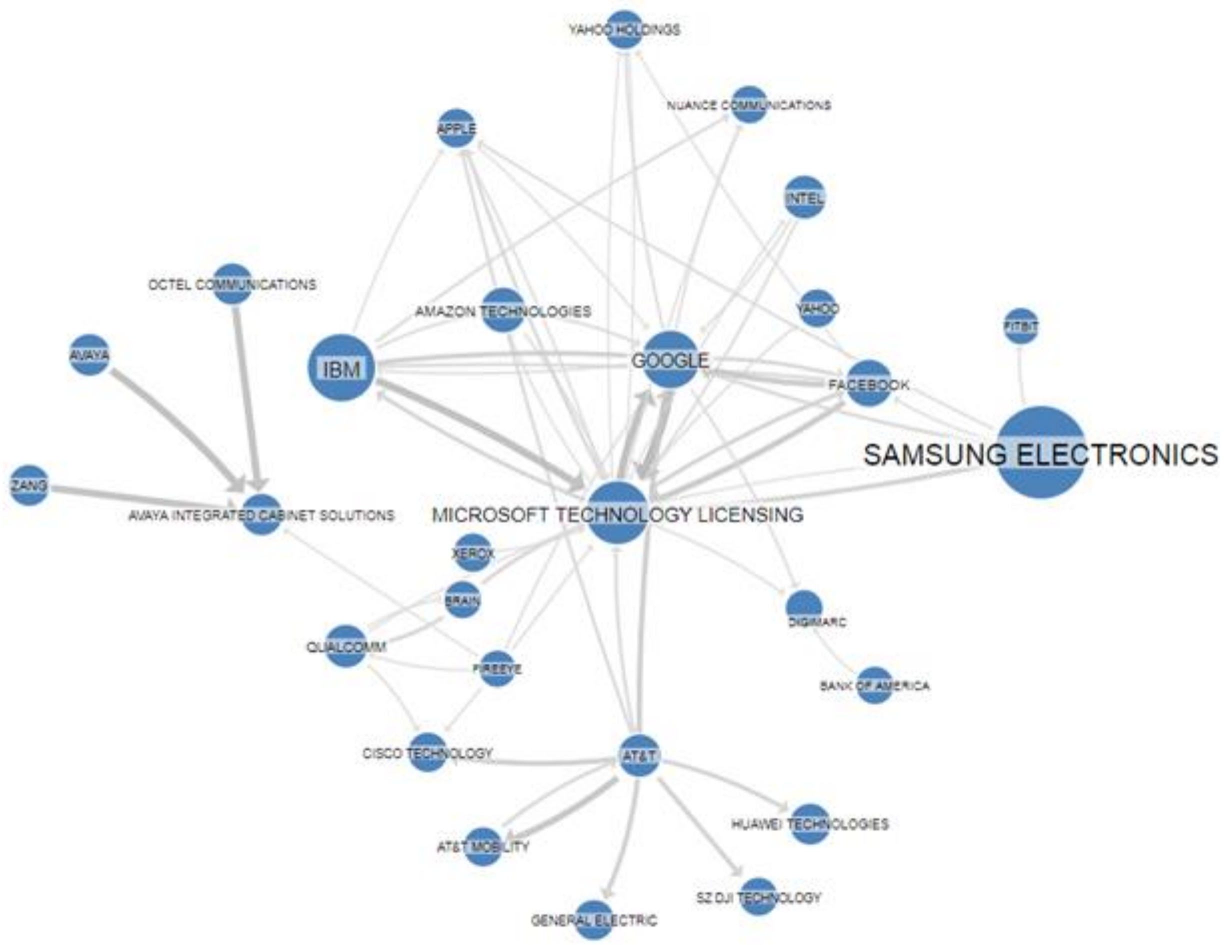
Appendix B GRAPHS OF CO-ASSIGNEE

This graph illustrates the interactions between applicants. A connecting row indicates that the two assignee have least one patent co-filing. The grater the width of the line, the higher the number of co-filings.

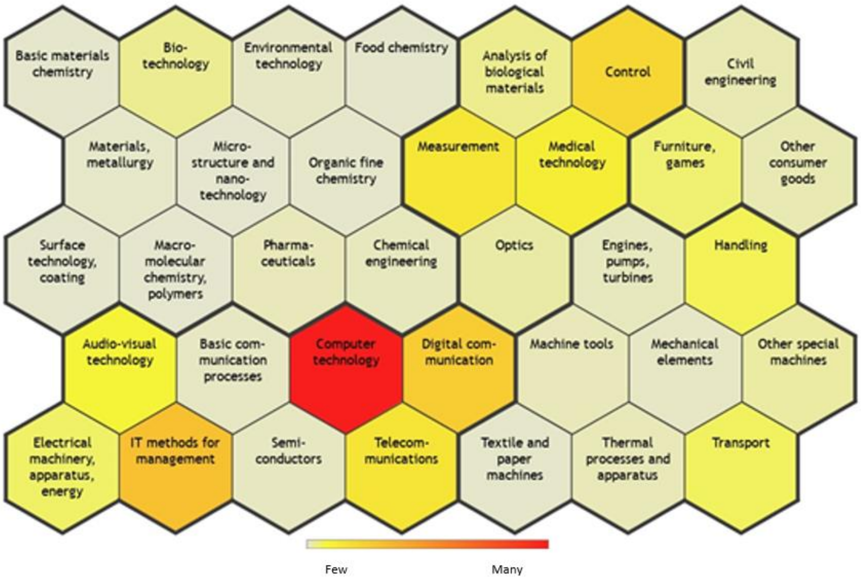


Appendix C GRAPH OF MOST RELEVANT ASSIGNEE CITATION

This graph illustrates citations between applicants. A connecting row form node A to node B indicates that at least one patent belonging to assignee A cited a patent belonging to B. The grater the width of the line, the higher the number of citation.



Appendix D PATENTS FAMILIES BY TECHNOLOGY DOMAIN (IPC)



In this figure, the IPC codes have been grouped in 35 technology fields.

Appendix E PATENTS FAMILIES BY CPC CLASSIFICATION

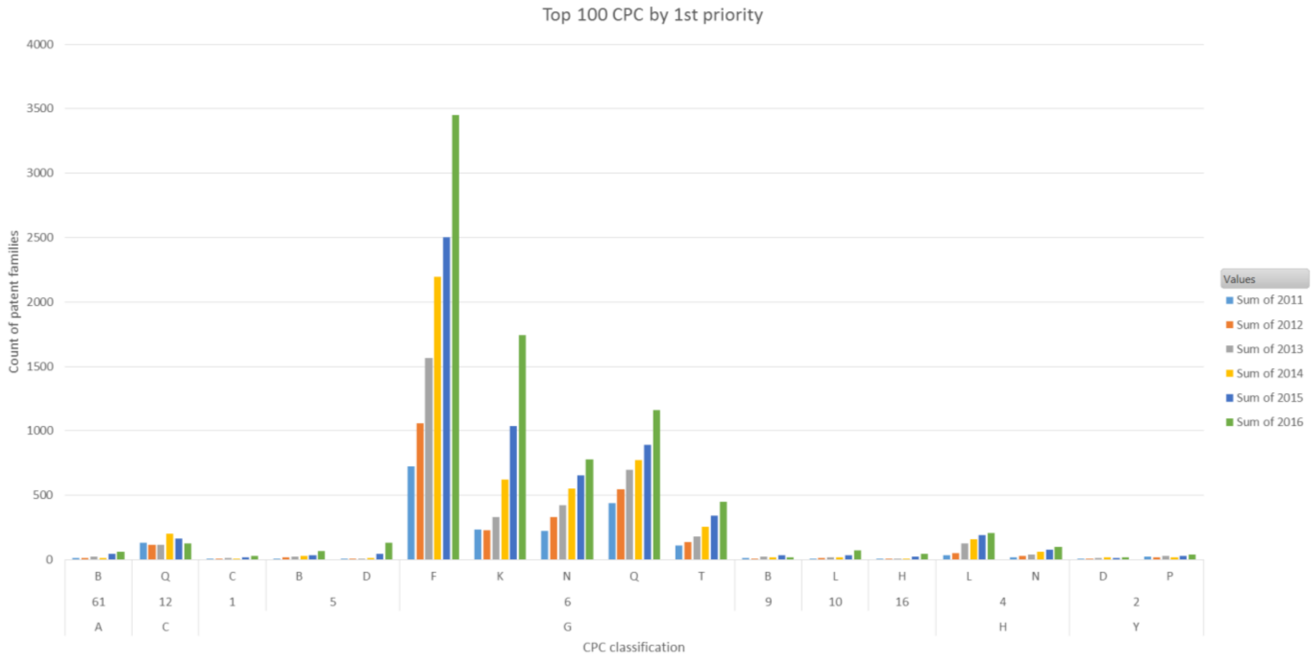


Table 2: Brief description of the CPC classes included in the graph

1ST LEVEL	2ND LEVEL	3RD LEVEL	DESCRIPTION
A			Human necessities health; amusement
	61		Medical or veterinary science; hygiene
		B	Diagnosis; surgery; identification
C			Chemistry; metallurgy
	12		Biochemistry; beer; spirits; wine; vinegar; microbiology; enzymology; mutation or genetic engineering
		Q	Measuring or testing processes involving enzymes or microorganisms; compositions or test papers therefor; processes of preparing such compositions; condition responsive control in microbiological or enzymological processes
G			Physics
	1		Measuring; counting; testing
		C	Measuring distances, levels or bearings; surveying; navigation; gyroscopic instruments; photogrammetry or videogrammetry
	5		Controlling; regulating (specially adapted to a particular field of use)
		B	Control or regulating systems in general; functional elements of such systems; monitoring or testing arrangements for such systems or elements
		D	Systems for controlling or regulating non-electric
	6		Computing; calculating; counting
		F	Electrical digital data processing
		K	Recognition of data; presentation of data; record carriers; handling record carriers
		N	Computer systems based on specific computational models
		Q	Data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes; systems or methods specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes, not otherwise provided for
		T	Image data processing or generation, in general
	9		Education; cryptography; display; advertising; seals

		B	Educational or demonstration appliances; appliances for teaching, or communicating with, the blind, deaf or mute; models; planetaria; globes; maps; diagrams
	10		Musical instruments;acoustics
		L	Speech analysis or synthesis; speech recognition; speech or voice processing; speech or audio coding or decoding
	16		Information and communication technology [ICT] specially adapted for specific application fields
		H	Healthcare informatics, i.e. Information and communication technology [ICT] specially adapted for the handling or processing of medical or healthcare data
H			Electricity
	4		Electric communication technique
		L	Transmission of digital information, e.g. Telegraphic communication
		N	Pictorial communication, e.g. Television
Y			General tagging of new technological developments; general tagging of cross-sectional technologies spanning over several sections of the IPC; technical subjects covered by former USPC cross-reference art collections and digests
	2		Technologies or applications for mitigation or adaptation against climate change
		D	Climate change mitigation technologies in information and communication technologies [ICT], i.e. Information and communication technologies aiming at the reduction of their own energy use
		P	Climate change mitigation technologies in the production or processing of goods