World Patent Information 55 (2018) 37–51

Contents lists available at ScienceDirect

World Patent Information

journal homepage: www.elsevier.com/locate/worpatin

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

Intellectual Property and Innovation Management (IIPM) research group, Centre for Technology Management (CTM), Institute for Manufacturing (IfM), Department of Engineering, University of Cambridge, United Kingdom

A R T I C L E I N F O

Keywords:
Intellectual property analytics
Patent analytics
Machine learning
Deep learning
Artificial intelligence

A B S T R A C T

Big data is increasingly available in all areas of manufacturing and operations, which presents an opportunity for better decision making and discovery of the next generation of innovative technologies. Recently, there have been substantial developments in the field of patent analytics, which describes the science of analysing large amounts of patent information to discover trends. We define Intellectual Property Analytics (IPA) as the data science of analysing large amount of IP information, to discover relationships, trends and patterns for decision making. In this paper, we contribute to the ongoing discussion on the use of intellectual property analytics methods, i.e artificial intelligence methods, machine learning and deep learning approaches, to analyse intellectual property data. This literature review follows a narrative approach with search strategy, where we present the state-of-the-art in intellectual property analytics by reviewing 57 recent articles. The bibliographic information of the articles are analysed, followed by a discussion of the articles divided in four main categories: knowledge management, technology management, economic value, and extraction and effective management of information. We hope research scholars and industrial users, may find this review helpful when searching for the latest research efforts pertaining to intellectual property analytics.

1. Research background

Big data is increasingly available in all areas of manufacturing and operations [1]. Data as such presents value for enabling a competitive data-driven economy, which is at the heart of the Internet of things and Industry 4.0 [2,3]. Increased data availability presents an opportunity for better decision making and strategy development [4], to introduce the next generation of innovative and disruptive technologies [5].

Over the last two decades, there have been substantial developments in the field of patent analytics. Patent analytics describes the science of analysing large amounts of intellectual property information, in relation to other data sources, to discover relationships and trends [6–9]. With the digitization of patent data, the world's largest repository of technical information has become accessible for rapidly decreasing costs. While patent data has long been considered the world's largest repository of technological information, and only with its digitization since the BACON project in 1984 [10] and numerous gradual and cumulative improvements of tbe data quality and analytical techniques over the last decades, patent data has become increasingly accessible to and useful for a non-specialist audience [11]. 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 IP data [6,12–14].

In a recent study, we have used the technology roadmapping approach [15] to explore the future of patent analytics [16,17]. We identify 11 priority technologies, such as artificial intelligence and artificial neural networks, that industry experts believe to be important to be adopted at a higher rate in the patent analytics domain [12]. While other domains have adopted such technologies already widely, the patent analytics domain seems to be catching up. We identify the need of adoption of these computer science techniques, to complement decision processes and provide decision support [11,12,18]. This is very much in line with the propositions by Refs. [19–22].

In this paper, we contribute to the ongoing discussion on the use of machine learning and deep learning approaches to analyse intellectual property data [12], by presenting the outcomes of a literature review on the state-of-the-art on intellectual property analytics. In particular, we focus the literature review on the use of artificial intelligence

* Corresponding author.
E-mail addresses: la324@cam.ac.uk (L. Aristodemou), frank.tietze@eng.cam.ac.uk (F. Tietze).

https://doi.org/10.1016/j.wpi.2018.07.002
Received 1 February 2018; Received in revised form 16 April 2018; Accepted 7 July 2018

techniques, such as machine learning and deep learning, to analyse intellectual property data. We create a taxonomy on the segmentation of the different approaches and methods to analyse the data, which builds on the work by Refs. [6,17,18].

The literature review is organized as follows. Section 2 defines intellectual property analytics. Section 3 presents the methodology adopted, followed by section 4, which presents the background and significance of the patent analysis process. Section 5 presents the bibliographic analysis results, followed by section 6, which discusses the reviewed papers. Section 7 concludes the literature review and highlights some future research directions.

2. Intellectual property analytics

Intellectual Property Analytics (IPA) is the data science of analysing large amount of intellectual property information, to discover relationships, trends and patterns in the data for decisioning 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. We make use of this definition, as there is no widely accepted definition of IPA; however, this is very much in line with the definition of *Patinformatics* [8,9].

3. Intellectual property analytics process

It is important to understand the process of analyzing patent data when discussing patent analytics and intellectual property analytics. Trippe [14] presents a WIPO guide, which identifies and explains a large number of concepts on patent analysis and the methodology on how to run the different types of analysis on patent data. With the recent advancements of artificial intelligence, there has been a positive amount of activity around the different methodologies involved that could be applied to intellectual property data [11,12].

Most of the literature makes use of the process as defined by Moehrle et al. [8] in a business context, and consists of three main stages: the pre-processing stage, the processing stage and the post processing stage. In the pre-processing stage, the data are collected, after information extraction, cleaned and well prepared, with the purpose in providing these data in high quality, correctness and completeness. In the processing stage, the analysis of the data extracted in the pre-processing stage takes place using different methods to classify, cluster, and identify meaningful insights from the information. In the post processing stage, also known as discovered knowledge, the results and information from the processing stage are visualized and evaluated to support strategic decision making. In this literature review, we give particular focus on the processing and post-processing stage, and the specific algorithms used to analyse the data.

This is similar to Abbas et al. [6], who presents a generic patent analysis work-flow, with the distinction that every analysis made has a specific purpose. Raturi et al. [23] argues that this process is a complementary process to the innovation cycle, and that the analysis of intellectual property data has many application in many fields. Bonino et al. [24] link the patent life cycle to the patent related information sources and the different tasks along the patent analytics tasks. They argue that a patent analytics process is a purpose-driven process, which consists of search tasks (patent ability, validity, infringement, portfolio survey, technology survey), analysis tasks (micro and macro assessment of business value, technical assessment and technology suggestions), and monitoring tasks (early sign monitoring, technology monitoring, portfolio monitoring, single patent monitoring). Similarly, Baglieri and Cesaroni [7] argue that patent analysis is a form of patent intelligence to support decision making. They argue that there are two meanings to the term of patent analysis, the process that considers all of the above, and the actual analysis of the patent data. They use the research by Bonino et al. [24] to define the three patent analysis tasks, patent searching, patent analysing and patent monitoring, and link value of information from these to the open innovation funnel.

4. Methodology

The paper aims to summarise the existing work, especially when it comes to the application of artificial intelligence methods, such as machine learning, artificial neural networks and deep learning, in the intellectual property domain [6,12]. To carry out the literature review, the narrative and scoping literature review approaches have been adopted [25,26], and a research search strategy has been developed [27,28]. The detailed search strategy is found in Appendix A. Fig. 1 shows the process flow for the comprehensive literature review.

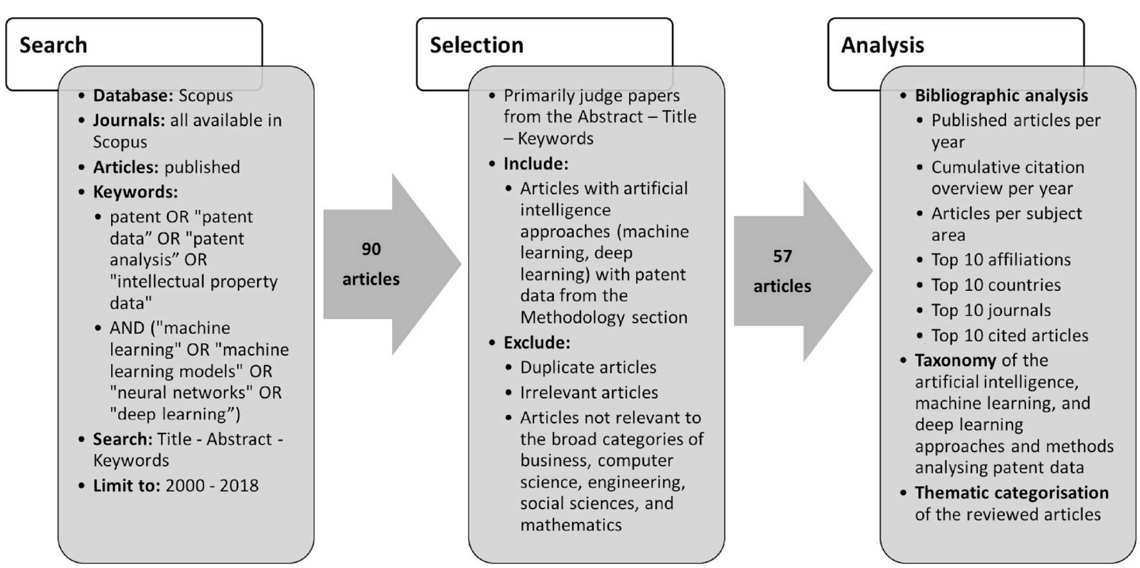

Fig. 1. Process flow of the search, selection and analysis of the comprehensive literature review.

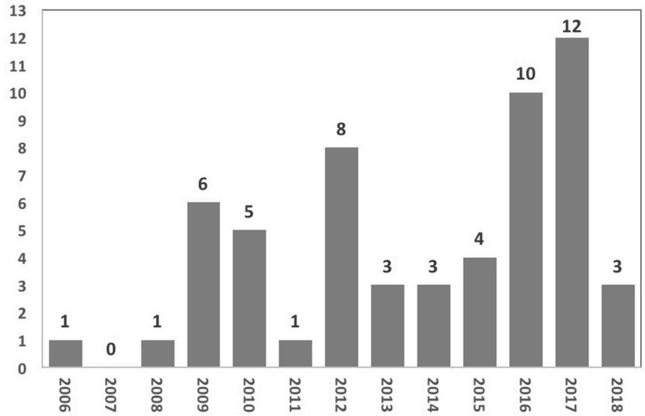

Fig. 2. Number of articles per year ($n_1 = 57$) since 2006 (< 2006 = 0 articles).

The articles on intellectual property analytics and patent analytics were identified from the Scopus database to find the most relevant published articles or in press articles. We search within the tittle, abstract and key words for various terms such as "patent", "patent data", "patent analysis" and "intellectual property data". The search is then narrowed to documents that also contain either in the tittle or the abstract or in the key words, the terms "machine learning", "machine learning models", "neural networks", "deep learning" and "artificial intelligence". In order to focus on recent literature, the search is limited to articles published after the year 2000, and to the fields of business, computer science, engineering, social science and mathematics. The search, effective the 8th January 2018, retrieved 90 documents. After manual screening, e.g. removing duplicate or irrelevant articles, 57 articles remained, which form the core of this review. The purpose of presenting the research articles in detail is to provide the readers with the latest research on the use of artificial intelligence methods, such as machine learning and deep learning, in the intellectual property domain, which analyse patent data. Results are presented in section 5, with the first level focusing on the bibliographic information, followed by the discussion of the methods in section 6 and the emerging themes of application.

5. Bibliographic analysis results

The first level of the analysis of the reviewed papers focuses on the analysis of the bibliographic information from the 57 articles (n_1). The detailed review of the aim, data, pre-processing and algorithms of each article is found in Appendix B. The number of articles have increased over the last few year, reaching a peak of 12 articles published in 2017. Fig. 2 shows the number of papers per year since the year 2000. There is an upward trend with the number of publications in recent years indicating an increasing interest in this particular field. This is further enhance by Fig. 3, which shows the cumulative number of citations received per article per year, rising with an upward trend and reaching the peak of 153 citations in the year 2017. Fig. 4 illustrates the percentage of articles by subject area. The majority of articles are concentrate in the subject area of computer science (29%), followed by social sciences (14%), business, management and accounting (12%), and engineering (8%).

Table 1 shows the top 10 affiliations of the article authors. It is evident from the information that Asia is the leading continent in the application of machine learning techniques to patent data. This is also supported by Table 2, where the top 3 countries are Taiwan, South Korea and China, with 18, 12 and 8 articles respectively. However, from Table 2 we can see that contributions are also made by US scholars (8% of the total articles) despite no US affiliation appearing in Table 1. European countries also are shown to have a strong influence in this domain, with countries like Germany, Serbia, accounting for the majority of articles, followed by Spain and Belgium.

Moreover, Table 3 reveals the top 10 journals in which relevant articles are published. The top two journals, which account for 30% of the articles (15% each) are Technological Forecasting and Social Change and Scientrometrics. These are followed by Expert Systems With Applications with 7%, and World Patent Information and

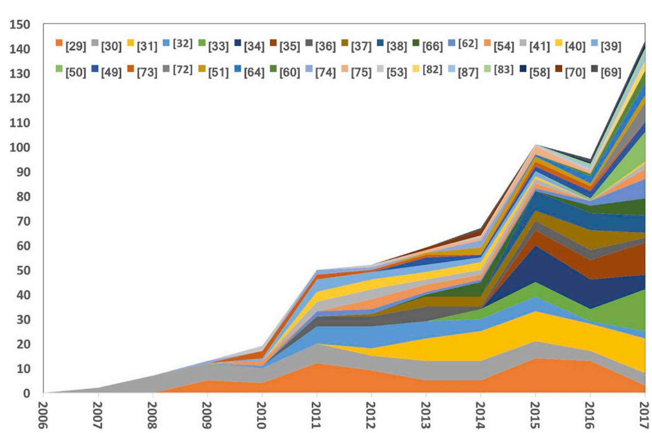

Fig. 3. Cumulatitve citation overview per article per year, for articles with more than 2 citations.

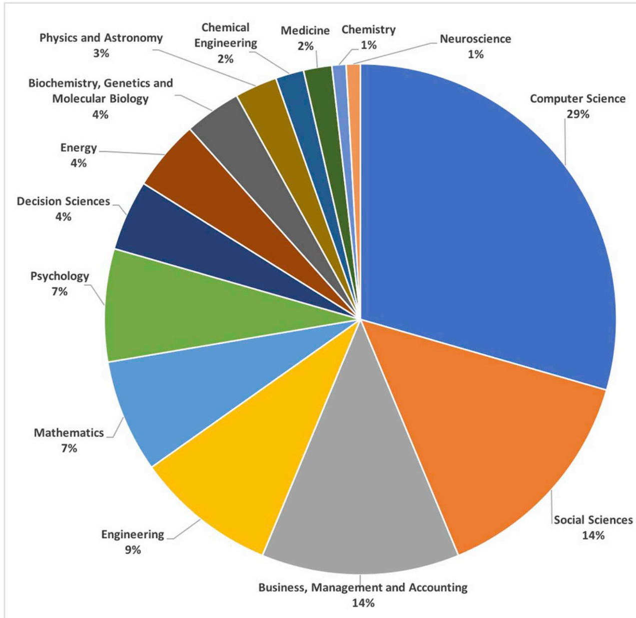

Fig. 4. Articles by subject area ($n_1 = 57$).

Table 1
Top 10 affiliations ($n_1 = 57$ articles, n_2 = 128 observations).

Affiliation	No. of observations	Share of total (%)
National Tsing Hua University	7	5%
National Chiao Tung University Taiwan	6	5%
Korea University	5	4%
Cheongju University	5	4%
National Yunlin University of Science and Technology	5	4%
University of Niš	4	3%
Korea Institute of Science and Technology Information	3	2%
Gainia Intellectual Asset Services, Inc.	2	2%
Chung Hua University	2	2%
Beijing Institute of Technology	2	2%
Total	41	33%

Note: Articles with one or more affiliations are multi-counted.

Table 2
Top 10 countries ($n_1 = 57$ articles, n_2 = 71 observations).

Country	No. of observations	Share of total (%)
Taiwan	18	25%
South Korea	12	17%
China	8	11%
United States	6	8%
Germany	4	6%
Serbia	4	6%
Spain	3	4%
Belgium	2	3%
Japan	2	3%
Hong Kong	1	1%
Total	60	84%

Note: Articles with one or more countries are multi-counted.

Sustainability Switzerland with 5% each. The top 10 journals account for a total of 33 articles. This indicates that published articles in this field are fragmented in a total of 33 journals. In both Tables 1 and 2, any article with one or more afflication from different countries is multi-counted. For example, if an article has 3 different affiliations from 2 different countries, is counted 3 times in Tables 1 and 2 times in Table 2.

Table 4 shows the top 10 cited articles, their citations and the citation frequency. The most cited article is Klinger et al. [29] with 70 citations, followed by Trappey et al. [30] with 68 citations, and Trappey et al. [31] with 61 citations. However, the article with the highest citation frequency (defined as the total number of citations over the age of the article) is Krallinger et al. [34] with 11.00, followed by Trappey et al. [31] with 10.17, and Klinger et al. [29] with 7.00. 8 out of the 10 articles are published since 2010, which indicates an increase in hte field's importance, as well as the importance of analysing patent data.

Table 3
Top 10 journals ($n_1 = 57$ articles).

Journal	No. of articles	Share of total (%)
Technological Forecasting And Social Change	8	14%
Scientometrics	8	14%
Expert Systems With Applications	4	7%
World Patent Information	3	5%
Sustainability Switzerland	3	5%
Database The Journal Of Biological Databases And Curation	2	4%
International Journal Of Applied Engineering Research	2	4%
Physica A Statistical Mechanics And Its Applications	2	4%
Advanced Engineering Informatics	1	2%
Applied Soft Computing Journal	1	2%
Total	33	61%

Note: The 57 articles have been published in 33 journals.

Table 4
Top 10 cited articles ($n_1 = 57$ articles).

Article	Cited by	Citation freq.
Klinger et al. [29]	70	7.00
Trappey et al. [30]	68	5.67
Trappey et al. [31]	61	10.17
Thorleuchter et al. [32]	39	4.88
Jun et al. [33]	34	8.50
Krallinger et al. [34]	33	11.00
Jun et al. [35]	27	5.40
Chen et al. [36]	25	3.13
Jun et al. [37]	23	3.83
Zhang et al. [38]	22	5.50

6. Intellectual property analytics methods

There are several analytic methods in the literature that have been used with intellectual property data, and specifically patent data [6,14]. Abbas et al. [6] provide a comprehensive literature review on the patent analytics techniques, where they distinguish between text mining and visualization approaches and the applicability to structured and unstructured data [24]. In this literature review, we specifically review articles that use intellectual property analytics methods, i.e artificial intelligence methods and approaches, to analyse intellectual property data and more specific patent data.

Table 5 presents a summary of the intellectual property analytics methods, identified in the literature. It is evident that 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). Moreover, the majority of articles focus on classification methods, with some combining both clustering and classification methods.

From the review, four categories emerge in which the use of artificial intelligence methods with intellectual property data is implemented: the first category is knowledge managmenet, with a number of scholars focusing on patent evaluation and patent quality classification (e.g Refs. [30,31,60]. The second category is technology management, which includes technological patentability, R&D planning within organisations, technology intelligence including monitoring technological changes, identification and forecasting of emerging technologies (e.g. Ref. [32,35,46]. The third category is the economic value of intellectual property(in this case patents), and its impact in other areas for example law (e.g. Refs. [41,57,62]. The fourth category is a hybrid category, which includes extraction of information and effective management of the information. This concentrates on extraction of features from patents, such as chemical formulae and figures, or the effective classification of patents in technological areas (e.g Refs. [29,65,75]. The boundaries between these categories are permeable, if not overlapping, as the categories are inter-related and one article could be part of more than one category. Further details on these categories are provided in the following subsections, with the focus being on AI, ML, and DL publications. In each category, the articles are reviewed by order of citations, i.e. starting with the article that has the highest number of citations, and then grouped into themes.

Table 5
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]

6.1. Knowledge management

Knowledge management is important for firms, as it supports both the internal and external organisations, and improves their ability to solve problems better, adapt, and evolve to meet changing business requirements [78]. To support explicit knowledge management, Trappey et al. [30] develop a document classification and search methodology based on neural network technology that helps companies manage patent documents more effectively. In an overlapping area with innovation management, Trappey et al. [31] help firms to evaluate intellectual property rights and the quality of patent documents for innovative product development and discovery of state-of-the-art technology trends. Using patent transactions with the back propagation neural network, the classify patents according to their quality, with an accuracy of 85%. Moreover, Trappey et al. [49] propose a knowledge management approach using ontology-based artificial neural network algorithm to automatically classify and search knowledge documents, stimulating new product development innovation for effective collaborative management [42].

The focus on patent quality is also evident in the work by Wu et al. [60], who develop an automatic patent quality analysis and qualification system, which is based on a combination of self-organising maps, kernel principal component analysis and support vector machine. To improve the incorporation of prior knowledge in incremental innovation, Lu et al. [74] use a hybrid min-max modular (M3) and support vector machine classifier to improve learning performance on Japanese patents. In addition, Hido et al. [69] assess the quality of patent applications, with a combined machine learning and text mining approach, which computes a patentability score. The patentability score gives the likelihood that a patent application is approved by a patent office.

6.2. Technology management

Technology management is a set of management disciplines that allows organisations to manage their technologies to create competitive advantage [79].

Identification of technological trends is important to decision makers for R&D management. Thorleuchter et al. [32] propose a methodology to make the technology impact more transparent, which is based on a quantitive cross impact analysis. This method shows current technology impact and trends from the R&D of an organization's competitors, comparing these to the technology impact and impact trends from the organization's own R&D, and estimating the impact across technologies. Similarly, Jun et al. [33] use a hollistic approach to analyse published articles, papers and patents on developing technologies, to identify scientific and technological trends. Suominen et al. [56] discuss the benefits and constraints of machine learning approaches in industry level patent analysis. They also propose a classification using full-text descriptions with Latent Dirichlet Allocation, to create an overall view of patenting within the industry. In this way, they are able to identify technology trends and forecast future trends. Moreover, the visual impact of the evolution of technological trends can be very useful. Sung et al. [80] use a growing cell structure neural network, paired with the Girvan-Newman algorithm, to construct a map that visualises technological evolution and shows the developmental trend in a technological field. In addition, technology life cycle analysis is important for firm related investment strategies, as technological trend monitoring. Lee et al. [72] propose a stochastic technology life cycle analysis that uses multiple patent indicators to examine a technology's progression through its life cycle. The authors employ a hidden Markov model to estimate the probability of a technology being at a certain stage of its life cycle and identify patterns. Govindarajan et al. [71] propose a topic modelling approach, based on latent Dirichlet allocationp (LDA) algorithm to construct a domain ontology and identify technical and functional development trends for Industry 4.0.

Closely related to the identification of technological trends, is the ability to forecast technological innovation. To study the technological innovation and forecasting of Apple, Jun and Park [35] firstly build a statistical models using the time series regression and multiple linear regression methods to create a technology map, followed by clustering to find Apple's vacant technology domains. They then use social network analysis to search for technologies central to Apple's future. Having a new technology opportunity is a significant variable that can lead to dominance in a competitive market. In that context, accurately understanding the state of development of technology convergence and forecasting promising technology convergence can determine the success of a firm. Kim and Lee [44] propose a forecasting methodology for multi-technology convergence, based on a patent-citation analysis, a dependency-structure matrix, and a neural-network analysis. The methodology enables planning for technology development of future technology combinations. Forecasting the number of patent applications is also an important factor to see the development of a technological field, where Zhang et al. [76] propose a support vector machine approach of doing that, overcoming the sparsity problem mainly found in patents. Jun [81] build a combined clustering method using dimension reduction and K-means clustering, which is based on support vector clustering and Silhouette measure, to identify clusters for technology forecasting by patent analysis. Furthermore, Tenorio-González and Morales [61] describe a system, called Automatic Discovery of Concepts that combines techniques from inductive logic programming with predicate invention and reinforcement learning with intrinsic motivation to discover new concepts.

Understanding current technological changes is the basis for better forecasting of technological changes. Trappey et al. [59] develop an Intellectual Property (IP) analytical methodology to explore the portfolios and evolution of patents in the bio additive manufacturing domain, for decision support and strategic planning. Momeni and Rost [82] suggest a method based on patent-development paths, k-core analysis and topic modelling of past and current trends of technological development to identify technologies that have the potential to become disruptive technologies [83]. Emerging technologies drive technological development and innovation in diverse fields [52,84]. Kyebambe et al. [45] propose an approach, which focuses on citation data, to automatically label data and train learners to forecast emerging technologies, within a one year window. Moreover, Lee et al. [46] propose a neural network approach, to identifying emerging technologies at early stages using multiple patent indicators, which can be defined immediately after the relevant patents are issued. Moreover, a purpose of R&D is technology commercialization and technology transfer. Jun and Lee [37] propose a patent information analysis, which combines statistical inference and neural networks to construct an emerging technology forecasting model, where as Choi et al. [64] construct a predictive patent analysis model, based on social network analysis and decision tress, for technology transfer.

6.3. Economic value of intellectual property

Economic development can be achieved through science and technology. By applying computational intelligence methodologies, such as artificial neural networks, economic development estimation can be determined based on different science and technology factors [43,47,85]. Markovic et al. [55]d evelop and apply the extreme learning machine (ELM) to forecast the gross domestic product (GDP) growth rate.

Moreover, Lee et al. [72] propose a quantitative corporate performance prediction model that applies the support vector regression (SVR) algorithm to predict corporate performance, from both financial and technical data. Lee et al. [57] construct a deep neural network-based corporate performance prediction model that uses a company's financial and patent indicators as predictors. The model includes an unsupervised learning phase, which uses a restricted Boltzmann machine, and a fine-tuning phase, with a backpropagation algorithm.

Furthermore, in a group of studies in the US pharmaceutical industry, a number of scholars apply artificial neural networks to explore the influences of the quantitative and qualitative patent indicators upon corporate market value, showing that US pharmaceutical companies should not concentrate most of their R&D resources on one particular technological field, but create wider technological capabilities to avoid missing new technological opportunities [36,39]. In a similar way, Chen and Chang [40] explore the nonlinear effects of firm size, profitability, and employee productivity on patent citations, whereas Chen and Chang [41] explore the relationship between the Herfindahl-Hirschmann Index (HHI) of patents and the relative patent position in the most important technological field of the firm [51]. Bass and Kurgan [62] analyse nanotechnology patents with a number of different classification algorithms, to identify the impact of different factors on patent value, determing which of these differentiate between the top-performing and the remaining nanotechnology patents. Lai and Che [54] focus their research on patent law and proposed a valuation model, based on an extension neural network, for the monetary legal value of patents, which uses the damage award of a patent infringement lawsuit as the legal value of a patent [53].

6.4. Extraction of information and effective management of information

Extraction of information and effective management of information are fundamental systems of any firm's management system. Research in this area has mainly focused on three theme: (1) the extraction and chemical name recognition, (2) the extraction and identification of figures, and (3) the effective management of collective information.

Klinger et al. [29] present a machine learning approach based on conditional random fields to find mentions of IUPAC and IUPAC-like names in patents. An evaluation of hand-selected patent sections containing large enumerations and terms with mixed nomenclature shows a good performance on these cases, with an accuracy of 81.5%. Krallinger et al. [34] describe the chemical compound and drug name recognition community challenge (27 teams took part), which promoted the development of novel, competitive and accessible chemical text mining systems [58,63,70]. This task allowed a comparative assessment of the performance of various methodologies using a carefully prepared collection of manually labeled text prepared by specially trained chemists as Gold Standard data. They evaluated two main tasks: one covered the indexing of documents with chemicals (chemical document indexing), with accuracy of 88.20%, and the other was concerned with finding the exact mentions of chemicals in text (chemical entity mention recognition), with accuracy of 87.39%. The main strategies used to detect chemicals include machine learning methods (e.g. conditional random fields) using a variety of features, chemistry and drug lexica, and domain-specific rules.

Furthermore, Vrochidis et al. [75] present an approach for automatically extracting concept information describing the patent image content to support searchers during patent retrieval tasks. The approach is based on a supervised machine learning framework, which relies upon image and text analysis techniques. Similarly, Riedl et al. [48] suggest a number of algorithm for graphical recognition of patent figures.

Patent classifications are important for effective patent analysis and innovation analysis. An interactive patent classification algorithm based on multi-classifier fusion and active learning is constructed by Zhang [38]. Similarly, Venugopalan and Rai [50] present a natural language processing based hierarchical method, in combination with a support vector machine algorithm that enables the automatic identification and classification of patent datasets into technology areas. Zhu et al. [65] propose an automatic requirement-oriented patent classification scheme as a complementary method using supervised machine learning techniques to classify patent dataset into a user-defined taxonomy. Callaert et al. [66]develop a machine learning method that allows the automated identification of scientific references. Wu and Yao [77] propose a novel patent analysis method, called the intelligent patent network analysis method, to make a visual network, which provides an automated procedure for searching patent documents, extracting patent keywords, and determining the weight of each patent keyword in order to generate a sophisticated visualization of the patent network. Moreover, due to the information overload problem, and the critical challenges faced by managers in utilizing the data in organisations, Li et al. [73] present a knowledge evolution processes with patent citations and introduce a labeled citation graph kernel to classify patents under a kernel-based machine learning framework. Trappey and Trappey [86] develop a methodology for discovering evolutions and linkages between litigations and disputed patents, using the modified formal concept analysis (MFCA) approach in the 4G telecommunications domain.

7. Conclusion

In this paper, we define intellectual property analytics and contribute by reviewing the literature on the use of intellectual property analytics methods, i.e. artificial intelligence, machine learning, deep learning and artificial neural network methods, for analysing IP data. In particular, we review 57 articles, which fall within 4 categories: (1) knowledge management, (2) technology management, (3) economic value, and (4) extraction and management of information. There has been an increase in the interest to the field, shown by the increase in the number of articles in recent years, and the increase in the total number of citations of the papers. We hope this literature review would be helpful for research scholars and industrial users, in finding the latest research efforts pertaining to intellectual property analytics in a unified form. This ensures the development of the field in both a research and industrial context. Further research is required in this field to identify use cases of intellectual property methods within the innovation process and apply these methods in firms.

Acknowledgement

The authors would like to acknowledge support of the Engineering and Physical Sciences Research Council (EPSRC).

Appendix

A. Search strategy

The detailed search strategy of key terms is found below:

- (TITLE-ABS-KEY (patent OR ″patent data" OR ″patent analysis" OR ″intellectual property data") AND TITLE-ABS-KEY ("machine learning" OR ″machine learning models" OR ″neural networks" OR ″deep learning")) AND (LIMIT-TO (DOCTYPE, "ar ") OR LIMIT-TO (DOCTYPE, "ip ")) AND (LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010) OR LIMIT-TO (PUBYEAR, 2009) OR LIMIT-TO (PUBYEAR, 2008) OR LIMIT-TO (PUBYEAR, 2007) OR LIMIT-TO (PUBYEAR, 2006) OR LIMIT-TO (PUBYEAR, 2005) OR LIMIT-TO (PUBYEAR, 2004) OR LIMIT-TO (PUBYEAR, 2003) OR LIMIT-TO (PUBYEAR, 2002) OR LIMIT-TO (PUBYEAR, 2001) OR LIMIT-TO (PUBYEAR, 2000)) AND (LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017) OR LIMIT-TO (PUBYEAR, 2016) OR LIMIT-TO (PUBYEAR, 2015) OR LIMIT-TO (PUBYEAR, 2014) OR LIMIT-TO (PUBYEAR, 2013) OR LIMIT-TO (PUBYEAR, 2012) OR LIMIT-TO (PUBYEAR, 2011) OR LIMIT-TO (PUBYEAR, 2010) OR LIMIT-TO (PUBYEAR, 2009) OR LIMIT-TO (PUBYEAR, 2008)) AND (LIMIT-TO (SRCTYPE, "j")) AND (EXCLUDE (EXACTSRCTITLE, "Recent Patents On Computer Science") OR EXCLUDE (EXACTSRCTITLE, "Recent Patents On Engineering") OR EXCLUDE (EXACTSRCTITLE, "Recent Patents On Electrical And Electronic Engineering") OR EXCLUDE (EXACTSRCTITLE, "Recent Advances In Electrical And Electronic Engineering") OR EXCLUDE (EXACTSRCTITLE, "Recent Patents On Chemical Engineering") OR EXCLUDE (EXACTSRCTITLE, "Recent Patents On DNA And Gene Sequences") OR EXCLUDE (EXACTSRCTITLE, "Recent Patents On Electrical Engineering")) AND (LIMIT-TO (SUBJAREA, "BUSI″) OR LIMIT-TO (SUBJAREA, "COMP″) OR LIMIT-TO (SUBJAREA, "ENGI″) OR LIMIT-TO (SUBJAREA, "SOCI″) OR LIMIT-TO (SUBJAREA, "MATH″))

B. Reviewed articles

Table 6 shows the 57 reviewed articles, included in this review, in order of highest cited article. The authors reviewed the article's aim, data, preprocessing methods to prepare the data, and the main method to analyse the data, linked to artificial intelligence approaches.

Table 6
Literature reviewed articles ($n_1 = 57$ articles), extracted from Scopus (last accessed on 8th January 2018).

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
1	[29]	Detection of IUPAC and IUPAC-like chemical names	Bioinformatics	70	To present an approach to find mentions of IUPAC and IUPAC-like names in scientific text and patents	463 abstracts, 26 patents	Bootstrapping	Conditional random fields
2	[30]	Development of a patent document classification and search platform using a back-propagation network	Expert Systems with Applications	68	To develop an automatic document classificiation module and a search module to find relevant and related patent documents	424 patents	Correlation matrix	Back propagation neural network
3	[31]	A patent quality analysis for innovative technology and product development	Advanced Engineering Informatics	61	To improve the analysis and ranking of patent quality, ultimately shortening the time required to determine and rank the quality of patents for new product R&D and innovation management	435 patents	Kaiser-Meyer-Olkin (KMO) approach, Principal Component Analysis (PCA)	Back propagation neural network
4	[32]	A compared R&D-based and patent-based cross impact analysis for identifying relationships between technologies	Technological Forecasting and Social Change	39	To support an organization's strategy and R&D planning, a methodology to make the technology impact more transparent is introduced	182,928 patents	Cross impact analysis, vector space modelling, multi-label text classification	Centroid-based classifier
5	[33]	Document clustering method using dimension reduction and support vector clustering to overcome sparseness	Expert Systems with Applications	34	To develop a clustering method of document data analysis, that also overcomes the sparsity problem	200 documents	Singular value decomposition (SVD), principal component analysis (PCA)	Kernel mapping support vector clustering, Silhouette measure, K-means clustering
6	[34]	CHEMDNER: The drugs and chemical names extraction challenge	Journal of Cheminformatics	33	To develop a novel, competitie and accessible chemical text mining system for chemical compount and drug name recognition	CHEMDNER corpus 10,000 abstracts	Rule-based approach, natural language processing, dictionary look-up	Conditional random fields, support vector machines
7	[35]	Examining technological innovation of Apple using patent analysis	Industrial Management and Data Systems	27	To study the technological innovation by analysing its patent applications	8119 patents	Time series regression, multiple linear regression	Silhoutette width, K-means algorithm, social network analysis
8	[36]	Exploring the nonlinear effects of patent citations, patent share and relative patent position on market value in the US pharmaceutical industry	Technology Analysis and Strategic Management	25	To explore the influences of the quantitative and qualitative patent indicators upon corporate market value in the US pharmaceutical industry	472 patents	Descriptive statistics and correlation coefficients	Back propagation neural network, self organising maps (SOM)
9	[37]	Emerging technology forecasting using new patent information analysis	International Journal of Software Engineering and its Applications	23	To construct an emerging technology forecasting model, which combines statistical inference and neural networks for new patent information analysis	2482 patents	Text mining techniques, multiple regression, pearson correlation analysis	Gradient-descent neural network
10	[38]	Interactive patent classification based on multi-classifier fusion and active learning	Neurocomputing	22	To construct an interactive patent classificiation method	5500 patents	Vector space model, multi-classifier fustion - linear fusion, super-kernel fusion	Support vector machine, active learning, dynamic certainty propagation (DCP)

(continued on next page)

Table 6 (*continued*)

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
11	[66]	Delineating the scientific footprint in technology: Identifying scientific publications within non-patent references	Scientometrics	18	To develop a method of automatically identifying scientific reference for patents	10 sets consisting of: 26,000 patents	TF-IDF, Monte Carlo process	Linear discriminant analysis
12	[62]	Discovery of factors influencing patent value based on machine learning in patents in the field of nanotechnology	Scientometrics	17	To discover the factors that influence patent value	132,670 patents	Feature selection, x-square method, gain ratio method, reliefF method	Probabilistic (Naive Bayes), logistic regression, decision trees (C4.5 and random forest)
13	[54]	Modelling patent legal value by Extension Neural Network	Expert Systems with Applications	17	To propose a valuation model for the monetary legal value of patents	163 patents	Factor analysis, principal component method, Kaiser normalised varimax rotation, multi-regression analysis	Extension neural network
14	[41]	The nonlinear nature of the relationships between the patent traits and corporate performance	Scientometrics	16	To explore the non-linear relationships between corporate performance and the patent trais measured from Herfindahl-Hirschman Index of Patents, patent citations, and relative patent position in the most important technological field in the US pharmaceutical industry	375 patents	Herfindahl-Hirschman Index of Patents, Relative Patent Position in the most important technological field	Back propagation neural network
15	[40]	Analyzing the nonlinear effects of firm size, profitability, and employee productivity on patent citations of the US pharmaceutical companies by using artificial neural network	Scientometrics	16	To explore the non-linear effect of firm size, profitability, and employee productivity on patent citations	430 patents	Descriptive statistics and correlation coefficients	Back propagation neural network, self organising maps (SOM)
16	[39]	Using neural network to analyse the influence of the patent performance upon the market value of the US pharmaceutical companies	Scientometrics	16	To analyse the influence of patent performance on market value	442 patents	Descriptive statistics and correlation coefficients	Back propagation neural network, self organising maps (SOM)
17	[50]	Topic based classification and pattern identification in patents	Technological Forecasting and Social Change	13	To automatically identify and classifiy patent datasets into technology areas	10,201 patents	Topic modelling, Linear discriminant analysis, quadratic discriminant analysis	Feed forward neural network, support verctor machine with a radial kernel
18	[49]	Ontology-based neural network for patent knowledge management in design collaboration	International Journal of Production Research	13	To develop a novel knoweldge management approach to automatically classify and search knowledge documents stored in patent corpuses	493 patents	TF-IDF, ontology schema	Back propagation neural network
19	[73]	Managing knowledge in light of its evolution process: An empirical study on citation network-based patent classification	Journal of Management Information Systems	11	To classify patents in a knowledge management system	18,271 patents	Kernel-based approaches	Support vector machine

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
20	[72]	Stochastic technology life cycle analysis using multiple patent indicators	Technological Forecasting and Social Change	10	To estimate the probability of a techology being at a certain stage of its life cycle	9488 patents	Patent indicator analysis, time-series analysis	Hidden Markov chain
21	[51]	Exploring the nonlinear effects of patent H index, patent citations, and essential technological strength on corporate performance by using artificial neural network	Journal of Informetrics	10	To explore the nonlinear relationships between patent performance and the corporate performance of the pharmaceutical companies	42 public firms with 679 firm-year observations	Descriptive statistics, ANOVA, Ordinary least squares regression analysis	Back propagation neural network
22	[64]	A predictive model of technology transfer using patent analysis	Sustainability (Switzerland)	9	To predict technology transfer using patent analysis	–	Text mining techniques	Social network analysis, Decision tree, Regression
23	[60]	A patent quality analysis and classification system using self-organising maps with support vector machine	Applied Soft Computing Journal	7	To automatically classify patents according to quality	17,971 patents	Self organising maps, Kernel principal component analysis	Support vector machine
24	[74]	Incorporating prior knowledge into learning by dividing training data	Frontiers of Computer Science in China	7	To propose a framework for incorporating prior knownedge in patent classification	8 sets of 1,700,000 patents	Task decomposition	Min-max (M3) modular network support vector machine
25	[75]	Concept-based patent image retrieval	World Patent Information	6	To automatically extract concept information describing the patent image content	300 patents	Adaptive hierarchical density histograms, text mining techniques	Supprot vector machine
26	[53]	Evaluating patents using damage awards of infringement lawsuits: A case study	Journal of Engineering and Technology Management	6	To propose a patent valuation model based on damage award for infrignement cases, for effective patent management	163 patents	Factor analysis, principal component method, Kaiser normalised varimax rotation, multi-regression analysis	Extension neural network
27	[82]	Identification and monitoring of possible disruptive technologies by patent-development paths and topic modelling	Technological Forecasting and Social Change	5	To identify technology paths of disruptive technologies	9328 patents	Patent-citation network	K-core and topic modelling
28	[87]	Patent analysis for technology development of artificial intelligence: A country-level comparative study	Innovation	5	To investigate the technological development of artificial intelligence	5228 patents	Technology indicators and citation indicators	Citation analysis, Clustering
29	[83]	Topic discovery and future trend forecasting for texts	Journal of Big Data	4	To discovery topis and forecast future trends from texts	6122 papers	Text mining techniques, association analysis, temporal correlation analysis	Ensemble forecasting approach
30	[58]	Chemical entity recognition in patents by combining dictionary-based and statistical approaches	Database: the journal of biological databases and curation	3	To develop a chemical entity recongition system	21000 patents	Dictionary-based approach, term inclusion and extraction, word2vec	K-means, tmChem conditional random fields

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
31	[70]	Chemical named entity recognition in patents by domain knowledge and unsupervised feature learning	Database: the journal of biological databases and curation	3	To develop a chemical entity mention recognition in patents and chemical passage detection system	21,000 patents	Domain-specific knowledge sources, Brown clusteringg	Conditional random fields, Semantic support vector machine
32	[69]	Modelling patent quality: A system for large-scale patentability analysis using text mining	Journal of Information Processing	3	To model patent quality using text mining	300,000 patents	TF-IDF, synactic complexity	Logistic regression
33	[52]	Non-linear multiclassifier model based on Artificial Intelligence to predict research and development performance in European countries	Technological Forecasting and Social Change	3	To predict R&D performance in European conntries	100 patents	k-means clustering	Evolutionary Sigmoidal Unit Neural Networks, Evolutionary Product Unit Neural Networks
34	[45]	Forecasting emerging technologies: A supervised learning approach through patent analysis	Technological Forecasting and Social Change	2	To forecast emerging technologies	Utility patents 1979–2010	Patent feature vectors, clustering	Naïve Bayes, Artificial Neural Networks (ANN), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), and Random Tree (RT)
35	[55]	Soft computing prediction of economic growth based in science and technology factors	Physica A: Statistical Mechanics and its Applications	2	To develop a forecasting model for GDP	28 country level patent data	–	Extreme learning machine
36	[65]	A Supervised Requirement-oriented Patent Classification Scheme Based on the Combination of Metadata and Citation Information	International Journal of Computational Intelligence Systems	2	To develop a patent classification scheme to classify patent datasets	14,414 patents	Requirement-oriented taxonomy, information gain, TF-IDF	Support vector machine, decision tree, Naives-Bayes
37	[42]	An intelligent system for automated binary knowledge document classification and content analysis	Jounral of Universal Computer Science	2	To develop an intelligent system for binary knowledge document classification and content analysis	170 patents	Hierarchical ontology technique, normalised term frequency	Back propagation neural network
38	[56]	Firms' knowledge profiles: Mapping patent data with unsupervised learning	Technological Forecasting and Social Change	1	To discuss the benefits and constranits of machine learning approaches in industry level patent analysis	160,000 patents	Text mining techniques	Latent dirichlet allocation
39	[88]	Hybrid corporate performance prediction model considering technical capability	Sustainability (Switzerland)	1	To predict corporate performance from technological capability	307,555 patents	–	Genetic algorithm support vector regression
40	[63]	Mining chemical patents with an ensemble of open systems	Database: the journal of biological databases and curation	1	To develop a system for named entity recognition of chemicals and genes/proteins in patents	21000 patents	Entity recongintion models	Ensemble classifier

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
41	[76]	Application research of robust LS-SVM regression model in forecasting patent application counts	Journal of Beijing Institute of Technology (English Edition)	1	To predict the number of patent applications	–	Cross validation	Support vector machine
42	[61]	Automatic discovery of concepts and actions	Expert Systems with Applications	0	To develop a fully autonomous artificial intelligence for new concept discovery	–	Concept formation algorithm, behavior policies learning algorithm	Inductive logic programming and reinforcement learning
43	[46]	Early identification of emerging technologies: A machine learning approach using multiple patent indicators	Technological Forecasting and Social Change	0	To develop an approach that identifies emerging technologies at early stages using multiple patent indications	35356 patents	Text mining techniques, patent indicator analysis	Back propagation neural network
44	[47]	Appraisal of Science and Economic Factors on Total Number of Granted Patents	Networks and Spatial Economics	0	To apply computational intelligence methodology for economic development estimation based on different science and technology factors	Total number of granted European patents	–	Neural network, fuzzy inference system
45	[43]	Economic development evaluation based on science and patents	Physica A: Statistical Mechanics and its Applications	0	To apply computational intelligence methodology, artificial neural network approach, for economic development estimation based on different science and technology factors	All European Union countries	–	Back propagation extreme learning machine
46	[80]	A visualization tool of patent topic evolution using a growing cell structure neural network	Scientometrics	0	To visualize technological evolution and development trend in a technological field	1215 patents	Text mining techniques, social network analysis	Growing cell structures neural network, paired with Girvan-Newman clustering algorithm
47	[68]	Visual patent trend analysis for informed decision making in technology management	World Patent Information	0	To provide decision support in technology management through visual patent trend analysis	2460 patents	Concept extraction, natural language processing	Regression
48	[57]	Deep learning-based corporate performance prediction model considering technical capability	Sustainability (Switzerland)	0	To predict the future performance of companies for the purpose of making investment decisions	59,740 patents	Restricted Boltzmann machines, Back propagation	Deep belief neural network
49	[85]	Prediction of economic growth by extreme learning approach based on science and technology transfer	Quality and Quantity	0	To analyse the influence of number of granted European patents on the economic growth by field of technology	28 countries in the European Union	–	Extreme learning approach
50	[44]	Forecasting and identifying multi-technology convergence based on patent data: the case of IT and BT industries in 2020	Scientometrics	0	To forecast multi-technology convergence	387703 patents and 353785 citations	Patent citation analysis, dependency structure matrix	Neural network
51	[48]	Detecting figures and part labels in patents: competition-based development of graphics recognition algorithms	International Journal on Document Analysis and Recognition	0	To detect figures in patents	–	Text detection, optical charater recognition	Graphic recognition

(*continued on next page*)

Table 6 (*continued*)

No.	Author	Title	Journal	Cited	Aim	Data	Pre-processing method	Main analytical method
52	[81]	Time series clustering model based on complexity for apple technology forecasting	International Journal of Applied Engineering Research	0	To propose a technology forecasting of Apple according to time trend of each technology	All Apple's patent applications	–	Time series clustering
53	[84]	Detection of technology opportunities from patents	International Journal of Applied Engineering Research	0	To detect and provide opportunities for the new technologies	All published patents in the last 20 years	Similarity-based named entity recognition, pattern-based relation extraction	Machine learning-based filtering
54	[77]	Constructing an intelligent patent network analysis method	Data Science Journal	0	To construct an intelligent patent network analysis method	–	Enhanced term frequency- inverse document frequency, patent network analysis, ontology	Association algorithm
55	[59]	IP portfolios and evolution of biomedical additive manufacturing applications	Scientometrics	0	To develop an Intellectual Property (IP) analytical methodology to explore the portfolios and evolution of patents in bio-Additive Manufacturing domain	58 patents	Key term extraction, NTF-IDF, Similarity matrix	K-Means, K-Medoids, Partitioning Around Medoids (PAM), Concept Lattice algorithm
56	[71]	Immersive Technology for Human-Centric Cyberphysical Systems in Complex Manufacturing Processes: A Comprehensive Overview of the Global Patent Profile Using Collective Intelligence	Complexity	0	To provide a thorough review literature, develop a domain ontology, and highlight technical and functional development trends of Industry 4.0	2672 patents	Technology Function Matrix (TFM) based on NTF	Latent Dirichlet allocation (LDA), Topic Modelling
57	[86]	Exploring 4G patent and litigation informatics in the mobile telecommunications industry	World Patent Information	0	To develop a computer-supported generic methodology for discovering evolutions and linkages between litigations and disputed patents	16 patents and 28 litigation cases	Key term frequency	Modified formal concept analysis, hierarchical concept lattice

References

[1] OECD, Enabling the Next Production Revolution: the Future of Manufacturing and Services - Interim Report. Technical Report June, OECD, 2016.
[2] EPO, India and Europe Explore the Impact of Industry 4.0 on the Patent System. Technical Report, European Patent Office, Munich, Germany, 2016.
[3] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): a vision, architectural elements, and future directions, Future Generat. Comput. Syst. 29 (7) (2013) 1645–1660.
[4] EPSRC, Delivery Plan 2016-2020 Top Ten Messages. Technical Report, (2016).
[5] W.A. Gunther, M.H. Rezazade Mehrizi, M. Huysman, F. Feldberg, Debating big data: a literature review on realizing value from big data, J. Strat. Inf. Syst. 26 (3) (2017) 191–209.
[6] A. Abbas, L. Zhang, S.U. Khan, A literature review on the state-of-the-art in patent analysis, World Patent Inf. 37 (2014) 3–13.
[7] D. Baglieri, F. Cesaroni, Capturing the real value of patent analysis for R&D strategies, Technol. Anal. Strat. Manag. 25 (8) (2013) 971–986.
[8] M.G. Moehrle, L. Walter, I. Bergmann, S. Bobe, S. Skrzipale, Patinformatics as a business process: a guideline through patent research tasks and tools, World Patent Inf. 32 (4) (2010) 291–299.
[9] A.J. Trippe, Patinformatics: tasks to tools, World Patent Inf. 25 (3) (2003) 211–221.
[10] J.P. Dintzner, J. Van Thieleny, Image handling at the european patent office: BACON and first page, World Patent Inf. 13 (3) (1991) 152–154.
[11] L. Aristodemou, F. Tietze, N. Athanassopoulou, T. Minshall, Exploring the future of patent analytics: a technology roadmapping approach, R&D Management Conference 2017, Leuven, Belgium, 2017, pp. 1–9.
[12] M. Lupu, Information retrieval, machine learning, and Natural Language Processing for intellectual property information, World Patent Inf. 49 (2017) A1–A3.
[13] G.R. Oldham, Y. Fried, Job design research and theory: past, present and future, Organ. Behav. Hum. Decis. Process. 136 (2016) 20–35.
[14] A. Trippe, Guidelines for Preparing Patent Landscape Reports. Technical Report, World Intellectual Property Organisation, 2015.
[15] R. Phaal, M. Routley, N. Athanassopoulou, D. Probert, Charting exploitation strategies for emerging technology, Res. Technol. Manag. 55 (2) (2012) 34–42.
[16] L. Aristodemou, F. Tietze, Exploring the Future of Patent Analytics. Technical Report, Institute for Manufacturing, University of Cambridge, Cambridge, UK, 2017.
[17] L. Aristodemou, F. Tietze, Exploring the Future of Patent Analytics: a Technology Roadmapping Approach, Centre for Technology Management working paper series, 2017 November(5).
[18] L. Aristodemou, F. Tietze, A Literature Review on the State-of-the-art on Intellectual Property Analytics (IPA), Centre for Technology Management working paper series, 2017November(2).
[19] A. Agrawal, J. Gans, A. Goldfarb, How AI will change the way we make decisions, Harv. Bus. Rev. (2017) July:1–7 https://hbr.org/2017/07/how-ai-will-change-the-way-we-make-decisions.
[20] A. Ciccatelli, The Future of Big Data and Intellectual Property, (2017).
[21] T. Stading, The Role of Artificial Intelligence in Intellectual Property, (2017).
[22] T. Stading, Using Big Data to Make Intellectual Property a Strategic Weapon, (2017).
[23] M.K. Raturi, P.K. Sahoo, S. Mukherjee, A.K. Tiwari, Patinformatics – an Emerging Scientific Discipline, (2010).
[24] D. Bonino, A. Ciaramella, F. Corno, Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics, World Patent Inf. 32 (1) (2010) 30–38.
[25] P. Cronin, F. Ryan, M. Coughlan, Undertaking a Literature Review: a Step-by-step Approach vol 17, (2008), pp. 38–43 1.
[26] G. Pare, M.C. Trudel, M. Jaana, S. Kitsiou, Synthesizing information systems knowledge: a typology of literature reviews, Inf. Manag. 52 (2) (2015) 183–199.
[27] J.W. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, fourth ed. edition, (2013).
[28] C. Robson, Real World Research, Edition, Blackwell Publishing, Malden, 2011, pp. 1–608.
[29] R. Klinger, C. Kolářik, J. Fluck, M. Hofmann-Apitius, C. Friedrich, Detection of iupac and iupac-like chemical names, Bioinformatics 24 (13) (2008) i268–i276.
[30] A. Trappey, F.-C. Hsu, C. Trappey, C.-I. Lin, Development of a patent document classification and search platform using a back-propagation network, Expert Syst. Appl. 31 (4) (2006) 755–765.
[31] A. Trappey, C. Trappey, C.-Y. Wu, C.-W. Lin, A patent quality analysis for innovative technology and product development, Adv. Eng. Inf. 26 (1) (2012) 26–34.
[32] D. Thorleuchter, D. den Poel, A. Prinzie, A compared r&d-based and patent-based cross impact analysis for identifying relationships between technologies, Technol. Forecast. Soc. Change 77 (7) (2010) 1037–1050.
[33] S. Jun, S.-S. Park, D.-S. Jang, Document clustering method using dimension reduction and support vector clustering to overcome sparseness, Expert Syst. Appl. 41 (7) (2014) 3204–3212.
[34] M. Krallinger, O. Rabal, F. Leitner, et al., The CHEMDNER corpus of chemicals and drugs and its annotation principles, J. Cheminf. 7 (Suppl 1) (2015) S2, https://doi.org/10.1186/1758-2946-7-S1-S2.
[35] S. Jun, S. Park, Examining technological innovation of apple using patent analysis, Ind. Manag. Data Syst. 113 (6) (2013) 890–907.
[36] Y. Chen, K. Chang, Exploring the nonlinear effects of patent citations, patent share and relative patent position on market value in the us pharmaceutical industry, Technol. Anal. Strat. Manag. 22 (2) (2010) 153–169.
[37] S. Jun, S.-J. Lee, Emerging technology forecasting using new patent information analysis, Int. J. Software Eng..Appl 6 (3) (2012) 107–116.
[38] X. Zhang, Interactive patent classification based on multi-classifier fusion and active learning, Neurocomputing 127 (2014) 200–205.
[39] Y.-S. Chen, K.-C. Chang, Using neural network to analyze the influence of the patent performance upon the market value of the us pharmaceutical companies, Scientometrics 80 (3) (2009) 637–655.
[40] Y.-S. Chen, K.-C. Chang, Analyzing the nonlinear effects of firm size, profitability, and employee productivity on patent citations of the us pharmaceutical companies by using artificial neural network, Scientometrics 82 (1) (2010) 75–82.
[41] Y.-S. Chen, K.-C. Chang, The nonlinear nature of the relationships between the patent traits and corporate performance, Scientometrics 82 (1) (2010) 201–210.
[42] T.-A. Chiang, C.-Y. Wu, C. Trappey, A. Trappey, An intelligent system for automated binary knowledge document classification and content analysis, J. Univers. Comput. Sci. 17 (14) (2011) 1991–2008.
[43] B. Jokanovic, B. Lalic, M. Milovancevic, N. Simeunovic, D. Markovic, Economic development evaluation based on science and patents, Phys. Stat. Mech. Appl. 481 (2017) 141–145.
[44] J. Kim, S. Lee, Forecasting and identifying multi-technology convergence based on patent data: the case of it and bt industries in 2020, Scientometrics 111 (1) (2017) 47–65.
[45] M. Kyebambe, G. Cheng, Y. Huang, C. He, Z. Zhang, Forecasting emerging technologies: a supervised learning approach through patent analysis, Technol. Forecast. Soc. Change 125 (2017) 236–244.
[46] C. Lee, O. Kwon, M. Kim, D. Kwon, Early identification of emerging technologies: a machine learning approach using multiple patent indicators, Technol. Forecast. Soc. Change 127 (2018) 291–303.
[47] D. Markovic, Appraisal of science and economic factors on total number of granted patents, Network. Spatial Econ. (2017) 1–8.
[48] C. Riedl, R. Zanibbi, M. Hearst, S. Zhu, M. Menietti, J. Crusan, I. Metelsky, K. Lakhani, Detecting figures and part labels in patents: competition-based development of graphics recognition algorithms, Int. J. Doc. Anal. Recogn. 19 (2) (2016) 155–172.
[49] A. Trappey, C. Trappey, T.-A. Chiang, Y.-H. Huang, Ontology-based neural network for patent knowledge management in design collaboration, Int. J. Prod. Res. 51 (7) (2013) 1992–2005.
[50] S. Venugopalan, V. Rai, Topic based classification and pattern identification in patents, Technol. Forecast. Soc. Change 94 (2015) 236–250.
[51] S. Zhang, C.-C. Yuan, K.-C. Chang, Y. Ken, Exploring the nonlinear effects of patent h index, patent citations, and essential technological strength on corporate performance by using artificial neural network, J.Informetrics 6 (4) (2012) 485–495.
[52] M. de la Paz-Marín, P. Campoy-Muñoz, C. Hervás-Martínez, Non-linear multi-classifier model based on artificial intelligence to predict research and development performance in european countries, Technol. Forecast. Soc. Change 79 (9) (2012) 1731–1745.
[53] Y.-H. Lai, H.-C. Che, Evaluating patents using damage awards of infringement lawsuits: a case study, J. Eng. Technol. Manag.- JET-M 26 (3) (2009) 167–180.
[54] Y.-H. Lai, H.-C. Che, Modeling patent legal value by extension neural network, Expert Syst. Appl. 36 (7) (2009) 10520–10528.
[55] D. Markovic, D. Petkovic, V. Nikolic, M. Milovancevic, B. Petkovic, Soft computing prediction of economic growth based in science and technology factors, Phys. Stat. Mech. Appl. 465 (2017) 217–220.
[56] A. Suominen, H. Toivanen, M. Seppänen, Firms' knowledge profiles: mapping patent data with unsupervised learning, Technol. Forecast. Soc. Change 115 (2017) 131–142.
[57] J. Lee, D. Jang, S. Park, Deep learning-based corporate performance prediction model considering technical capability, Sustainability 9 (6) (2017).
[58] S. Akhondi, E. Pons, Z. Afzal, H. van Haagen, B. Becker, K. Hettne, E. van Mulligen, J. Kors, Chemical entity recognition in patents by combining dictionary-based and statistical approaches, Database: The Journal of Biological Databases and Curation 2016 (2016).
[59] A. Trappey, C. Trappey, C. Chung, Ip portfolios and evolution of biomedical additive manufacturing applications, Scientometrics 111 (1) (2017) 139–157.
[60] J.-L. Wu, P.-C. Chang, C.-C. Tsao, C.-Y. Fan, A patent quality analysis and classification system using self-organizing maps with support vector machine, Appl. Soft Comput.J 41 (2016) 305–316.
[61] A. Tenorio-González, E. Morales, Automatic discovery of concepts and actions, Expert Syst. Appl. 92 (2018) 192–205.
[62] S. Bass, L. Kurgan, Discovery of factors influencing patent value based on machine learning in patents in the field of nanotechnology, Scientometrics 82 (2) (2010) 217–241.
[63] R. Leaman, C.-H. Wei, C. Zou, Z. Lu, Mining chemical patents with an ensemble of open systems, Database: The Journal of Biological Databases and Curation 2016 (2016).
[64] J. Choi, D. Jang, S. Jun, S. Park, A predictive model of technology transfer using patent analysis, Sustainability 7 (12) (2015) 16175–16195.
[65] F. Zhu, X. Wang, D. Zhu, Y. Liu, A supervised requirement-oriented patent classification scheme based on the combination of metadata and citation information, Int. J. Comput. Intell. Syst. 8 (3) (2015) 502–516.
[66] J. Callaert, J. Grouwels, B. van Looy, Delineating the scientific footprint in technology: identifying scientific publications within non-patent references, Scientometrics 91 (2) (2012) 383–398.
[67] J.-C. Lamirel, S. Al Shehabi, M. Hoffmann, C. François, Intelligent patent analysis through the use of a neural network, Proceedings of the ACL-2003 Workshop on Patent Corpus Processing, vol 20, 2003, pp. 7–23.
[68] Q. Han, F. Heimerl, J. Codina-Filba, S. Lohmann, L. Wanner, T. Ertl, Visual patent trend analysis for informed decision making in technology management, World

Patent Inf. 49 (2017) 34–42.
[69] S. Hido, S. Suzuki, R. Nishiyama, T. Imamichi, R. Takahashi, T. Nasukawa, T. Idé, Y. Kanehira, R. Yohda, T. Ueno, A. Tajima, T. Watanabe, Modeling patent quality: a system for large-scale patentability analysis using text mining, J. Inf. Process. 20 (3) (2012) 655–666.
[70] Y. Zhang, J. Xu, H. Chen, J. Wang, Y. Wu, M. Prakasam, H. Xu, Chemical named entity recognition in patents by domain knowledge and unsupervised feature learning, Database 2016 (2016).
[71] U. Govindarajan, A. Trappey, C. Trappey, Immersive technology for human-centric cyberphysical systems in complex manufacturing processes: a comprehensive overview of the global patent profile using collective intelligence, Complexity 2018 (2018).
[72] C. Lee, J. Kim, O. Kwon, H.-G. Woo, Stochastic technology life cycle analysis using multiple patent indicators, Technol. Forecast. Soc. Change 106 (2016) 53–64.
[73] X. Li, H. Chen, Z. Zhang, J. Li, J. Nunamaker, Managing knowledge in light of its evolution process: an empirical study on citation network-based patent classification, J. Manag. Inf. Syst. 26 (1) (2009) 129–153.
[74] B. Lu, X. Wang, M. Utiyama, Incorporating prior knowledge into learning by dividing training data, Front. Comput. Sci. China 3 (1) (2009) 109–122.
[75] S. Vrochidis, A. Moumtzidou, I. Kompatsiaris, Concept-based patent image retrieval, World Patent Inf. 34 (4) (2012) 292–303.
[76] L.-W. Zhang, Q. Zhang, X.-F. Wang, D.-H. Zhu, Application research of robust ls-svm regression model in forecasting patent application counts, J. Beijing Inst. Technol. (Soc. Sci. Ed.) 18 (4) (2009) 497–501.
[77] C.-C. Wu, C.-B. Yao, Constructing an intelligent patent network analysis method, Data Sci. J. 11 (2012) 110–125.
[78] M. Alavi, D.E. Leidner, Review: knowledge management and knowledge management systems: conceptual foundations and research issues, MIS Q. 25 (1) (2001) 107–136.
[79] M. Gregory, Technology management: a process approach, Proc. IME B J. Eng. Manufact. 209 (5) (1995) 347–356.
[80] H.-Y. Sung, H.-Y. Yeh, J.-K. Lin, S.-H. Chen, A visualization tool of patent topic evolution using a growing cell structure neural network, Scientometrics 111 (3) (2017) 1267–1285.
[81] S. Jun, Time series clustering model based on complexity for apple technology forecasting, Int. J. Appl. Eng. Res. 11 (20) (2016) 10343–10347.
[82] A. Momeni, K. Rost, Identification and monitoring of possible disruptive technologies by patent-development paths and topic modeling, Technol. Forecast. Soc. Change 104 (2016) 16–29.
[83] J. Hurtado, A. Agarwal, X. Zhu, Topic discovery and future trend forecasting for texts, J.Big Data 3 (1) (2016).
[84] H.-W. Chun, J.-M. Lee, W. Yeo, S. Kim, H.-S. Yoon, I. Song, S.-W. Hong, B.-Y. Coh, Detection of technology opportunities from patents, Int. J. Appl. Eng. Res. 9 (21) (2014) 8731–8736.
[85] P. Karanikic, I. Mladenovic, S. Sokolov-Mladenovic, M. Alizamir, Prediction of economic growth by extreme learning approach based on science and technology transfer, Qual. Quantity 51 (3) (2017) 1395–1401.
[86] C. Trappey, A. Trappey, Exploring 4g patent and litigation informatics in the mobile telecommunications industry, World Patent Inf. 50 (2017) 38–51.
[87] C.-Y. Tseng, P.-H. Ting, Patent analysis for technology development of artificial intelligence: a country-level comparative study, Innovat. Manag. Pol. Pract. 15 (4) (2013) 463–475.
[88] J. Lee, G. Kim, S. Park, D. Jang, Hybrid corporate performance prediction model considering technical capability, Sustainability 8 (7) (2016).