Insights from a text mining survey on Expert Systems research from 2000 to 2016

Paulo Cortez [1*]
Sérgio Moro [2,1]
Paulo Rita [3,4]
David King [5]
Jon Hall [5]

Abstract

This study presents a literature analysis using a semi-automated text mining and topic modelling approach of the body of knowledge encompassed in seventeen years (2000-2016) of literature published in the Wiley's Expert Systems journal, a key reference in Expert Systems (ESs) research, in a total of 488 research articles.

The methodological approach included analysing countries from authors' affiliations, with results emphasising the relevance of both US and UK researchers, with Chinese, Turkish and Spanish holding also a significant relevance. As a result of the sparsity found on the keywords, one of our goals became to devise a taxonomy for future submissions under two core dimensions: ESs´ methods and ESs´ applications. Finally, through topic modelling, data-driven methods were unveiled as the most relevant, pairing with evaluation methods in its application to managerial sciences, arts and humanities. Findings also show that most of the application domains are well represented, including health, engineering, energy, and social sciences.

Conflicts of interest: none

Keywords

Expert systems; literature analysis; research categorization; research evolution; text mining.

[1] ALGORITMI Research Centre, University of Minho, Guimarães, Portugal
[*] Corresponding author. Email: pcortez@dsi.uminho.pt
[2] Instituto Universitário de Lisboa (ISCTE-IUL), ISTAR-IUL, Lisboa, Portugal
[3] Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, Lisboa, Portugal
[4] NOVA Information Management School (NOVA IMS), Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
[5] The Open University, United Kingdom

1. Introduction

Expert Systems (ESs) have been at the centre of decision support for managerial decision making. In the seventies and eighties, ESs were focused on mimicking human experts and separating explicit knowledge (stored in a knowledge base) from the Artificial Intelligence inference machine (Buchanan,1986). Yet, since then, and in particular after the 2000s, much has changed due to the explosion of data, evolution of the Internet (e.g., Internet of Things), mobile and social stances. Thus, there has been a pressure to extract as useful knowledge from past data and incorporate such knowledge in ESs, leading to an increase in data related fields, such as Business Intelligence, Data Mining, Big Data and Data Science (Cortez and Santos, 2015).

Due to such ESs related research and technological evolution, it is relevant to perform a review of what has been recently published in the ESs domain. As such, this paper focuses on analysing literature published in Wiley's Expert Systems journal (EXSY), emphasising its relevance and evolution through a recent timeline of 17 years, from 2000 to 2016. This journal, established more than 30 years ago, has always been at the forefront of investigation on expert systems, thus constituting one of the main sources for research in this major area. Single-source literature analyses provide an historical picture of the main topics addressed by that source, helping to guide the board of editors' future strategies while at the same time providing a thorough perspective over the addressed topics (e.g., Moro et al., 2017). With this in mind, we set out to perform this task by analysing all EXSY research articles published within the 17-year timeline, in a total of 488 articles. Given the sparsity of keywords used by authors to classify articles, we aimed to develop a taxonomy aggregating the main methods and applications of EXSY research. Finally, due to the large volume of research, we conducted a semi-automated literature analysis using text mining (Moro et al., 2015) to assess the main research trends. In particular, the text analysis includes an assessment of authors´ affiliations and the analysis of research topics based on both ESs' methods and ESs' application areas.

2. Background

Expert systems emerged in the 1970s to emulate human expert behaviour in decision making process, in an attempt to benefit from artificial intelligence in performing computational reasoning tasks for solving complex real-world problems (Jackson, 1986). These systems typically used a knowledge-based architecture to infer decisions grounded on a knowledge base of established facts relevant to the problem being addressed (Reid, 1985; Buchanan,1986).

The relevance of ESs is widely recognized by both scholars and practitioners, with serious managerial implications for organizational efficiency (Yoon et al., 1995). The scientific literature has kept pace with this innovative subject, with a large number of sources publishing research during the past three decades. Specialized journals devoted to ESs have emerged, namely Expert Systems: The Journal of Knowledge Engineering (referred here as EXSY) in 1984, published by Wiley-Blackwell, and Expert Systems with Applications in 1990, published by Elsevier. Those journals have further contributed to the dissemination of applied research on ESs. The late 1990s and early 2000s saw the worldwide spread of ESs to become a pivotal domain which benefitted from an overflow of information, making them intelligent systems with data mining capabilities to handle such volumes of data (Shim et al., 2002). The Internet revolution with the rise of Web 2.0 and social media has led to a data explosion as any Internet user is now a data producer (Moro et al., 2016). This Big Data tendency is showing no signs of slowing (Cortez & Santos, 2015). More recently, the new paradigm of the Internet of Things implies that virtually all devices can be Internet data generators, with devices working together for increased efficacy (Costa et al., 2017). ESs have evolved to keep pace with the rise of Big Data, in order to provide better assistance in the decision making process (Abbasi et al., 2016). As such, ESs have embedded distinct Artificial Intelligence approaches, such as control, data-driven, monitoring, and knowledge representation methods, among others.

ESs can be virtually applied to any domain and science, including, for example, agricultural and earth sciences: Tocatlidou et al. (2002) devised a reasoning system for identifying plant disease, while Kutbay and Hardalaç (2017) developed a tomography system. One of the most prolific and studied domains for ESs´ application is business and management. This includes a wide range of managerial sciences such as marketing (Moro et al., 2014), economics and finance (Zhang et al., 2016), planning (Surma, 2015), and auditing (Gray et al., 2014). Another domain that has recently been subjected to highly innovative research practices is health, with distinct examples of applications such as nutrition (Espín et al., 2016), patient care (James et al., 2017), depression (Chattopadhyay, 2014), and cancer (Acharya et al., 2014). Psychology and social sciences have also benefited from ES´s implementations (e.g., Jiménez et al., 2016). Such diversity shows the broad applicability of ESs, providing solid grounded evidence of its relevance.

The length of time between ESs first emergence and their dissemination to the vast array of applications justifies periodical literature analyses to assess the current state-of-the-art. The study by Liao (2004) analysed research published in the 1995-2004 timeframe to assess ES´s development under two categories: methodologies and applications. Their findings suggested that social sciences could benefit from ES´s implementations. This is certainly a path followed subsequently by researchers (e.g., Mumpower et al., 2012). It is also interesting to note that, back then, ESs were already being implemented in a large number of categories, including biology, health, and managerial sciences. More recently, Sahin et al. (2013) presented an EXSY editorial of ESs´ literature published in 170 journals from 1989 to 2012. Their study introduced a manual categorization of the collected literature, relating the methodological approach to the country of the first author's affiliation. However, the study did not evaluate the relationship between applications and methods. This is a gap the present research attempts to address by using a semi-automated text mining and topic modelling approach to specifically summarize 17 years of ESs´ literature in the EXSY journal, which is specifically devoted to ESs´ research.

3. Materials and methods

This study sought to analyse the body of knowledge of seventeen years of EXSY literature, encompassing a total of 646 publications, from which 50 are editorials, 488 are full research articles, and the remaining 108 include other types of publications, such as information items and acknowledgement notes. Only the research articles were analysed, as those are the ones that communicate innovative research on ESs.

Figure 1 shows the semi-automated approach followed to cover such a large volume of literature. Initially, the articles were converted into text format to ease textual content extraction and analysis. From the dataset of articles, an approach comprising three stages of analysis emerged. First, country affiliations were assessed through text mining to reveal the worldwide nature of ESs´ literature. This task was accomplished by obtaining a document-term matrix showing the intersection of authors' nationalities per article. Authors from the same country were counted only once, thus avoiding to overweight articles with several authors from the same country. The list of countries used was compiled in July 2014 by Mottershead (2014) and includes 193 United Nations member states.

In the second stage, a categorization task took place to reveal the aggregate domains of ESs. Given the subjectivity inherent in such task, all article keywords were gathered to ensure independent article classification, for a total of 1,597 different keywords. We note that this keyword list is very sparse: only 13 terms appeared more than 10 times (e.g. “expert systems”, “neural networks”) and most of the terms (1,395 keywords) only occurred once. Next, a manual analysis was completed by the authors of this article to categorize each keyword as belonging to an ES´s specific method (e.g., “backpropagation neural network” was categorized into the “neural networks” method) or to an application of ESs (e.g., “direct mailing” into “marketing”). Such manual labelling was conducted iteratively, during several rounds and also guided by a sorting of the keywords, which allowed an easier detection of terminology overlaps (e.g., “feedforward neural network” and “feed-forward neural network”). After preprocessing the keywords, a total of 105 methods and 59 applications were identified. These were still considered too specific to obtain a manageable picture of the literature. Thus, a taxonomy[6] of categories was developed (see Section 4.2) in an iterative procedure to further aggregate categories. At the highest level, a tuned list of eleven methods and fourteen applications was achieved. These two high level categories were compared to the respective methods and application categories used by Sahin et al. (2013) and also with two widely known library classification systems: the Dewey Decimal Classification and the Library of Congress Classification.

Finally, the last stage comprised topic modelling using, as an input, the dictionary of both methods and application categories obtained in the previous stage. The latent Dirichlet allocation (LDA) algorithm was adopted, as it is the most popular and widely used topic modelling technique (Calheiros et al., 2017). LDA enables to group text documents by classifying them using computed measures that represent the distance of each document to a given topic and from each document to each of the terms, thus providing a characterisation of the topics by the terms that are more closely related to the topic. This enabled the discovery of topics and highlighted the main findings in a procedure similar to the one followed by Moro et al. (2015) for the

[6] https://fenix.iscte-iul.pt/homepage/smcmo@iscte.pt/expert-systems-taxonomy

business intelligence in banking literature, and by Moro and Rita (2018) for branding strategies in social media in the hospitality and tourism literature.

All the experiments were conducted using the R statistical tool, which is open source and offers a simple scripting language designed for data analysis with a wide number of packages developed by an enthusiastic community (Cortez, 2014). Specifically, the “tm” and “topicmodels” packages were chosen for the text mining and topic modelling tasks.

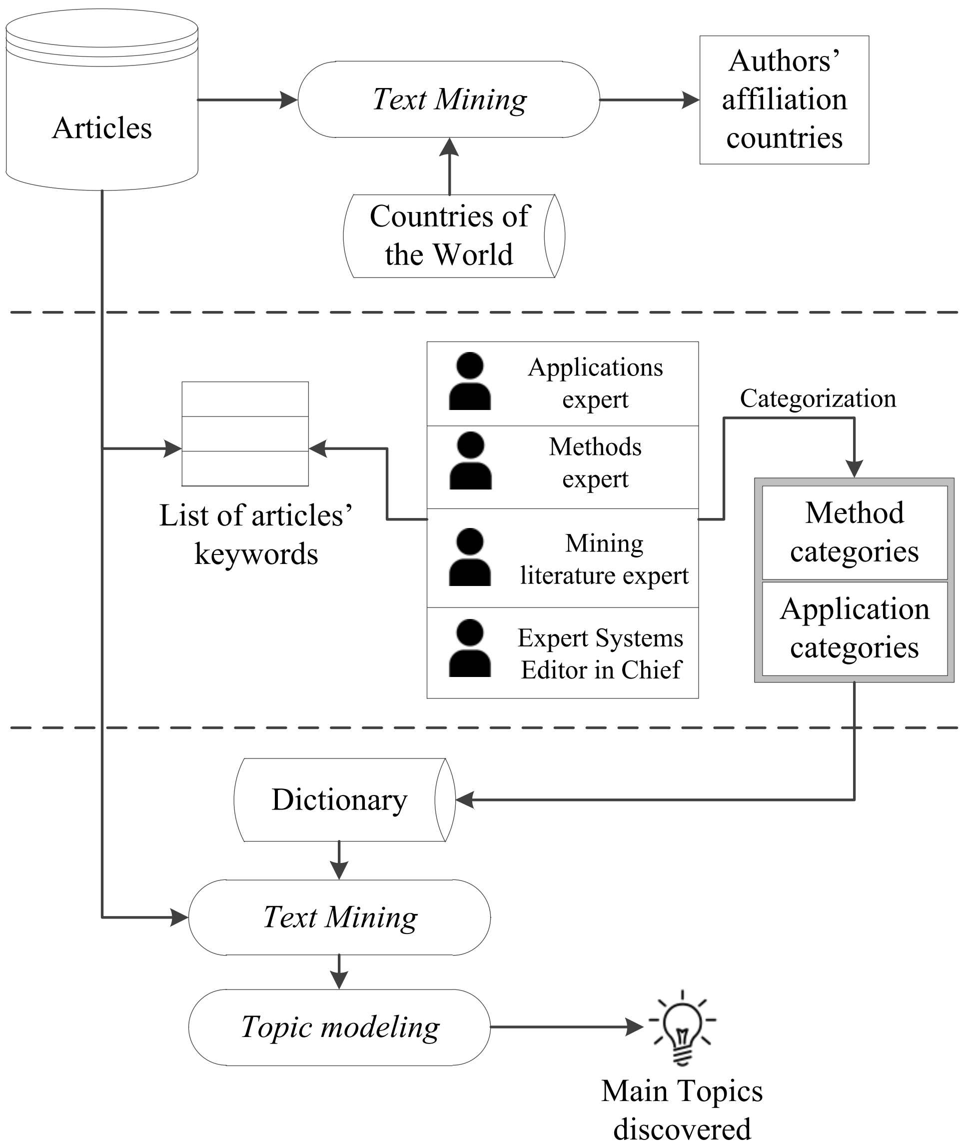

Figure 1 - Experimental procedure.

4. Experiments and results

4.1. Authors country's affiliations

Articles published in EXSY (Table 1) are predominantly from authors affiliated in US universities (79), followed by British (55) and Chinese (54) higher education institutions, and then by Turkish (45) and Spanish ones (42). Twelve other countries are also represented by ten or more papers in this journal.

Table 1 - Authors' affiliation country.

Position	Country	Frequency	Scimago Country rankings			
			Computer Science		Engineering	Mathematics
			Artificial Intelligence	Computational Theory and Mathematics	Control and Systems Engineering	Theoretical Computer Science
1	USA	79	2	2	2	1
2	UK	55	4	3	6	4
3	China	54	1	1	1	2
4	Turkey	45	21	29	24	34
5	Spain	42	8	11	12	8
6	Taiwan	29	11	10	11	14
7	Korea	22	12	13	8	10
8	Iran	21	16	19	17	33
9	Italy	18	9	9	9	7
10	Greece	17	20	28	30	26
11	Poland	16	17	14	18	15
12	Hong Kong	15	19	18	21	25
13	India	14	7	8	7	13
14	Portugal	14	24	30	25	24
15	Australia	13	13	12	13	11
16	Canada	12	10	7	10	9
17	France	10	6	5	4	5

Out of the seventeen countries shown in Table 1, eight are European and six are from Asia. Those remaining are from North America (2) and Oceania (1), with none from Latin American or Africa. Cumulatively, Europe accounts for 217 papers whereas Asia has 155 articles. North America reaches 91 and Oceania 13 publications.

Another interesting perspective is that countries from the so-called Anglo-Saxon world (specifically US, UK, Australia, Canada) account for 159 papers with most contributions coming from US and UK. Moreover, four other large countries (in particular, China, Turkey, Iran and India) associated with emerging economies have together 134 papers.

It is also worth noting the fact that relatively small countries, such as Greece and Portugal, are disproportionally represented and lead, even if slightly, much larger countries as is the case of Australia, Canada and France.

For comparison purposes, the Scimago country rankings published by Scopus for the four categories in which the ES journal is indexed are also shown on Table 1 (the last four columns). The presented results show some misalignment with authors publishing in EXSY journal. Most notably, Turkish authors are publishing significantly more in EXSY when compared to the general Scimago categories. Iranian, Greek and Portuguese authors also seem to publish proportionally more in EXSY. These observations raise an interesting hypothesis related to possible academic networks that may be disseminating EXSY as a reference journal in some countries. Further studies specifically addressing authors' motivations are required to answer such question.

4.2. Categorization

As explained in Section 3, we adopt two categorization dimensions in this work, focused on ES´s methods and ES´s applications based on the surveyed ES´s article keywords. The respective proposed categories include eleven main ES´s methods (Table 2) and fourteen ES´s application domains (Table 3). The high sparsity found on EXSY's articles keywords justified the purpose of building a taxonomy based on past EXSY publications that can be useful for future submissions. As such, we compared the proposed taxonomy[6] with others, including not only a specialized ES taxonomy (Sahin et al., 2013) but also the two general library classification systems: the Dewey Decimal Classifications (DDC) (Scott, 1998), and the Library of Congress Classifications (LCC)

(Chan, 1990). Particularly, the comparison with the two widely used library systems is quite interesting when analysing the ES' applications dimension, showing a high alignment, thus backing our taxonomy in this dimension. We further note that, following our research, the EXSY journal recently (from 2018) adopted our taxonomy system. The eleven ES´s method categories (Table 2) cover distinct aspects of ES´s systems, namely: classical ES´s problem solving approaches – {"control", "decision support", "knowledge representation", "monitoring and optimization"} (Russel & Norvig, 1995); ES technical implementation issues – {"distribution and infrastructure", "software"}; mathematical methods – {"mathematics"}; and ES´s methodological aspects – {"research methodology", "evaluation"}. Each main category will be automatically assigned by a text mining procedure to any ESs´ paper that includes any of its associated lower granularity terms (set using the articles keywords, see Section 3). The full list of ES´s method and application terms is made available at: https://fenix.iscte-iul.pt/homepage/smcmo@iscte.pt/expert-systems-taxonomy. As an illustrative example, the full list of terms for the "control" category is: {"control", "control system", "linear control", "smith predictor", "multi-tier control", "non-linear control", "self-tuning scaling factors", "state space model"}.

Table 2 – List of main ES method categories.

ES´s methods	**(Sahin et al., 2013)**	**Dewey Decimal Classification**	**Library of Congress Classification**
control			
data-driven	Artificial Intelligence – Fuzzy Expert Systems (AI-F-ES); Artificial Neural Network – Expert Systems (ANN-ES); Fuzzy Expert Systems (F-ES); Neural Expert Systems (N-ES); Neuro -Fuzzy Expert Systems (NF-ES); Rough Set-based Expert Systems (RS-ES); Web-based Expert Systems (WB-ES)	004 Data processing & computer science; 005 Computer programming, programs & data	QA (sub-range)
decision support	Artificial Intelligence –Decision Support Systems (AI-DSS-ES); Decision Support Systems (DSS-ES); Fuzzy – Decision Support Expert Systems (F-DSS-ES); Multi-Criteria Decision Making Expert Systems (MCDM-ES)		
distribution and infrastructure	Multi-agent Expert Systems (MA-ES)	003 Systems	

evaluation			
knowledge representation	Classical Expert Systems (CES)	000 Computer science, knowledge and general works	
mathematics		510 Mathematics	QA - Mathematics
monitoring			
optimization		510 Mathematics	QA - Mathematics
research methodology		607 Research	T175-178
software		005 Computer programming, programs & data	QA (sub-range)

As shown in Table 2, none of the comparison classification systems covers all proposed eleven categories, in particular, the terms “control”, “evaluation” and “monitoring” are unique to our proposal. The ESs´ method classification of Sahin et al. (2013) is often too specific and in some cases, there seems to be an overlap of categories. For instance, Sahin et al. (2013) define three distinct neural network and four fuzzy methods but no control or optimization method. Moreover, Sahin et al. (2013) does not explain the distinction between some methods, such as Artificial Intelligence – Fuzzy Expert Systems (AI-F-ES) and Fuzzy Expert Systems (F-ES). Regarding the comparison with DDC and LCC, both classification systems are too high-level and thus are not that useful for classifying ES´s methods. For instance, both our data-driven and software categories are mapped by more genetic categories, namely 005 Computer programming, programs & data for DCC and a sub-range of the QA Mathematics LCC system.

Turning to ES´s applications, we selected fourteen main categories that range through distinct real-world domains, including agriculture, arts, business, engineering, health and social sciences (Table 3). As an example (see https://fenix.iscte-iul.pt/homepage/smcmo@iscte.pt/expert-systems-taxonomy for the full list), the list of keywords for “decision sciences” is: {“forecast”, “forecasting”, “audit & quality”, “auditing”, “risk assessment”, “risk control_”, “ISO 9000”, “ISO 9241-210”, “quality”, “risk analysis”, “risk mitigation”, “risk prediction”, “risk regulations”, “service selection”}.

In terms of application domains, our list of main terms is well aligned with the classical DDC and LCC classification systems, with all terms having an equivalent classification except for “decision sciences”. Sahin et al.’s list (2013) does not have a match for eight of the adopted ES´s applications (e.g., “arts and humanities”, “engineering”). Nevertheless, Sahin et al.’s list includes the broader “Other” term that includes everything else.

Table 3 - List of main ES´s application categories.

ESs´ application	(Sahin et al., 2013)	Dewey Decimal Classification	Library of Congress Classification
agricultural science	Agriculture	630 Agriculture & related technologies	S - Agriculture
arts and humanities		700-705 Arts; 001 Knowledge	N - Fine Arts; AZ - Humanities
biology and chemistry		570/572/574... 600 - "Biology"; 540...549 - "Chemistry"	QH Natural History/Biology; QD - Chemistry
business and management	Finance-Business; Human Resources Management	650-659 "Management and Accounting"	HB/HC - "Economic"; HG - Finance;
computer science	Telecommunication	000 Computer science, knowledge and general works	QA (sub-range)
decision sciences			
earth sciences		550 Earth sciences	QE - Geology
energy	Energy-Power	333 Economics of land & energy	HD9502-9502.5 Energy industries. Energy policy. Fuel trade
engineering		620 Engineering & allied operations	TA - Technology
environmental sciences		333.7 Environmental sciences	GE - Environmental Sciences
health	Medical-Heath	610 Medicine & health	R - Medicine
physics		530 Physics	QC - Physics
psychology		100 Philosophy & psychology	BF - Psychology
social sciences	Education; Law	300 Social Sciences	H - Social Sciences

4.3. Expert Systems topics

A total of thirteen topics were obtained by LDA modelling. In Table 4, the topics are presented in decreasing order (column **Topics**) of the total number of articles (column **Total**) for the time period analysed. For each topic, the table shows the four most relevant terms, where LDA relevance is measured by the absolute value of the **β** coefficient, i.e., the closer the **β** value is to zero the more relevant is the term for the topic (Moro et al, 2015). In the table, and to distinguish between ES´s methods and application terms, the latter terms are highlighted in grey. We also present the number of papers assigned to each topic according to publication time in terms of four year blocks (2000-2003; 2004-2007; 2008-2011; and 2012-2016). The last two rows present the total number of papers and average impact factor Journal Citation Reports (JCR)[7] for each time period.

Figure 2 offers a visual picture of the results presented in Table 4 through a topical map (Moro & Rita, 2018). For simplicity purposes, only the method and application that best characterise each topic are shown. The thickness of each connection represents the β coefficient shown in Table 4. Both Figure 2 and Table 4 provide some interesting insights into the surveyed ES journal research. All ES´s topics have both method and application terms, which confirms the area of ESs as the application of artificial intelligence methods to the real-world. Also, when analysing both topic relevance and the number of associated articles, data-driven methods is ranked as the most relevant method (used in 278 articles), followed by software (144 papers), knowledge representation (83 papers) and optimization (much lower value of 17 papers). Regarding the applications, arts and humanities is a strong domain, appearing in six topics (related with 326 articles), followed by business and management (four topics and 282 papers), engineering (seven topics and 253 papers), health (two topics and 48 papers) and environmental science (two topics and 36 papers). Moreover, evaluation seems to be a strong ESs´ methodological issue, appearing in six topics and often linked with data-driven methods. In fact, the association between data-driven methods and evaluation occurs in the most popular ESs´ journal topic, which are mostly applied to business and management (e.g., administration, banking, marketing) and arts and humanities (e.g., entertainment, multimedia). The second most popular topic is also interesting as it combines an ESs´ approach ("knowledge representation") with an implementation issue ("software"). This was expected since the category "software" includes terms such as "logic programming" or "modal logic", which are often used for knowledge representation. The second topic shows that knowledge representation has been particularly relevant within the domains of the arts and humanities (e.g., entertainment) and engineering (e.g., fault detection, ship design). Finally, Table 4 shows that both the total number of papers and average impact factor have increased during the analysed time period, an effect that generally occurs in all topics. In particular, the total number of ES´s papers have almost doubled when considering the two last periods (from the 2008-2011 to 2012-2016).

[7] From Clarivate Analytics: https://jcr.incites.thomsonreuters.com/

Table 4 – ES´s topics discovered (application terms are highlighted in grey)

Topic	1st Term: β	2nd Term: β	3rd Term: β	4th Term: β	2000-2003	2004-2007	2008-2011	2012-2016	Total
1	data-driven:0.1	evaluation: 3.1	business and management: 4.6	arts and humanities: 4.6	24	22	38	86	170
2	knowledge representation: 0.2	software: 2.7	arts and humanities: 3.5	engineering: 3.7	14	14	19	17	64
3	business and management: 0.1	software: 3.4	engineering: 3.6	data-driven: 3.6	11	13	9	27	60
4	engineering: 0.1	data-driven: 3.5	business and management: 3.9	arts and humanities: 5.0	6	13	6	13	38
5	health: 0.1	evaluation: 3.3	data-driven:3.4	engineering: 4.5	6	6	22	4	38
6	social sciences: 0.5	software: 1.5	arts and humanities: 3.4	engineering: 3.4	2	5	3	10	20
7	psychology: 0.9	environmental science: 1.3	engineering: 2.6	knowledge representation: 3.2	5	3	3	8	19
8	optimization: 0.4	arts and humanities: 2.8	environmental science: 2.8	evaluation: 3.0	4	5	3	5	17
9	arts and humanities: 0.4	biology and chemistry: 1.7	data-driven: 2.7	evaluation: 3.6	5	1	2	9	17
10	computer science: 0.2	arts and humanities: 3.2	distribution and infrastructure: 3.4	data-driven: 3.5	3	1	3	8	15
11	evaluation: 0.4	decision support: 2.0	business and management: 2.1	engineering: 3.6	3	1	2	8	14
12	mathematics: 0.4	data-driven: 1.3	physics: 4.2	health: 4.3	0	2	2	6	10
13	decision sciences: 1.0	energy: 1.4	data-driven: 2.0	evaluation: 2.5	0	2	0	4	6
				Total	83	88	112	205	488
				Average Impact Factor	0.28	0.32	0.84	0.88	

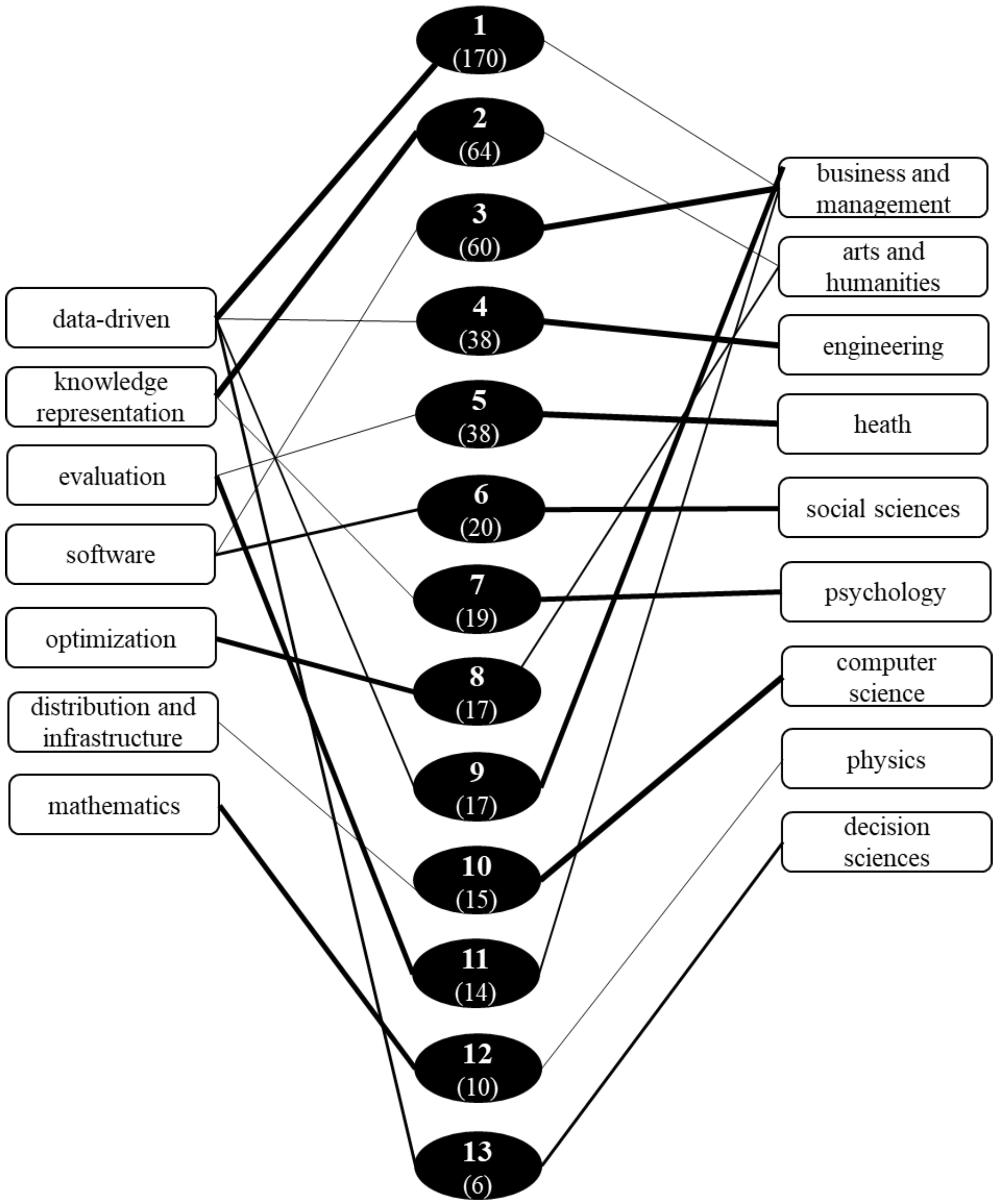

Figure 2 - Topical map.

In addition, some specific temporal trends can be observed. "Knowledge representation" is a steady trend since 2000, while "data-driven" articles have exponentially increased after 2012. Indeed, Big Data is emerging from a multitude of devices (potentiated by the Internet-of-Things) or sources (most notably, social media) demands more research on how to extract knowledge from data, leveraging data-driven methods (Song & Zhu, 2016). On an opposite direction, health-related articles have flourished in the 2008-2011 period (22, from topic 5) but have significantly decreased in the latter period. This is an interesting gap, especially since health have long been an innovative application area for expert systems (Lucas, 2008). Such finding can lead the EXSY editorial board to promote a special issue to revitalize publications in this domain. Table 4 shows there is a prevalence of data-driven methods applied to engineering (topics 3, 4, 5) and arts and humanities (topics 1, 4, and 9). While the former represents a typical domain where data emerges that needs to be handled, the latter is a surprising result. Examples of articles encompassed on the abovementioned arts and humanities' topics include music recommendation systems (Chen et al., 2016) and language translation (Al- Shawakfa & Evens, 2001). Such finding may indicate data-driven solutions can be virtually applied to any domain knowledge, given we live in a data-driven age flooded by Big Data.

5. Conclusions and Discussion

In this Expert System (ES) literature analysis, we have surveyed all research articles published in the Wiley's Expert Systems (EXSY) reference journal from 2000 to 2016. A total of 488 articles were analysed using a semi-automated text mining approach that included an assessment of authors´ affiliation countries and the analysis of research topics based on eleven ES´s methods and fourteen ES´s application categories. These categories were chosen by the authors of this paper, by analysing all articles keywords and performing several pre-processing steps, leading to a two dimension categorization system (methods and applications) that is made publicly available at: https://fenix.iscte-iul.pt/homepage/smcmo@iscte.pt/expert-systems-taxonomy. Also, the high level categories were compared with three other classification systems. As a result, we defined a taxonomy for EXSY which has already been adopted by the journal to classify future articles, helping to structure the body of knowledge published in EXSY.

The literature analysis revealed interesting insights. Both US and UK researchers play a leading role in EXSY, followed by Chinese, Turkish and Spanish authors. Cumulatively, Europe accounts for 217 papers whereas Asia has 155 articles. North America reaches 91 and Oceania 13 publications. Also, EXSY keywords are too sparse and authors submitting to this journal should analyse previously used keywords in order to increase coherence and facilitate article searches. Moreover, all identified topics combine both methods and applications, confirming ESs is an applied Artificial Intelligence field. Research and technology has evolved greatly in the last decades and EXSY research has adapted to the evolution. Indeed, data-driven methods are the most popular, followed by software, knowledge representation and optimization. Evaluation seems to be a strong ESs´ methodological issue, often associated with data-driven methods. ESs are also applied in a large range of real-world domains, in particular in arts and humanities, the most popular domain, followed by business and management, engineering, health and environmental science. Moreover, EXSY research is increasing in terms of quantity and impact, as there were only 83 research articles from 2000 to 2004 (average impact factor of 0.28), while 205 articles were published from 2012 to 2016 (average impact factor of 0.88).

All these findings confirm that the ESs´ domain is a vibrant field of research that has evolved in the last couple of decades. It is our contention that the classical ESs´ definition of Buchanan (1986) needs to be updated. One possibility could be a broader definition that does not necessary separate explicit knowledge from inference, such as: "*ESs are computerized systems that use Artificial Intelligence techniques to solve a specific real-world domain application task"*. Finally, the proposed method can be applied to other journals, helping to categorize future submissions and align them with the past in terms of journal's publications.

References

Abbasi, A., Sarker, S., & Chiang, R. H. (2016). Big Data Research in Information Systems: Toward an Inclusive Research Agenda. *Journal of the Association for Information Systems*, **17**(2), 1-33.

Acharya, U. R., Ng, E. Y. K., Sree, S. V., Chua, C. K., & Chattopadhyay, S. (2014). Higher order spectra analysis of breast thermograms for the automated identification of breast cancer. *Expert Systems*, **31**(1), 37-47.

Al- Shawakfa, E., & Evens, M. (2001). The dialoguer: An interactive bilingual interface to a network operating system. *Expert Systems*, **18**(3), 131-149.

Buchanan, B. G. (1986). Expert systems: working systems and the research literature. *Expert systems*, **3**(1):32–50.

Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment Classification of Consumer-Generated Online Reviews Using Topic Modeling. *Journal of Hospitality Marketing & Management*, **26**(7), 675-693.

Chan, L. M. (1990). *Immroth's guide to the Library of Congress Classification*. Libraries Unlimited, Englewood, Col., USA.

Chattopadhyay, S. (2014). Neurofuzzy models to automate the grading of old- age depression. *Expert Systems*, **31**(1), 48-55.

Chen, Y. S., Cheng, C. H., Chen, D. R., & Lai, C. H. (2016). A mood- and situation- based model for developing intuitive Pop music recommendation systems. *Expert Systems*, **33**(1), 77-91.

Cortez, P. (2014). *Modern optimization with R*. Springer.

Cortez, P., & Santos, M. F. (2015). Recent advances on knowledge discovery and business intelligence. *Expert Systems*, **32**(3), 433-434.

Costa, A., Julián, V., & Novais, P. (2017). Advances and trends for the development of ambient- assisted living platforms. Expert Systems, 34(2), DOI: 10.1111/exsy.12163.

Espín, V., Hurtado, M. V., & Noguera, M. (2016). Nutrition for elder care: A nutritional semantic recommender system for the elderly. *Expert Systems*, **33**(2), 201-210.

Gray, G. L., Chiu, V., Liu, Q., & Li, P. (2014). The expert systems life cycle in AIS research: What does it mean for future AIS research?. *International Journal of Accounting Information Systems*, **15**(4), 423-451.

Jackson, P. (1986). *Introduction to expert systems*. Addison-Wesley.

James, T. L., Calderon, E. D. V., & Cook, D. F. (2017). Exploring patient perceptions of healthcare service quality through analysis of unstructured feedback. *Expert Systems with Applications*, **71**, 479-492.

Jiménez, F., Jódar, R., Martín, M. D. P., Sánchez, G., & Sciavicco, G. (2016). Unsupervised feature selection for interpretable classification in behavioral assessment of children. *Expert Systems*. DOI: 10.1111/exsy.12173.

Kutbay, U., & Hardalaç, F. (2017). Development of a multiprobe electrical resistivity tomography prototype system and robust underground clustering. *Expert Systems*. DOI: 10.1111/exsy.12206.

Liao, S. H. (2005). Expert system methodologies and applications—a decade review from 1995 to 2004. *Expert Systems with Applications*, **28**(1), 93-103.

Lucas, H. (2008). Information and communications technology for future health systems in developing countries. *Social Science & Medicine*, **66**(10), 2122-2132.

Moro, S., Cortez, P., & Rita, P. (2014). A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, **62**, 22-31.

Moro, S., Cortez, P., & Rita, P. (2015). Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent Dirichlet allocation. *Expert Systems with Applications*, **42**, 3 1314-1324.

Moro, S., Rita, P., & Cortez, P. (2017). A text mining approach to analyzing Annals literature. *Annals of Tourism Research*, **66**, 208-210.

Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research*, **69**(9), 3341-3351.

Moro, S., & Rita, P. (2018). Brand strategies in social media in hospitality and tourism. *International Journal of Contemporary Hospitality Management*, **30**(1), 343-364.

Mottershead, D. (2014). Download an Excel spreadsheet and CSV file listing all the countries in the world. Retrieved in 30/June/2017 from http://www.davidmottershead.com/articles/excel-csv-countries/.

Mumpower, J. L., Phillips, L. D., Renn, O., & Uppuluri, V. R. R. (Eds.). (2012). Expert judgment and expert systems (Vol. 35). Springer Science & Business Media.

Reid, S. (1985). Knowledge-based systems concepts, Techniques, Examples. Canadian High Technology, 3(22), 238-281.

Russell, S., Norvig, P., & Intelligence, A. (1995). *Artificial Intelligence: A modern approach*. Prentice-Hall, Englewood Cliffs, 25, 27.

Sahin, S., Tolun, M. R., & Hall, J. G. (2013). A valedictory for Expert Systems print edition. *Expert Systems*, **30**(5), 381-384.

Scott, M. L. (1998). *Dewey Decimal Classification*. Libraries Unlimited, Englewoord, Col.. USA.

Shim, J. P., Warkentin, M., Courtney, J. F., Power, D. J., Sharda, R., & Carlsson, C. (2002). Past, present, and future of decision support technology. *Decision Support Systems*, **33**(2), 111-126.

Song, I. Y., & Zhu, Y. (2016). Big data and data science: what should we teach?. *Expert Systems*, **33**(4), 364-373.

Surma, J. (2015). Case- based approach for supporting strategy decision making. *Expert Systems*, **32**(4), 546-554.

Tocatlidou, A., Passam, H. C., Sideridis, A. B., & Yialouris, C. P. (2002). Reasoning under uncertainty for plant disease diagnosis. *Expert Systems*, **19**(1), 46-52.

Yoon, Y., Guimaraes, T., & O'Neal, Q. (1995). Exploring the factors associated with expert systems success. *MIS Quarterly*, 83-106.

Zhang, G., Zhang, X., & Feng, H. (2016). Forecasting financial time series using a methodology based on autoregressive integrated moving average and Taylor expansion. *Expert Systems*, **33**(5), 501-516.