Engineering Applications of Artificial Intelligence 92 (2020) 103625

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

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

A bibliometric analysis and cutting-edge overview on fuzzy techniques in Big Data

Amit K. Shukla [a,*], Pranab K. Muhuri [a], Ajith Abraham [b]

[a] Department of Computer Science, South Asian University, New Delhi 110021, India
[b] Machine Intelligence Research Labs (MIR Labs), 3rd Street NW, P.O. Box 2259, Auburn, WA 98071, USA

ARTICLE INFO

Keywords:
Big data
Fuzzy sets
Type-2 fuzzy sets
Bibliometric study
Web of science
Scopus

ABSTRACT

Over the last few years, Big Data has gained a tremendous attention from the research community. The data being generated in huge quantity from almost every field is unstructured and unprocessed. Extracting knowledge base and useful information from the big raw data is one of the major challenges, present today. Various computational intelligence and soft computing techniques have been proposed for efficient big data analytics. Fuzzy techniques are one of the soft computing approaches which can play a very crucial role in current big data challenges by pre-processing and reconstructing data. There is a wide spread application domains where traditional fuzzy sets (type-1 fuzzy sets) and higher order fuzzy sets (type-2 fuzzy sets) have shown remarkable outcomes. Although, this research domain of "fuzzy techniques in Big Data" is gaining some attention, there is a strong need for a motivation to encourage researchers to explore more in this area. In this paper, we have conducted bibliometric study on recent development in the field of "fuzzy techniques in big data". In bibliometric study, various performance metrics including total papers, total citations, and citation per paper are calculated. Further, top 10 of most productive and highly cited authors, discipline, source journals, countries, institutions, and highly influential papers are also evaluated. Later, a comparative analysis is performed on the fuzzy techniques in big data after analysing the most influential works in this field.

1. Introduction

Big data these days have become a buzz word that is used to describe data with huge volume that cannot be processed using the conventional databases and software techniques. Data is emerging at the exponential speed and it is generated from various sectors such as: social network (Facebook, Twitter), YouTube, e-commerce (Amazon, Ali Express etc.). Various methodologies have been developed in the literature to handle and retrieve better results from these ever-growing data sources (Giacalone et al., 2018; Meng et al., 2016; Li et al., 2014b; Sun et al., 2018; Elshawi et al., 2018). The volume is getting bigger and bigger day by day. Apart from the volume, which has been a major focus of the research community over the years, there are other characteristics of Big Data which also needs to be focused. Laney in 2001 (Laney, 2001) classified the properties for the Big Data that separates it from the simple data.

According to his theory, the three dimensions of Big Data are Volume, Variety, and Velocity which can be termed as Three V's. Later on, three other dimensions such as veracity, variability and value are added and collectively all these six dimensions of the Big Data are termed as 6 Vs. Fig. 1 shows the pictorial representation of the 6 V's.

All these 6 V's with their own characteristic completes the characterization of Big Data. Volume denotes the enormous size of the dataset from various sectors. The data that is flowing today is of low value density as expressed by Oracle; however, effective analysis of data is needed for the high quality value. This is collectively categorized into Value. Variety is the various formats of data structure generated from different sources. Variability is the rate of flow of data at different density. Veracity is the uncertainty in the dataset and Velocity is the rate of speed with which data is being generated. All these properties of big data possess individual challenges. From the problem to deal with the storage and the unstructured data, we are also facing bottlenecks in big data velocity which keeps changing significantly and requires different techniques which have high processing speed. Moreover, there is problem of data security and pre-processing of the ambiguous data.

A data that is generated from many real world applications is associated with uncertainty, missing values, ambiguous information, and noises in the dataset. This is due to the unpredictability of the environment, imbalanced environment parameters, error through device readings, unstructured architecture of the databases and other unnecessary factors. The field of big data is nearly decade old and over the years it has been seen that fuzzy techniques have significantly contributed

* Corresponding author.
E-mail address: amitkshukla@live.com (A.K. Shukla).

https://doi.org/10.1016/j.engappai.2020.103625
Received 12 May 2019; Received in revised form 7 February 2020; Accepted 22 March 2020
Available online 29 April 2020

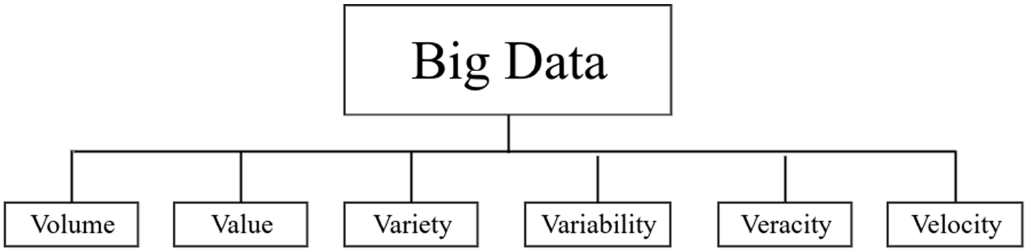

Fig. 1. Pictorial representation of the 6 V's.

to provide solutions to various big data problems in several complex applications. First incepted in 1965 by Zadeh (1965), Fuzzy Sets (FSs) have been used in various applications over the years due to their capability to handle the uncertain and missing information (Li et al., 2014a; Li and Tong, 2014; Zhou et al., 2017a; Saha et al., 2017; Feng and Chen, 2016). There have been various methodologies already available in the literature to model these uncertain parameters depending on the specific applications (Muhuri and Shukla, 2017; Colchester et al., 2017; Wang and Mendel, 2016; Vluymans et al., 2016). However, the knowledge representation capabilities of FSs enable them to efficiently handle all types of uncertainty in every kind of datasets. Also, FSs could play a very crucial role in current big data challenges by pre-processing and reconstructing data. In the past, there are only a few papers which address or motivate the use of fuzzy techniques in big data. This paper not only provides the intrinsic structure of the publications in this domain but also discusses the contribution from the top publications.

Bibliometrics analysis is a study to understand the inner structure in the research of a particular area. Sometimes, it is also referred as Scientometrics analysis. The literal meaning of this research field was provided by Pritchard (1969) and Broadus (1987). Then, further work on this area was done by Heck and Bremser (1986) in 1986, where they performed the authors and institutional analysis for the accounting field. Van Fleet et al. (2006) gave the Scientometric analysis and a concise review of Journal of Managements for first 30 years. Similar type of previous noticeable works were (Alonso et al., 2009; Hirsch, 2005; Franceschini and Maisano, 2010; Li et al., 2017; Levitt and Thelwall, 2011). In recent times, Yu and Shi (2015) discussed the growth of the Atanassov's intuitionistic fuzzy set with the help of citation analysis. Shukla et al. (2018) provided the progress of publication in real-time operating systems. An extended survey with bibliometric aspect on Industry 4.0 was studies by Muhuri et al. (2019). The bibliometric review on type-2 fuzzy sets and systems was provided by Shukla et al. (2020a). Fernández et al. (2017) studied the analysis on prion disease from 1973–2002. Not only domain specific analysis, there has been lot of publications with bibliometric analysis on the several journals such as: Knowledge-Based Cobo et al. (2015), IEEE Transactions on Fuzzy Systems (Xu et al., 2017), European Journal of Operational Research (Laengle et al., 2017), Information Sciences (Yu et al., 2017), Neurocomputing (Janmaijaya et al., 2018a), International Journal of Intelligent Systems (Merigó et al., 2017), Engineering Applications of Artificial Intelligence (Janmaijaya et al., 2018b), Applied Soft Computing (Muhuri et al., 2018) etc. This study helps the reader to get the overall structure of the growth in the field using the accessed knowledge structure.

The major contribution of this paper is highlighted as follows:

(a) The bibliometric study of "fuzzy techniques in Big Data" is analysed to explore the recent development and essential structure in this field. It will serve as the starting source for the researchers to work in this unexplored yet potentially significant domain.
(b) Bibliometric analysis covers the research growth in terms of number of publications and total citations received over the years.
(c) Further, best of 10 entities are extracted in terms of: Authors (productive and influential), discipline, source (productive and influential), countries, institutions (productive and influential), and highly influential papers in the field of "fuzzy techniques in big data".
(d) On the basis of the bibliometric analysis, the recent and influential works are summarized and explained in detail which emphasizes majorly in the use of fuzzy techniques for big data.
(e) From this study, the researchers can infer the inner structure and get a broad picture of this area.
(f) This is one of the kind study which underlines the bibliometric as well as detailed overview of the fuzzy techniques in big data.

The rest of the paper is organized as follows: Section 2 describes the process of data collection and methodology used in this paper. The detailed and extensive bibliometric analysis including research growth, productive and highly cited authors, major subject areas, top journals, contributing countries and institutions is performed in Section 3. In Section 4, we have provided a detailed bibliographic coupling between authors and countries. Section 5 discusses the most common keywords used and a comparative analysis of the various techniques for Fuzzy techniques in big data. Finally, Section 6 concludes the paper with the summarized results.

2. Data collection and methodology

The data collection is performed on the Web of Science (WoS) and Scopus repository, which are the largest databases for the bibliometric analysis. Our prime concern is to perform bibliometric study on "fuzzy techniques in big data", thus the keywords used in the search query is as: TOPIC: ("big data") AND TOPIC: (fuzzy) Timespan: 2000–2019. Indexes: SCI-EXPANDED, SSCI. We took 2000 as the starting year for this search because the big data emerged in late 2000's. However, first publication came in year 2009 and 2006 as indexed by WoS and Scopus, respectively. The search query is performed on 24th April, 2019. The Indexes are: Science Citation Index-Expanded (SCI-E), the Social Science Citation Index (SSCI) and Emerging Social Science Citation Index (E-SCI) which are the standard indexes used in the computer science and engineering community. There are several tags which are retrieved (from WoS and Scopus) such as author, title, abstract, country, citation record, author affiliation.

In WoS, there were a total of 375 publications were extracted with five different document types in which articles are 164, reviews are 11, proceeding papers are 2 and book chapters and editorial are one each. In Scopus, we have nine kinds of document types i.e. articles (370), reviews (15), conference paper (724), conference review (196), article in press (22), book chapters (34), book (3), Editorial (1) and Note (2). All the document types are compiled in Table 1. Here, '%' represents the percentage of contribution of a particular document type.

The extracted data are analysed by the various performance indicators available in the literature for the bibliometric analysis. Out of many, we have used Total Papers (TP), which is the total number of publications from the source, Total Citations (TC), which is the total number of citations received by the publication, and the Citations Per

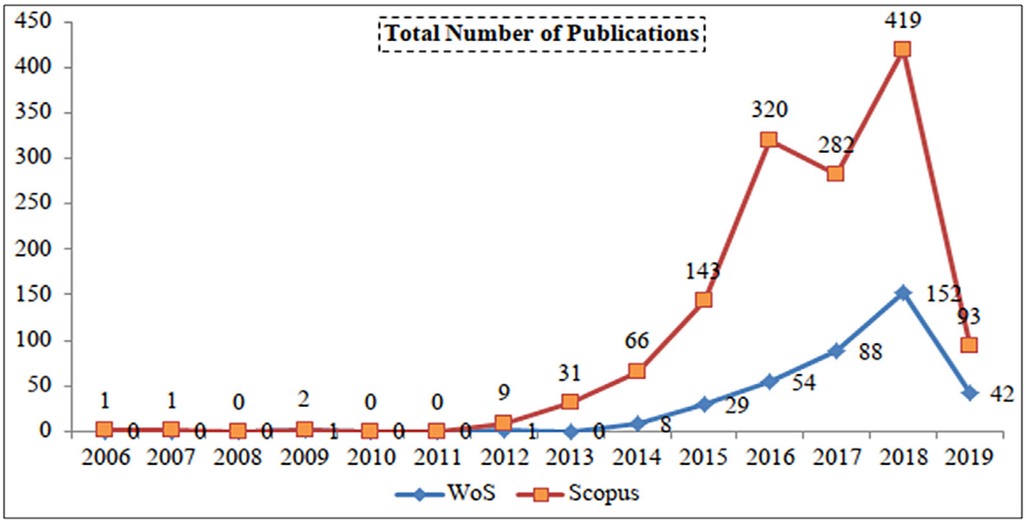

Fig. 2. Publication growth over the years in WoS and Scopus.

Table 1
Document types in WoS and Scopus.

WoS			Scopus	
Document types	Total number	%	Total number	%
Article	356	91.99	370	27.07
Review	18	4.65	15	1.10
Proceedings Paper	11	2.84	–	–
Conference Paper	–	–	724	52.96
Conference Review	–	–	196	14.34
Article in Press	–	–	22	1.61
Book Chapter	1	0.26	34	2.49
Book	–	–	3	0.22
Editorial	1	0.26	1	0.07
Note	–	–	2	0.15

Table 2
Top 10 most productive authors.

WoS				Scopus			
Authors	TP	TC	CPP	Authors	TP	TC	CPP
Herrera, F.	7	195	27.86	Wang Y.	22	84	3.818
Fernandez, A	5	87	17.4	Zhang Y.	16	28	1.75
Marcelloni, F.	4	19	4.75	Wang X.	15	54	3.6
Chang, E.	4	11	2.75	Chen Y.	13	21	1.615
Saberi, M.	4	11	2.75	Liu Y.	13	7	0.538
Rani, R.	4	10	2.5	Lu J.	12	15	1.25
Havens, T.C.	3	174	58	Herrera F.	11	278	25.27
Lopez, V.	3	163	54.33	Li Y.	11	16	1.455
Del Rio, S.	3	117	39	Liu J.	11	9	0.818
Chang, V.	3	26	8.667	Liu X.	11	8	0.727

Paper (CPP), that is total number of received citations count divided by total publications. For assessing a journal, Impact Factor (IF) is usually used as a generally accepted metric. It is calculated as the average citations of the publications in that journal for last two or five years.

3. Bibliometric analysis

This section is divided into various sub-sections including: research growth, most productive and highly cited authors, topmost subject areas, top source referenced, country-wise analysis, institution wise analysis, and highly influential papers in the field of "fuzzy techniques in big data".

3.1. Research growth

The field of Big Data has gained reputation in recent years due to its huge significance in various fields. The potential of fuzzy techniques were exploited to the big data domain which resulted in a few publications in the starting years. However, recent years have seen a tremendous growth in this field. Fig. 2 shows the research evolution in this field since 2006, as we could see the first publication in that year indexed in Scopus. The first publication indexed by WoS in 2009.

Now, according to the Scopus, there were only 4 publications till 2011. In 2012, Scopus saw the indexing of 9 papers, which eventually increased after that in subsequent years. The highest number of publications could be seen in year 2018 with 419 papers. Interestingly, 2017 saw a decline in the number of publications with only 282 papers as compared to the previous year's 320 publications.

If we examine WoS, only two contributions could be found till 2013. Then, there were exponential growth in terms of publications. In 2014, 8 paper came which increased to 262% in 2015 to 29 papers. Further, 54 and 88 publications came in consecutive years. We have the highest publications in 2018 (TP=152) and we hope this trend will increase continuously over the coming years. Quite interestingly, around 64% (240) of publications came only in the last two years (2017–18). Stating the obvious that growth in this domain is remarkable over the years, the number of citations per paper over the years in Fig. 3 justifies the same. Till 2011, none of the indexing platforms shows any major citation counts.

Scopus has the increasing growth of the citation count. It got maximum of 1538 citations in 2018. Till now, Scopus shows the total citation count of 3883 in the field of fuzzy techniques in big data. The citation count of 699 in 2019 will supposedly increase over the time.

In case of WoS, the year 2014 which was the first year when the actual publications (TP=8, TC=27) started, sees the increase in the citation count as it was only 5 in previous 4 years. Similar to the growth behaviour in total publications, more than 67% of citations came in the two years (2017–18) with a total of 1467 citations.

3.2. Most productive and highly cited authors

The most productive authors list is extracted and ranked on the basis of total number of papers from both the indexing databases. The top 10 most productive authors are shown in Table 2 and a side-by-side listing from both the databases is presented.

According to the WoS, Herrera is the highest contributor with 7 total publications in this field of "fuzzy techniques in Big Data". He is followed by Fernandez with total of 5 publications. Then there are four authors Marcelloni, Chang, Saberi, and Rani with TP of 4. In

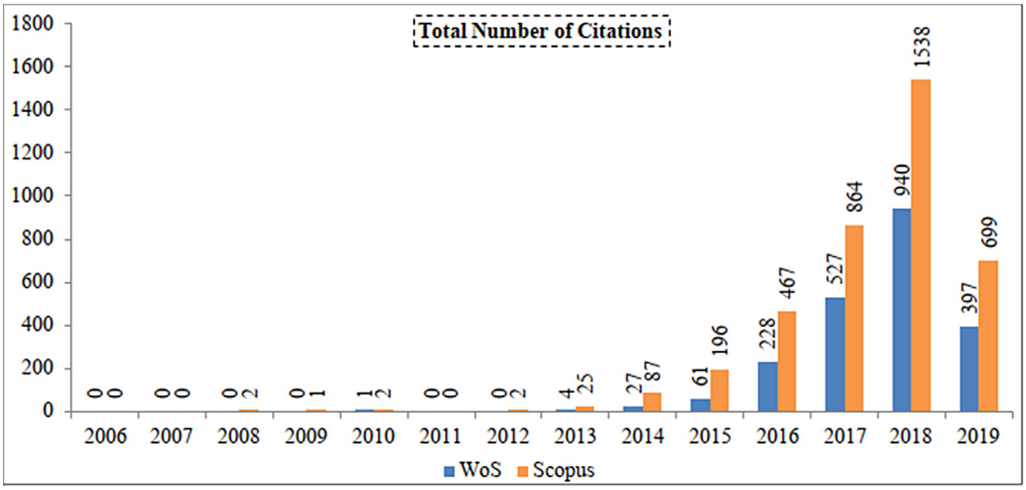

Fig. 3. Citation trend over the years.

Table 3
Top 10 most influential authors.

WoS				Scopus			
Authors	TC	TP	CPP	Authors	TC	TP	CPP
Herrera, F.	195	7	27.86	Herrera F.	278	11	25.27
Havens, T.C.	174	3	58	Lopez, V.	225	4	56.25
Lopez, V.	163	3	54.33	Del Rio, S.	171	4	42.75
Del Rio, S.	117	3	39	Jose Del Jesus, M.	107	5	21.4
Benitez, J.M.	108	2	54	Fernandez, A	103	4	25.75
Fernandez, A	87	5	17.4	Li C.	95	9	10.56
Jose Del Jesus, M.	76	2	38	Li T.	94	7	13.43
Chen, H.	67	2	33.5	Wang Y.	84	22	3.818
Li, T.	67	2	33.5	Alsubaiee S.	81	4	20.25
Zeng, A.	67	2	33.5	Zhang J.	78	7	11.14

Scopus, Wang Y is the most productive author with 22 publications. He is followed by Zhang Y (TP=16), Wang X (TP=15), Chen Y (TP=13), and Liu Y (TP=13). If we see the top 10 list of Scopus, the minimum number of publication count is 11, while the highest publication count in WoS is only 7. This is because of high quality Journals listing from the WoS, whereas Scopus index from various other sources including conferences.

In WoS, Havens has the highest total citations with 195 (TP=7) in the list of top 10 productive authors. There are four other authors with more than 100 citations viz. Havens (TC=174), Lopez (TC=163), Del Rio (TC=117), and Benitez (TC=108). Even though Havens has higher citations, the citation per paper is highest for Lopez with 54.33. Table 3 shows the list of authors ordered by the total citations.

Quite interestingly, Marcelloni, Chang E, Saberi, Rani and Chang V, who were present in list of top 10 productive authors, do not stand at any rank in the most influential authors. In Scopus, Herrera stands at the top of the list with most number of citations of 278 in 11 publications. The highest CPP of 56.25 is again held by Lopez. He is followed by Del Rio (CPP=42.75), Fernandez (CPP=25.75), and Herrera (CPP=25.27). Since, this field is more dynamic and demanding, we believe that this ranking of the authors will change considerately, over the next 5 years.

3.3. Discipline wise analysis

The WoS and Scopus repository assigns subject category to the papers indexed by them. We have extracted the top 10 discipline with applications in fuzzy techniques in Big Data and shown in Table 4. Two research areas i.e. Computer Science and Engineering are the top research areas in both the databases. The total publication count for these top 2 areas in WoS is 246 and 113 while Scopus counts for 1150 and 451 for these two subject areas. One important thing here to point out is that there may be more than one category where the paper may be assigned. Thus, it is quite possible that the total percentage can add up to more than 100%. Since, Big Data is a very lucrative field in terms of knowledge retrieval; we can see researchers from Biotechnology and business economics have also started publishing with 10 papers each, according to WoS. Scopus also indexes diversified research areas such as Social Sciences (TP=103) and Medicine (TP=43).

3.4. Top source journal

In the academic community, a journal is a timely (monthly, yearly etc.) publication with the objective to enhance the progress of the subject area, it is intended to publish. We have extracted top 10 productive journals that are publishing works on fuzzy techniques in big data and shown in Table 5. All these journals are sorted by the number of publication counts.

In the top 10 listing by WoS, Expert Systems with Applications and Information Sciences have taken the top spot with 15 publications each followed by Journal of Intelligent & Fuzzy Systems (TP=13) and IEEE Access (TP=12). Interestingly, in terms of total citation count, all of the 9 journals lacks behind the Fuzzy sets and systems, which is at 8th position and has only 9 publications but highest of 163 citation count. Information sciences (TC=134) and Knowledge based systems (TC=105, TP=9) are next in terms of citation count. As expected, citation per count of Fuzzy sets and systems is also highest among all with 18.11. The Impact Factor (IF) is highest for the Future generation computer Systems (4.639) followed by Knowledge based systems (4.396), Information Sciences (4.305), and applied soft computing (3.907).

In Scopus, almost all the top 10 sources are either proceedings or international conferences. Only two journals make the list with 8th and 9th positions: Journal of Intelligent & Fuzzy Systems and Cluster Computing. The highest of 124 publications is indexed in Advances in Intelligent Systems and Computing followed by Lecture Notes in Computer Science (75), Communications in Computer and Information Science (40) and IEEE International Conference on Fuzzy Systems (29). Quite interestingly, none of the source citation count from Scopus is close to the highest citation count of the source from WoS. This happens as the researchers would prefer to cite the publications from the good quality journal.

Table 6 shows the top 10 journals in WoS and Scopus with the highest influence (citations) on the research community. According to WoS, IEEE Transactions on Fuzzy Systems has the highest number of citation count of 213 with only 7 publications. Second spot is hold by Fuzzy sets and Systems with 163 citations followed by Information Sciences with 134 citations. Then we have the Knowledge Based Systems with

Table 4
Top 10 subject area of fuzzy techniques in big data.

WoS			Scopus		
Research areas	TP	%	Research areas	TP	%
Computer Science	246	65.6	Computer Science	1150	44.1
Engineering	113	30.13	Engineering	451	17.3
Operations Research Management Science	27	7.2	Mathematics	398	15.3
Automation Control Systems	25	6.67	Decision Sciences	192	7.4
Mathematics	25	6.67	Social Sciences	103	3.9
Telecommunications	24	6.4	Business, Management and Accounting	44	1.7
Science Technology Other Topics	19	5.07	Medicine	43	1.6
Environmental Sciences Ecology	11	2.93	Physics and Astronomy	41	1.6
Biotechnology Applied Microbiology	10	2.67	Energy	31	1.2
Business Economics	10	2.67	Materials Science	21	0.8

Table 5
Top 10 productive journals publishing works on fuzzy techniques in big data.

WoS			Scopus		
Journal (IF)	TP	TC	Journal (IF) /Conferences	TP	TC
Expert Systems With Applications (3.768)	15	86	Advances in Intelligent Systems and Computing	124	47
Information Sciences (4.305)	15	134	Lecture Notes In Computer Science	75	36
Journal of Intelligent & Fuzzy Systems (1.426)	13	34	Communications in computer and information science	40	24
IEEE Access (3.557)	12	16	IEEE international conference on fuzzy systems	29	101
Future Generation Computer Systems (4.639)	11	87	ACM international conference proceeding series	28	25
Applied Soft Computing (3.907)	10	76	Procedia Computer Science	25	79
Neurocomputing (3.241)	10	55	ICNC-FSKD 2017 — 13th International Conference On Natural Computation, Fuzzy Systems And Knowledge Discovery	23	4
Fuzzy Sets and Systems (2.675)	9	163	Proceedings — 2015 International Conference On Intelligent Transportation, Big Data And Smart City, ICITBS 2015	22	11
International Journal Of Fuzzy Systems (2.396)	9	16	Journal of Intelligent & Fuzzy Systems (1.426)	13	46
Knowledge-Based Systems (4.396)	9	105	Cluster Computing (1.601)	12	18

105 citations. These journals are the only four journals with more than 100 citations.

Unlike the behaviour in the most productive sources, Scopus list good quality journals in the list of top 10 influential journals other than international conferences. Similar to WoS, IEEE Transactions on Fuzzy Systems has the highest citations of 305 with a CPP of 50.8. Then there is Fuzzy Sets and Systems (TC=187), Expert Systems with Applications (TC=130), and Future Generation Computer Systems (125).

3.5. Country-wise analysis

We have extracted results on the basis of work distribution over several countries. The related work on fuzzy techniques in Big Data by several countries is shown in Table 7. The top 10 countries are mentioned and ranked on the basis of total publications.

According to WoS, around 36% of total publications are produced by China only, with TP of 137. China is followed by USA (TP=55), India (TP=39), Australia (TP=23) and Spain (TP=23). Although, Spain is on the fifth spot, it has highest CPP of 12.86 followed by Saudi Arabia (CPP=12.41). If we see the trend as extracted from Scopus, top three countries are same as that of WoS except India is 2nd and USA is 3rd. Also, they differ by the number of publications and citation count. While China has 479 publications, India and USA have 138 and 119 publications, respectively. Here, Australia has the highest CPP of 9 followed by Spain (7.5) and USA (7.28).

3.6. Institution wise analysis

This section discusses the institutions wise analysis who has contributed to the fuzzy techniques in big data research. Table 8 shows top 10 leading universities/ organizations who are publishing in this field ordered on the basis of total publications. In WoS, University of Granada in Spain tops the list with total of 10 papers followed by the Central South University, China (TP=8) and City University of Hong Kong, Hong Kong (TP=7). Moreover, University of Granada hold the highest CPP of 22 with 220 total citations. The next five institutes have TP of 6 which are: Universidad De Jaen, Spain, Ton Duc Thang University, Vietnam, Vellore Institute of Technology, India, Hong Kong Polytech University, Hong Kong, and Northeastern University, USA. Since TP of these institutions is 6, the ranking is then decided on the basic of higher TC. Here, Universidad De Jaen has highest citation of 87 among all of them. Country wise, we have 2 institutes from China, Hong Kong and Spain.

The top 4 productive institutes as listed by Scopus have 4 publications each, but the citation of University of Granada, Spain is higher among all with 15 citations. The other three institutes are Guangzhou

Table 6
Top 10 influential journals.

WoS			Scopus		
Journal (IF)	TP	TC	Journal (IF) /Conferences	TP	TC
IEEE Transactions on Fuzzy Systems (8.415)	7	213	IEEE Transactions on Fuzzy Systems (8.415)	6	305
Fuzzy Sets and Systems (2.675)	9	163	Fuzzy Sets and Systems (2.675)	7	187
Information Sciences (4.305)	15	134	Expert Systems with Applications (3.768)	11	130
Knowledge-Based Systems (4.396)	9	105	Future Generation Computer Systems (4.639)	10	125
Future Generation Computer Systems (4.639)	11	87	Information Sciences (4.305)	10	125
Expert Systems with Applications (3.768)	15	86	Knowledge-Based Systems (4.396)	6	116
Applied Soft Computing (3.541)	10	76	2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2015	6	108
Soft Computing (2.367)	8	64	IEEE International Conference on Fuzzy Systems	29	101
International Journal of Computational Intelligence Systems	3	58	Soft Computing (2.367)	8	101
Neurocomputing (3.317)	10	55	Applied Soft Computing (3.907)	10	84

Table 7
Top 10 leading countries in publications.

WoS			Scopus		
Country	TP	TC	Country	TP	TC
China	137	821	China	479	1129
USA	55	449	India	138	337
India	39	162	United States	119	867
Australia	23	270	United Kingdom	55	306
Spain	23	296	Spain	52	390
England	22	216	Australia	47	423
Taiwan	16	87	Italy	34	154
Iran	15	88	Taiwan	32	144
Italy	15	84	Poland	24	139
Saudi Arabia	12	149	South Korea	23	30

University, China, Dalian Maritime University, China and Chongqing University, China. Further, six institutes have 3 publications each, among which HP labs has higher citation of 85 followed by Southwest Jiaotong University (TC=63) and Indian Statistical Institute (TC=40). Interestingly, there are five institutes from China and a total of 7 countries from Asia, India and Vietnam being the other two countries.

In term of highly influential institutions, University of Granada has got citation count of 220, according to the WoS. Michigan State University, USA, University of Melbourne, Australia and University of South Florida, USA hold the second position with TC of 165 each. Highest CPP is also accounted for these three universities. King Abdulaziz University of Saudi Arabia stands at 5th spot with total citations of 131. Among the top 10, there are 2 universities from China, Spain and USA each. In Scopus, top position is again taken by University of Granada with citation count of 149. A multi-national IT company HP labs stands at 2nd position with citation count of 85, which is followed by Google, United States and University of California, Riverside, United States each with 76 citations each. Table 9 compiles the top 10 influential institutions as extracted from WoS and Scopus.

3.7. Top 10 highly influential papers

Tables 10 and 11 shows the top 10 papers with maximum citations as indexed by WoS and Scopus, simultaneously. These tables also include the author's details and publication year. The paper is influential if the total citation count is increasing gradually. In both the tables, there are six papers which are in common (Havens et al., 2012; López et al., 2015; Zeng et al., 2015; Fernandez et al., 2015; Rajan, 2015; Azar and Hassanien, 2015). Havens et al. (2012) have got the highest citation in both the databases (165 in WoS and 225 in Scopus). This paper aptly satisfies the criteria of fuzzy techniques in big data since authors have proposed a novel algorithm for clustering using fuzzy c-means in very large datasets.

Second position is acquired by López et al. (2015) in WoS with about 60% less citation count (TC=73). This paper has discussed about the fuzzy rule based classification system for the big datasets. The paper from Mittelstadt and Floridi (2016) comes next with citation count of 69 followed by the publications of Son (2015) and Zeng et al. (2015) each with 63 and 55 citations, respectively. There are total of six papers in the top 10 list which were published in 2015.

Now, according to Scopus, there are three papers which are not in WoS top 10 list. The paper from Zhou et al. (2015) got 102 citations and López et al. (2015) received 101 citations. Other two papers in Scopus are from Mittelstadt and Floridi (2016) and Azar and Hassanien (2015) with citation counts of 78 and 74, respectively.

4. Bibliographic landscape

In this section, we have demonstrated the bibliographic coupling between the authors and countries using the VOSviewer, which is widely used visualization tool used in the literature (van Eck and Waltman, 2010; Xu et al., 2017). It is used to visualize a network of publications, authors, institutions, subject area, countries etc. Bibliographic coupling is the number of times two entities (either authors or countries) cite the same entity. It specifies the disciplinary links between the nodes.

Figs. 4 and 5 show the bibliographic coupling between the authors according to WoS and Scopus, respectively. In the figures, there are nodes (represented as authors) and links between the nodes. The width of these links represents the overlap between the common references in the papers as indexed by the respective repository. Also, nodes are marked with different colours. The nodes with same colour forms a cluster with similar elements. A cluster is a set of items included in a cohesive group. The numbers of authors in one cluster have the more co-cited work and related research area.

Table 8
Top 10 leading institutions in publications.

WoS			Scopus		
Institutions	TP	TC	Institutions	TP	TC
University of Granada, Spain	10	220	University of Granada, Spain	4	15
Central South University, China	8	43	Guangzhou University, China	4	6
City University of Hong Kong, Hong Kong	7	86	Dalian Maritime University, China	4	0
King Abdulaziz University, Saudi Arabia	7	131	Chongqing University, China	4	0
Hong Kong Polytech University, Hong Kong	6	53	HP Labs, United States	3	85
Northeastern University, USA	6	11	Southwest Jiaotong University, China	3	63
Ton Duc Thang University, Vietnam	6	60	Indian Statistical Institute, India	3	40
Universidad De Jaen, Spain	6	87	University of Texas, United States	3	40
Vellore Institute of Technology, India	6	59	Ton Duc Thang University, Vietnam	3	38
Tsinghua University, China	5	46	Beijing Institute Of Technology, China	3	36

Table 9
Top 10 influential institutions in publications.

WoS			Scopus		
Institutions	TP	TC	Institutions	TP	TC
University of Granada, Spain	10	220	University of Granada, Spain	2	149
Michigan State University, USA	1	165	HP Labs, United States	3	85
University of Melbourne, Australia	1	165	Google, United States	2	76
University of South Florida, USA	1	165	University of California, United States	2	76
King Abdulaziz University, Saudi Arabia	7	131	University of Jaén, Spain	2	68
Universidad De Jaen, Spain	6	87	University of California, United States	2	64
City University of Hong Kong	7	86	Southwest Jiaotong University, China	3	63
University of Oxford, UK	1	69	Nanyang Technological University, Singapore	2	59
Southwest Jiaotong University, China	2	67	King Abdulaziz University, Saudi Arabia	2	58
Yibin University, China	2	67	Polish Academy of Sciences, Poland	2	54

Table 10
Top 10 highly cited papers on fuzzy techniques in big data in WoS.

Authors	Title	TC	Year
Havens, T. C.; Bezdek, J. C.; Leckie, C.; Hall, L. O.; Palaniswami, M.	Fuzzy c-Means Algorithms for Very Large Data (Havens et al., 2012)	165	2012
Lopez, V.; del Rio, S.; Manuel B., J.; Herrera, F.	Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data (López et al., 2015)	73	2015
Mittelstadt, B. D.; Floridi, L.	The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts (Mittelstadt and Floridi, 2016)	69	2016
Rajan, K.	Materials Informatics: The Materials Gene and Big Data (Rajan, 2015)	63	2015
Fernandez, A.; Lopez, V.; Jose del J., M.; Herrera, F.	Revisiting Evolutionary Fuzzy Systems: Taxonomy, applications, new trends and challenges (Fernandez et al., 2015)	55	2015
Zeng, A.; Li, T.; Liu, D.; Zhang, J.; Chen, H.	A fuzzy rough set approach for incremental feature selection on hybrid information systems (Zeng et al., 2015)	54	2015
Azar, A. T.; Hassanien, A. E.	Dimensionality reduction of medical big data using neural-fuzzy classifier (Azar and Hassanien, 2015)	46	2015
Wu, Kuo-J.; Liao, Ching-J.; Tseng, Ming-L.; Lim, M. K.; Hu, J.; Tan, K.	Toward sustainability: using big data to explore the decisive attributes of supply chain risks and uncertainties	42	2017
Fotovatikhah, F.; Herrera, M.; Shamshirband, S.; Chau, Kwok-W.; Ardabili, Sina F.; Piran, Md. J.	Survey of computational intelligence as basis to big flood management: challenges, research directions and future work	41	2018
Zhang, D.; Song, Xiao-D.; Wang, X,; Li, K.; Li, W.; Ma, Z.	New agent-based proactive migration method and system for Big Data Environment (BDE) (Zhang et al., 2015)	36	2015

In Fig. 4, we can clearly see three different clusters with different colours. The cluster with blue colour has Herrera, Fernandez, and Benitez, among which Herrera seems to have contributed highly and his work has been cited mostly. The other cluster consists of Havens, Liu J, Xu WH, Kundu, and De Maio. The last cluster shown in green colour is formed by Liu Y, Wang R, Wang C, Chen HM, and Li FC. We have used the initials of author names only for the Chinese authors as they may have common last name.

Due to the wider reachability of Scopus, various conferences and journals papers are indexed, and thus a different group of authors are visualized in Fig. 5, which is the bibliographic coupling of the authors as indexed in Scopus. There are six clusters in total and each cluster have different group of authors as compared with the group of authors in WoS. The cluster with yellow colour has three authors namely: Herrera, Wang X, and Marcelloni. The prominent closer is of red colour with authors such as: Wang Y, Zhang Y, Lui X, Liu H, and Wang H. The

Table 11
Top 10 highly cited papers on fuzzy techniques in big data in Scopus.

Authors	Title	TC	Year
Havens T.C., Bezdek J.C., Leckie C., Hall L.O., Palaniswami M.	Fuzzy c-Means Algorithms for Very Large Data (Havens et al., 2012)	225	2012
Zhou K., Liu T., Zhou L.	Industry 4.0: Towards future industrial opportunities and challenges (Zhou et al., 2015)	102	2016
López V., Del Río S., Benítez J.M., Herrera F.	Cost-sensitive linguistic fuzzy rule based classification systems under the MapReduce framework for imbalanced big data (López et al., 2015)	101	2015
Mittelstadt B.D., Floridi L.	The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts (Mittelstadt and Floridi, 2016)	78	2016
Azar A.T., Hassanien A.E.	Dimensionality reduction of medical big data using neural-fuzzy classifier (Azar and Hassanien, 2015)	74	2015
Fernández A., López V., Del Jesus M.J., Herrera F.	Revisiting Evolutionary Fuzzy Systems: Taxonomy, applications, new trends and challenges (Fernandez et al., 2015)	66	2015
Rajan K.	Materials Informatics: The Materials Gene and Big Data (Rajan, 2015)	65	2015
Son L.H.	DPFCM: A novel distributed picture fuzzy clustering method on picture fuzzy sets (Son, 2015)	62	2015
Alsubaiee S., Altowim Y., Altwaijry H., Behm A., Borkar V., Bu Y., Carey M., Cetindil I., Cheelangi M., Faraaz K., Gabrielova E., Grover R., Heilbron Z., Kim Y.-S., Li C., Li G., Ok J.M., Onose N., Pirzadeh P., Tsotras V., Vernica R., Wen J., Westmann T.	AsterixDB: A scalable, open source BDMS (Alsubaiee et al., 2014)	61	2014
Zeng A., Li T., Liu D., Zhang J., Chen H.	A fuzzy rough set approach for incremental feature selection on hybrid information systems (Zeng et al., 2015)	60	2015

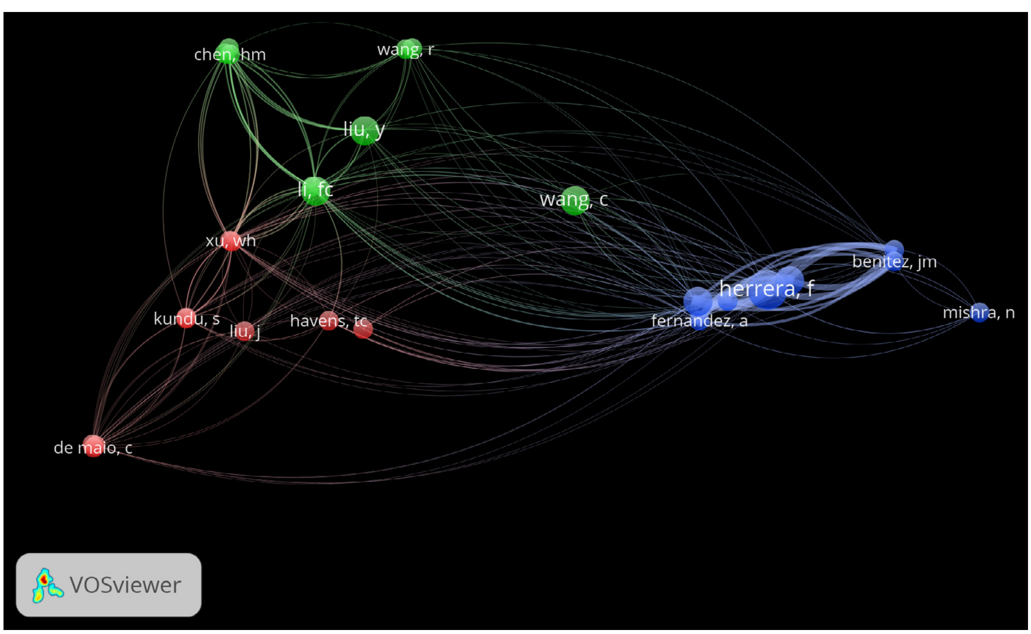

Fig. 4. Bibliographic coupling of the authors in WOS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

other important cluster is formed of blue colour and consists of Zhang G, Liu W, Zhang Q, Chen Y, Wu D, and Lu J.

Figs. 6 and 7 shows the bibliographic coupling of the contributing countries/territories publishing in the field of fuzzy techniques in Big Data. From both the outcomes of WoS and Scopus, China is the prominent node followed by USA, India and United Kingdom (England). According to the WoS, VOSviewer image (Fig. 6), China, Taiwan, and Brazil have common collaborations or common referencing, thus form a cluster with yellow colour. The other cluster of red colour is formed by USA, India, Australia, Singapore and Iran. The cluster with blue colour is formed from Saudi Arabia, Poland, South Korea, and Canada. Interestingly, Saudi Arabia has a strong link with the node of the other cluster of green colour (Spain, Greece, and Italy) which represents the more co-citations between them.

On the other hand, Fig. 7 shows a different picture with the bibliographic coupling of authors, as indexed by Scopus. The cluster with

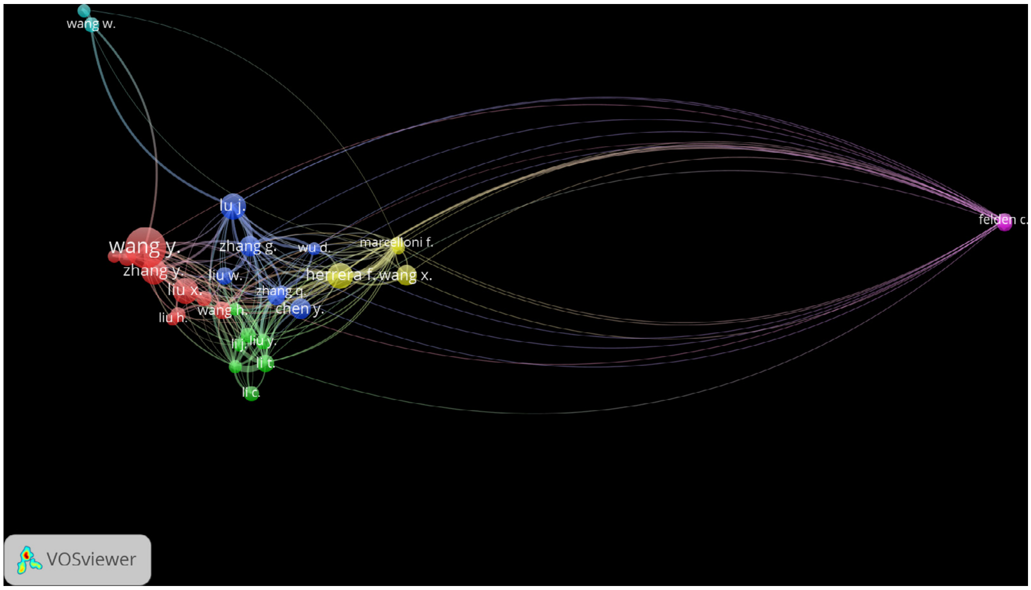

Fig. 5. Bibliographic coupling of the authors in Scopus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

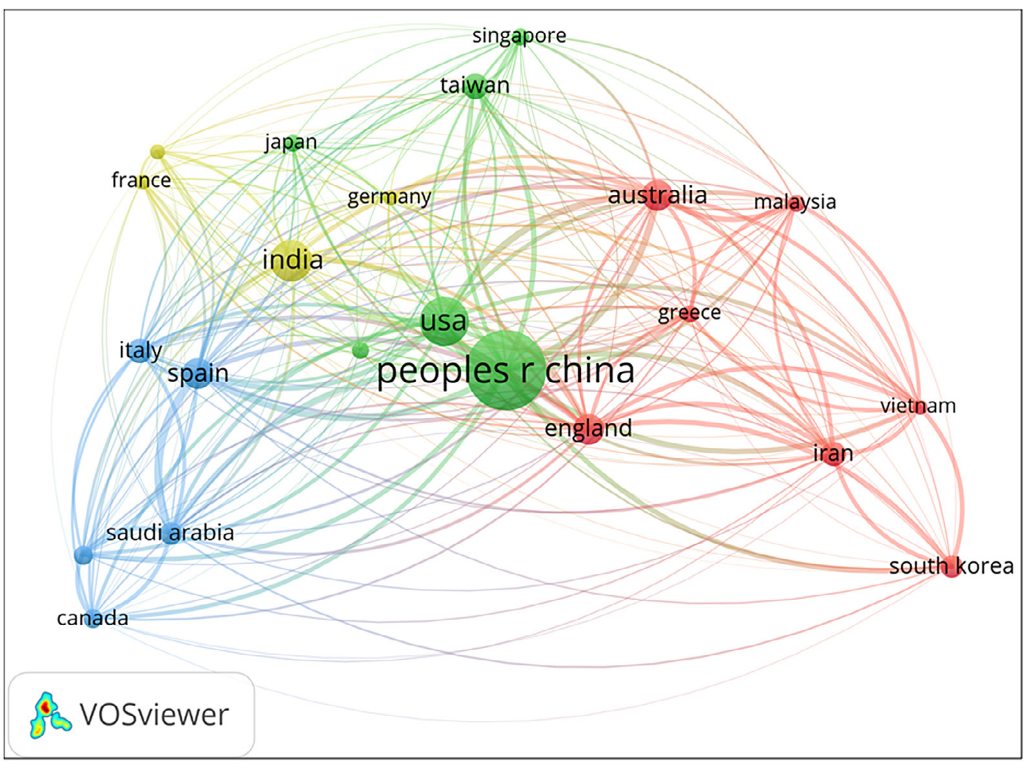

Fig. 6. Bibliographic coupling of the countries in WOS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sky blue colour has only China and Hong Kong. Here, Spain, India, and Italy are in one cluster with green colour. The other important cluster is of blue colour with countries such as: USA, Germany, Russia, Finland, and Iran.

5. Overview of the research works on fuzzy techniques in Big Data

In this section, we have visualized the most common keywords with the help of VOSviewer and discussed from the most influential papers what authors have tried to achieve in the field of "fuzzy techniques in Big Data". This section also gives the reader the rationale used in recent times to tackle big data problems and various areas to explore in future, thus providing a clearer and better perspective for the further works.

5.1. Research considering traditional Fuzzy techniques in Big Data

This sub-section further divides the research works on traditional fuzzy techniques in big data into six categories. It also shows the visualization of the most common keywords used by the researchers.

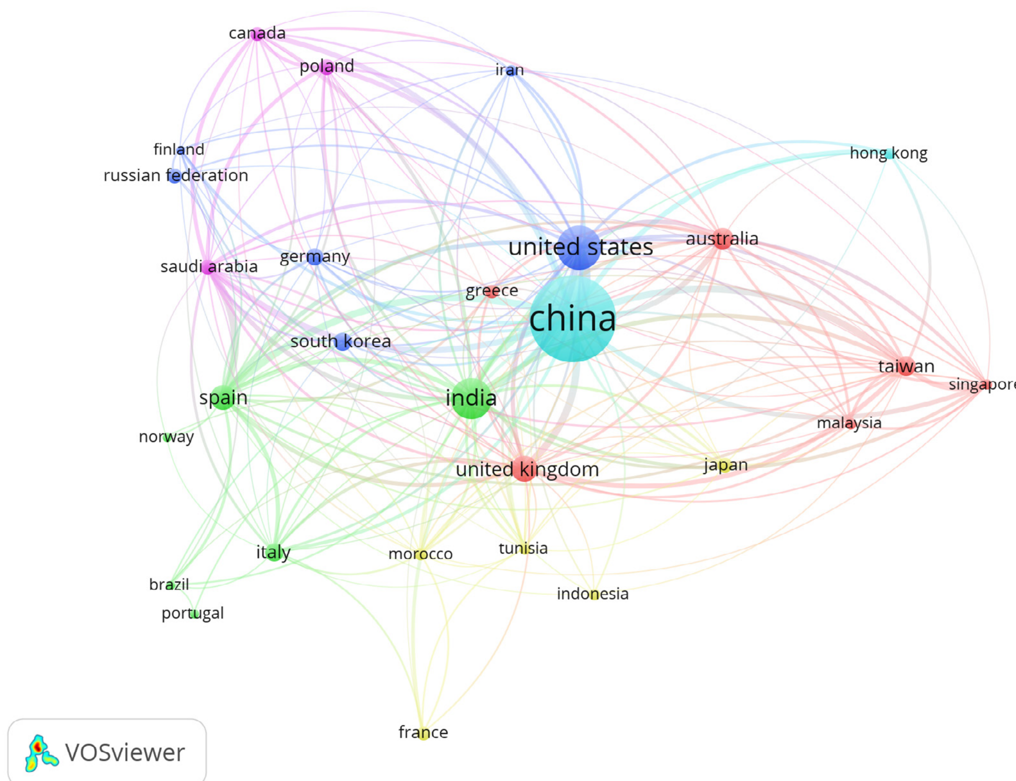

Fig. 7. Bibliographic coupling of the countries in Scopus. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

5.1.1. Fuzzy clustering

Havens et al. (2012) proposed few novel algorithms for clustering the big data. According to their arguments, the sampling method of the generalized extensible fast FCM (geFFCM) algorithm was incompetent for the very large dataset. Thus, geFCM was enhanced to the simple random sampling plus extension FCM (rseFCM) algorithm. The authors devised the kernelized algorithms of the single-pass FCM (spFCM) and online FCM (oFCM) which were named as spkFCM and okFCM, respectively. These algorithms were used to calculate the approximate FCM solutions. Moreover, the comparison between the rseFCM, single-pass FCM (spFCM), online FCM (oFCM), and bit-reduced FCM (brFCM) algorithms was performed for computing the fuzzy partition for the large datasets. Along with these algorithms, few other kernel FCM algorithms: rsekFCM, akFCM, spkFCM, and okFCM were also compared with each other.

The big data analytics to extract deeper information was performed with the help of already available Matlab tools by Tannahill and Jamshidi (2014). The Big dataset used here was in the domain of micro-grid. Authors predicted the amount of solar power generated by a micro-grid. Various techniques were tested such as: training input parameter selection, dataset sanitation, and model generation with the help of tools including, Neural Network training using back propagation (NFTOOL), Fuzzy C-Means Clustering and Rule Inference (GENFIS3) etc. For dimensionality reduction of the training data, PCA was utilized while maintaining the necessary information. Also, genetic algorithms were used to optimize the reduced trained network. A special case on social sensing to extract some necessary features of consumer behaviour in mobile network was proposed by Wang et al. (2014). This process was carried out by judiciously analysing the big data. Various factors of consumer were taken into account such as: movement, data consumption, communication etc. and the relationship among them was exhibited. Fuzzy c-means clustering were used to study the behaviour of the consumers. Bharill et al. (2016) proposed a Scalable Random Sampling with Iterative Optimization Fuzzy c-Means algorithm (SRSIO-FCM) algorithm to successfully mine the relevant and necessary information from the big dataset. Their proposed algorithm was implemented on the Apache Spark Cluster. The algorithm was compared with Literal Fuzzy c-Means (LFCM) and Random Sampling plus Extension Fuzzy c-Means (rseFCM) to verify the efficiency. Results showed that the proposed approach provides better result in less time.

Su et al. (2015) utilized the various available cluster ensemble approaches to propose a novel and enhanced approach for fuzzy clustering in big data. The combination of FCM and hierarchical clustering was used to form the base cluster. del Río et al. (2015) provided two novel approaches to develop an algorithm with various fusion techniques. The self-collected CDR data of about 4 months with fuzzy clustering was studied to analyse the mobility of users and communication behaviours by Wang et al. (2014). A total of one million users provided the input and it was concluded that ARPU level and user behaviours possess some sort of relationship between them. A novel architecture for the agriculture decision support system was proposed by Dutta et al. (2014), where they considered the machine learning approaches and unified resource description framework (RDF). A unique clustering approach was used which comprises of PCA, FCM and SOM. Son (2015) proposed a distributed fuzzy clustering algorithm viz. DPFCM, which hold the fundamental principles of the traditional fuzzy approaches and could perform efficiently for the big datasets in distributed environments.

Zhou et al. (2017b) analysed the routine electricity consumption of a local province in China. A novel structure viz. State Grid Corporation of China (SGCC) was presented for extracting the information of the daily consumption by the users. For the analysis, two of the widely used fuzzy clustering approaches were used i.e. FCM and fuzzy cluster validity index (PBMF). Morshed et al. (2013) discussed about the progress in the field of knowledge recommendation system and integration of very huge amount of environmental knowledge. The combined methods of principal component analysis (PCA), FCM and Self-organizing map (SOM) clustering were used for feature extrusion in the various environmental big datas such as AWAP, ASRIS, MODIS etc.

5.1.2. Fuzzy Rule Based Systems (FRBSs)

A more focused approach on the fuzzy techniques for handling uncertainty in big data was given by López et al. (2015). They proposed a fuzzy rule based classification system which was capable to handle uncertainly in the very large datasets. The algorithm was named as Chi-FRBCS-BigDataCS algorithm that is loosely based on the methodology of Pratama et al. (2015), a traditional FRBCS learning method adapted to deal with big datasets. The proposed approach took MapReduce framework to allocate the computational operations of the fuzzy model to tackle the imbalanced and big data, simultaneously. The empirical results on the twenty-four imbalanced big data cases of study showed the effectiveness of the algorithm. A complete and extensive review on the Evolutionary Fuzzy Systems (EFS) was carried out by Fernandez et al. (2015). EFS could effectively handle the uncertainty with the incorporation of FRBSs and the problems with evolutionary optimization. These techniques make EFS best suitable to efficiently be applied in the various Big Data analytics approaches. Furthermore, latest trend and challenges with the scope of improvement were also discussed.

5.1.3. MapReduce framework

Since the hard clustering algorithms for the big datasets were not accurate, Ludwig (2015) proposed an approach to parallelize the fuzzy c-means (FCM) algorithm with the MapReduce for better clustering results. The approach was compared with the already available methods using the validity analysis and scalability analysis. A novel clustering algorithm for efficient customer churn management in the industry framework was proposed by Bi et al. (2016) and named as: semantic-driven subtractive clustering method (SDSCM). From the experimental results, the authors concluded that the proposed SDSCM method provides has more resilient clustering semantic strength as compared to the subtractive clustering method (SCM) and FCM. A faster version of SDSCM was also devised, which was parallelized with MapReduce framework. Ye et al. (2012) proposed a scalable MapReduce platform based framework viz. vHadoop, which had the potential of performing efficiently on the large scale datasets. Authors explained the inner structure of this platform and utilized various fuzzy algorithms such as Fuzzy k-means, Meanshift etc. for the clustering process.

5.1.4. Fuzzy classifier and machine learning approaches

A novel classifier for the very large and complex dataset is proposed Azar and Hassanien (2015). They named the classifier as linguistic hedges neuro-fuzzy classifier with selected features (LHNFCSF). Not only this performs classification, but also can efficiently implement dimensionality reduction and feature selection. Authors considered medical big data for their analysis. However, medical big data involves various challenges regarding storage and their structure. The classifier used is neuro-fuzzy classifier and the proposed approach was tested for four real-world datasets to test the performance. The results showed that LHNFCSF efficiently reduces the data dimensionality and supports faster computation time.

Wang et al. (2015) proposed an uncertainty reduction technique and introduced an extreme learning machine tree (ELM-Tree), a classification technique for the Big Datasets. Information entropy and ambiguity were the uncertainty measures used in this model. These measures were used for splitting decision tree (DT) nodes where to solve the problem of the over-partitioning, ELMs was used as the leaf nodes. Results showed that using this approach one could easily reduce the computational time. A novel agent-based proactive migration technique for Big Data Environment (BDE) in Mobile services was proposed by Zhang et al. (2015). The Big Data scope of media services was composed of various communication, multimedia and computing techniques. The decision making process was solely based on the fuzzy neural network which ultimately improves the fusion belief degree. From the usability point of view, this approach had the ability to follow the mobile user from one node to other. Khatib et al. (2015) gave a method to extract relevant information from the knowledge based system (KBS) using solved troubleshooting cases. Also, the rules for diagnosis method in learning troubleshooting were generated using on Fuzzy Logic Controllers (FLCs). The approach was applied on the LTE network which was considered as the very large scaled network thus a Big Data problem. The data generation from the social media has expanded in exponential rate. Iglesias et al. (2016) discussed the big data problem of web news reading. This huge data was classified into several categories based on the similar content. To deal with accumulating the emerging data in this field, authors employed Evolving Fuzzy Systems. The proposed approach did better classification on the real-time data as well.

5.1.5. Fuzzy sets and fuzzy rough sets

He et al. (2016) for the first time applied the artificial intelligence techniques of fuzzy TOPSIS and Rough Set in extracting the actual reasons of the product infant failure. The authors proposed the connotation and intelligent analysis method of product infant failure mechanism. One of the major concerns that the authors also tried to address was the analysis of the product infant failures with respect to the big data and AI technologies. A novel fuzzy rough community approach for the community detection algorithm in which a node could be signified with various memberships for different groups was proposed by Kundu and Pal (2015). The fuzzy granular theory was applied in the big data domain of social network dataset and a unique index of normalized fuzzy mutual information for quality assessment was utilized. One of the spot light of the paper was that this proposed approach showed the superiority to other algorithms in the case of overlapping communities.

Fan et al. (2017) proposed a novel double-quantitative decision-theoretic rough fuzzy set (Dq-DTRFS) model which would rely on the operations of logical conjunction and disjunction. The big dataset on the medical diagnosis field was considered from the UCI repository and a comparison with two non-rough fuzzy sets was also provided. The generation of huge amount of messages during resource allocation was a major concern. To tackle this issue, Wang and Su (2015) proposed a dynamically hierarchical, resource-allocation (DHRA) algorithm which utilizes fuzzy pattern recognition to divide the task into several levels. This algorithm had a scalable capability for the large cloud computing sources. An extensive survey on the advancement of the sensing technology in big data domain was discussed by Zhu et al. (2015). Further, the application areas, challenges and future scopes were explored.

Balan et al. (2015) proposed a novel technique which could detect the suspicious characteristics of a node using the intrusion detection system combined with a fuzzy logic method. Not only this technique could detect most widely transmitted attacks such as black hole, grey hole etc., but also provided an unique and protected connection between two nodes. The security and intrusion concern of Multi-factor authentication (MFA) was addressed by Liu et al. (2014). They proposed an effective multi-factor authentication system which utilized the MACA feature of the big data. Further, for protecting users profile and important information, authors used fuzzy hashing and fully homomorphic encryption (FHE). A feasible security system model for the smart city environment was studied by Zhou and Luo (2017) using the combination of fuzzy logic and entropy weight methods. De Maio et al. (2016) proposed the Time Aware Knowledge Extraction (TAKE) which applies the concept of fuzzy formal concept analysis to extract useful information from the emerging data on the social network platforms. Zeng et al. (2017) analysed the variations in the attribute values of the fuzzy rough approximations. Further, they also discussed about the fuzzy equivalence relation in the fuzzy rough sets (FRSs). For experimentations they considered the eight datasets from the UCI repository. Wang et al. (2016a) utilized the heuristic approach of genetic algorithms combined with fuzzy control to solve the problems in big data based intelligent transport systems. Their method named as NeverStop basically models the traffic control system with the integration of fuzzy logic system.

5.1.6. Big data analytics, database and applications

Zhai et al. (2014) presented a quite remarkable description of the term 'Big Dimensionality' from the empirical point of view. They pointed out the extensive difference between the Big Data Volume and Big Data Dimensionality. Furthermore, authors argued on the fact that this emerging area of Big Dimensionality could be a curse as well as blessing to the society, which basically depends on how the latest computational intelligence techniques were able to handle the threat. To remove the unnecessary features for the big data analytics, various feature selection methods were used in the literature. Hybrid Information System (HIS) is one such application domain for Big Data analysis which has gained lot of attraction from the researchers. Zeng et al. (2015) proposed a novel mixture of a new Hybrid Distance (HD) in HIS and fuzzy rough set approach for feature selection. The experimental results from the various datasets on the UCI repository had shown the significance of the proposed incremental approach over the traditional non-incremental approaches. The big data aspect of the material informatics was explored for the new standard of the materials by Rajan (2015). The large volumes of data related to gene structure are generated through various experiments and computations. The major challenge in mining these material informatics is that there are no predefined theories available in the literature and extraction is saturated among the wide set of possibilities.

The drawbacks of the tag-based personalized searches in the past was tackled by Cai et al. (2014), where they proposed a novel collaborative tagging system to model user and resource profiles. Collaborative tagging or folksonomy was the efficient method to organize, share, and retrieve different types of social resources. A novel model for retrieval systems, termed as: Folksonomy-based Multimedia Retrieval System (FMRS) was devised. The data from the FMRS dataset and the MovieLens dataset were utilized as big dataset and results showed the dominance of the proposed approach over the traditional methods. Mittelstadt and Floridi (2016) argued that only the exponentially growing size of the Big Data remains the prime focus of the research community over so long. They emphasizes on the ethical inference from the analysis of big data, which according to them was a next big thing. First, the literature works in this under explored field were pointed out in the paper and later the future implications were suggested. The data is growing from various sources including social networks; internet etc. and their users have become a prime source of information. Although a lot of information is accessible and inferences are significant, these information's are not structured and badly maintained. Thus most of this information is garbage. To solve this problem, Morente-Molinera et al. (2016), proposed an automatized technique to retrieve only subjective and necessary information with the help of the knowledge data bases. The collected information was stored through fuzzy ontologies.

Alsubaiee et al. (2012) proposed an ASTERIX parallel database system, capable of managing the problem of big data and providing solutions to the fuzzy queries. Later on, Alsubaiee et al. (2014) developed an enhanced ASTERIX for big data management called as AsterixDB, which covers a wide range of applications including web-data warehousing, social data analysis etc. Pospiech and Felden (2012) correctly argued that fuzzy nature of the term Big Data and its growing research needed to be classified correctly so as to emphasis only on the relevant findings. In their significant work, authors provided the extensive review of the big data literature with the potential area of future work. In the book by Szmidt (2014), the distance and similarity measures of intuitionistic fuzzy sets in the big data domain was discussed thoroughly.

We have displayed the most common keywords used in the above explained literature overview with the help of a VOSviewer. VOSviewer is a bibliometric tool for graphical visualization of network of certain entity (keywords) (van Eck and Waltman, 2010). The commonly used keywords form a cluster of same colour. In Fig. 8, the keyword with the higher weight (most frequency) is shown more prominently as compared to the others. As in, Big Data in the centre of Fig. 8 is comparatively big in size as compared to other keywords. The keywords represented with same colour are used mostly in same work. The cluster with red colour is formed with keywords such as: big data, fuzzy sets, fuzzy logic, algorithms, computation theory, and social network. There are two other clusters with green and blue colour. Now, each cluster and every element of each cluster is connected with a link, which shows the collaboration among the works that used these keywords.

5.2. Non-traditional fuzzy techniques in Big Data

This sub-section discusses the research works in the big data using the higher order fuzzy techniques such as: T2 FSs, T2 Fuzzy logic system, IT2 FSs, GT2 FSs etc.

A distributed type-2 fuzzy logic system was implemented for the mobile agent based distributed computing platform by Benchara et al. (2015). This approach was implemented on a cooperative mobile agent model which was referred as the Single Program Multiple Data (SPMD). The big dataset which was considered here was the image dataset of MRI (Magnetic Resonance Images) images. An efficient learning approach to handle big data applications was proposed by Pratama et al. (2015) and named as Evolving Type-2 Extreme Learning Machine (eT2ELM). It was inclined to solve the problems of uncertainty, complexity and curse of dimensionality, which were the bottleneck for the standard ELM. The IT2 multivariate Gaussian MFs were used in the hidden layer of the generalized IT2 Fuzzy Neural Network (FNN) and nonlinear Chebyshev function is implemented in the output layer. To overcome the drawback of the int index limitation and data resource management in Big Data a new fuzzy model was proposed in Wang (2016). This model was based on type-2 fuzzy event parallel computing system with novel type-reduction computational optimization technique. However, the proposed model was quite different as it accumulates statistical inference with T2 fuzzy events rather than the classical If-then-else rules of the fuzzy logic system. The application used was Wireless Soft-Switch CAPS (Call Attempt Per Second).

A detailed review on the Atanassov intuitionistic fuzzy set theory was compiled by Biswas (2016). The discussions about the soft computing techniques including FS theory, T2 FS theory, L-FS theory etc. were also included with the perspective of decision making problem. Most importantly, the applicability of fuzzy set theory in solving large decision making problem was questioned and researched in the literature. An Intervals Numbers (INs) was discussed in Kaburlasos and Papakostas (2015) by Kaburlasos and Papakostas, which was based on a fuzzy neural classifier or fuzzy lattice reasoning (flr) fuzzy-ARTMAP (FAM), or flrFAM and on lattices of (Type-1) Intervals' Numbers. The application of pattern recognition was taken and solved using higher-level (meta-) representations of numerical data and a unique learning representation.

A novel clustering approach was proposed by Nouri et al. (2014) to enhance the accuracy of data pre-processing in classification problems, since big data clustering was a tedious process and computationally expensive. A hybrid clustering approach was proposed which combines general type-2 fuzzy c-means (GT2 FCM) with the k-means algorithm and termed as KGT2FCM. This has proved to be an efficient data pre-processing for the classification problems. Five benchmark datasets from the UCI repository has been taken for experimental purposes. The statistically complex problems of MRI and forecasting used type-2 fuzzy logic in Benchara et al. (2015) and Wang (2016). The variant of T2 theory in decision making problem (Biswas, 2016), and T2 interval numbers in pattern recognition (Kaburlasos and Papakostas, 2015) were explored by researchers. T2 FSs were modelled in machine learning problem of eT2ELM in Pratama et al. (2015). Finally, a more complex variant of T2 FSs, GT2 FSs were used in data processing approaches before clustering in big data domain. Ayed et al. (2014) provided a detailed overview on the current fuzzy clustering techniques with respect to the big data domain. Moreover, authors also presented a novel type-2 fuzzy clustering method with scalability to the big data.

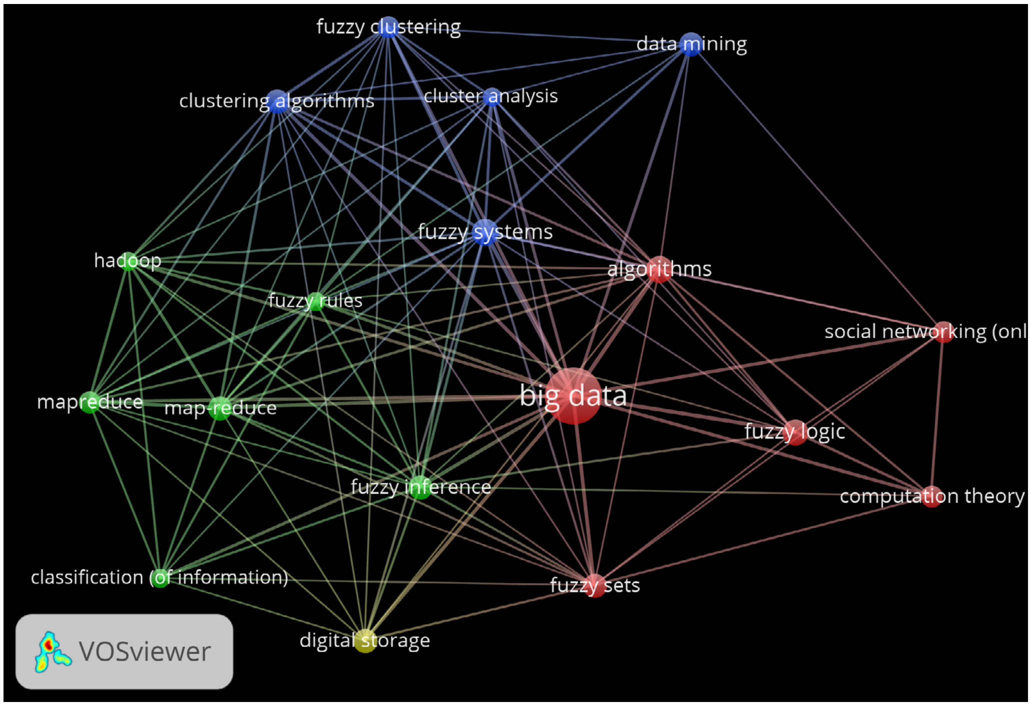

Fig. 8. Most commonly used keywords for traditional fuzzy techniques in Big Data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

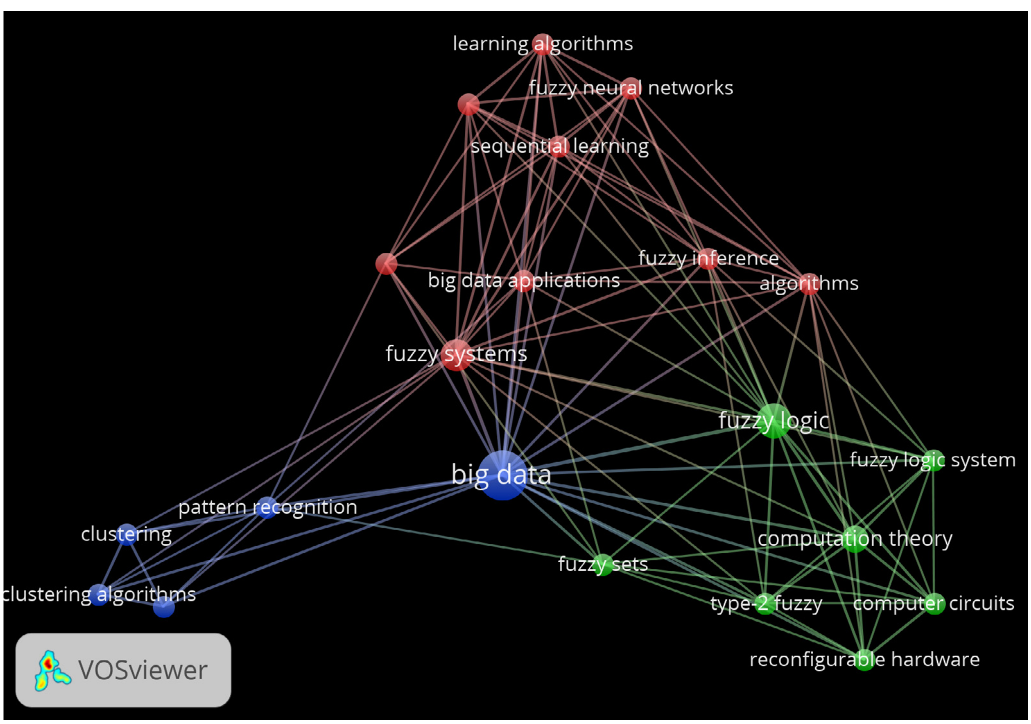

Fig. 9. Most commonly used keywords for higher order fuzzy techniques in Big Data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Moreno et al. (2020) utilized the granular computing theory to define the limits of uncertainty within IT2 membership function i.e. to explain the amount of uncertainty in the model to represent the dataset. Melin and Sánchez (2018) applied the novel granulation technique on the modular neural network integrated with multi-objective hierarchical genetic algorithm in the application domain of pattern recognition. Thus, resulting in modular granular NN on which the optimization is performed. Granularity is used as an intelligent control mechanism by Castillo et al. (2016) with the help of alpha planes theory of GT2 FSs. For that, several small controllers of a global individual controller in a non-linear aerospace control application are optimized efficiently for improved control performance. Sanchez et al. (2014) introduced the new method to find the information granules from multivariate data with the help of gravitational clustering techniques. Later, Sanchez et al. (2015) also proposed the novel approach for creating information granules using the theory of uncertainty based information. IT2 FSs were used to represent these information granules due to its proximity with the theory of granular computing and applicability to support various measures of uncertainty. Shukla et al. (2020b) addressed the scalability of big datasets by modelling it with footprint of uncertainty in the IT2 FSs.

Fig. 9 shows the most common used keywords for the type-2 fuzzy techniques in Big Data. Here again, we have three clusters. In the cluster with green colour, we have 'fuzzy logic' as the most prominent keyword. The other keywords in this cluster are fuzzy sets, type-2 fuzzy, computation theory, fuzzy logic system, computer circuits, and reconfigurable hardware. In the red colour cluster, we have a similar group among fuzzy systems, big data applications, fuzzy inference, algorithms, fuzzy neural networks, and learning algorithms. Apart from these two clusters, the blue colour cluster gas 'big data' as the most weighted keyword in the whole figure. Other keywords in this cluster are: pattern recognition, clustering, and clustering algorithm.

5.3. Conclusion: Fuzzy techniques in Big Data

As volume is one of the major issues in big data, the use of fuzzy techniques is more feasible if there are some granular computing techniques already being employed for the big data (Wang et al., 2017). In the framework of modelling uncertain data, these granular techniques consist of technologies/tools such as: Fuzzy set theory, fuzzy logic systems, interval analysis, interval type-2 fuzzy sets, type-2 fuzzy sets, rough sets etc. All these technologies and fuzzy logic system could help in the big data analytics in following ways:

- Before processing with the any established machine learning or artificial intelligence approach, model uncertainties in the raw data itself at the data pre-processing stage.
- FS techniques have found extensive application in the big data clustering. Fuzzy c-means (FCM) and extended FCM are the recent example in the advancement.
- The FSs with different membership values helps in the sampling process of the big data analytics for the division of the big data into small samples.
- In an artificial neural network and more complex deep learning architectures, fuzzy set techniques could help in the classification process for enhanced prediction models.

In conclusion, traditional fuzzy techniques in big data involve the utilization of the traditional fuzzy sets, rule based systems and machine learning etc. The datasets varying from social media, network sensor data, mobile computing, real-time big data were considered by several authors for the analytics. The basic overview on dimensionally reduction in big data and application for instance selection can be found in Zhai et al. (2014) and Zeng et al. (2015), respectively. Tannahill and Jamshidi (2014), applied PCA for dimensionality reduction and fusion of neural networks, FCM and rule based inference in power grid. Rule based inference systems were also used by López et al. (2015) for imbalanced big data and (Fernandez et al., 2015) in evolutionary optimization. In extension to rule based systems, fuzzy logic controllers in LTE networks were handled in Khatib et al. (2015). FCM was very popular for analytics in big data.

Authors in Havens et al. (2012) and Bi et al. (2016) discussed several variant of FCM, a parallelized FCM was proposed in Ludwig (2015) and Wang et al. (2014) used FCM for analysing consumer behaviour. FCM was combined with other approaches for data analysis in various applications (Su et al., 2015; Son, 2015; Dutta et al., 2014; Zhou et al., 2017b; Morshed et al., 2013). Various machine learning approaches including neuro fuzzy systems (Azar and Hassanien, 2015), ELM (Wang et al., 2015), fuzzy neural networks (Zhang et al., 2015), and fuzzy TOPSIS (He et al., 2016) were employed for the big data domain. While Mittelstadt and Floridi (2016) addresses the volume aspect of big data, fuzzy granular computing was proposed in Kundu and Pal (2015). Further, EFS and fuzzy ontologies were also used in Iglesias et al. (2016) and Morente-Molinera et al. (2016). Tools such as ASTERIX (Alsubaiee et al., 2012) and AsterixDB (Alsubaiee et al., 2014) were the parallel database systems which were used for database management. Apart from the above mentioned synopsis, there were quite remarkable related works on the overview of the fuzzy techniques in the big data (Wang et al., 2017, 2016b; Kreinovich and Ouncharoen, 2015; Ding et al., 2015; Wong et al., 2016; Sun and Wang, 2017).

Moreover, for the non-traditional techniques in big data, we can conclude that, although there is less number of publications with T2 FSs domain, researchers have applied this higher order uncertainty modelling tool for the complex applications of big data. The higher order fuzzy sets such as GT2 FSs has the potential to model more complex uncertainties in the dataset. It can further be predicted after analysing the literature that, modelling uncertainty, incomplete, and missing information with T2 FSs can provide better interpretation and more insights to the outcomes when handling the big datasets.

Further, we took the most influential papers and classified the various techniques used by several authors in this area. Table 12 compiles the overview of the latest work done on fuzzy techniques in big data. From this table, one can easily infer the latest area to be explored, big datasets to consider and fuzzy techniques to apply for their research. Since the problems of unbalanced data, missing data and various uncertainties issues are growing with the big data, the higher order FSs have been utilized by several authors to deal with these issues. This section classifies the use of IT2 FSs, T2 Interval numbers, and GT2 FSs in various big data analytics.

5.4. Future trends and impact of this area

The field of artificial intelligence and machine learning have been adopting fuzzy sets for the development of new algorithms as they can efficiently handle the uncertainty in all kind of datasets. Looking at the current scenario, this trend is expected to follow in near future. With the problem of handling big datasets, fuzzy algorithms such as FCM and Type-2 based granulation may either be used for high level of granules or for the purpose of clustering.

Other fuzzy based approach which is gaining attention is computing with words (CWW). These are the linguistic classifiers which utilized IT2 FSs for modelling the membership functions of multiple interpretations. Moreover, fuzzy sets may help curb the computational complexity associated with the computation of big data analytics. Additionally, fuzzy rough sets, hesitant fuzzy sets and Atanassov's Intuitionistic fuzzy sets may also impact immensely in the analytics of the big data domain.

6. Discussion and conclusion

This paper is a unique collection of the bibliometric study and recent development in the field of "fuzzy techniques in big data". Bibliometric study helped to discover the hidden structures of the publications in this area. Over the years, fuzzy techniques in big data have gained tremendous attention from the research community. Out of 375 publications since 2009, around 75% (282) of publications came only in these last three years (2016 and 2019). Similarly, more than 85% of citations per paper came in these three years with total of 1864 citations. The most productive authors are Herrera, Fernandez, and Marcelloni while the most influential authors are Herrera, Havens and Lopez. The most of the work is produced in Computer Science and Engineering discipline. Expert Systems With Applications is the journal with maximum publication. However, researchers are following IEEE Transactions on Fuzzy Systems journal more as it has highest total citation count and citations per paper. China remains at the top position in terms of number of publication in country wise analysis. In the Institution wise analysis, University of Granada stands at the

Table 12
Overview of the fuzzy techniques in big data.

Paper	Fuzzy technique used	Big data	Application framework	Proposed algorithm
Havens et al. (2012)	FCM	Magnetic resonance (MR) images of the brain	MATLAB	spkFCM and okFCM
Zhai et al. (2014)	– – – – – – – – – – – – – – – – –Discussions on Big Data Dimensionality– – – – – – – – – – – – – –			
López et al. (2015)	Fuzzy rule based classification system	KDD Cup 1999 dataset RLCP dataset and the Poker Hand dataset.	MapReduce framework	Chi-FRBCS-BigDataCS
Zeng et al. (2015)	Fuzzy rough set	Various datasets on the UCI repository	Hybrid Information System (HIS)	–
Fernandez et al. (2015)	– – – – – – – – – – – – – – – – – –Survey on Evolutionary Fuzzy Systems (EFS)– – – – – – – – – – – –			
Rajan (2015)	Traditional fuzzy logic	Gene structure dataset	–	–
Azar and Hassanien (2015)	Neuro-fuzzy classifier for dimensionality reduction and feature selection	Erythemato-squamous diseases dataset	MATLAB	linguistic hedges neuro-fuzzy classifier with selected features (LHNFCSF)
Wang et al. (2015)	Information entropy and ambiguity are used as the uncertainty measures	Magic Telescope, Image Segment, Page Blocks and Wine Quality-White	MATLAB and C	ELM-tree
Zhang et al. (2015)	Fuzzy neural network	Mobile services data	Mobile network services	Agent-based proactive migrating method
Tannahill and Jamshidi (2014)	GENFIS3	Micro-grid data	Fuzzy interference, neural networks, PCA, and genetic algorithms are used.	–
He et al. (2016)	Fuzzy TOPSIS and Rough Set	–	product infant failure	Mechanism analysis of product infant failure
Kundu and Pal (2015)	Fuzzy rough community and fuzzy granular theory	Zachary Karate Club and Dolphin Social Network	–	FRC-FGSN
Cai et al. (2014)	Fuzzy satisfaction problem	FMRS dataset and the MovieLens dataset	Folksonomy-based Multimedia Retrieval System (FMRS).	–
Morente-Molinera et al. (2016)	Fuzzy ontologies	Social network dataset	–	–
Ludwig (2015)	Fuzzy c-means (FCM)	Datasets in GB	MapReduce	MR-FCM algorithm
Khatib et al. (2015)	Fuzzy Logic Controllers	LTE network dataset	LTE network	Knowledge Acquisition learning algorithm
Bi et al. (2016)	FCM	Telco Big Data of China	MapReduce framework	Semantic-driven subtractive clustering method
Wang et al. (2014)	FCM	Mobile mobility data	–	Algorithm of user segmentation
Iglesias et al. (2016)	Evolving Fuzzy Systems	web news reading dataset and real-time data	NYT API	Evolving web news classification
Bharill et al. (2016)	FCM	MNIST8m USPS, Replicated Skin, Monarch, SUSY dim32, and Replicated	Apache Spark Cluster	Scalable Random Sampling with Iterative Optimization Fuzzy c-Means algorithm (SRSIO-FCM)
Benchara et al. (2015)	Type-2 fuzzy logic	image dataset of MRI	Single Program Multiple Data (SPMD).	Distributed clustering approach
Pratama et al. (2015)	IT2 Fuzzy Neural Network (FNN) with IT2 multivariate Gaussian MFs	ROCB, Yeast, ROSP, and 10dplane	–	Evolving Type-2 Extreme Learning Machine (eT2ELM).
Wang (2016)	Statistical inference with T2 fuzzy	WSS CAPS forecasting dataset	Wireless Soft-Switch CAPS (Call Attempt Per Second).	Type-2 fuzzy event parallel computing system
Biswas (2016)	Atanassov intuitionistic fuzzy set theory , FS theory, T2 FS theory	–	Decision making problem	Theory of CISF
Kaburlasos and Papakostas (2015)	Type-2 Intervals' Numbers	YALE, TERRAVIC, JAFFE, and TRIESCH I	Pattern recognition	–
Nouri et al. (2014)	General type-2 FCM (GT2 FCM)	Five benchmark datasets from the UCI repository	Data pre-processing before clustering	KGT2FCM

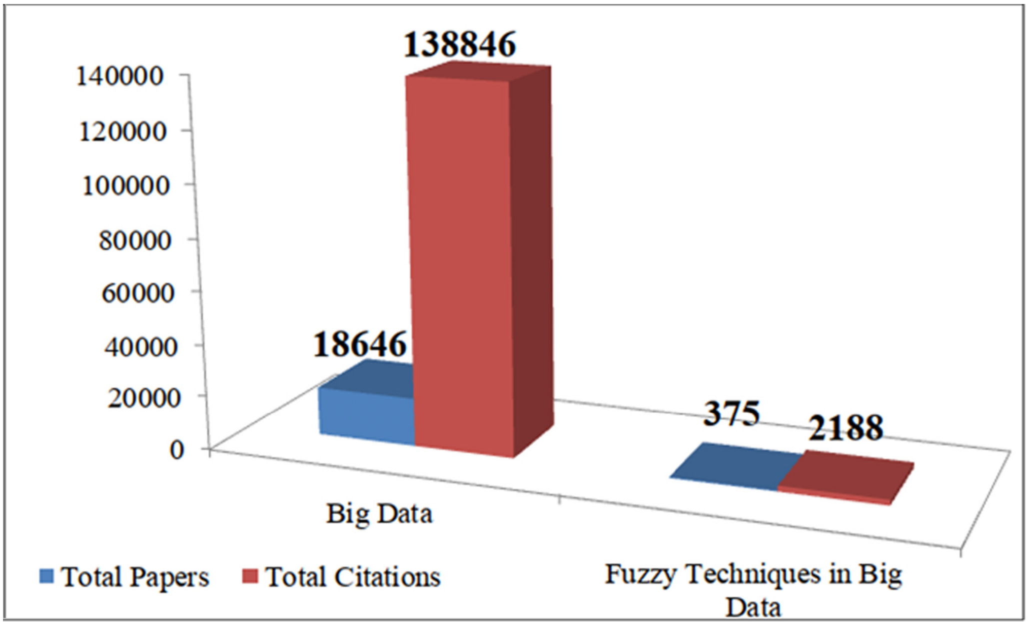

Fig. 10. WoS records on 'Big Data' versus 'Fuzzy Techniques in Big Data'.

top. Then, the comparative analysis of fuzzy techniques in the big data domain is performed from the context of the most influential papers in this field. This analysis could provide a clearer and better perspective for the new researchers.

Research works on Big Data has evolved over the years as it can be seen from the indexing by WoS. Till now (from 2009), there are 18,646 publications in Big Data as indexed by WoS. These publications have received a total of 1,38,846 citations, which specifies the significance and wider acceptability of the big data domain. However, in the case of "Fuzzy techniques in big data" publications, there are only 375 publications with only 2188 citation counts. Fig. 10 shows the respective comparison between the publications and citations count of "Big Data" and "Fuzzy Techniques in big data".

We can clearly see that there is a huge gap between the publications on general big data research and fuzzy technique based big data research. Therefore, one of the major limitations of this study is the available numbers of papers in "fuzzy techniques in big data". However, this also limelight's the immense scope and need for more and more research in the domain of fuzzy techniques in Big Data. The future scope of this study may entail the more depth analysis with other indexing databases such as Google Scholar.

Acknowledgments

Authors gratefully acknowledge the valuable comments received from the reviewers and the editors which have helped them in improving the paper significantly. First author is grateful to the Department of Science and Technology, Government of India for the financial support in the form of INSPIRE Fellowship.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., Herrera, F., 2009. H-index: A review focused in its variants, computation and standardization for different scientific fields. J. Informetr. 3 (4), 273–289.
Alsubaiee, S., Altowim, Y., Altwaijry, H., Behm, A., Borkar, V., Bu, Y., et al., 2012. ASTERIX: an open source system for big data management and analysis. Proc. VLDB Endow. 5 (12), 1898–1901.
Alsubaiee, S., Altowim, Y., Altwaijry, H., Behm, A., Borkar, V., Bu, Y., et al., 2014. AsterixDB: A scalable, open source BDMS. Proc. VLDB Endow. 7 (14), 1905–1916.
Ayed, A.B., Halima, M.B., Alimi, A.M., 2014. Survey on clustering methods: Towards fuzzy clustering for big data. In: Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of, August. IEEE, pp. 331–336.
Azar, A.T., Hassanien, A.E., 2015. Dimensionality reduction of medical big data using neural-fuzzy classifier. Soft Comput. 19 (4), 1115–1127.
Balan, E.V., Priyan, M.K., Gokulnath, C., Devi, G.U., 2015. Fuzzy based intrusion detection systems in MANET. Procedia Comput. Sci. 50, 109–114.
Benchara, F.Z., Youssfi, M., Bouattane, O., Ouajji, H., 2015. A mobile agent team works model for HPC big data analysis: Fuzzy logic application. In: Information & Communication Technology and Accessibility (ICTA), 2015 5th International Conference on, December. pp. 1–6.
Bharill, N., Tiwari, A., Malviya, A., 2016. Fuzzy based scalable clustering algorithms for handling big data using apache spark. IEEE Trans. Big Data 2 (4), 339–352.
Bi, W., Cai, M., Liu, M., Li, G., 2016. A big data clustering algorithm for mitigating the risk of customer churn. IEEE Trans. Ind. Inf. 12 (3), 1270–1281.
Biswas, R., 2016. Is 'fuzzy theory' an appropriate tool for large size decision problems? In: Imprecision and Uncertainty in Information Representation and Processing. Springer International Publishing, pp. 93–118.
Broadus, R., 1987. Toward a definition of bibliometrics. Scientometrics 12 (5–6), 373–379.
Cai, Y., Li, Q., Xie, H., Min, H., 2014. Exploring personalized searches using tag-based user profiles and resource profiles in folksonomy. Neural Netw. 58, 98–110.
Castillo, O., Cervantes, L., Soria, J., Sanchez, M., Castro, J.R., 2016. A generalized type-2 fuzzy granular approach with applications to aerospace. Inform. Sci. 354, 165–177.
Cobo, M.J., Martínez, M.A., Gutiérrez-Salcedo, M., Fujita, H., Herrera-Viedma, E., 2015. 25years at knowledge-based systems: A bibliometric analysis. Knowl.-Based Syst. 80, 3–13.
Colchester, K., Hagras, H., Alghazzawi, D., Aldabbagh, G., 2017. A survey of artificial intelligence techniques employed for adaptive educational systems within E-learning platforms. J. Artif. Intell. Soft Comput. Res. 7 (1), 47–64.
De Maio, C., Fenza, G., Loia, V., Parente, M., 2016. Time aware knowledge extraction for microblog summarization on twitter. Inf. Fusion 28, 60–74.
Ding, S., Du, M., Zhu, H., 2015. Survey on granularity clustering. Cogn. Neurodynamics 9 (6), 561–572.
Dutta, R., Morshed, A., Aryal, J., D'este, C., Das, A., 2014. Development of an intelligent environmental knowledge system for sustainable agricultural decision support. Environ. Model. Softw. 52, 264–272.
Elshawi, R., Sakr, S., Talia, D., Trunfio, P., 2018. Big data systems meet machine learning challenges: Towards big data science as a service. Big Data Res.
Fan, B., Tsang, E.C., Xu, W., Yu, J., 2017. Double-quantitative rough fuzzy set based decisions: A logical operations method. Inform. Sci. 378, 264–281.
Feng, S., Chen, C.L.P., 2016. A fuzzy restricted Boltzmann machine: Novel learning algorithms based on crisp possibilistic mean value of fuzzy numbers. IEEE Trans. Fuzzy Syst.

Fernandez, A., Lopez, V., del Jesus, M.J., Herrera, F., 2015. Revisiting evolutionary fuzzy systems: Taxonomy, applications, new trends and challenges. Knowl.-Based Syst. 80, 109–121.
Fernández, A., del Río, S., Bawakid, A., Herrera, F., 2017. Fuzzy rule based classification systems for big data with MapReduce: granularity analysis. Adv. Data Anal. Classif. 11 (4), 711–730.
Franceschini, F., Maisano, D.A., 2010. Analysis of the Hirsch index's operational properties. European J. Oper. Res. 203 (2), 494–504.
Giacalone, M., Cusatelli, C., Santarcangelo, V., 2018. Big data compliance for innovative clinical models. Big Data Res. 12, 35–40.
Havens, T.C., Bezdek, J.C., Leckie, C., Hall, L.O., Palaniswami, M., 2012. Fuzzy c-means algorithms for very large data. IEEE Trans. Fuzzy Syst. 20 (6), 1130–1146.
He, Y.H., Wang, L.B., He, Z.Z., Xie, M., 2016. A fuzzy TOPSIS and rough set based approach for mechanism analysis of product infant failure. Eng. Appl. Artif. Intell. 47, 25–37.
Heck, J.L., Bremser, W.G., 1986. Six decades of the accounting review: a summary of author and institutional contributors. Account. Rev. 73, 5–744.
Hirsch, J.E., 2005. An index to quantify an individual's scientific research output. Proc. Natl. Acad. Sci. USA 102 (46), 16569.
Iglesias, J.A., Tiemblo, A., Ledezma, A., Sanchis, A., 2016. Web news mining in an evolving framework. Inf. Fusion 28, 90–98.
Janmaijaya, M., Shukla, A., Abraham, A., Muhuri, P., 2018a. A scientometric study of neurocomputing publications (1992–2018): An aerial overview of intrinsic structure. Publications 6 (3), 32.
Janmaijaya, M., Shukla, A.K., Abraham, A., Muhuri, P.K., 2018b. The bibliometric analysis of the engineering applications of artificial intelligence (1992-2017). Eng. Appl. Artif. Intell..
Kaburlasos, V.G., Papakostas, G.A., 2015. Learning distributions of image features by interactive fuzzy lattice reasoning in pattern recognition applications. IEEE Comput. Intell. Mag. 10 (3), 42–51.
Khatib, E.J., Barco, R., Gómez-Andrades, A., Muñoz, P., Serrano, I., 2015. Data mining for fuzzy diagnosis systems in LTE networks. Expert Syst. Appl. 42 (21), 7549–7559.
Kreinovich, V., Ouncharoen, R., 2015. Fuzzy (and interval) techniques in the age of big data: An overview with applications to environmental science, geosciences, engineering, and medicine. Internat. J. Uncertain. Fuzziness Knowledge-Based Systems 23 (Suppl. 1), 75–89.
Kundu, S., Pal, S.K., 2015. Fuzzy-rough community in social networks. Pattern Recognit. Lett. 67, 145–152.
Laengle, S., Merigó, J.M., Miranda, J., Słowiński, R., Bomze, I., Borgonovo, E., et al., 2017. Forty years of the european journal of operational research: A bibliometric overview. European J. Oper. Res..
Laney, D., 2001. 3D data management: Controlling data volume, velocity and variety. META Group Res. Note 6, 70.
Levitt, J.M., Thelwall, M., 2011. A combined bibliometric indicator to predict article impact. Inf. Process. Manage. 47 (2), 300–308.
Li, F., Shi, P., Wu, L., Zhang, X., 2014a. Fuzzy-model-based ${{\cal D}}$-stability and nonfragile control for discrete-time descriptor systems with multiple delays. IEEE Trans. Fuzzy Syst. 22 (4), 1019–1025.
Li, Y., Tong, S., 2014. Adaptive fuzzy output-feedback control of pure-feedback uncertain nonlinear systems with unknown dead zone. IEEE Trans. Fuzzy Syst. 22 (5), 1341–1347.
Li, X., Wang, Y., Li, X., Wang, X., Yu, J., 2014b. GDPS: an efficient approach for skyline queries over distributed uncertain data. Big Data Res. 1, 23–36.
Li, J., Wu, D., Li, J., Li, M., 2017. A comparison of 17 article-level bibliometric indicators of institutional research productivity: Evidence from the information management literature of China. Inf. Process. Manage. 53 (5), 1156–1170.
Liu, W., Uluagac, A.S., Beyah, R., 2014. MACA: A privacy-preserving multi-factor cloud authentication system utilizing big data. In: Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on, April. IEEE, pp. 518–523.
López, V., del Río, S., Benítez, J.M., Herrera, F., 2015. Cost-sensitive linguistic fuzzy rule based classification systems under the mapreduce framework for imbalanced big data. Fuzzy Sets and Systems 258, 5–38.
Ludwig, S.A., 2015. Mapreduce-based fuzzy c-means clustering algorithm: implementation and scalability. Int. J. Mach. Learn. Cybern. 6 (6), 923–934.
Melin, P., Sánchez, D., 2018. Multi-objective optimization for modular granular neural networks applied to pattern recognition. Inform. Sci. 460, 594–610.
Meng, L., Tan, A.H., Wunsch, D.C., 2016. Adaptive scaling of cluster boundaries for large-scale social media data clustering. IEEE Trans. Neural Netw. Learn. Syst. 27 (12), 2656–2669.
Merigó, J.M., Blanco-Mesa, F., Gil-Lafuente, A.M., Yager, R.R., 2017. Thirty years of the international journal of intelligent systems: A bibliometric review. Int. J. Intell. Syst. 32 (5), 526–554.
Mittelstadt, B.D., Floridi, L., 2016. The ethics of big data: Current and foreseeable issues in biomedical contexts. In: The Ethics of Biomedical Big Data. Springer International Publishing, pp. 445–480.
Moreno, J.E., Sanchez, M.A., Mendoza, O., Rodríguez-Díaz, A., Castillo, O., Melin, P., Castro, J.R., 2020. Design of an interval type-2 fuzzy model with justifiable uncertainty. Inform. Sci. 513, 206–221.
Morente-Molinera, J.A., Pérez, I.J., Ureña, M.R., Herrera-Viedma, E., 2016. Creating knowledge databases for storing and sharing people knowledge automatically using group decision making and fuzzy ontologies. Inform. Sci. 328, 418–434.
Morshed, A., Dutta, R., Aryal, J., 2013. Recommending environmental knowledge as linked open data cloud using semantic machine learning. In: Data Engineering Workshops (ICDEW), 2013 IEEE 29th International Conference on, April. IEEE, pp. 27–28.
Muhuri, P.K., Shukla, A.K., 2017. Semi-elliptic membership function: Representation, generation, operations, defuzzification, ranking and its application to the real-time task scheduling problem. Eng. Appl. Artif. Intell. 60, 71–82.
Muhuri, P.K., Shukla, A.K., Abraham, A., 2019. Industry 4.0: A bibliometric analysis and detailed overview. Eng. Appl. Artif. Intell. 78, 218–235.
Muhuri, P.K., Shukla, A.K., Janmaijaya, M., Basu, A., 2018. Applied soft computing: A bibliometric analysis of the publications and citations during (2004–2016). Appl. Soft Comput. 69, 381–392.
Nouri, V., Akbarzadeh-T, M.R., Rowhanimanesh, A., 2014. A hybrid type-2 fuzzy clustering technique for input data preprocessing of classification algorithms. In: Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on, July. IEEE, pp. 1131–1138.
Pospiech, M., Felden, C., 2012. Big data–a state-of-the-art.
Pratama, M., Lu, J., Zhang, G., 2015. A novel meta-cognitive extreme learning machine to learning from data streams. In: Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on, October. IEEE, pp. 2792–2797.
Pritchard, J., 1969. Statistical bibliography or bibliometrics? J. Doc. 25 (4), 348–349.
Rajan, K., 2015. Materials informatics: The materials gene and big data. Annu. Rev. Mater. Res. 45, 153–169.
del Río, S., López, V., Benítez, J.M., Herrera, F., 2015. A mapreduce approach to address big data classification problems based on the fusion of linguistic fuzzy rules. Int. J. Comput. Intell. Syst. 8 (3), 422–437.
Saha, A., Konar, A., Nagar, A.K., 2017. EEG analysis for cognitive failure detection in driving using type-2 fuzzy classifiers. IEEE Trans. Emerg. Topics Comput. Intell. 1 (6), 437–453.
Sanchez, M.A., Castillo, O., Castro, J.R., 2015. Information granule formation via the concept of uncertainty-based information with interval type-2 fuzzy sets representation and Takagi–Sugeno–Kang consequents optimized with Cuckoo search. Appl. Soft Comput. 27, 602–609.
Sanchez, M.A., Castillo, O., Castro, J.R., Melin, P., 2014. Fuzzy granular gravitational clustering algorithm for multivariate data. Inform. Sci. 279, 498–511.
Shukla, A.K., Banshal, S.K., Seth, T., Basu, A., John, R., Muhuri, P.K., 2020a. A bibliometric overview of the field of type-2 fuzzy sets and systems [discussion forum]. IEEE Comput. Intell. Mag. 15 (1), 89–98.
Shukla, A.K., Sharma, R., Muhuri, P.K., 2018. A review of the scopes and challenges of the modern real-time operating systems. Int. J. Embedded Real-Time Commun. Syst. 9 (1).
Shukla, A.K., Yadav, M., Kumar, S., Muhuri, P.K., 2020b. Veracity handling and instance reduction in big data using interval type-2 fuzzy sets. Eng. Appl. Artif. Intell. 88, 103315.
Son, L.H., 2015. DPFCM. Expert Syst. Appl.: Int. J. 42 (1), 51–66.
Su, P., Shang, C., Shen, Q., 2015. A hierarchical fuzzy cluster ensemble approach and its application to big data clustering. J. Intell. Fuzzy Systems 28 (6), 2409–2421.
Sun, N., Sun, B., Lin, J.D., Wu, M.Y.C., 2018. Lossless pruned naive Bayes for big data classifications. Big Data Res.
Sun, Z., Wang, P.P., 2017. A mathematical foundation of big data. New Math. Nat. Comput. 13 (02), 83–99.
Szmidt, E., 2014. Distances and Similarities in Intuitionistic Fuzzy Sets. Springer International Publishing.
Tannahill, B.K., Jamshidi, M., 2014. System of systems and big data analytics–bridging the gap. Comput. Electr. Eng. 40 (1), 2–15.
van Eck, N.J., Waltman, L., 2010. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 84 (2), 523–538.
Van Fleet, D.D., Ray, D.F., Bedeian, A.G., Downey, H.K., Hunt, J.G., Griffin, R.W., et al., 2006. The journal of management's first 30 years. J. Manag. 32 (4), 477–506.
Vluymans, S., Tarragó, .D.S., Saeys, Y., Cornelis, C., Herrera, F., 2016. Fuzzy multi-instance classifiers. IEEE Trans. Fuzzy Syst. 24 (6), 1395–1409.
Wang, Y., 2016. Type-2 fuzzy event parallel computing system: overcoming computer int index limitation in big data. Appl. Soft Comput. 38, 1076–1087.
Wang, R., He, Y.L., Chow, C.Y., Ou, F.F., Zhang, J., 2015. Learning ELM-tree from big data based on uncertainty reduction. Fuzzy Sets and Systems 258, 79–100.
Wang, C., Li, X., Zhou, X., Wang, A., Nedjah, N., 2016a. Soft computing in big data intelligent transportation systems. Appl. Soft Comput. 38, 1099–1108.
Wang, L.X., Mendel, J.M., 2016. Fuzzy opinion networks: A mathematical framework for the evolution of opinions and their uncertainties across social networks. IEEE Trans. Fuzzy Syst. 24 (4), 880–905.
Wang, Z., Su, X., 2015. Dynamically hierarchical resource-allocation algorithm in cloud computing environment. J. Supercomput. 71 (7), 2748–2766.
Wang, Z., Tu, L., Guo, Z., Yang, L.T., Huang, B., 2014. Analysis of user behaviors by mining large network data sets. Future Gener. Comput. Syst. 37, 429–437.
Wang, L., Wang, Y., Chang, Q., 2016b. Feature selection methods for big data bioinformatics: A survey from the search perspective. Methods 111, 21–31.
Wang, H., Xu, Z., Pedrycz, W., 2017. An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities. Knowl.-Based Syst. 118, 15–30.

Wong, H.T., Chiang, V.C.L., Choi, K.S., Loke, A.Y., 2016. The need for a definition of big data for nursing science: a case study of disaster preparedness. Int. J. Environ. Res. Publ. Health 13 (10), 1015.

Xu, Z., Yu, D., Kao, Y., Lin, C.T., 2017. The structure and citation landscape of IEEE transactions on fuzzy systems (1994-2015) IEEE Trans. Fuzzy Syst..

Ye, K., Jiang, X., He, Y., Li, X., Yan, H., Huang, P., 2012. Vhadoop: A scalable hadoop virtual cluster platform for mapreduce-based parallel machine learning with performance consideration. In: Cluster Computing Workshops (CLUSTER WORKSHOPS), 2012 IEEE International Conference on, September. IEEE, pp. 152–160.

Yu, D., Shi, S., 2015. Researching the development of Atanassov intuitionistic fuzzy set: Using a citation network analysis. Appl. Soft Comput. 32, 189–198.

Yu, D., Xu, Z., Pedrycz, W., Wang, W., 2017. Information sciences 1968-2016: A retrospective analysis with text mining and bibliometric Inform. Sci..

Zadeh, L.A., 1965. Fuzzy sets. Inf. Control 8 (3), 338–353.

Zeng, A., Li, T., Hu, J., Chen, H., Luo, C., 2017. Dynamical updating fuzzy rough approximations for hybrid data under the variation of attribute values. Inform. Sci. 378, 363–388.

Zeng, A., Li, T., Liu, D., Zhang, J., Chen, H., 2015. A fuzzy rough set approach for incremental feature selection on hybrid information systems. Fuzzy Sets and Systems 258, 39–60.

Zhai, Y., Ong, Y.S., Tsang, I.W., 2014. The emerging big dimensionality. IEEE Comput. Intell. Mag. 9 (3), 14–26.

Zhang, D.G., Song, X.D., Wang, X., Li, K., Li, W.B., Ma, Z., 2015. New agent-based proactive migration method and system for big data environment (BDE). Eng. Comput. 32 (8), 2443–2466.

Zhou, T., Ishibuchi, H., Wang, S., 2017a. Stacked-structure-based hierarchical Takagi-Sugeno-Kang fuzzy classification through feature augmentation. IEEE Trans. Emerg. Topics Comput. Intell. 1 (6), 421–436.

Zhou, K., Liu, T., Zhou, L., 2015. Industry 4.0: Towards future industrial opportunities and challenges. In: Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on, August. IEEE, pp. 2147–2152.

Zhou, Q., Luo, J., 2017. The study on evaluation method of urban network security in the big data era. Intell. Autom. Soft Comput. 1–6.

Zhou, K., Yang, C., Shen, J., 2017b. Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China. Util. Policy 44, 73–84.

Zhu, T., Xiao, S., Zhang, Q., Gu, Y., Yi, P., Li, Y., 2015. Emergent technologies in big data sensing: a survey. Int. J. Distrib. Sens. Netw. 11 (10), 902982.