

# Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016



Yi Zhang, Hongshu Chen\*, Jie Lu, Guangquan Zhang

Decision Systems & e-Service Intelligence Research Lab, Centre for Artificial Intelligence, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

## ARTICLE INFO

### Article history:

Received 9 March 2017

Revised 12 May 2017

Accepted 9 July 2017

Available online 10 July 2017

### Keywords:

Topic analysis

Topic detection and tracking

Bibliometrics

Text mining

Knowledge-based Systems

## ABSTRACT

The journal Knowledge-based Systems (KnoSys) has been published for over 25 years, during which time its main foci have been extended to a broad range of studies in computer science and artificial intelligence. Answering the questions: “What is the KnoSys community interested in?” and “How does such interest change over time?” are important to both the editorial board and audience of KnoSys. This paper conducts a topic-based bibliometric study to detect and predict the topic changes of KnoSys from 1991 to 2016. A Latent Dirichlet Allocation model is used to profile the hotspots of KnoSys and predict possible future trends from a probabilistic perspective. A model of scientific evolutionary pathways applies a learning-based process to detect the topic changes of KnoSys in sequential time slices. Six main research areas of KnoSys are identified, i.e., expert systems, machine learning, data mining, decision making, optimization, and fuzzy, and the results also indicate that the interest of KnoSys communities in the area of computational intelligence is raised, and the ability to construct practical systems through knowledge use and accurate prediction models is highly emphasized. Such empirical insights can be used as a guide for KnoSys submissions.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

Topic detection and tracking (TDT), as a representative approach of topic analysis, can be dated to the 1990s, highlighting the task of identifying topics from a collection of documents [1]. Using co-word, citation statistics, or topic models, TDT has become a significant bibliometric tool [2] and assists in the studies of science, technology, innovation, and policy (STIP), e.g., profiling research performance [3,4], exploring emergent scientific or technological topics [5,6], and tracing scientific activities and development trends [7,8]. However, the development of science and technology is a process with incremental change and disruptive revolution, and both the external representation and the internal content of a scientific topic changes over time [9]. Unfortunately, traditional approaches of topic analysis are relatively static; they apply fixed models to the entire dataset and ignore any possible change resulting from time. As an example, the research scope of Knowledge-based Systems is always changing, from expert systems in the 1990s to broad disciplines in computer science and artificial intelligence, e.g., machine learning, data mining, optimization,

and decision science. Under these circumstances, two questions are raised by the journal Knowledge-based Systems' (KnoSys) research community: “What is the KnoSys community interested in?” and “How does such interest change over time?”. The answers are significant to both the editorial board and the audience of KnoSys.

To address the above questions, we focused on topics derived from all articles in KnoSys and conducted a topic-based bibliometric study to discover the topic changes in KnoSys from 1991 to 2016. The main objectives of this paper are: 1) profiling hotspots to outline the research landscape of KnoSys; 2) detecting the topic changes of KnoSys from 1991 to 2016 by tracing the predecessors and descendants of topics in evolutionary routes; and 3) predicting research trends of significant topics in KnoSys. These objectives were achieved using a model based on latent Dirichlet allocation (LDA) [10] to profile the research landscape of KnoSys by identifying academic hotspots from a probabilistic perspective. Then a model of scientific evolutionary pathways (SEP) [11] applied a learning-based process to detect the topic changes in KnoSys in a given time period, in which offline data was simulated as a data stream and the function of “sleeping beauties” was specifically used to indicate emerging topics. A prediction model is then proposed to foresee possible topic trends of KnoSys by using a probability-based weighting approach.

\* Corresponding author.

E-mail addresses: [Yi.Zhang@uts.edu.au](mailto:Yi.Zhang@uts.edu.au) (Y. Zhang), [Hongshu.Chen@uts.edu.au](mailto:Hongshu.Chen@uts.edu.au) (H. Chen), [Jie.Lu@uts.edu.au](mailto:Jie.Lu@uts.edu.au) (J. Lu), [Guangquan.Zhang@uts.edu.au](mailto:Guangquan.Zhang@uts.edu.au) (G. Zhang).

**Table 1**  
Stepwise results of the term clumping process.

Step	Description	#Term
0	Natural language processing - to retrieve raw terms from abstract and title	49,780
1	Basic cleaning - to remove terms starting with non-alphabetic characters, e.g., "2 items"	46,148
2	Basic cleaning - to remove meaningless and common terms, e.g., pronouns, prepositions, and conjunctions	43,239
3	Basic cleaning - to remove common terms in scientific articles, e.g., "method" and "introduction"	42,593
4	Knowledge-based consolidation - to consolidate terms with the same stem, e.g., the singular and plural of a noun	37,788
5	Knowledge-based consolidation - to engage expert knowledge for de-duplicating <sup>a</sup> , e.g., "classification" and "classification analysis"	29,151
6	Rule-based cleaning - to remove single words <sup>b</sup> , e.g., "information" and "feasibility"	27,065
7	Rule-based cleaning - to remove terms appearing in only one article <sup>c</sup>	4265
8	Rule-based consolidation - to consolidate terms sharing more than 3 sequential words <sup>d</sup> , e.g., "Atanassov intuitionistic fuzzy set theory" and "intuitionistic fuzzy set"	3956

*Note.* (a) The authors and co-authors of this paper arranged certain meetings to browse the remaining terms, and identified certain patterns for such consolidation. (b) A thorough consolidation was given in Step 5, where almost all meaningful single words (e.g., "classification") had been associated with multi-word terms, and the remaining single words were very common and failed to represent exact meaning. Under this circumstance, we decided to remove all single words except those associated with multi-word terms despite possible over-cleaning. (c) A vector-based similarity measurement is one main approach for the model of SEP, and terms appearing in only one article will be useless then. (d) We have ever developed an algorithm of term cluster analysis, which is to group terms based on their sharing words [13]. Despite the fact that consolidating terms that share two sequential words is a default setting, KnoSys articles contain more specific terms, e.g., "fuzzy set," "intuitionistic fuzzy set," and "hesitant fuzzy set." At this stage, we increased the required number of sharing words to "three".

This paper is organized as follows: [Section 2](#) introduces the data, including related data sources, search strategy and pre-processing efforts. [Section 3](#) follows and profiles the research landscape of KnoSys by using the LDA-based topic model. In [Section 4](#), following the SEP model the evolutionary pathways of KnoSys topics from 1991 to 2016 are visualized in the form of science maps. Predictions of future topic trends of KnoSys are given in [Section 5](#). We summarize the key findings and draw an in-depth and open discussion on the development of KnoSys in [Section 6](#). [Section 7](#) concludes our study and addresses concerns on the limitations and future directions of our study.

## 2. Data

We retrieved 2657 KnoSys articles from the Web of Science<sup>1</sup> database using the search strategy "Publication Name = Knowledge-based Systems" and a time span from January 1, 1991 to September 30, 2016.

We prefer the benefit of terms<sup>2</sup> over single words, because they represent complete semantic meanings and provide extra dimensions for similarity measures [9]. A function of natural language processing (NLP) was applied to retrieve raw terms from combined abstracts and titles, and a term clumping process [12] was used to remove noise and consolidate technological synonyms in a semi-automatic way. The stepwise results of the term clumping process are given in [Table 1](#).

The term clumping process was designed with three main functions, i.e., basic cleaning, knowledge-based consolidation, and rule-based cleaning/consolidation, and we usually follow the sequence of the three functions to conduct the term clumping process. The steps of basic cleaning are based on thesauri and stopword lists, so the sequence of these steps (i.e., Steps 1–3) is flexible. We define "knowledge" and "rule" as the criteria that are summarized from previous studies and experiments, but "knowledge" is relatively well-approved or wide-applied while "rule" can be problematic sometimes and requires further assistance of expert knowl-

edge, e.g., configuring related parameters. In order to maximize the reduction of noisy terms but minimize possible negative influence, those knowledge and rule-based steps (i.e., Steps 4–8) are fixed. Two general considerations are: to remove noisy terms before consolidation (to increase the efficiency of the term clumping process), and to conservatively remove terms (to avoid over-cleaning).

The term clumping process reduces the scale of terms derived by NLP techniques and helps identify core technological terms to assist in further analyses. However, we aimed to avoid the issue of over-cleaning, i.e., a large number of articles will only contain one core term or less, so we ran descriptive statistics to demonstrate the efficiency of the term clumping process. As given in [Table 2](#), the remaining articles maintained a coverage of 96.6%, and 91 articles without any core technological terms were identified as irrelevant articles (most of them were earlier articles with missing abstracts). The average number of core terms contained in each article is 6.73, with a standard deviation of 3.28, indicating that the main semantic information of related articles has been well-represented. The number of remaining articles in each year is given in [Fig. 1](#).

## 3. Profiling topics in KnoSys

Before profiling topics in KnoSys, a co-occurrence map (shown in [Fig. 2](#)) was generated to generally describe the research hotspots of KnoSys at a macro-level. Based on the frequency of terms, we collected the top 100 terms<sup>3</sup> and visualized them via VOSviewer [14]. [Fig. 2](#) provides a general output of traditional bibliometric studies to address such research hotspots. Some core areas are classification, machine learning, optimization, decision making, and expert systems, and the main approaches are: Bayesian analysis, support vector machine, neural network, fuzzy set, ontology, etc. Apparently, the granularity of such maps is small, i.e., each term is treated as a topic and the relationships between terms are simply based on co-occurrence statistics which can only indirectly represent the semantic information of related terms.

Aiming to profile the hotspots of KnoSys in detail and further identify their relationships, we applied an LDA-based model to identify the research landscape of KnoSys in a probabilistic way. The 2566 remaining articles (shown in [Table 2](#)) were used as the

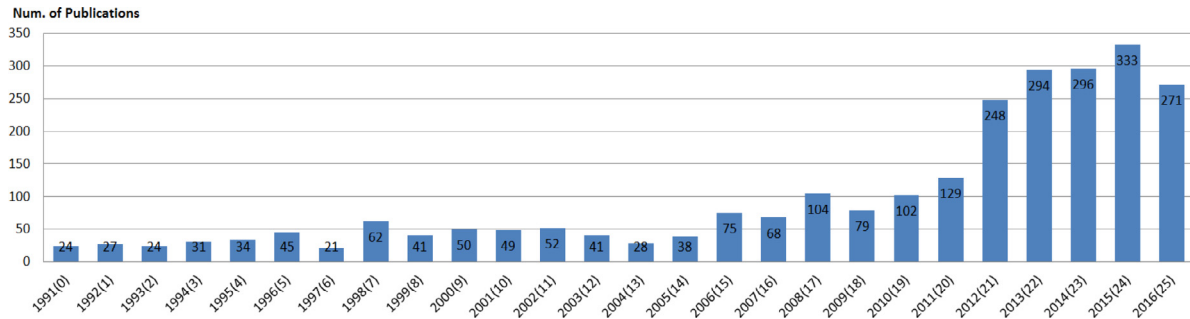
<sup>1</sup> Web of Science (WoS) is one of the most important bibliometric databases, provided by Thomson Reuters. More detail can be addressed on the website: <https://webofknowledge.com/>

<sup>2</sup> Although a lot of experiments have shown that LDA approaches work better with single words than terms, in this case we linked the words of core technological terms via dashes and treated them as single words in the LDA-based model, e.g., knowledge-based-systems and data-mining.

<sup>3</sup> Actually it is 103 terms, since seven terms at the end of the queue have the same frequency.

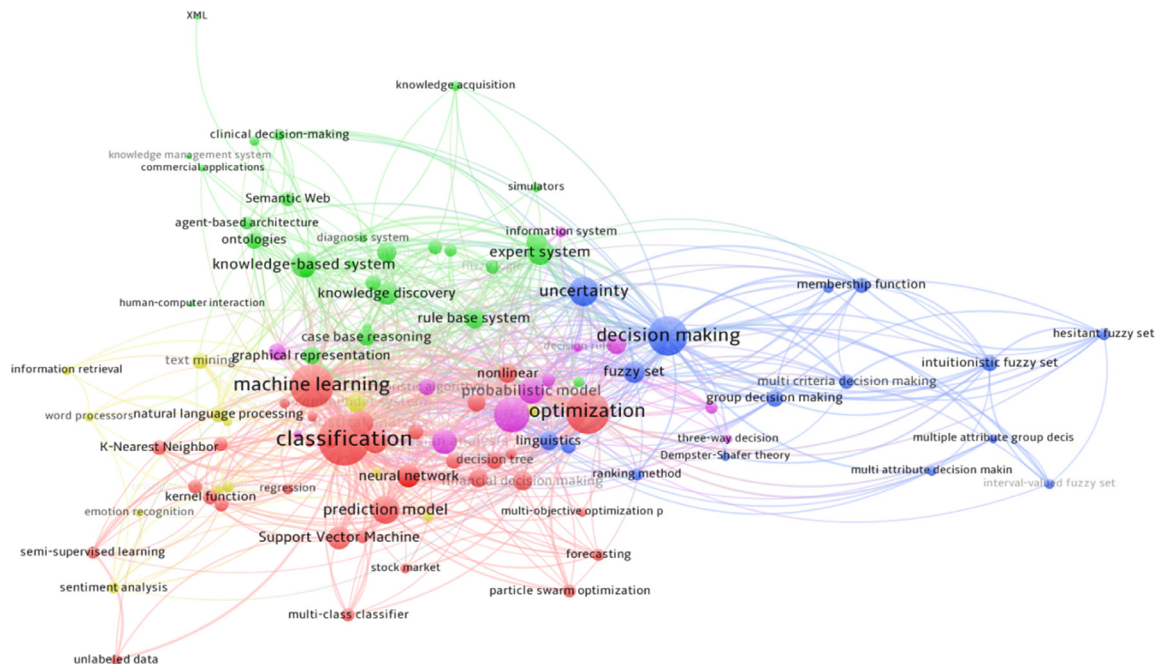
**Table 2**  
Descriptive statistics.

Num. of pub.	Coverage	Min #term	Max #term	Avg. #term	Std. dev.
2566	96.6%	1	25	6.73	3.28



**Fig. 1.** Number of articles from 1991 to 2016.

Note. The format of X axis is: Year (the sequence number of the related time slice).



**Fig. 2.** Co-occurrence map of the top 100 terms in KnoSys.

input. However, considering the preference of LDA models in single words and the benefit of the term clumping process, a pre-process was applied to the 3956 core technological terms (i.e., the outputs of the term clumping process) that linked all words of a term via short lines and treated them as single words. Then, focusing on a set of 6459 single words, we introduced an LDA-based topic model to identify topics.

### 3.1. Topic modeling

LDA approaches are based on a hierarchical structure with “word – topic – article”, where a topic is defined as an abstract semantic concept in a collection of articles and can be used to represent specific meanings by a combination of words [15]. The determination of the parameters in the LDA-based topic model is based on prior knowledge and experiments: the two hyper-parameters  $\alpha$  and  $\beta$ , controlling the amount of smoothing applied to the topic distributions for each article and the word distributions for each

topic, were 0.5 and 0.01, respectively; and the number of iterations of Gibbs sampling was set to 5000. In addition, despite the fact that in the frontier of machine learning areas non-parametric LDA models have been developed [16], which can automatically decide the most appropriate number of topics, such number sometimes can be still large for manually understanding. Under this circumstance, we ran a number of experiments for the number of topics in an interval [15,35], and selected the number of topics K as 25, with the aid of expert knowledge. This number presents comparatively lower misrepresentation of the words in multiple experiments, and better capture of the KnoSys topics with expert knowledge, which can help thoroughly understanding of the data set. In particular, considering the randomness introduced by the Gibbs sampling method, we ran multiple-round experiments and computed a perplexity score [17] to measure the level of ability of a trained model in generalizing unseen data. Perplexity is defined as the reciprocal geometric mean of the likelihood of a test corpus; a lower perplexity score indicates a lower misrepresentation

**Table 3**  
Top 25 hotspots in KnoSys.

Rank	Short labels	Topics	Detailed content
1	T10-clas	classification	classifier, machine learning, decision tree
2	T21-fuzz	fuzzy	intuitionistic fuzzy set, fuzzy logic
3	T12-gran	granulation	rough set, multi-granulation, formal concept
4	T09-eSys	expert system	architecture, real time, multimedia
5	T08-kMan	knowledge management	modeling, organization, enterprise
6	T03-kRep	knowledge representation	graphical, text mining, wordnet
7	T20-KBS	knowledge based system	collaboration, decision support, reusability
8	T22-dMin	data mining	association rule mining, knowledge discovery, KDD
9	T05-optim	optimization	convergence, particle swarm, fly optimization
10	T15-Bayes	Bayesian analysis	probability, uncertainty, learner
11	T17-sNet	social network	relationship, propagation, bandwidth
12	T01-clust	clustering	fuzzy c means, k means, vector space model
13	T06-NN	neural network	forecasting, nonlinear, stock market
14	T11-onto	ontology	case based reasoning, heterogeneous, semantic web
15	T14-pred	prediction model	regression, outliers, bankruptcy
16	T19-dMak	decision making	group decision, MADM, fuzzy preference relations
17	T25-recom	recommendation	personalization, collaborative filtering, factorization
18	T16-SVM	support vector machine	k-nearest neighbor, dimensionality, tw svm
19	T23-MVL	multi view learning	tuple, pareto, comprehensibility
20	T02-lAlg	learning algorithm	semi supervised, unlabeled, scalability
21	T13-robust	robustness	hyper heuristic, genetic algorithm, descriptors
22	T04-prog	programming	expression, boolean, microarray
23	T07-pRec	pattern recognition	discriminant analysis, discriminative, dimensionality
24	T24-maSys	multi agent system	trajectory, quality of service, trustworthiness
25	T18-topsis	TOPSIS	dictionary, reputation, simulations

Note. Here, we removed short lines that were used to link all words of a term.

of words in the corpus. The calculation of the perplexity score is:

$$\text{Perplexity}(D) = \exp - \frac{\sum_d^D \log(p(w))}{\sum_d^D N_d}$$

where  $N_d$  is the length of an article  $d$  in  $D$  and  $\sum_d^D \log(p(w))$  represents the likelihood of the corpus given the trained model.

The topic set with the lowest perplexity score is given in Table 3. The rank indicates the popularity of topics (i.e., hotspots), which is calculated by the total proportion of a topic in all articles.<sup>4</sup> We labeled topics according to expert knowledge by manually screening the detailed content of topics and selecting a representative word or word combination. A short label was used for the convenience of presentation, and its number was randomly assigned in the raw results of topic modeling.

Observations derived from Table 3 are outlined as follows: 1) classification is highlighted. As a general task in data mining, its solution strongly associates with techniques such as machine learning, classifier, Bayes, and decision tree [18,19]; 2) fuzzy is another hotspot in KnoSys and is also closely related to decision making [20,21]. Related terms include fuzzy logic, intuitionistic fuzzy set, uncertainty, group decision, etc.; 3) despite several separated topics, e.g., expert system, knowledge management, and knowledge-based system, we consider expert systems as one traditional hotspot in KnoSys [22,23]. In addition, topics such as social network [24], neural network [25], ontology [26], recommendation [27], and support vector machine (SVM) [19] are also popular for KnoSys communities, and stock market forecasting and bankruptcy prediction are two representative real-world applications in KnoSys [28,29].

### 3.2. Identifying the semantic relationships between topics

The semantic relationships between topics are identified, which are defined as the co-occurrence statistics between topics, i.e., two

<sup>4</sup> Topics are the decomposition of semantic meaning in a collection of articles, and an article can associate with various topics with different proportions, e.g., one article may introduce fuzzy logic to recommender system.

topics are discussed in the same article. The way to identify such relationships is described as follows: 1) assigning each article to two topics<sup>5</sup> with the highest probabilistic proportion in a topic proportion matrix; 2) measuring the relationships between two topics by using their co-occurrence statistics; and 3) generating a topic co-occurrence matrix and visualizing it via Circos [30]. Shown in Fig. 3, each segment represents a topic, with the short label given in Table 3. The ribbons between segments stand for their semantic relationships, i.e., a stronger relationship between the two linked segments is represented by a wider ribbon, which is distinguished by a more intense blue.

According to Fig. 3, decision making and fuzzy share a very strong semantic relationship in KnoSys, the same as what we observe in Table 3, and the relationship between granulation and fuzzy is impressive [31]. Close relationships also exist between classification and topics such as cluster, SVM, and learning algorithm. In some sense, machine learning techniques have been widely involved in solving classification issues [32]. Another interesting relationship is the one among Knowledge-based Systems, knowledge representation, and ontology, which is considered a combination of expert systems and text mining (or semantic web) [33]. In addition, the use of recommender systems in social network analysis can be seen [27].

## 4. Detecting topic change in KnoSys

### 4.1. Methodology of scientific evolutionary pathways

Science maps, highlighted in the field of bibliometrics, emphasize the representation of the relationships among scientific disciplines and portfolios [34,35], in which citation/co-citation analysis [36], co-word analysis [37], and bibliographic coupling [38] are widely used. A thorough review on the analytic techniques and software tools of science maps was given in [39], and a consensus map was developed to act as a way to validate science maps [40]. In addition, the engagement of network analytic techniques with

<sup>5</sup> In the default setting of topic models, an article will hold a probabilistic proportion with all topics.



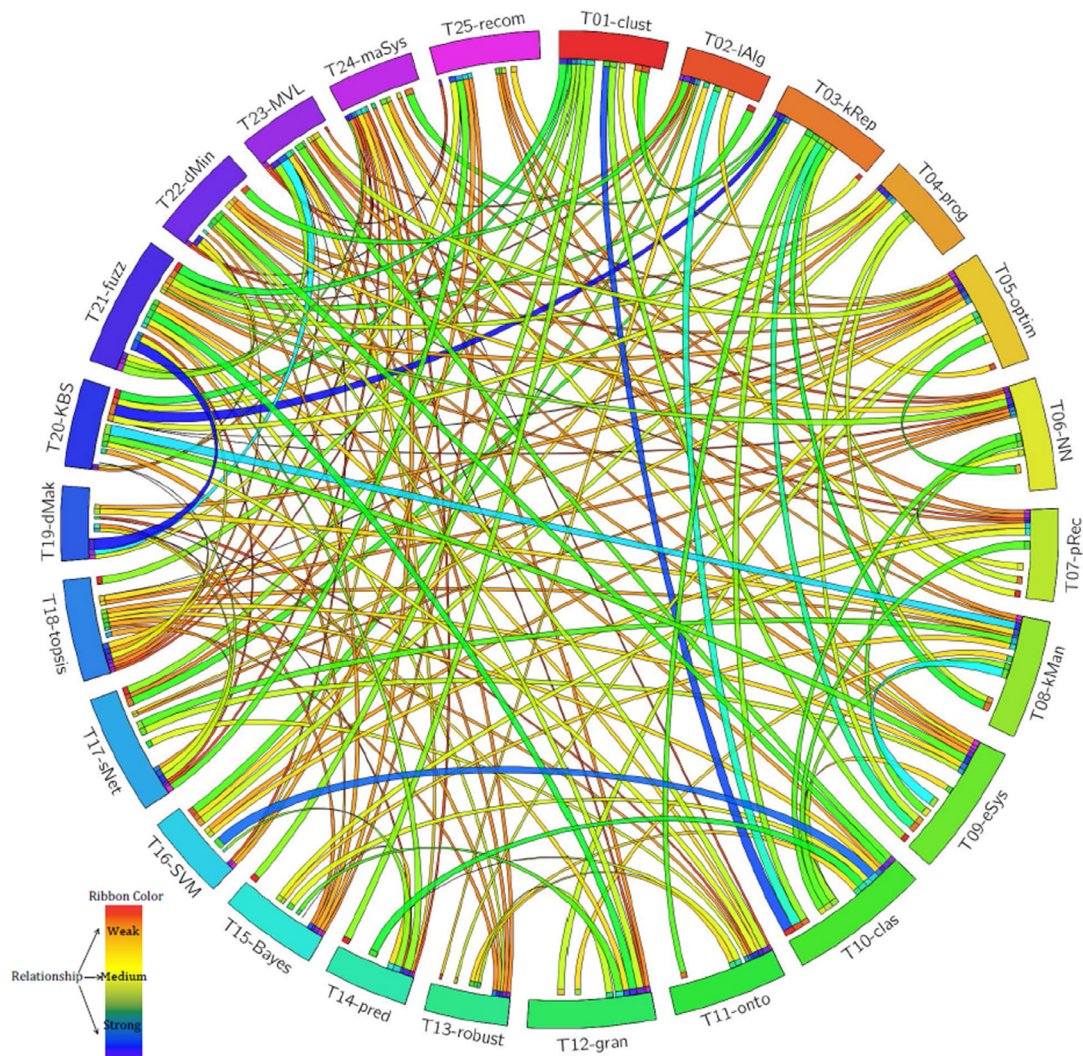


Fig. 3. Co-occurrence map of Top 25 hotspots in KnoSys.

science maps (especially for the analysis of collaboration networks) is also becoming an emergent direction in science maps [41,42].

The methodology of scientific evolutionary pathways (SEP), used as the main model of this section, is within the scope of science maps. The SEP method provides a learning process to identify topics and their relationships [11], in which topic change is defined as the dynamics of a topic in a sequential time period. Aiming to explore the underlying topic change in KnoSys, two kinds of topic change are highlighted in this paper:

- Topic evolution: the incremental change of either feature space (i.e., the composition of involved terms) or data distribution (i.e., the frequency of involved terms) in a topic, and such a change results in the appearance of new topics. In this circumstance, we set the existing topic as the predecessor and the new topics as the descendants;
- Topic death: if a topic has not been assigned any new articles for certain sequential time slices, we set this topic as a dead one. In particular, a function of detecting “sleeping beauties” is used to detect the situation where a new topic has a strong relationship with a dead topic; the new topic is added to the dead one and resurged.

Note that using citation analysis to detect “sleeping beauties” in scientific documents was fully discussed in [43], and the main idea is that: if an old topic (or a dead topic) is highlighted again

in a new time slice, such a topic can be of emerging importance. A function of detecting “sleeping beauties” by assessing the dynamic information of topics (rather than citation statistics) has been integrated with the SEP method, and we will follow this way to analyze dead topics.

The SEP method defines a topic as a collection of scientific articles and uses a circle to represent a topic. Thus, the centroid of a topic is represented by the article sharing the largest similarity with all other articles in the topic, and its boundary is the largest Euclidean distance between the centroid and the other articles. At this stage, the basic algorithm of the SEP method in our study can be described as follows:

Step 1: We simulate all 2566 articles as a data stream, consisting of 26 sequential time slices from 1991 to 2016, and group the 24 articles in Time Slice 0 (i.e., 1991) to one initial topic, i.e., knowledge-based system. Starting from Time Slice 1, we run an iterative process to analyze articles in each time slice, and re-configure related parameters (e.g., the centroid and boundary of each topic) before the end of one loop.

Step 2: A loop starts here. We use Salton’s cosine measure [44] to calculate the similarity between each forthcoming article of the time slice and all existing topics, and assign the article to the most similar topic;

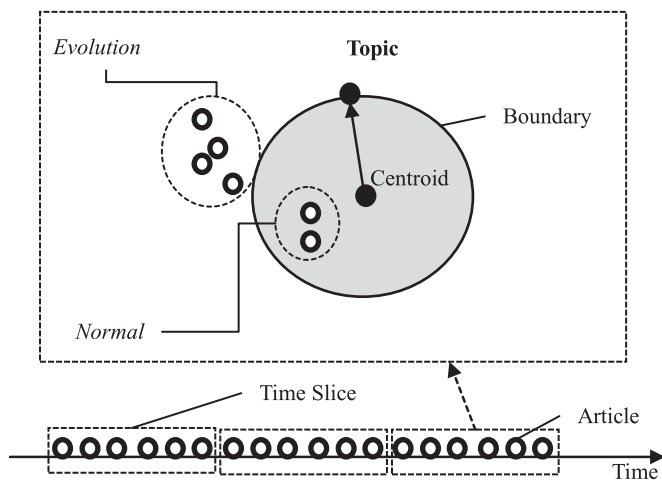


Fig. 4. Diagram of the SEP model.

- Step 3: We calculate the Euclidean distance between the article and the centroid of its assigned topic. If the distance is within the boundary, we set the article as “normal.” If it is without the boundary, we set the article as “evolution.”
- Step 4: Before ending a loop, we use a K-mean approach<sup>6</sup> [9] to group the articles labeled with “evolution” in each topic, and the new topics are set as the descendants of their assigned topic, which is considered their predecessor. At the same time, if one topic does not have descendants in two sequential time slices, we set this topic as dead.
- Step 5: The function of detecting “sleeping beauties” follows, in which we measure the similarity between new topics and all existing and dead topics. If a new topic shares the highest similarity with a topic that is not its predecessor, we will combine the new topic to the old one. If the old topic is dead, it will be resurged. Those resurged topics will be considered “sleeping beauties”.
- Step 6: A loop ends, and the centroid and boundary of each topic is recalculated. If the data stream does not end, we return to Step 2.

Note that in Step 2 the Salton’s cosine measure has been adopted to calculate the semantic similarity between two articles, which is proved a better performance in bibliometrics [45,46], and the Euclidean distance used in Step 3 is to measure the distance between a new-assigned article and the centroid of this topic in a geometric way and such distance then is used to classify the article to label “normal” or “evolution”. This design takes our assumption that a topic is a circle into consideration, and attempts to involve two semantic measures together for detecting the change of topics.

Based on the algorithm given above, a diagram of the SEP model is given in Fig. 4. Articles with time stamps are organized in sequential time slices, and the iteration will access article by article and time slice by time slice. When assuming a topic is a circle, two parameters can be given, i.e., centroid and boundary. An article is first assigned to the topic with which it shares the highest similarity value, and then, we label this article based on its distance to the centroid of the topic. Finally the K-means clustering approach will be applied to group articles labeled with “evolution” and the detection of “sleeping beauties” is addressed.

<sup>6</sup> The strategy of deciding the initial topic number  $K$  in the K-means approach is described as follows: if the number of publications is less than 10, we group them as one topic (i.e.,  $K=1$ ); if the number of publications is within the interval [10,30], we let  $K=2$ ; if the number of publications is more than 30, we set  $K$  as 3.

## 4.2. Topics and “sleeping beauties”

Fifty-three topics were generated, including the initial one in Time Slice 0. Information about the 53 topics is given in Table 4. The name of a topic is labeled by the term with a highest value of prevalence [13], and the calculation of term frequency inverse document frequency (TFIDF) analysis follows the classic formula given by Gerard Salton [47]. In addition, considering the function of detecting “sleeping beauties”, the value of survival length only counts the number of time slices in which the topic receives new articles.

Based on the topics’ information, in particular the values of survival length and death time, we classify the 53 topics into five categories:

- Category 1: *Always alive* (29 topics) – the topics are always alive;
- Category 2: *Alive with resurgence* (9 topics) – the topics are alive in Time Slice 25 (i.e., 2016) but were resurged in certain time slices before. These topics are the “sleeping beauties”, which might contain potential for innovation and have been re-discovered years after they were first claimed;
- Category 3: *Dead with resurgence* (1 topic) – this topic had been resurged, but was set as dead after the resurgence and is still dead in Time Slice 25. This topic is an extension of “sleeping beauties”, which might not be innovative any more or still needs time to be re-discovered;
- Category 4: *Dead without resurgence* (13 topics) – the topics are dead in Time Slice 25 and have never been resurged, which may or may not hold potential in innovation;
- Category 5: *Dead when born* (1 topic) – the topic died one time slice after it was generated, but they have descendants.

Certain observations obtained from Table 4 are concluded as follows: 1) the topics *always alive* may be the main interest of KnoSys, e.g., machine learning [48] and optimization [49]; 2) the topics *alive with resurgence* might be those that were not the main foci of KnoSys initially but are definitely within the scope now, e.g., data mining [50] and recommender system [27]; 3) the only topic *dead with resurgence* (i.e., expert system) can be considered as a relatively broad field, which is definitely of interest in KnoSys but can have certain more specific names in recent years, e.g., recommender system [27] and early warning system [51]; 4) it is intriguing to discuss the topics *dead without resurgence*. On one hand, it can be something interesting for KnoSys in a certain given time period and might be not the main foci any more, e.g., engineering design and linguistics. On the other hand, we notice certain topics in this category relate to either very broad research fields (e.g., information retrieval, artificial intelligence, and pattern recognition) or basic approaches (e.g., Bayesian analysis and decision tree). One possible insight here is that it is not popular to emphasize “old” terms or basic approaches in the title and abstract of an article in KnoSys, and certain detailed highlights would be preferred; and 5) regarding the topic *dead when born*, we believe it is not meaningless. In contrast, it is easy to link semantic web with ontologies [26], one of the main interests of KnoSys. In this situation, the topic semantic web is a closely-relevant term to ontologies, which then takes the place of semantic web with both horizontal and vertical advantages.

## 4.3. Topic evolutionary pathways of KnoSys

Referring to basic rules in complex network analysis, we identified the 53 topics as 53 nodes, with 52 directed arcs between predecessors and their descendants. In addition, we also calculated the similarities between the 53 topics (excluding predecessors and

**Table 4**  
Information of topics.

ID	Topic Name	Parent	Born TS	TFIDF	Death TS	Surv. Len.	Notes for Sleeping Beauties
262	Knowledge-based System	0	0	0.228	n/a	25	Alive with resurgence
447	expert system	262	1	0.139	16	14	Dead with resurgence
448	information retrieval	262	1	0.047	4	3	Dead without resurgence
450	agent based architecture	447	2	0.169	n/a	24	Always alive
453	intelligent system	448	4	0.052	8	4	Dead without resurgence
455	human computer interaction	448	4	0.105	n/a	16	Alive with resurgence
456	pattern recognition	262	4	0.020	8	4	Dead without resurgence
457	engineering design	447	5	0.028	7	2	Dead without resurgence
461	artificial intelligence	457	7	0.058	11	4	Dead without resurgence
467	knowledge discovery	453	8	0.6	n/a	12	Alive with resurgence
469	association rule mining	467	9	0.149	n/a	17	Always alive
471	data mining	461	11	0.178	n/a	13	Alive with resurgence
472	virtual reality	461	11	0.082	21	10	Dead without resurgence
478	linguistics	471	15	0.065	20	5	Dead without resurgence
480	decision tree	471	15	0.037	17	2	Dead without resurgence
481	rule base system	471	15	0.123	n/a	7	Alive with resurgence
484	recommender system	453	15	0.145	n/a	9	Alive with resurgence
485	security system	484	16	0.079	n/a	10	Always alive
486	machine learning	447	16	0.188	n/a	10	Always alive
487	Bayesian analysis	480	17	0.056	21	4	Dead without resurgence
489	outlier detection	486	17	0.026	n/a	9	Always alive
492	optimization	455	17	0.247	n/a	9	Always alive
493	robustness	486	17	0.058	21	4	Dead without resurgence
494	classification	480	17	0.264	n/a	9	Always alive
496	information system	481	17	0.017	20	3	Dead without resurgence
501	decision making	492	19	0.200	n/a	7	Always alive
505	neural network	492	19	0.058	n/a	7	Always alive
507	attribute reduction algorithms	494	19	0.074	n/a	7	Always alive
510	product design	262	20	0.078	n/a	6	Always alive
512	semantic web	478	20	0.046	21	1	Dead when born
513	uncertainty	487	20	0.161	n/a	6	Always alive
519	fuzzy logic	501	20	0.072	n/a	6	Always alive
523	computational intelligence	519	21	0.091	n/a	5	Always alive
526	cluster analysis	484	21	0.101	n/a	5	Always alive
531	fuzzy preference relations	519	21	0.004	23	2	Dead without resurgence
532	forecasting	487	21	0.027	23	2	Dead without resurgence
534	knowledge management system	262	21	0.054	n/a	5	Always alive
537	prediction model	493	21	0.109	n/a	5	Always alive
539	statistical analysis	487	21	0.101	n/a	5	Always alive
541	ontologies	512	21	0.069	n/a	5	Always alive
544	particle swarm optimization	523	22	0.074	n/a	4	Always alive
554	multiple attribute group decision making	501	22	0.054	n/a	4	Always alive
557	spatial information	537	22	0.055	n/a	4	Always alive
558	intuitionistic fuzzy set	519	22	0.052	n/a	4	Always alive
563	financial decision making	501	22	0.049	n/a	4	Always alive
568	stock market	532	23	0.056	n/a	3	Always alive
575	emotion recognition	539	23	0.018	n/a	2	Alive with resurgence
579	probabilistic model	513	23	0.082	n/a	3	Always alive
583	topic model	532	23	0.054	n/a	3	Always alive
588	group decision making	501	24	0.014	n/a	2	Always alive
596	social network	526	24	0.035	n/a	2	Always alive
597	support vector machine	494	24	0.004	n/a	1	Alive with resurgence
600	incremental algorithm	486	24	0.010	n/a	1	Alive with resurgence

Note. TS = Time Slice; Surv. Len. = Survival Length.

their descendants), and obtained the top 50 largest linkages between topics as undirected arcs. We imported information about the nodes and the arcs to Gephi [52], and generated the topic evolutionary pathways of KnoSys, given in Fig. 5.

Three research directions of KnoSys are observed and summarized in Fig. 5, and we discuss them in detail as follows:

*Expert systems* (including Knowledge-based Systems): Undoubtedly, this is one of the main foci of KnoSys, and its scope has been extended to a broad range of real-world problem-driven applications. On one hand, new practical areas keep appearing, such as engineering design, product design, security systems, and virtual reality [53–55]. Although these directions did not form relatively continuous clusters, it demonstrates well the target of KnoSys in pursuing novel applications of Knowledge-based Systems. On the other hand, novel concepts, methods, and algorithms have been widely engaged with traditional Knowledge-based Systems, and these novel techniques cover a large number of areas in the field

of computer science, e.g., artificial intelligence, pattern recognition, information retrieval, and knowledge discovery [33,56]. In addition, recommender systems, which are evolved from intelligent systems, can be considered as a highlight of KnoSys in recent years, and it further connects with the topic of social network, a hotspot associated with e-service personalization these days [27]. In particular, despite no closely-related nodes are shown in Fig. 5, medical health care systems (e.g., diagnosis, decision making, and prediction) have also brought substantial attention from different communities [57], and become an emergent interest in the area of expert systems.

*Decision making and computational intelligence*: Decision making has become one of the most important topics in KnoSys, which began from the topics of human computer interaction and optimization, and is currently closely related to computational intelligence. Related studies can be summarized as follows:



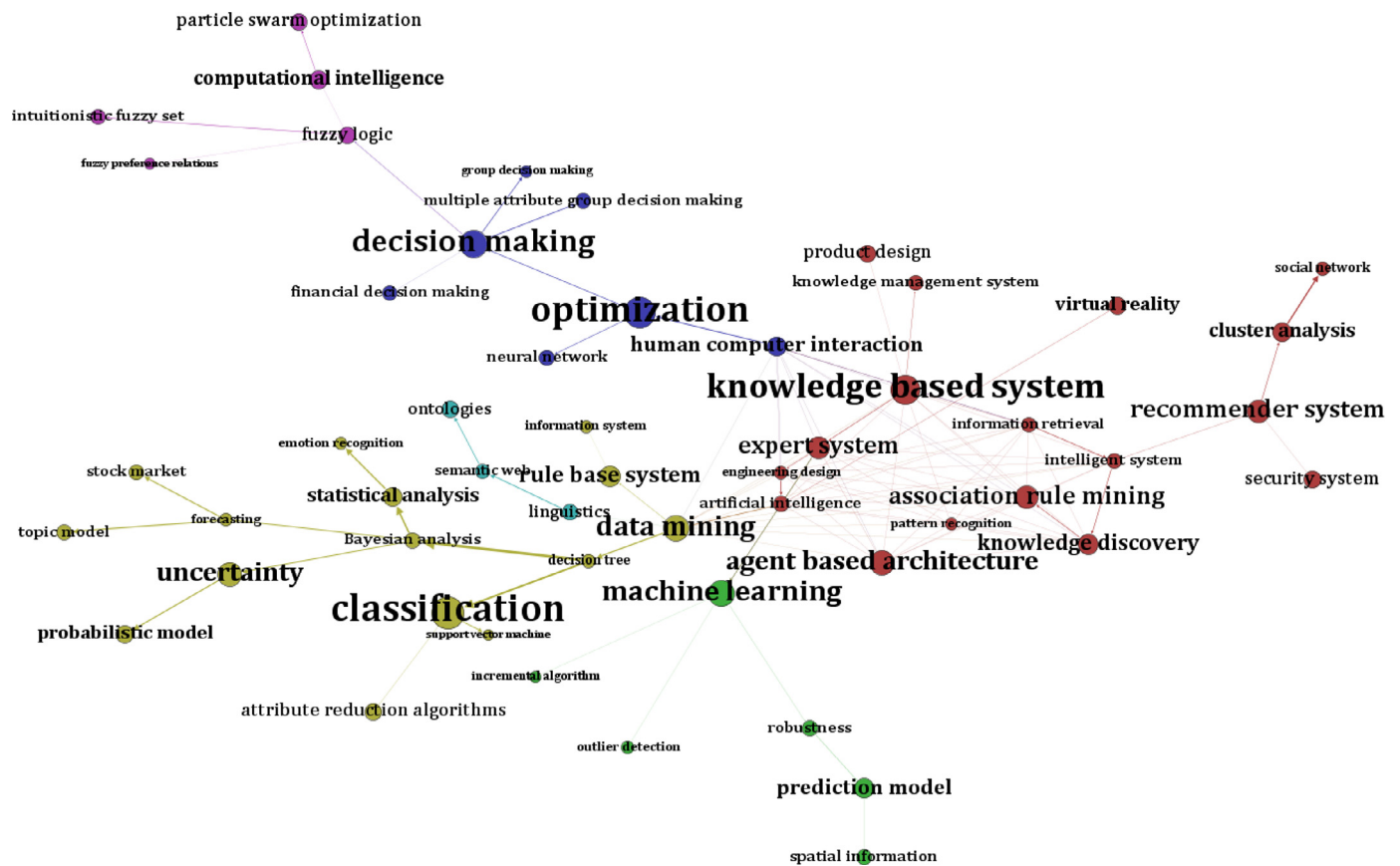


Fig. 5. Top of evolutionary pathways of KnoSys from 1991 to 2016.

- Studies oriented to diverse decision-making situations (e.g., group decision making, multi-attribute decision making, multi-criteria decision making, and three-way decision making) and actual decision-making applications (e.g., financial decision making and strategic decision making) - as a combination of decision making and an expert system, decision support systems may be a mainstream direction in KnoSys [58,59];
- The engagement of fuzzy logic, including intuitionistic fuzzy sets, interval valued fuzzy sets, and fuzzy preference relations - fuzzy logic has been widely used as a tool in a range of studies in artificial intelligence and information systems [21,60];
- The theoretical and practical studies of computational intelligence and optimization (including swarm intelligence and evolutionary computing) - this direction closely interacts with both decision making and artificial intelligence, and certain emergent topics also appear in this area, such as data analytics and business intelligence [25,61].

*Artificial intelligence* (including machine learning and data mining): despite the fact that two routes can be traced in Fig. 5 for machine learning and data mining respectively, we group them as one direction because these techniques act as basic tools and have been highly involved with the studies of the other two directions. Four main topics can be highlighted here:

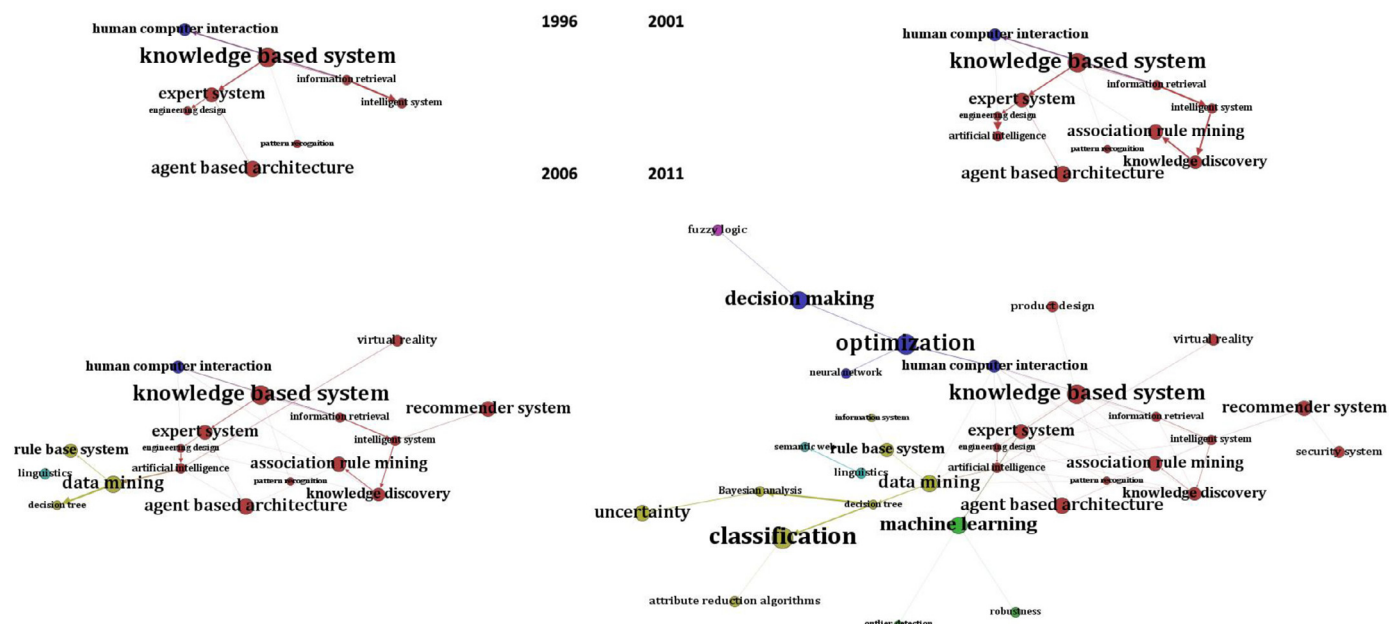
- Classification - the scope of data mining has been widely extended in the big data age, and data analytics or data science have become a more popular name for this field. As a traditional research question in data mining, new techniques are continuously involved in the task of classification, e.g., multi-class classification [62] and ensemble learning [48], and classification-related studies are playing a proficient role in KnoSys these days.

- Prediction model - oriented to actual issues in medical, social, or other services, using knowledge to develop prediction models has become an emerging topic in KnoSys [29,50]. Such studies highly relate to machine learning techniques and can be dated to case-based reasoning and logic programming-related articles published in the first decade of KnoSys. Enhanced classification techniques are also required here.
- Uncertain information processes - a broad range of analytic models have been introduced to handle the issue of uncertainty, including Bayesian analysis and some other probabilistic models [63], and one interesting actual application of such study is to forecast stock markets, which can be a hot case in KnoSys [28];
- Ontologies - despite not having a huge cluster in KnoSys, it is still necessary to trace the route of semantic web and ontologies as shown Fig. 5, which is definitely within the scope of artificial intelligence [26,56].

Fig. 5 provides a bird's eye view of the topic changes in KnoSys from 1991 to 2016, and aims to better detect such dynamics. We further divided the 26 time slices into five time periods (i.e., 1991 to 1996, 1997 to 2001, 2002 to 2007, 2008 to 2011, and 2012 to 2016) and present related topics in first four segments, given in Fig. 6 - the fifth segment can be viewed in Fig. 5. Considering the time dimension of the 53 topics, we can easily detect the topic changes in KnoSys by the following trends from 1991 to 2016:

Period 1 (1991 - 1996): Expert systems and related topics dominated KnoSys in this period. Related sub-topics included intelligent systems, agent-based architecture, and human computer interaction, and topic information retrieval and pattern recognition can be considered as two parallel emphases at that time;





**Fig. 6.** Topic evolutionary pathways of KnoSys in four time periods.

Note. The appearance of a new topic does not mean that its content has never been in KnoSys before, but it highlights the situation that the content of this topic has become more and more important and is able to attract the interest of a given research community.

Period 2 (1997 – 2001): The branch of information retrieval was largely enriched, and the main foci of this period were closely related to knowledge discovery and representation. In particular, the topic artificial intelligence appeared, despite it not being a dominant topic in this period;

Period 3 (2002 – 2006): The most significant change during this period was the growth of studies in data mining and rule-based systems. In parallel, the topic recommender systems were generated from the group of expert systems, and it could be another hotspot at that time;

Period 4 (2007 – 2011): In these five years, the topic diversity of KnoSys dramatically increased. On one hand, studies in data mining and machine learning boomed and reached a peak; on the other hand, the topics of decision making and optimization appeared, indicating that KnoSys had successfully involved research communities in these two areas;

Period 5 (2011 – 2016): The topics of artificial intelligence and decision making grew rapidly in the most recent five years, and related research evolved into two directions: 1) fuzzy logic and computational intelligence, and 2) information uncertainty-oriented studies, e.g., probabilistic models and prediction models. In addition, the topics of social network (aligned with recommender systems) and ontologies (the descendants of semantic web) might indicate certain emergent interests of KnoSys communities.

#### 4.4. Evolutionary trends of six hotspots in KnoSys

It is clear that Figs. 5 and 6 show the detailed topic changes of KnoSys from 1991 to 2016, and highlight certain main research directions and emergent groups in diverse time periods. Based on this knowledge, we selected six hotspots (expert systems, data mining, decision making, fuzzy logic, machine learning, and optimization), and following the criteria given in Table 5 to collect related core technological terms.

Based on the term frequency of the collected terms, we draw on the evolutionary trends of the six hotspots in a cross-comparative way, as shown in Fig. 7. Certain trends can be observed: 1) the hotspot expert system is in a downtrend; 2) the hotspot optimization is slowly growing but has been stable since 2014; 3) fuzzy-related studies increased rapidly and reached a peak around 2012, but this trend has slowed since 2014; 4) the hotspot decision making has entered a phase of rapid growth since 2006, and it is one of the largest topics in 2015; and 5) the division of data mining and machine learning might weaken the competitiveness of artificial intelligence, which might occupy the other half of KnoSys. The hotspot of machine learning is the most competitive topic in KnoSys. Its growth in particular accelerated after 2009, while the hotspot of data mining is relatively stable, maintaining a slow increase.

#### 5. Predicting the trend of six hotspots in KnoSys

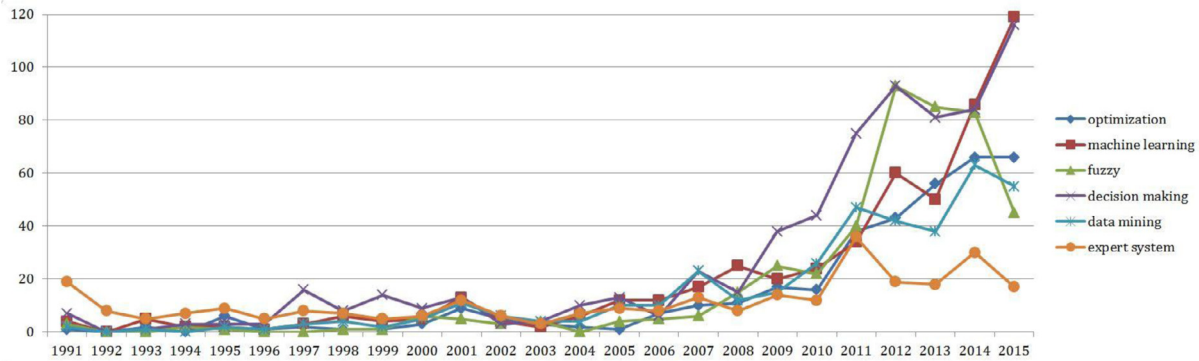
When addressing concerns of the two questions “What is the KnoSys community interested in?” and “How does such interest change over the time?” in the above sections, we push this study forward by predicting future trends of the six selected hotspots,<sup>7</sup> i.e., expert systems, data mining, decision making, fuzzy, machine learning, and optimization. A prediction model was developed by calculating the average proportion of related topics in all articles and applying a linear fit approach to estimate their future trends.

Aiming to measure the reliability of the prediction model, we first applied the model to a set of historical data from 1991 to 2013 to predict topic trends in the three forthcoming years (i.e., from 2014 to 2016), and then the real topic trend from 2014 to 2016 was used to validate the performance of our prediction. In traditional bibliometrics, the proportion that the number of articles related to one area occupies in the entire corpus is widely applied

<sup>7</sup> Reflected in the results of the LDA-based topic model, the six hotspots were matched with eight topics, i.e., T09-eSys, T20-KBS, T22-dMin, T19-dMak, T21-fuzz, T10-clas, T02-lAlg and T05-optim. Each topic represented one hotspot, except 1) Topics T09-eSys and T20-KBS were accessed as a package for expert systems, and 2) topics T10-clas and T02-lAlg were used to represent the hotspot machine learning.

**Table 5**  
Criteria of collecting relevant terms of the six hotspots.

No.	Hotspot	Criteria
1	Expert systems	expert system OR intelligent system OR information system OR multi agent system OR knowledge based systems OR diagnosis system OR warning system
2	Data mining	*mining* OR data driven OR data analysis OR big data
3	Decision making	*decision*
4	Fuzzy	*fuzzy*
5	Machine learning	learning OR learner
6	Optimization	*optim*



**Fig. 7.** Evolutionary trends of six hotspots in KnoSys.

Note. The Y axis is the total term frequency of related hotspots. The data of the year 2016 is not complete, which is due to September 30, 2016.

as an indicator, and here, we estimated the dynamics of the proportion along a time line as the reference of the real topic trend. Thus, the real topic trends of the six selected hotspots are given in Fig. 8(a), while Fig. 8(b) presents our predicted results. It is clear that both the four increasing trends (i.e., decision making, fuzzy, machine learning and optimization) and the two downward trends (i.e., expert system and data mining) are well-matched. Under this circumstance, it is reasonable to consider that the performance of the prediction model is reliable.

We then perform trend prediction using the data from year 1991 to 2016 for the 6 target topics. As shown in Fig. 9, the work on expert systems is decreasing significantly year by year, indicating the digressive popularity of this topic in KnoSys. However, as we discussed in Section 4, one persuasive reason can be the research of expert systems has been extensively developed and related terms have developed into some specific ones, e.g., recommender systems and decision support systems. A similar situation has occurred in data mining, but it is comparably relatively steady, with only a slight downward trend. One insight here is with the big data boom. Data mining-related techniques mostly described as data analytics, and the applications of data mining, especially those within the business and industry sectors have led to the name business intelligence increasing in popularity. Under this circumstance, we still consider the two hotspots have downward trends. In contrast, hotspots machine learning and fuzzy logic have dramatically increased, while decision making and optimization show slight upward trends. Computational intelligence [64], the combination of artificial neural networks, fuzzy systems, and evolutionary computation (including machine learning and probabilistic methods) has received impressive accomplishment in handling complex real-world problems. Undoubtedly, this is an emerging trend for KnoSys communities, and related studies will be widely explored in KnoSys [25,57,65–67].

## 6. Discussion and implication

In this section, we summarize the key findings of our study and conduct a comparative study with a bibliometric study given

for the 25-year ceremony of KnoSys [3]. An extensive discussion is given to indicate the possible benefits of our research to a broad range of audiences and fields.

### 6.1. Key findings

As an overview, KnoSys hotspots cover six main research areas, i.e., expert systems, machine learning, data mining, decision making, optimization, and fuzzy. As indicated in Figs. 7 and 9, the interests of KnoSys communities in the area of computational intelligence (including studies in machine learning, fuzzy, decision making, optimization, etc.) are highlighted. In particular, as observed from Figs. 5 and 6, the recombination of existing techniques to address real-world problems has become an emergent interest, which has well-established extensive interactions with expert system-related studies in KnoSys. The ability to construct better avenues for systems that provide social, medical, or other services through knowledge use and accurate prediction models is emphasized, where advanced data analytics and machine learning techniques are highly involved.

Regarding the 25-year time span and the topic change of KnoSys visualized in Figs. 5, 6 and 9, we outline certain key findings as follows: 1) expert systems are the only emphasis from 1991 to 2001, when information retrieval and knowledge discovery were two highlights; 2) increasing interest in machine learning, optimization and data mining appeared gradually from 2002 to 2011, and recommender systems became one of the most representative directions in the area of expert systems; 3) the six main research areas have been established and developed to a relatively mature phase in the second decade of KnoSys; 4) some emerging topics of KnoSys are identified, e.g., computational intelligence, ontology, prediction model and uncertainty processing, and social network analysis.

### 6.2. Comparative study

Since the first issue published in 1991, KnoSys has experienced a period of extensive and fruitful development. For the 25-year cer-

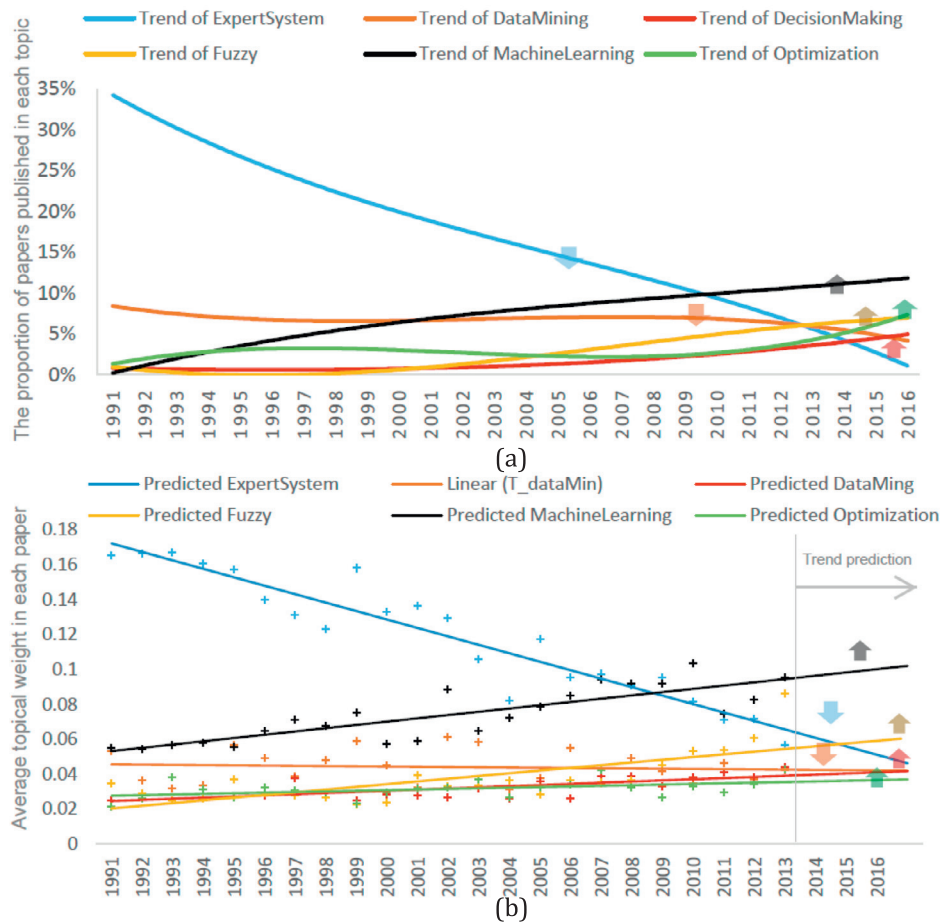


Fig. 8. Validation measure of the prediction model: (a) presents the real topic trends of the six selected hotspots; (b) presents the topic trends predicted by the prediction model.

emony, an in-depth bibliometric study was conducted in 2015 [3], and the comparison between our study and [3] will not only help discover extensive knowledge hidden behind the development of KnoSys, but also provide comparable evidence to demonstrate the reliability of our study.

- Breath (the engagement of multiple bibliometric indicators) – our study only uses words/terms and concentrates on the analysis of topics, while a number of extra bibliometric indicators are involved in [3], e.g., the number of articles, citation statistics, impact factor, h-index, the productivity of authors, and geographic distribution of articles. Based on these indicators, [3] conducted an overall evaluation on the development of KnoSys from multiple perspectives.
- Depth (the foci on topic analyses and science maps) – [3] developed an approach to detect and visualize the conceptual sub-domains of a research field, and the strategic diagrams and thematic evolution of KnoSys were generated. The impression includes: 1) a co-word analysis and the h-index was combined to identify the strategic features of themes (e.g., “highly developed”, “emerging or declining”, “motor”, and “basic and transversal”). 2) A technique of similarity measures was used to identify the evolutionary relationships between topics in different periods. Comparably, our study profiled topics and evaluated their popularity from a probabilistic perspective, and the detection of “sleeping beauties” specified the insights on the emerging importance of topics. Further, machine learning techniques were used to explore the evolutionary relationships (i.e.,

predecessor and descendants) by handling the issue of detecting the gradual change of topics in sequential time slices.

- Results – 1) compared with the eight great thematic areas given in [3], i.e., classification, data mining, rough sets, decision support systems, agent systems, information retrieval, group decision making, and semantic web, the six main research areas identified in our study cover all of them, despite some crossovers. 2) Considering the increasing/declining importance of topics, our observations were based on the dynamics among five time periods while [3] divided the entire 25-year dataset into two time periods. However, our findings are still the same in most parts, e.g., we highlight the growing interest of KnoSys communities to areas such as classification/machine learning, rough sets/fuzzy, and decision making, and the shrinking interest to areas such as information retrieval and semantic web. 3) It is also necessary to be aware that although the both studies use terms to label topics/themes, the granularity of topics in our study and that of themes in [3] are different sometimes, e.g., we grouped decision support system and group decision making to the topic decision making while [3] discussed the two themes separately, and we discussed the topic of data mining by involving classification-related studies but in particular set machine learning as an relatively independent topic.

Despite in-depth overlapping, our study and [3] analyzed articles published in KnoSys from diverse perspectives and with diverse analysis criteria, but the meaningful insights explored by the both studies demonstrate the great power of bibliometrics in the studies of research evaluation and performance analysis.





**Fig. 9.** Trend prediction of six hotspots in KnoSys.

Note. Subfigure (a-f) presents the trend prediction of topics expert systems, data mining, decision making, fuzzy, machine learning, and optimization respectively.

### 6.3. Benefits

This paper provides a solution to detect and predict the change of research topics in KnoSys, and the empirical insights explored from this case can benefit the KnoSys community (including both the audiences and editorial board of KnoSys) from the following aspects:

The empirical insights provide a systematic overview to track the trend of KnoSys topics and uncover related research frontiers, with the answers of questions such as “what are happening in this research area” and “what will happen in the near future”. Such multi-dimensional knowledge will help the audiences align with the majority of the research community and also quickly capture the core of related areas.

Our study addresses the concerns held by the editorial board of KnoSys and research administrations in related universities and institutions, e.g., how to evaluate the performance of related research at a macro-level, how to allocate or re-allocate resources to

researchers in diverse disciplines and research areas, and how to utilize such objective evidence to support research and development plan.

Our methodology can be used as a tool of bibliometric analysis to adapt to a broad range of actual requirements, e.g., profiling scientific articles in other disciplines, journals, or conference proceedings, and identifying competitive technical intelligence from other scientific documents such as patents, academic proposals, or technical reports. These efforts would then benefit studies in science policy, technology and innovation management, and entrepreneurship.

### 7. Conclusions

This paper specifically focuses on the change of these research topics in KnoSys by conducting a topic-based bibliometric study to address two questions “What is the KnoSys community interested in?” and “How does such interest change over the time?”.

An LDA-based model was used to identify the hotspots and their corresponding relations of KnoSys in a probabilistic way, and a SEP model introduced a learning process to detect topic changes in KnoSys from 1991 to 2016. The prediction model forecast possible future trends for selected hotspots by tracing the average proportion of topics in all articles over time. Generally, six main research areas of KnoSys are identified, i.e., expert systems, machine learning, data mining, decision making, optimization, and fuzzy, and the results also indicate that the interest of KnoSys communities in the area of computational intelligence is raised, and the ability to construct practical systems through knowledge use and accurate prediction models is highly emphasized.

The limitations and related further studies are summarized: 1) considering the complexity of semantic structures (e.g., words, terms, and phrases), the performance of topic analysis usually depends highly on the magnitude of data pre-processing of noise removal and synonym consolidation. Benefiting from the aid of experts, our study performed well and achieved its targets. However, an intelligent algorithm (e.g., word embedding, semantic web and ontology techniques) to accurately represent semantic structures may help further improve the performance and robustness of our methodology; and 2) compared with topics derived from terms, citation statistics is the other mainstream indicator in bibliometrics, and engagement of citation analysis may provide extra insights in this study.

## References

- [1] J. Allan, *Topic Detection and Tracking: Event-Based Information Organization*, Springer, US, 2002.
- [2] W. Ding, C. Chen, Dynamic topic detection and tracking: a comparison of HDP, C-word, and cocitation methods, *J. Assoc. Inf. Sci. Technol.* 65 (2014) 2084–2097.
- [3] M.J. Cobo, M. Martínez, M. Gutiérrez-Salcedo, H. Fujita, E. Herrera-Viedma, 25years at knowledge-based systems: a bibliometric analysis, *Knowl.-Based Syst.* 80 (2015) 3–13.
- [4] A.H. Goodall, Highly cited leaders and the performance of research universities, *Res. Policy* 38 (2009) 1079–1092.
- [5] H. Small, K.W. Boyack, R. Klavans, Identifying emerging topics in science and technology, *Res. Policy* 43 (2014) 1450–1467.
- [6] A. Suominen, H. Toivanen, Map of science with topic modeling: comparison of unsupervised learning and human-assigned subject classification, *J. Assoc. Inf. Sci. Technol.* 67 (2016) 2464–2476.
- [7] E. Garfield, Historiographic mapping of knowledge domains literature, *J. Inf. Sci.* 30 (2004) 119–145.
- [8] J. Yoon, K. Kim, TrendPerceptor: a property-function based technology intelligence system for identifying technology trends from patents, *Expert Syst. Appl.* 39 (2012) 2927–2938.
- [9] Y. Zhang, G. Zhang, H. Chen, A.L. Porter, D. Zhu, J. Lu, Topic analysis and forecasting for science, technology and innovation: methodology and a case study focusing on big data research, *Technol. Forecast. Social Change* 105 (2016) 179–191.
- [10] H. Chen, G. Zhang, Y. Zhang, J. Lu, D. Zhu, Identifying topic relations in scientific literature using topic modeling, *J. Assoc. Inf. Sci. Technol.* (2016) under review.
- [11] Y. Zhang, G. Zhang, D. Zhu, J. Lu, Science evolutionary pathways: identifying and visualizing relationships for scientific topics, *J. Assoc. Inf. Sci. Technol.* (2016) accepted.
- [12] Y. Zhang, A.L. Porter, Z. Hu, Y. Guo, N.C. Newman, Term clumping” for technical intelligence: a case study on dye-sensitized solar cells, *Technol. Forecast. Social Change* 85 (2014) 26–39.
- [13] Y. Zhang, X. Zhou, A.L. Porter, J.M.V. Gomila, How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: “Problem & Solution” pattern based semantic TRIZ tool and case study, *Scientometrics* 101 (2014) 1375–1389.
- [14] L. Waltman, N.J. van Eck, E.C. Noyons, A unified approach to mapping and clustering of bibliometric networks, *J. Informetrics* 4 (2010) 629–635.
- [15] D.M. Blei, J.D. Lafferty, Dynamic topic models, in: *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 113–120.
- [16] J. Xuan, J. Lu, G. Zhang, R.Y. Da Xu, X. Luo, Doubly nonparametric sparse non-negative matrix factorization based on dependent Indian buffet processes, *IEEE Trans. Neural Networks Learn. Syst.* (2017).
- [17] F. De Battisti, A. Ferrara, S. Salini, A decade of research in statistics: a topic model approach, *Scientometrics* 103 (2015) 413–433.
- [18] M. Hall, A decision tree-based attribute weighting filter for naive Bayes, *Knowl.-Based Syst.* 20 (2007) 120–126.
- [19] W. Zhang, T. Yoshida, X. Tang, Text classification based on multi-word with support vector machine, *Knowl.-Based Syst.* 21 (2008) 879–886.
- [20] F.J. Cabrerizo, I.J. Pérez, E. Herrera-Viedma, Managing the consensus in group decision making in an unbalanced fuzzy linguistic context with incomplete information, *Knowl.-Based Syst.* 23 (2010) 169–181.
- [21] G. Wei, Hesitant fuzzy prioritized operators and their application to multiple attribute decision making, *Knowl.-Based Syst.* 31 (2012) 176–182.
- [22] K.L. Choy, W. Lee, H.C. Lau, L. Choy, A knowledge-based supplier intelligence retrieval system for outsource manufacturing, *Knowl.-Based Syst.* 18 (2005) 1–17.
- [23] P. Gervás, An expert system for the composition of formal spanish poetry, *Knowl.-Based Syst.* 14 (2001) 181–188.
- [24] J. Wu, F. Chiclana, A social network analysis trust-consensus based approach to group decision-making problems with interval-valued fuzzy reciprocal preference relations, *Knowl.-Based Syst.* 59 (2014) 97–107.
- [25] H.-Z. Li, S. Guo, C.-J. Li, J.-Q. Sun, A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm, *Knowl.-Based Syst.* 37 (2013) 378–387.
- [26] D. Sánchez, M. Batet, D. Isern, Ontology-based information content computation, *Knowl.-Based Syst.* 24 (2011) 297–303.
- [27] J. Bobadilla, F. Ortega, A. Hernando, A. Gutiérrez, Recommender systems survey, *Knowl.-Based Syst.* 46 (2013) 109–132.
- [28] E. Hadavandi, H. Shavandi, A. Ghanbari, Integration of genetic fuzzy systems and artificial neural networks for stock price forecasting, *Knowl.-Based Syst.* 23 (2010) 800–808.
- [29] C.-F. Tsai, Feature selection in bankruptcy prediction, *Knowl.-Based Syst.* 22 (2009) 120–127.
- [30] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, et al., Circos: an information aesthetic for comparative genomics, *Genome Res.* 19 (2009) 1639–1645.
- [31] S.-K. Oh, W. Pedrycz, H.-S. Park, Rule-based multi-FNN identification with the aid of evolutionary fuzzy granulation, *Knowl.-Based Syst.* 17 (2004) 1–13.
- [32] B. Yu, Z.-b. Xu, A comparative study for content-based dynamic spam classification using four machine learning algorithms, *Knowl.-Based Syst.* 21 (2008) 355–362.
- [33] M. Gaeta, F. Orciuoli, P. Ritrovato, Advanced ontology management system for personalised e-learning, *Knowl.-Based Syst.* 22 (2009) 292–301.
- [34] P. Shapira, J. Wang, Follow the money, *Nature* 468 (2010) 627–628.
- [35] H. Small, Visualizing science by citation mapping, *J. Am. Soc. Inf. Sci.* 50 (1999) 799–813.
- [36] H. Small, Co-citation in the scientific literature: A new measure of the relationship between two documents, *J. Am. Soc. Inf. Sci.* 24 (1973) 265–269.
- [37] M. Callon, J.-P. Courtial, W.A. Turner, S. Bauin, From translations to problematic networks: an introduction to co-word analysis, *Social Science Information* 2 (1983) 191–235.
- [38] M.M. Kessler, Bibliographic coupling between scientific papers, *Am. Documentation* 14 (1963) 10–25.
- [39] M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, F. Herrera, Science mapping software tools: review, analysis, and cooperative study among tools, *J. Am. Soc. Inf. Sci. Technol.* 62 (2011) 1382–1402.
- [40] R. Klavans, K.W. Boyack, Toward a consensus map of science, *J. Am. Soc. Inf. Sci. Technol.* 60 (2009) 455–476.
- [41] S.L. Scott, Multi-armed bandit experiments in the online service economy, *Appl. Stochastic Models Bus. Ind.* 31 (2015) 37–45.
- [42] E. Yan, Y. Ding, Applying centrality measures to impact analysis: A coauthorship network analysis, *J. Assoc. Inf. Sci. Technol.* 60 (2009) 2107–2118.
- [43] A.F. van Raan, Sleeping beauties in science, *Scientometrics* 59 (2004) 467–472.
- [44] G. Salton, M.J. McGill, *Introduction to Modern Information Retrieval*, McGraw-Hill, Auckland, 1986.
- [45] L. Leydesdorff, On the normalization and visualization of author co-citation data: Salton’s Cosine versus the Jaccard index, *J. Am. Soc. Inf. Sci. Technol.* 59 (2008) 77–85.
- [46] M.G. Moehrl, Measures for textual patent similarities: A guided way to select appropriate approaches, *Scientometrics* 85 (2010) 95–109.
- [47] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, *Inf. Process. Manage.* 24 (1988) 513–523.
- [48] J. Hu, T. Li, H. Wang, H. Fujita, Hierarchical cluster ensemble model based on knowledge granulation, *Knowl.-Based Syst.* 91 (2016) 179–188.
- [49] W.-T. Pan, A new fruit fly optimization algorithm: taking the financial distress model as an example, *Knowl.-Based Syst.* 26 (2012) 69–74.
- [50] J. Sun, H. Li, Data mining method for listed companies’ financial distress prediction, *Knowl.-Based Syst.* 21 (2008) 1–5.
- [51] J. Ma, G. Zhang, J. Lu, A state-based knowledge representation approach for information logical inconsistency detection in warning systems, *Knowl.-Based Syst.* 23 (2010) 125–131.
- [52] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks, in: *Proceedings of International AAAI Conference on Web and Social Media*, 8, 2009, pp. 361–362.
- [53] C.B. Chapman, M. Pinfold, Design engineering—a need to rethink the solution using knowledge based engineering, *Knowl.-Based Syst.* 12 (1999) 257–267.
- [54] R.M. Moraes, A.V. Rocha, L.S. Machado, Intelligent assessment based on beta regression for realistic training on medical simulators, *Knowl.-Based Syst.* 32 (2012) 3–8.
- [55] B. Yang, L.X. Li, H. Ji, J. Xu, An early warning system for loan risk assessment using artificial neural networks, *Knowl.-Based Syst.* 14 (2001) 303–306.

- [56] R.J. Gil, M.J. Martin-Bautista, A novel integrated knowledge support system based on ontology learning: model specification and a case study, *Knowl.-Based Syst.* 36 (2012) 340–352.
- [57] N.T. Thong, Intuitionistic fuzzy recommender systems: an effective tool for medical diagnosis, *Knowl.-Based Syst.* 74 (2015) 133–150.
- [58] J. Ma, J. Lu, G. Zhang, Decider: a fuzzy multi-criteria group decision support system, *Knowl.-Based Syst.* 23 (2010) 23–31.
- [59] R. Urena, F. Chiclana, H. Fujita, E. Herrera-Viedma, Confidence-consistency driven group decision making approach with incomplete reciprocal intuitionistic preference relations, *Knowl.-Based Syst.* 89 (2015) 86–96.
- [60] N. Chen, Z. Xu, M. Xia, Interval-valued hesitant preference relations and their applications to group decision making, *Knowl.-Based Syst.* 37 (2013) 528–540.
- [61] W. Shen, X. Guo, C. Wu, D. Wu, Forecasting stock indices using radial basis function neural networks optimized by artificial fish swarm algorithm, *Knowl.-Based Syst.* 24 (2011) 378–385.
- [62] D. Tomar, S. Agarwal, A comparison on multi-class classification methods based on least squares twin support vector machine, *Knowl.-Based Syst.* 81 (2015) 131–147.
- [63] N. Fenton, M. Neil, Making decisions: using Bayesian nets and MCDA, *Knowl.-Based Syst.* 14 (2001) 307–325.
- [64] A.P. Engelbrecht, *Computational Intelligence: An Introduction*, John Wiley & Sons, 2007.
- [65] J. Lu, V. Behbood, P. Hao, H. Zuo, S. Xue, G. Zhang, Transfer learning using computational intelligence: a survey, *Knowl.-Based Syst.* 80 (2015) 14–23.
- [66] A. Fernandez, V. Lopez, M.J. del Jesus, F. Herrera, Revisiting evolutionary fuzzy systems: taxonomy, applications, new trends and challenges, *Knowl.-Based Syst.* 80 (2015) 109–121.
- [67] S. Salehi, A. Selamat, M.R. Mashinchi, H. Fujita, The synergistic combination of particle swarm optimization and fuzzy sets to design granular classifier, *Knowl.-Based Syst.* 76 (2015) 200–218.