zhang_2017_detecting_and_predicting_the_topic_change_of_knowledge_based_systems_a_topic_based_bibliometric_analysis_from_1991_to_2016

Year

2017

Author(s)

Yi Zhang and Hongshu Chen and Jie Lu and Guangquan Zhang

Title

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

Venue

Knowledge-Based Systems

Topic labeling

Manual

Focus

Secondary

Type of contribution

Established approach

Underlying technique

Manual labeling

Topic labeling parameters

\

Label generation

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.

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.

The rank indicates the popularity of topics (i.e., hotspots), which is calculated by the total proportion of a topic in all articles.

Motivation

\

Topic modeling

LDA

Topic modeling parameters

α: 0.5
β: 0.01
Iterations of Gibbs sampling: 5000
Nr of topics (k): 25

Nr. of topics

25

Label

Manually selected representative word or word combination

Label selection

\

Label quality evaluation

\

Assessors

\

Domain

Paper: Bibliometrics
Dataset: Computer science and AI

Problem statement

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 (research areas) 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.

Corpus

Origin: Knowledge-Based Systems (WoS database)
Nr. of documents: 2657
Details:

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

Document

Article extracted from the WoS DB and published in the Knowledge-Based Systems journal

Pre-processing

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 [a], e.g., "classification" and "classification analysis"	29,151
6	Rule-based cleaning - to remove single words [b], e.g., "information" and "feasibility"	27,065
7	Rule-based cleaning - to remove terms appearing in only one article [c]	4265
8	Rule-based consolidation - to consolidate terms sharing more than 3 sequential words [d], 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".

```
@article{zhang_2017_detecting_and_predicting_the_topic_change_of_knowledge_base
d_systems_a_topic_based_bibliometric_analysis_from_1991_to_2016,
  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.},
  author = {Yi Zhang and Hongshu Chen and Jie Lu and Guangquan Zhang},
  date-added = {2023-03-17 14:59:15 +0100},
  date-modified = {2023-03-17 14:59:15 +0100},
  doi = {https://doi.org/10.1016/j.knosys.2017.07.011},
  issn = {0950-7051},
  journal = {Knowledge-Based Systems},
  keywords = {Topic analysis, Topic detection and tracking, Bibliometrics, Text mining, Knowledge-based Systems},
  pages = {255-268},
  title = {Detecting and predicting the topic change of Knowledge-based Systems: A topic-based bibliometric analysis from 1991 to 2016},
  url = {https://www.sciencedirect.com/science/article/pii/S0950705117303271},
  volume = {133},
  year = {2017}}

#Thesis/Papers/Initial