Trend Analysis of Machine Learning Research Using Topic Network Analysis

Deepak Sharma[1(✉)] , Bijendra Kumar[1], and Satish Chand[2]

[1] Department of Computer Engineering, Netaji Subash Institute of Technology, New Delhi, India
`deepak.btg@gmail.com, bizender@gmail.com`
[2] School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India
`schand20@gmail.com`

Abstract. In this paper, a topic network analysis approach is proposed which integrates topic modeling and social network analysis. We collected 16,855 scientific papers from six top journals in the field of machine learning published from 1997 to 2016 and analyzed them with the topic network. The dataset is break down into 4 intervals to identify topic trends and performed the time-series analysis of topic network. Our experimental results show centralization of the topic network has the highest score from 2002 to 2006, and decreases for next 5 years and increases again. For last 5 years, centralization of the degree centrality and closeness centrality increases, while centralization of the betweenness centrality decreases again. Also, data analytic and computer vision are identified as the most interrelated topic among other topics. Topics with the highest degree centrality evolve component analysis, text mining, biometric and computer vision according to time. Our approach extracts the interrelationships of topics, which cannot be detected with conventional topic modeling approaches, and provides topical trends of machine learning research.

Keywords: Topic network analysis · Social network analysis · Topic modeling
Latent Dirichlet Allocation · Research trend analysis

1   Introduction

As the Big Data interest has exploded recently, machine learning techniques, which are key technologies for big data analysis, are getting attention by data scientist. Machine learning is an interdisciplinary area of research in various disciplines such as statistics, artificial intelligence, and databases, has been studied steadily since the word 'machine learning' coined by Arthur Samuel. In particular, since last two decades, existing machine learning techniques have been applied to large-scale data processing environments or have been extended to various application areas such as fraud detection, stock market, weather forecasting etc. Also, the algorithm is changed according to newly emerged technology. So understanding the machine learning research themes of the past two decades will help, studying current machine learning trends and applies it to practical applications.

Text mining techniques have been commonly used to automatically classify documents [1, 2]. In particular, topic modeling [3] used to understand trend analysis in research from research papers and its many subjects of interest by applying the transform algorithm i.e. hot topic and less interested subject i.e. cold topic in finding research trends. In [4], a study of trends in the field of library and information science literature has performed using Latent Dirichlet Allocation (LDA) in order to analyze literature of information management. In [5], again LDA was applied to discovered topics in the field of law literature about contract law and discrimination.

The topic modeling technique is effective in analyzing research articles and trends according to research topics and time, but there are few studies considering the relationship between them. In particular, in research fields where interdisciplinary interactions such as machine learning are active, other topics may coexist within one topic rather than being independent, and the relationship between topics may be important information. Therefore, it is essential to re-analyze the topic that has a great influence on other topics, such as the relationship between topics, and the topic that coexists in a topic.

Social Network Analysis [6] is a popular technique for analyzing the relationships among interested entities. For text documents, keyword network analysis is used to find relationships between keywords in the document. In [7], a keyword is considered as a node in a research article and the connection established between the concurrent words found in a research article. The weight of each connection has assigned in a keyword network with the number of words appearing in another article. In the keyword network analysis, since each node is a single word, it is difficult to know a topic composed of a plurality of words as in the topic modeling, so there is a limitation in grasping macroscopic research trends.

In this study, a topic network analysis method that applies the network analysis technique to the topics generated as a result of topic modeling to analyze the relationship between topics. In order to construct a topic network, correlation amongst the topics discovered by LDA depending on the terms probability that appeared in common terms instead of the common number of terms between two different topics as weights. By analyzing the constructed topic network, we found the topics with the highest degree, closeness, and betweenness.

Next, to understand the trends of topics over time, we divided the articles published over two decades into four sets for 5 years to analyzed the topic network and analyzed the trends changes of the network structure over time.

In order to analyze the research topic, six prominent journals in the field of machine learning were selected, the subject term was designated as machine learning field and the research papers published in six journals for 20 years from 1997 to 2016. Almost 16,855, research articles were collected and applied to the topic network analysis. Table 1 shows, in particular, the summary of articles collected as 2988 papers published between 1997–2001, 3825 papers published between 2002–2006, 4641 papers published between 2007–2011, and 5401 papers published between 2012–2016 are analyzed.

Table 1. Distribution of number of articles in our corpus

Journals/Year	1997–2001	2002–2006	2007–2011	2012–2016
IEEE-PAMI	656	814	905	973
IEEE-NN	727	729	857	1047
JMLR	25	358	509	740
ScD-PR	821	1132	1507	1596
ScD-NN	547	563	618	742
Springer-ML	212	229	245	303
#ArticlesCount = 16855	2988	3825	4641	5401

In order to identify the trends of the most influential topics, the topics with the highest degree, closeness, and betweenness were analyzed. In the case of degree centrality, the topics are identified as component analysis, text mining, biometric, and computer vision. The topics of betweenness centrality are changed to image or video processing, text mining, biometric, and data analysis. Finally, topics of high closeness centrality are same as degree centrality.

It is noteworthy that the topic trends of text mining and biometric has emerged as an influential topic in several sections and these influential topics in research articles published between 2002–2006 and 2007–2011 respectively. In addition, since the patterns of degree centrality and closeness centrality are similar to each other, as shown in the topic network structure.

The contribution of this paper is as follows. First, we discussed a network analysis technique that combines topic modeling and network analysis to find topics. Then, established the link between topics that have influenced on other topics and topics that are closest to other topics. Secondly, we divide the papers published over the past 20 years into four sets, and then analyzed the topic network to identify changes in network structure over time and trends in influential topics.

This paper is composed as follows. In Sect. 2, the related research is reviewed. In Sect. 3, the topic network analysis technique discussed in this study is explained in detail and experimental results are presented. In Sect. 4, we review the research trends of topics that have been selected as influential topics and how the themes change with time. Finally, in Sect. 5 present conclusions and future research.

2   Related Work

The LDA algorithm [3] used in this study is a based on Bayesian probabilistic topic model. As the document consists of several topics and each topic is made up of the distribution of words, the topic is generated based on the probability of words. In [8], topic modeling by using LDA algorithm has been successfully applied to analyzing research trends. Also in [9], LDA has used in issue tracking systems that became a very popular subject in Twitter. However, since the LDA algorithm does not take into account the relationship between the topics, there is a limitation in that it is impossible to grasp the relationship between the research topics and the relationship between them.

In [10], the limitation of the LDA is overcome by introducing Correlated Topic Models (CTM) that takes into account the correlation between topics. CTM is a more realistic model because it generates a topic considering the case where another topic comes together in one topic. Also, considering the correlation between the topics is similar to the problem of grasping the relationship between the topics covered in this study. The topic network analysis proposed in this study uses the LDA algorithm to find independent topics first, and then builds a topic network by generating the correlation matrix amongst the topics discovered by LDA depending on the number of words that appeared in common between two different topics as weights to find degree centrality, closeness centrality, and betweenness to other topics.

In [11], a social network was constructed based on the authorship of their research articles, and then applied topic modeling to interested nodes to find out topics in the network structure. This study is most similar to the present study in that it integrates social network analysis and topic modeling techniques. The main point of this study is to analyze the social network structure of the formed topics. However, research on forming and analyzing a topic network based on correlation among topics generated as a result of topic modeling has been rarely performed.

The topic generated by topic modeling can consist of several subtopics. In [12], proposed a framework for integrating hierarchical LDA and latent topics to represent the hierarchy of topics. Also, in [13] proposed a non-Markov model for finding topics over time. The most significant difference between these studies and the present study is that the topic network analysis is based on the network structure that these topics can form rather than the topics found through topic modeling. Similar study was also being performed in [14] on trend analysis of data mining research, where only the common words between the topics were considered as weight for topic network graph. In our present study, we are generating the correlation amongst the topics discovered by LDA depending on the terms probability that appeared in common terms instead of the common number of terms between two different topics as weights.

3 Analysis of Topic Network

In this section, we first look at the methodology for analysis of topic network. Then, discussed the algorithm for topic network construction. Finally, we present the time-series analysis of the topic network.

3.1 Methodology for Analysis of Topic Network

Figure 1 shows the method for topic network analysis. We started collecting data with the preparation of a list of appropriate well-known journals for publishing high-quality research in machine learning. We include established journals like Journal of machine learning research (JMLR), Springer machine learning (Sp-ML), Science Direct pattern recognition (ScD-PR), Science Direct neural networks (ScD-NN), IEEE transactions on pattern analysis and machine intelligence (IEEE-PAMI) and IEEE transactions on neural networks (IEEE-NN). Titles and abstracts of research papers are taken from the

electronic library of mentioned journal articles. We have considered only journal articles in our study. Table 1 lists the number of articles included in our study.
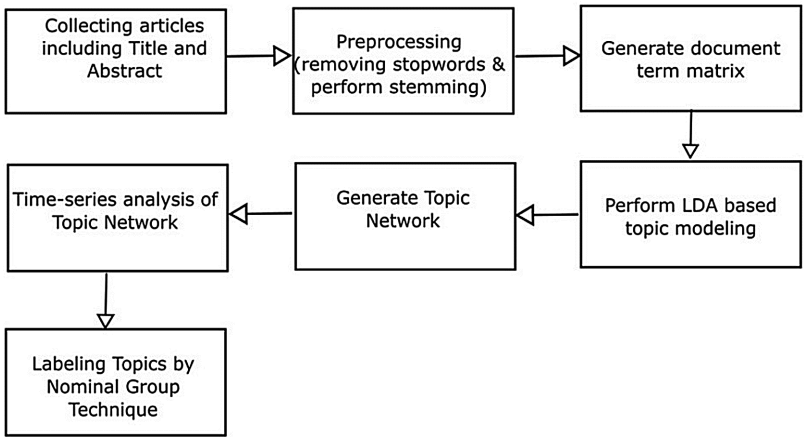

Fig. 1. Topic network analysis method

Next, the statistical analysis software environment R [15] and topicmodels package [16] were used to apply the LDA algorithm to the titles and abstracts of the collected articles. Preprocessing of the articles is done by filtering terms (words) by applying a custom stop word list including "title" and "abstract" along with the original stop words of the SMART system (available for download from ftp://ftp.cs.cornell.edu/pub/smart/english.stop) is a common pre-processing step in text mining. Next, applying stemming [17] to words before generating Document Term Matrix (DTM). Thereafter, perform LDA based topic modeling by using DTM. In order to select the most appropriate number of topics, we investigated with 10, 20, 30, and 40 topics. After this investigation, 30 is the number of topics which mostly suits for our study.

Based on the results of the topic modeling, we construct a topic network and compare the degree centrality, betweenness centrality, and closeness centrality defined in [6] as important measures of the network for the generated topic network. In order to examine the change process of topic network in the field of machine learning research for the past 20 years, we analyzed topic network by dividing it into 4 sets for 5 years from 1997–2001 refers as S1, 2002–2006 refers as S2, 2007–2011 refers as S3, and finally 2012–2016 refer as S4.

Finally, we applied Nominal Group techniques (NGT) [18] for labeling topics which are highly subjective. We have consulted two senior researchers and two peer researchers working in area of machine learning. The results in form of topic name for each topic have been ranked and selected through voting.

3.2 Topic Network Construction

Algorithm 1 is an algorithm for building a topic network. The LDA algorithm used in the experiment was the LDA () function of the topicmodels package that implements the model in [1].

In this paper, we generate a document-term matrix DTM [f_{ij}] by inputting the title and abstract refer as TA, where f_{ij} is the frequency at which the j_{th} word appears in the i_{th} document. Here, considering only the frequency of terms. The most commonly terms found in machine learning field such as model, data, network, learn and algorithm are ranked at the top. Therefore, we selected terms whose frequency is greater than 2 and its word length is greater than 4.

```
Algorithm 1. Algorithm for Topic Network Construction
Input: a set of titles and abstracts TA
Output: Topic Weights Matrix TWM[i,j]
Step1. generate document-term matrix DTM[fi,j]
Step2. extract important terms by filtering out
fi,j>= 3 and len(word) >=4
Step3. create LDA model asldamodel with 30 topics
Step4. generate Topic Weights Matrix TWM based on the
correlation among topics
for each i from 0 to 29
  for each j from 0 to 29
    if i == j then
      TWM[i,j] <- 0
    else
      TWM[i,j] <- cor(ti,tj)
    endif
  endfor
endfor
```

The topic modeling was performed by applying the LDA algorithm to the generated document-word matrix. The hyper parameter value of alpha represents document-topic density is calculated as $50/k$ when k is the number of topics, and is set to 30. The beta represents topic-word density and its values for the word distribution for the topic can be estimated by fixing the topic distribution. In this study, we are filtered out the correlation coefficient among topics having value less than or equal to 0.02.

Table 2 shows the results of topic modeling by applying the LDA algorithm to the top ten keywords for 30 topics.

Table 2. Top 10 keywords in 30 topics from latent Dirichlet Allocation (1997–2016)

Topic0		Topic1		Topic2		Topic3		Topic4	
data,	number,	error,	show,	measur,	evalu,	recognit,	text,	approach,	improv,
cluster,	propos,	general,	loss,	differ,	rank,	face,	rate,	techniqu,	combin,
fuzzi,	real,	bound,	case,	similar,	metric,	databas,	facial,	result,	present,
group,	base,	regress,	result,	distanc,	compar,	document,	word,	perform,	strategi,
partit,	unsupervis	regular,	squar	perform,	correl	charact,	system	search,	compar
Topic5		**Topic6**		**Topic7**		**Topic8**		**Topic9**	
problem,	convex,	function,	general,	structur,	represent,	implement,	applic,	match,	invari,
optim,	formul,	approxim,	continu,	rule,	repres,	process,	provid,	transform,	code,
graph,	constraint,	paramet,	deriv,	tree,	concept,	oper,	present,	pattern,	base,
solut,	minim,	valu,	properti,	relat,	describ,	comput,	system,	scheme,	fingerprint,
solv,	program	basi,	discret	hierarch,	present	design,	power	propos,	paper
Topic 10		**Topic11**		**Topic12**		**Topic13**		**Topic14**	
featur,	base,	object,	human,	method,	result,	local,	global,	analysi,	research,
select,	relev,	detect,	target,	propos,	novel,	robust,	effect,	applic,	work,
inform,	subset,	visual,	motion,	base,	effect,	nois,	scale,	studi,	provid,
extract,	import,	track,	appear,	paper,	exist,	propos,	enhanc,	develop,	recent,
propos,	effect	video,	multipl	compar,	experiment	filter,	show	import,	mani
Topic15		**Topic16**		**Topic17**		**Topic18**		**Topic19**	
kernel,	matrix,	point,	contour,	imag,	retriev,	view,	estim,	model,	probabilist,
space,	data,	shape,	boundari,	segment,	pixel,	imag,	motion,	variabl,	process,
analysi,	spars,	curv,	comput,	region,	spatial,	surfac,	scene,	bayesian,	markov,
linear,	dimension,	object,	geometr,	color,	base,	reconstruct,	correspond,	infer,	paramet,
discrimin,	represent	line,	correspond	textur,	level	camera,	illumin	prior,	framework
Topic20		**Topic21**		**Topic22**		**Topic23**		**Topic24**	
network,	output,	estim,	gaussian,	network,	delay,	learn,	problem,	vector,	complex,
neural,	layer,	distribut,	compon,	neural,	global,	task,	knowledg,	comput,	larg,
train,	architectur,	probabl,	densiti,	condit,	numer,	domain,	supervis,	machin,	reduc,
input,	unit,	statist,	random,	stabil,	paper,	framework,	exampl,	number,	size,
weight,	recurr	independ,	sourc	result,	exampl	multipl,	reinforc	support,	effici
Topic25		**Topic26**		**Topic27**		**Topic28**		**Topic29**	
time,	seri,	algorithm,	gradient,	neuron,	cell,	classif,	perform,	system,	design,
predict,	show,	converg,	present,	activ,	behavior,	classifi,	sampl,	control,	robot,
sequenc,	differ,	iter,	base,	model,	spike,	class,	accuraci,	adapt,	propos,
chang,	result,	onlin,	updat,	memori,	respons,	label,	ensembl,	dynam,	simul,
state,	process	effici,	step	mechan,	associ	train,	instanc	nonlinear,	develop

The topic network was constructed based on the correlation between the terms among the topics. The nodes of the network are topic, and the edge connecting the node uses the correlation coefficient between the two topics as the weight. The number of words considered per topic is the top 10 terms in topic. In order to calculate the weight value, the correlation of the i_{th} topic and the j_{th} topic is obtained, and the result of this correlation as its correlation coefficient is determined as the weight. In this way, a topic weighting matrix $TWM\big[cor(t_i, t_j)\big]$ is generated that takes into consideration the relationship that 30 topics share the terms.

Figure 2 is a topical network graph generated by applying Algorithm 1 to the titles and abstracts of research articles published over the 20 years from 1997 to 2016. The size of each node is expressed as a degree centrality of the topic so that the node with high degree centrality can be seen at a glance.
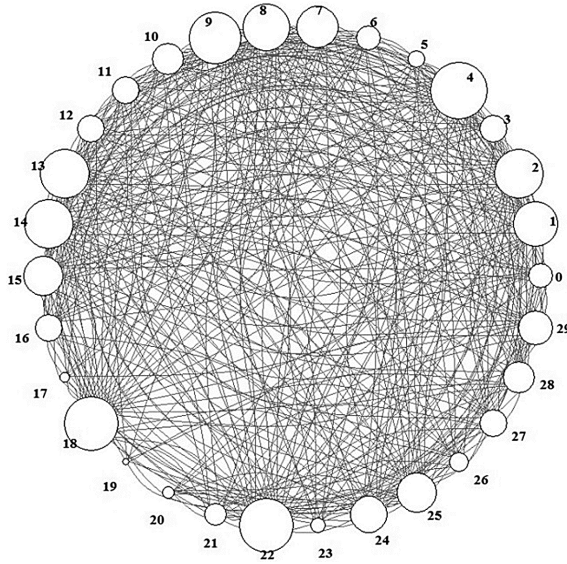
Fig. 2. Topic network graph (1997–2016)

3.3   Time-Series Analysis of Topic Network

To analyze the generated topic network, the centralization measures like centrality of the degree, the betweenness, and the closeness are commonly [19] used in social network analysis. Particularly, in order to investigate the change of topic networks generated over the past twenty years, we divided the topic network into five years to rebuild the topic network.

Connectivity between nodes is measured by degree centrality, as the number of nodes connecting to each node. Even if the number of connection edges is not small, this node also plays an important role if one node is at the shortest distance connecting the other two nodes. This node can be determined by measuring the closeness centrality. Finally, the node with the shortest neighbor distance to all nodes can also be regarded as an important node, which can be determined by measuring the betweenness centrality.

Figure 3(a) is a topic network analyzing the titles and abstracts of research articles from 1997–2001, in the same way, Fig. 3(b) from 2002–2006, Fig. 3(c) from 2007–2011, and Fig. 3(d) from 2012–2016.

In order to grasp the structural characteristics of the four topic networks, we have grasped the centrality of each network. Centrality is a concept of identifying the most important vertex in the whole network and is used to hold the overall characteristics of the network. The centrality of each measure of centrality can be determined by Freeman [19] and defined as Eq. (1).

$$C_A = \frac{\sum_{i=1}^{g}[C_A(n^*) - C_A(n_j)]}{\max \sum_{i=1}^{g}\left[C_A(n^*) - C_A(n_j)\right]} \tag{1}$$
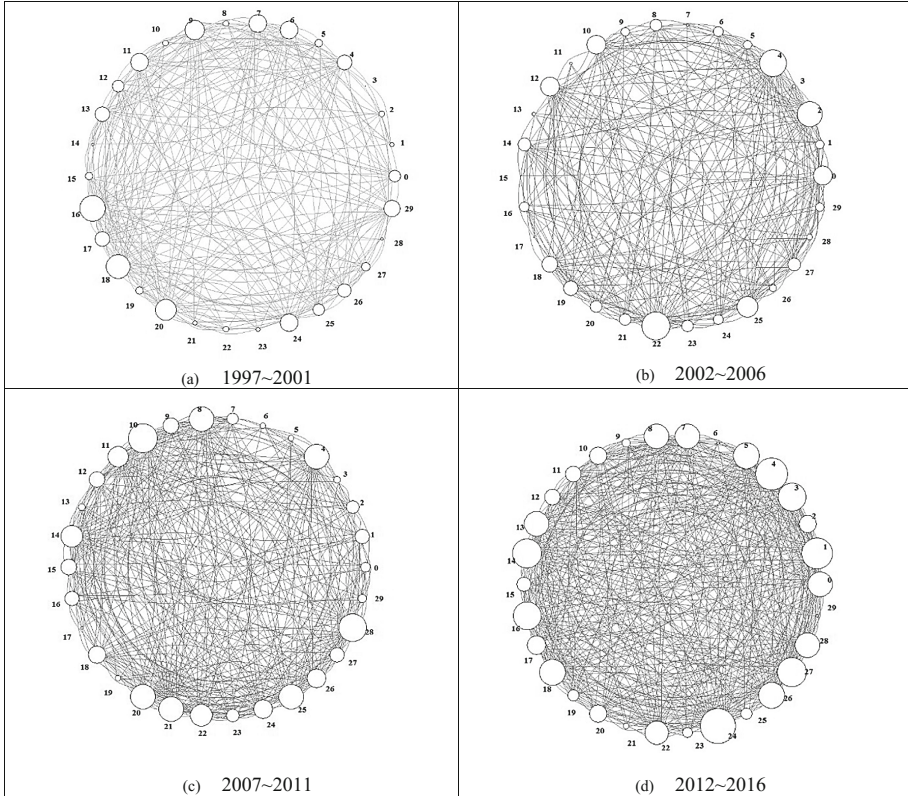
Fig. 3. (a) Topic network of 1997–2001, (b) Topic network of 2002–2006, (c) Topic network of 2007–2011, (d) Topic network of 2012–2016

whereas, $C_A(n_j)$ is called the center of a node, and $C_A(n^*)$ is the maximum value of the centroid of all the nodes belonging to the network. $\sum_{i=1}^{g} [C_A(n^*) - C_A(n_j)]$ is the sum of the differences between the maximum centroid value and the centroid of all nodes and finally, $\max \sum_{i=1}^{g} [C_A(n^*) - C_A(n_j)]$ is logically the maximum difference. The central value is the sum of the differences between the maximum centrality value and the centrality of all nodes logically divided by the sum of the largest differences.

Figure 4 shows the result of analyzing the centrality of the topic network over time. The values of degree centrality, closeness centrality, and betweenness centrality are expressed as normalization values. First, in the case of the degree centrality is the highest value in 2002–2006, which was the time rise of machine learning research, decreased in the next five years because of the transition in the research of machine learning, this is the era when deep learning and computer vision just emerging area of research. Then, the trend is gradually increasing over the next five years. Also, the closeness centrality shows the same pattern as degree centrality. However, the betweenness centrality is the highest value in 2002–2006 and then its shows gradually decreasing trend since then.
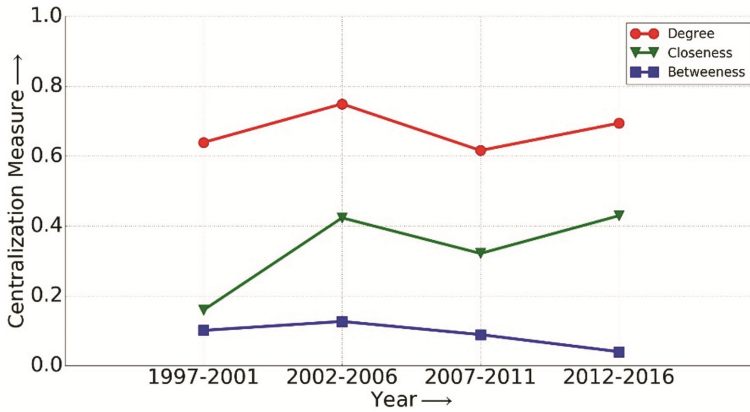
Fig. 4. Time series analysis of centralization

4 Trend Analysis of Machine Learning Research

In this section, we will look specifically at topics that are selected as important topics by topic network analysis. Table 3 shows the top ten keywords of the topic with the highest degree centrality. In all tables, S1 refer as time period of 1997–2001, S2 refer as time period of 2002–2006, S3 refer as time period of 2007–2011, and S4 refer as time period of 2012–2016.

Table 3. Top 10 keywords in topics with highest degree centrality

Year	Topic	Keywords
S1: 1997–2001	Component analysis	analysi, compon, paper, applic, result, provid, independ, theoret, mani, present
S2: 2002–2006	Text mining	robust, nois, filter, result, show, document, text, effect, paper, threshold
S3: 2007–2011	Biometric	propos, perform, system, paper, match, fingerprint, identif, scheme, level, fusion
S4: 2012–2016	Computer Vision	recognit, face, propos, databas, express, facial, differ, result, variat, base

Figure 5 shows the degree centrality of 30 topics. The following topics with highest degree centrality are Topic16, Topic22, Topic10, Topic24, and Topic4 in time period 1997–2001, 2002–2006, 2007–2011, 2012–2016 and 1997–2016 respectively.
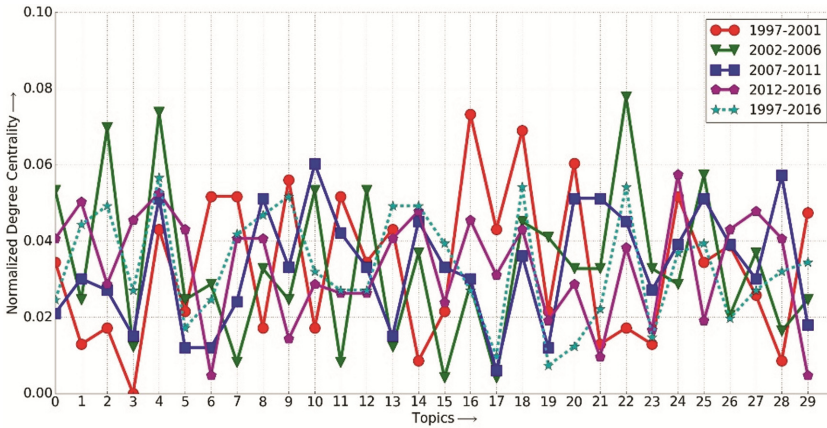
Fig. 5. Normalized degree centrality of 30 topics

Table 4 shows the top ten keywords of the topic with highest betweenness centrality. First of all, the topic with highest betweenness centrality during time period S1 is *Image or Video Processing*, which includes the words like *motion, camera, image, and track.* The *text* and *documents* words along with *filter* and *threshold* words appeared during time period S2 which gave intuition of *Text Mining* topic. Similarly, the topics identified during the time period of S3 and S4 are *Biometric* and *Data Analysis* respectively.

Table 4. Top 10 keywords in topics with highest betweenness centrality

Year	Topic	Keywords
S1: 1997–2001	Image or video processing	motion, sequenc, camera, project, track, imag, estim, correspond, constraint, paramet
S2: 2002–2006	Text mining	robust, nois, filter, result, show, document, text, effect, paper, threshold
S3: 2007–2011	Biometric	propos, perform, system, paper, match, fingerprint, identif, scheme, level, fusion
S4: 2012–2016	Data analysis	discrimin, analysi, data, space, compon, project, dimension, subspac, manifold, linear

Figure 6 shows the betweenness centrality of 30 topics. The following topics with highest betweenness centrality are Topic18, Topic4, Topic10, Topic24, and Topic4 in time period 1997–2001, 2002–2006, 2007–2011, 2012–2016 and 1997–2016 respectively.
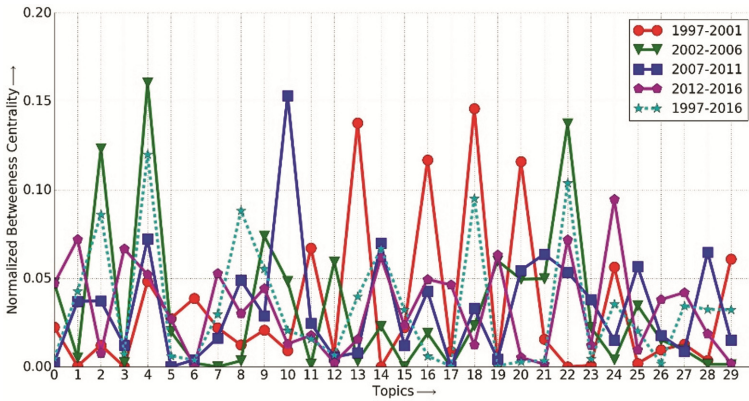
Fig. 6. Normalized betweenness centrality of 30 topics

Table 5 shows the top ten keywords of the topic with highest closeness centrality. A closeness centered topic is a topic that connects two independent topics. The topics identified with closeness centrality are same topics as with degree centrality. In the time period of S1, the *Component Analysis* topic is most relevant due to following words like component, analysis, and independent. In the time period of S2, the *Text Mining* topic comes out to be related due to words like *document, text, threshold, and filter*.

Table 5. Top 10 keywords in topics with highest closeness centrality

Year	Topic	Keywords
S1: 1997–2001	Component analysis	analysi, compon, paper, applic, result, provid, independ, theoret, mani, present
S2: 2002–2006	Text mining	robust, nois, filter, result, show, document, text, effect, paper, threshold
S3: 2007–2011	Biometric	propos, perform, system, paper, match, fingerprint, identif, scheme, level, fusion
S4: 2012–2016	Computer vision	recognit, face, propos, databas, express, facial, differ, result, variat, base

Similarly, the topics in the time period of S3 and S4 are come out as *Biometric and Computer Vision* respectively.

Figure 7 shows the closeness centrality of 30 topics. The following topics with highest degree centrality are Topic16, Topic22, Topic10, Topic24, and Topic4 in time period of 1997–2001, 2002–2006, 2007–2011, 2012–2016 and 1997–2016 respectively.

It is noteworthy that computer vision and text mining topics are emerged as a topic with a high degree of degree centrality, closeness centrality, and betweenness centrality in the research articles published in period S2 and S4 respectively. *Component analysis and Image or video processing* is also included in high centrality topics during the S1 time period of machine learning research.
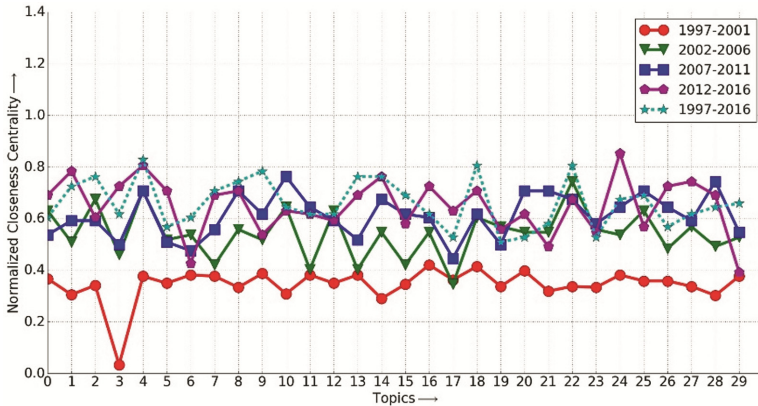
Fig. 7. Normalized closeness centrality of 30 topics

In the most recent 5-year interval computer vision and data analytics are high centrality topics.

5   Conclusions and Future Work

In this study, we performed the trend analysis of the last two decades of machine learning research articles published in six well-known journals for the period of 1997–2016. In order to understand the trends of the research topic for 20 years, we divided topic network into 4 sets and conducted topic network analysis. Topic modeling was performed to select 30 topics and related words corresponding to topics, and a topic network was constructed using the correlation amongst the topics discovered by LDA depending on the terms probability that appeared in common terms instead of the common number of terms between two different topics as weights. By applying social network analysis techniques to established topic networks, we found topics with high degree, closeness, and betweenness.

The contribution of this study was to construct a topic network and analyze it to find out the correlation between topics and identify the influential topics. Also, in order to overcome the limitation of LDA topic model, that cannot consider relation between topics. In addition, by analyzing the topic network analysis in a time series, we have found out what trends the most influential topics show over time.

Future research will build a web application where the interested researcher who is newly venturing into this field can run the model to understand the effectiveness of the proposed topic network analysis results and have directed the latest trending topics pertaining to research area. We collect titles and abstracts of machine learning research articles. Similar study can also be performed on other newly emerging areas such as 'Internet of Things' as well as 'cloud computing' and use them as test cases.

References

1. Kim, C., Hong, Y.-S.: Classification techniques for XML document using text mining. J. Korea Soc. Comput. Inf. **11**(2), 15–23 (2006)
2. Moon, J.-P., Lee, W.-S., Chang, J.-H.: A proper folder recommendation technique using frequent item sets for efficient e-mail classification. J. Korea Soc. Comput. Inf. **16**(2), 33–46 (2011)
3. Blei, D.M., Andrew, Y.N., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
4. Park, J.H., Song, M.: A study on the research trends in library & information science in Korea using topic modeling. J. Korean Soc. Inf. Manag. **30**(1), 7–32 (2013)
5. Blei, D.M.: Probabilistic topic models. Commun. ACM **55**(4), 77–84 (2012)
6. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)
7. Duvvuru, A., Kamarthi, S., Sultornsanee, S.: Undercovering research trends: network analysis of keywords in scholarly articles. In: Proceedings of the 9th International Joint Conference on Computer Science and Software Engineering, pp. 265–270 (2012)
8. Griffiths, T.L., Steyvers, M.: Finding scientific topics. In: Proceedings of the National Academy of Sciences of the USA, vol. 101, no. 1, pp. 5228–5235, April 2004
9. Bae, J., Han, N., Song, M.: Twitter issue tracking system by topic modeling techniques. J. Intell. Inf. Syst. **20**(2), 109–122 (2014)
10. Blei, D.M., Lafferty, J.D.: Correlated topic models. In: Proceedings of Neural Information Processing Systems, pp. 147–154 (2005)
11. Mei, Q., et al.: Topic modeling with network regularization. In: Proceedings of International Conference on World Wide Web, pp. 101–110 (2008)
12. Mao, X.-L., et al.: SSHLDA: a semi-supervised hierarchical topic model. In: Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pp. 800–809 (2012)
13. Wang, X., McCallum, A.: Topics over time: a non-Markov continuous-time model of topical trends. In: Proceedings of the 12th International Conference on Knowledge Discovery and Data Mining, pp. 424–433 (2006)
14. Kim, C., Hong, Y.-S.: Trend analysis of data mining research using topic network analysis. J. Korea Soc. Comput. Inf. **11**(5), 141–148 (2016)
15. R: The R Project for Statistical Computing. https://www.r-project.org/
16. Gruen, B., Hornik, K.: topicmodels: an R package for fitting topic models. J. Stat. Softw. **40**(13), 1–29 (2011)
17. Porter, M.: An algorithm for suffix stripping. Program **14**(3), 130–137 (1980)
18. Delp, P., Thesen, A., Motiwalla, J., Seshardi, N. (eds.) System Tools for Project Planning. International Development Institute, Bloomington (1977)
19. Freeman, L.C.: Centrality in social networks: conceptual clarification. Soc. Netw. **1**, 215–239 (1979)