I.J. Modern Education and Computer Science, **2017, 7, 50-62**
Published Online July 2017 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijmecs.2017.07.06

A Survey on Journey of Topic Modeling Techniques from SVD to Deep Learning

Deepak Sharma[1]
[1]Department of Computer Engineering, Netaji Subash Institute of Technology, Sector-3, Dwarka, New Delhi, 110078, India
Email: deepak.btg@gmail.com

Bijendra Kumar[1], **Satish Chand**[2]
[1]Department of Computer Engineering, Netaji Subash Institute of Technology, Sector-3, Dwarka, New Delhi, 110078, India
[2]School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi, 110067, India
Email: {bizender, schand20}@gmail.com

Abstract—Topic modeling techniques have been primarily being used to mine the topics from text corpora. These techniques reveal the hidden thematic structure in a collection of documents and facilitate to build up new ways to browse, search and summarize large archive of texts. A topic is a group of words that frequently occur together. A topic modeling can connect words with similar meanings and make a distinction between uses of words with several meanings. Here we present a survey on journey of topic modeling techniques comprising Latent Dirichlet Allocation (LDA) and non-LDA based techniques and the reason for classify the techniques into LDA and non-LDA is that LDA has ruled the topic modeling techniques since its inception. We have used the three hierarchical classification criteria's for classifying topic models that include LDA and non-LDA based, bag-of-words or sequence-of-words approach and unsupervised or supervised learning for our survey. Purpose of this survey is to explore the topic modeling techniques since Singular Value Decomposition (SVD) topic model to the latest topic models in deep learning. Also, provide the brief summary of current probabilistic topic models as well as a motivation for future research.

Index Terms—Topic Modeling, Latent Semantic Analysis, Latent Dirichlet Allocation, Deep Learning, Survey.

I. Introduction

In this era of information technology, the combined knowledge is being generated in digital form that is stored in numerous forms such as blogs, news, research articles, social networks, web pages etc. There are fairly sophisticated search techniques to extract meaningful information from these collections of digitized documents. Traditional search engines would search for terms and retrieve the list of relevant documents whereas each document has thematic structure related to many topics. Topic modeling does the same in which it models groups of word collectively representing the topic for each document and each document across a mixture of latent topics or themes. The intuition behind topic modeling is that each document is comprised of some 'topics' or 'themes'. A "topic" is understood as a group of words that represent the topic as a whole.

In the related work on some past surveys on topic modeling have been done that include [1, 2, 3, 4]. In survey paper [1], presents a classification of directed probabilistic topic models and explains a broader view on graphical models. It may be considered as an enormous initial point for venture in the field of topic modeling. In [2] and [3] an introductory discussion about topic modeling is presented. In paper [4] discusses the classification of probabilistic topic modeling algorithms.

In this survey, we provide survey on topic modeling techniques since Singular Value Decomposition (SVD) topic model to the latest topic models in deep learning techniques for generating topics from collection of documents. The earlier surveys were very rudimentary and to our best knowledge there doesn't exit any survey that considers state of the art algorithm in topic modeling. The rest of the paper is organized as follows. Section 2 gives a brief overview of classifications for topic modeling of our survey. Section 3 describes the topic modeling techniques. Section 4 describes the applications of topic modeling. Section 5 concludes the paper.

II. Classification

In this paper, Fig. 1. shows the classified topic modeling techniques according to three criteria's. The first criterion is based on Latent Dirichlet Allocation (LDA) (L) based techniques and non-LDA (NL) based techniques as LDA has been ruled the topic modeling techniques since its inception. Each of the categories are classified based on bag-of-words and sequence-of-words approaches, which in turn are further classified based on unsupervised or supervised learning. A second criterion is based on ordering of word and representation of document. The most commonly used solution known as bag-of-words (B) approach, which neglects the word

ordering to enable the global semantic structures whereas another approach doesn't neglect the word ordering known as sequence-of-words (S). Although the bag-of-words approach is frequently used for its simplicity as compared to sequence-of-words approach which requires additional information direct to better results in some problem domains. A third criterion is based on the dependability of labeled data. The intuition behind topic modeling was to apply unsupervised learning (U) where unlabeled data can be provided.

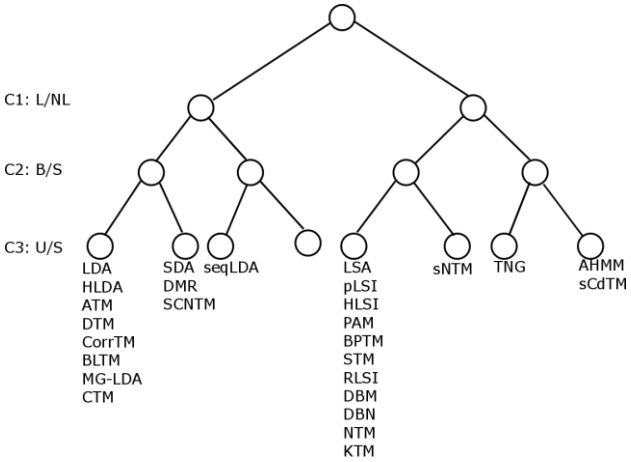

Fig.1. The classification tree of topic modeling techniques

In unsupervised learning topics consist of collection of words with no data labels are provided and human expertise is requiring to represent a topic name to the collection of words. Most of the topics modeling techniques are fully unsupervised; there are few models used semi-supervised or supervised learning (S) for classify topics based on labeled data. Fig. 1. shows the classification tree of topic modeling techniques. Legend: L/NL – LDA based vs. Non-LDA based techniques; B/S - bag of words vs. sequence of words; U/S – unsupervised vs. supervised. The class of LDA based sequence of word models with supervised learning has no known implementations. In the following figures, filled circle shows for the given input variable and the white circle shows for the observed variables.

III. Topic Modeling Techniques

In this section, we would focus on discussing topic modeling techniques based on the classification criteria. These techniques are very frequently used in many applications to extract the topics from collections of documents which help to tag the similar document under topics.

A. LDA based bag of words topic models with unsupervised learning

a. Latent Dirichlet Allocation (LDA)

In [5], introduced the generative probabilistic model i.e. LDA becomes foundation for numerous latent factors discovery algorithms which was collectively known as Probabilistic topic models. This model upgrades into Bayesian graphical model by introducing priors on document-topic distributions than pLSI. By introducing the Dirichlet prior distribution in LDA resolves the problematic issues of increasing number of estimation parameters in pLSI. Fig. 2. shows the graphical model for LDA.

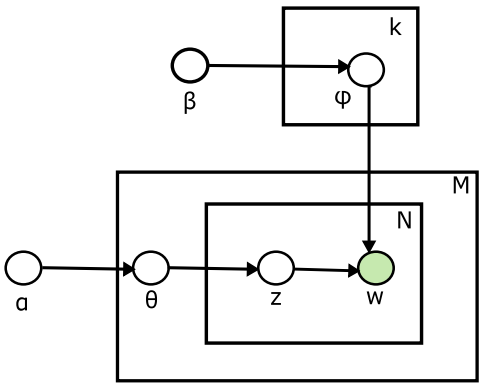

Fig.2. LDA Graphical Model

b. Hierarchical Latent Dirichlet Allocation (HLDA)

In [6], HLDA introduced as an expansion to LDA which models a tree of topic instead of flat topic proposed by LDA. HLDA uses a non-parametric Bayesian model to generate a prior via distribution on partitions using Chinese restaurant process [7] to build topical hierarchies. Each node in the tree is associated as a topic whereas topic is distribution of words. The document can be generated by traversing a path from root to leaf, repetitively sampling topics down the path and sampling the words from the nested topics. Fig. 3. shows the graphical model for HLDA.

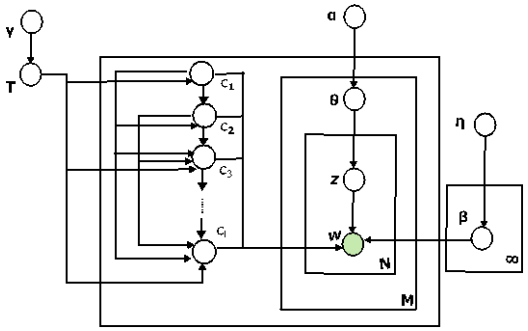

Fig.3. HLDA Graphical Model

c. Author-Topic Model (ATM)

In [8], ATM was first proposed as a generative probabilistic model expansion of LDA and later this was expanded in [9]. This model resulting from LDA and used metadata present in each document for extracting topic distribution with respect to author in corpora. In this model, each word is connected with two latent variables as an author and a topic. Alike LDA, each topic has distribution over words and each author has distribution over topics. ATM goes step advance to models author-

topic distribution along with model only document-topic and topic-word distribution. MCMC algorithm has used to learn the author-topic and topic-word distribution. Also, in ATM author and words are observable variables. Fig. 4. shows the graphical model for ATM.

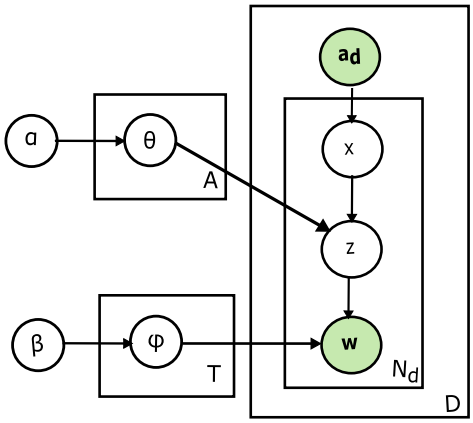

Fig.4. ATM Graphical Model

d. Dynamic Topic Model (DTM)

In [10], DTM was introduced as an extension on LDA based on detecting the topics as an evolution of time in sequentially organized documents. The document metadata has used as notion of time that describes the word-topic distribution which make intuition to develop topic trends. A topic is a sequence of distributions over words instead single distribution over words. DTM yield more complicated inferences because of sampling methods and non-conjugacy, which are overcome by using Variational Wavelet Regression or Variational Kalman filtering variational methods. DTM has an ability to track fixed number of topics through discrete notion of time. Fig. 5. shows the graphical model for DTM.

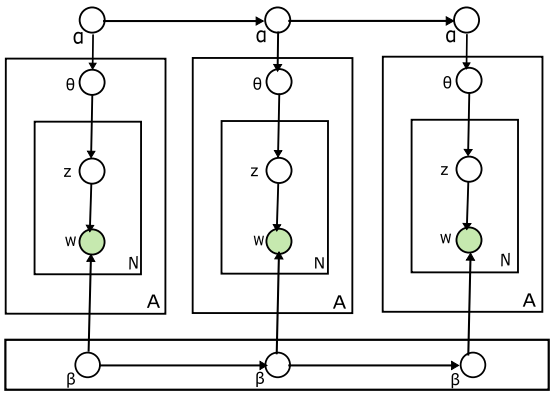

Fig.5. DTM Graphical Model

e. Correlated Topic Model (CorrTM)

In [11], CorrTM was introduced to overcome the inability of LDA to model correlation among the topics. The logistic normal distribution [12] exhibits correlation among the topic proportions instead of Dirichlet distribution in LDA. CorrTM able to model complex structure of underlying topics and gives a covariance matrix which form a topic graph, as compared to LDA model that compels mutual independence on topics and therefore making it more expressive. Mean variational algorithm [13] has used for inference in this model to form a factorized distribution of the topics. CorrTM has always proves to be more expressive than LDA and very effective for topic exploration and visualization. Fig. 6. shows the graphical model for CorrTM.

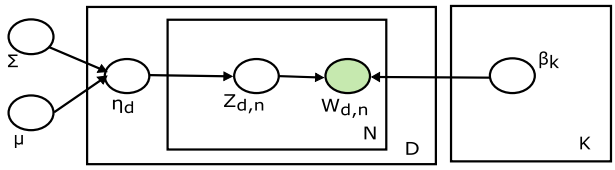

Fig.6. CorrTM Graphical Model

f. Bigram LDA Topic Model (BLTM)

In [14], BLTM was introduced to overcome the bag-of-word approach to construct the model based on N-gram model. BLTM has predicted each word based on the measurement of its previous word. BLTM generation extends over LDA by defining the distribution of words over topic and context. BLTM has incorporated Dirichlet prior with hyper-parameters into biterm conditional distribution matrix and did marginalization. These hyper-parameters are inferred using Gibbs EM algorithm. The properties of a hierarchical Dirichlet bigram language model [15] were used to explore the hierarchical generative probabilistic model for latent topic variables and n-gram statistics. Instead of drawing distribution over words for topics this model draw distribution over words for a particular context from one of two priors. Fig. 7. shows the graphical model for BLTM.

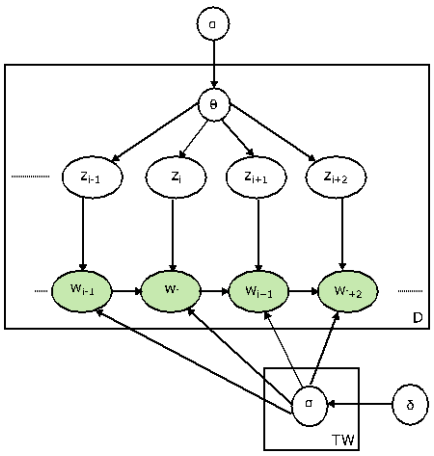

Fig.7. BLTM Graphical Model

g. Multi-grain LDA (MG-LDA)

In [16], MG-LDA was introduced as an extension to LDA and PLSA to induce multi-grain topics. In this model, topics are uniquely identified as global topics and local topics. The distribution of local topics is varying across document. Sampling of the word in the document is from the mixture of local topics specific for the local

context of the word or from the mixture of global topics. In MG-LDA document are represent as a set of sliding windows. Gibbs Sampling has used for inference in this model. MG-LDA model has showed promising results when applied for extracting rating aspects from online reviews. This model is well suited for online reviews problem since it identifies the significant topics along with cluster them into coherent groups. Fig. 8. shows the graphical model for MG-LDA.

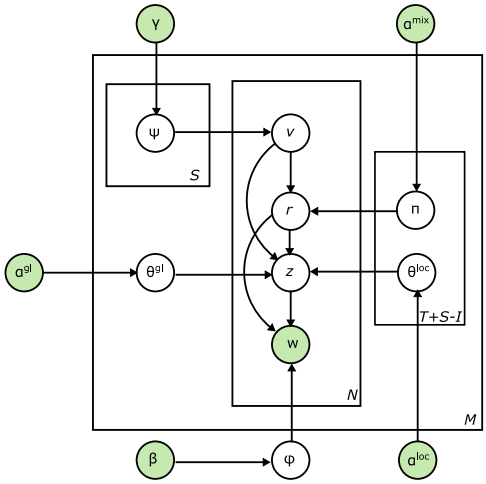

Fig.8. BLTM Graphical Model

h. Concept Topic Model (CTM)

In [17], CTM was introduced as probabilistic model which combining the hierarchy of human defined semantic concepts along with a statistical topic model to induce semantically rich concepts. In this topic model, concepts have added to the topics of the topic model which produce an effective set of topic and concept word distribution for each document. Hierarchical Concept Topic Model (hCTM) has incorporated the hierarchy structure of the concept set. hCTM used generative process in which word tokens would be generated from the concept part of this model by sampling a pathway from root of the concept tree to some distribution over word types related with the concept. Fig. 9. shows the graphical model for CTM.

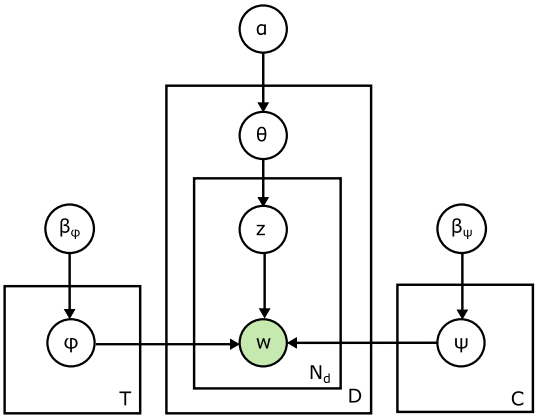

Fig.9. CTM Graphical Model

B. LDA based bag of words topic models with supervised learning

a. Supervised Latent Dirichlet allocation (SLDA)

In [18], SLDA was introduced as probabilistic model an extension to LDA of labeled documents. In SLDA, a response variable has added for each document. In order to find latent topics that would best predict the response variables for future unlabeled documents by jointly model the responses and documents. The appropriate Mean field variational inference method has used for posterior inference which incorporated in EM algorithm for maximum-likelihood for estimating the unknown parameters. Fig. 10. shows the graphical model for SLDA.

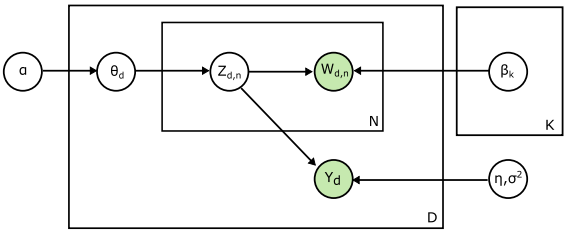

Fig.10. SLDA Graphical Model

b. Dirichlet Multinomial Regression (DMR)

In [19], DMR was introduced as an extension to LDA which incorporates arbitrary features of documents i.e. metadata. The metadata includes the author, publication venue, references and dates as arbitrary features of document. DMR has able to incorporate arbitrary types of observed discrete, continuous and categorical features with no additional coding. This model has trained using stochastic EM sampling algorithm. DMR model has incurred additional complexity results in a larger number of variables to sample and convoluted sampling distribution. Fig. 11. shows the graphical model for DMR.

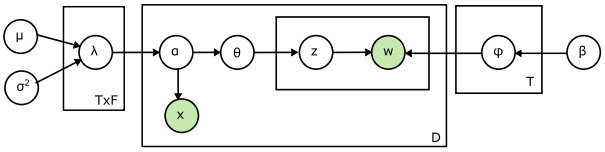

Fig.11. DMR Graphical Model

c. Supervised Citation Network Topic Model (SCNTM)

In [20], SCNTM was introduced as non-parametric extension topic model of a mixture of the author-topic model and the Poisson mixed-topic link model that incorporate bibliographic analysis of authors, topics and documents. SCNTM generates probability vectors represents as counts by using Griffiths-Engen-McCloskey (GEM) distribution [21] and the base distribution as Pitman-Yor process (PYP) [22] which rise to hierarchical Pitman-Yor process (HPYP). In this model the categorical labels of each document as well as author information used for supervised learning. The learning of this model has used Markov chain Monte Carlo (MCMC)

algorithms, which make use of the conjugacy of the Multinomial distribution and the Dirichlet distribution, which allowed the sampling of the citation networks to be of similar form to the collapsed sampler of a topic model. The clustering task has been improved by incorporating supervised learning to this model. Fig. 12. shows the graphical model for SCNTM.

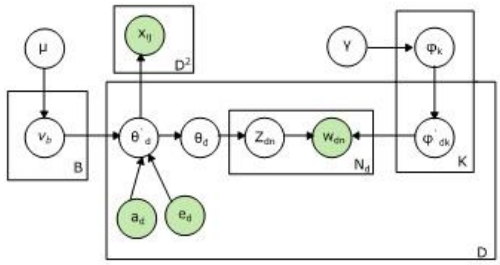

Fig.12. SCNTM Graphical Model

C. LDA based sequence of words topic models with unsupervised learning

a. Sequential Latent Dirichlet Allocation (seqLDA)

In [23], seqLDA was introduced as novel variant of LDA underlying sequential structure. Hierarchy modeling was applied to documents which were considered as a multiple segments and each segment is associated to its predecessor and subsequent segments. The first-order Markov chain was used to bind a sequence of LDA model. The two-parameter Poisson-Dirichlet process (PDP) [24,25] was used to capture the progressive topical dependency in a hierarchical way. An efficient collapsed Gibbs sampling algorithm was proposed based on the hierarchical two-parameter Poisson-Dirichlet process (HPDP) on top of the corresponding Chinese Restaurant Process. This model has showed higher fidelity over LDA in terms of perplexity. Fig. 13. shows the graphical model for seqLDA.

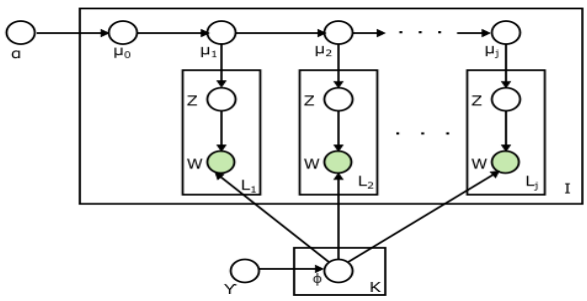

Fig.13. seqLDA Graphical Model

D. LDA based sequence of words topic models with supervised learning

This class of LDA based sequence of word topic models with supervised learning has no known implementations.

E. non-LDA based bag of words topic models with unsupervised learning

a. Latent Semantic Analysis (LSA)

LSA was pioneered by [26] in early 1990s. In establishment phase it was used as a technique in information retrieval for evaluating the performance of search engine query [27, 28, 29, 30, and 31] and indexing library. Later, LSA was adopted by psychology researchers [32] and it has been deployed in the research of education, information systems, cognitive sciences and artificial intelligence.

The primitive idea of LSA is to extract hidden meaning of text from a set of documents. LSA processes a text document from a corpus of documents, identifies term and helps to pull out latent factor i.e. topic from these extracted terms. LSA creates a term-document matrix where term is the most frequently term occurring in a document. A Singular Value Decomposition (SVD) is applied on the matrix which further decomposes into three matrices as U, S, and V; whereas U denotes the term eigenvectors matrix, V denotes the document eigenvectors matrix and S denotes the diagonal matrix of singular values. To avoid the overfitting of factors, the dimensionality of SVD matrices are truncated by keeping the first *k* dimensions. Fig. 14. shows the graphical model of LSA.

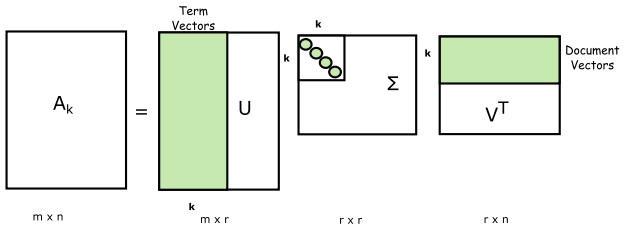

Fig.14. Dimensionality reduction on SVD matrices by keeping first k dimensions in LSA

b. Probabilistic Latent Semantic Indexing (pLSI)

In [33], pLSI was introduced as a probabilistic alternate of its predecessor that shows a proper generative model along with sound statistical foundation. The core of pLSI is a statistical model which has been called aspect model [34, 35]. The generative model prepares each document in corpus and topics can be interpreted as a sampling of each word in a document by applying mixture of multinomial distribution. An inference algorithm is used for inferring the document-topic distributions as well as topic-word distribution from text corpus. An Expectation Maximization [36] is used for computing the word-topic and topic-document distribution. Fig. 15. shows the generative model to infer the equation of E and M steps.

The pLSI as compared to its non-probabilistic predecessor resolves issues of capturing polysemy and has well-built statistical foundation. The limitation of pLSI is its offline nature which unable to applied incrementally to unseen documents and also, issue with overfitting due to huge number of estimation parameters depends on size of corpus.

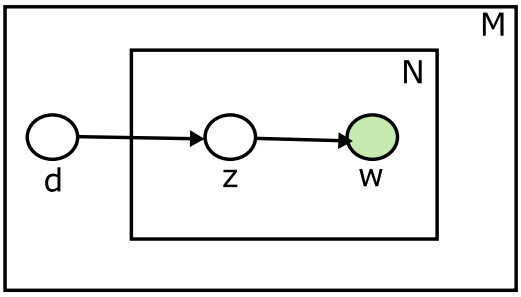

Fig.15. pLSI generative model

c. Hierarchy-Regularized Latent Semantic Indexing (HLSI)

In [37], HLSI was introduced to build hierarchical taxonomy into similarity graph of documents by optimizing the mapping of each document into a low space dimensional vector. A class hierarchy is maintained to describe the similarity between the documents. All the documents are connecting within the class which acts as a bridge. HLSI incorporates the proximity of classes within the hierarchy which implies a connection between the documents belonging to the same class. HLSI integrate the information within a class hierarchy into a variety of learning and retrieval tasks. Fig. 16. shows the graphical model for HLSI.

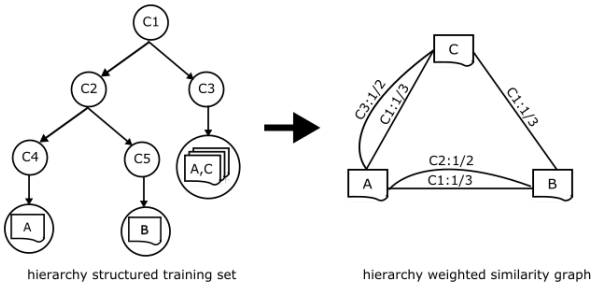

Fig.16. A hierarchy structured training set and the corresponding hierarchy induced similarity graph of HLSI Model

d. Pachinko Allocation Model (PAM)

In [38], PAM was introduced as an alternative to CTM in which topic correlations were identified using covariance matrix representing pairwise correlation among the topics except multiple topics. PAM reinstates the topic distribution over words and other topics too. Directed acyclic graph (DAG) topic structure were used to recognize correlation between topics. Gibbs sampling were used to perform parameter learning and inference. In this model topic has been distributed over other topics including words. Fig. 17. shows the graphical model for PAM.

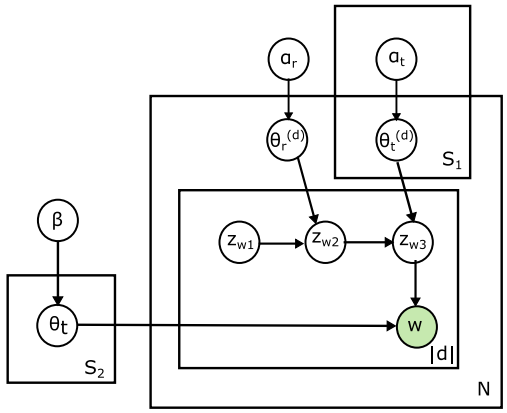

Fig.17. PAM Graphical Model

e. Bigram PLSA Topic Model (BPTM)

In [39], BPTM was introduced as an extension to PLSA in which a modified training strategy that unevenly assigns latent topics to context words according to an estimation of their latent semantic complexities. The training procedure of BPTM is similar to PLSA, whereas few modifications has introduced for EM algorithm for maximizing the log-likelihood and estimating the posterior probabilities of latent variables. PLSA has some issues of performance because of latent variable assignment in which matrix decomposition would influenced by the number of latent topics and the complexity of matrix to be decom-posed. These issues have resolved by refining the BPTM by introducing a unique variable set for every context word. Fig. 18. shows the graphical model for BPTM.

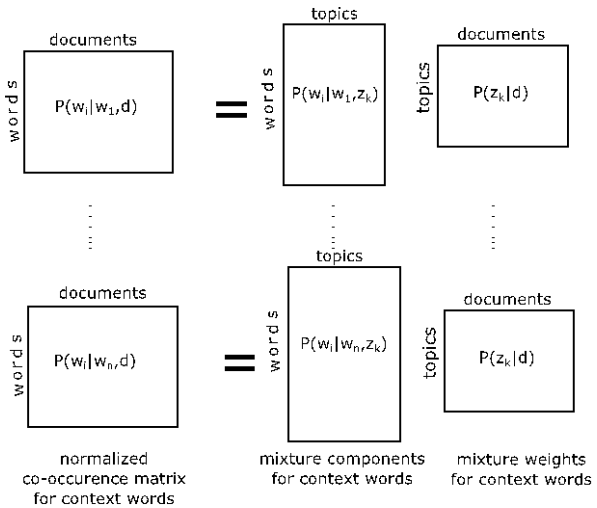

Fig.18. BPTM Graphical Model

f. Syntactic Topic Model (STM)

In [40], STM was introduced as a Bayesian nonparametric model that determines latent distributions of words i.e. topics that are both semantically and syntactically coherent. The STM models dependency parsed corpora where sentences are grouped into documents. It assumes that each word is drawn from a latent topic chosen by combining document-level features and the local syntactic context. Each document has a distribution over latent topics, as in topic models, which provides the semantic consistency. Each element in the dependency parse tree also has a distribution over the topics of its children, as in latent-state syntax models, which provides the syntactic consistency. These distributions are convolved so that the topic of each word is likely under both its document and syntactic context. A faster posterior inference algorithm has derived based on variational methods. The STM has more predictive model of language than current models based only on syntax or only on topics. Fig. 19. shows the graphical model for STM.

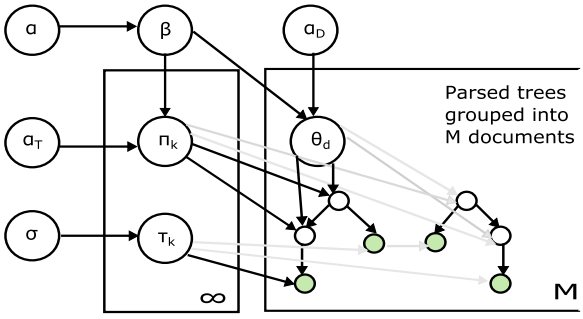

Fig.19. STM Graphical Model

g. Kernel Topic Model (KTM)

In [41], KTM was introduced as a nonparametric regression for modeling topic on document metadata as spatial, hierarchical, social, temporal and other structure between documents. KTM has also referred as a topic model conditional on document features or Gaussian process latent variable model. A lightweight Laplace approximation was used efficiently linked as bridge on a conditionally independent mixture of latent Dirichlet allocation and Gaussian process regression. The strength of kernel formulation in KTM was its applicability to non-Euclidian feature spaces. An inference in KTM was cubic in the number of documents which incurs high computational cost. Fig. 20. shows the graphical model for KTM.

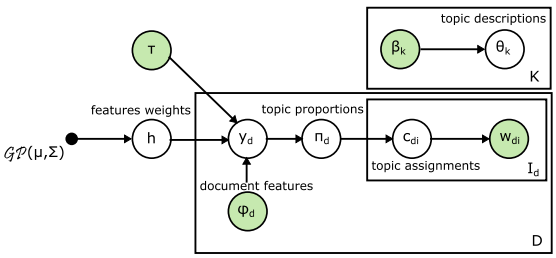

Fig.20. KTM Graphical Model

h. Regularized Latent Semantic Indexing (RLSI)

In [42], RLSI was introduced as topic modeling techniques applicable for larger dataset. RLSI has adopted a new methodology which was designed for scaling to larger document collections via parallelization. The parallelization would be achieved by formulating parallel execution of learning process by decomposing into multiple sub-optimization problems via MapReduce. RLSI has developed topic modeling as a minimization of a quadratic loss function regularized by ℓ1 and/or ℓ2 norm. This regularization by ℓ1 norm applied on topics and ℓ2 norm applied on document representation to create model with readable topics useful for retrieval. RLSI has applied regularization i.e. ℓ1 and/or ℓ2 norm instead of orthogonality and probability distribution used in earlier topic models. Fig. 21. shows the graphical model for RLSI.

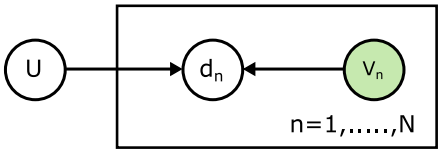

Fig.21. RLSI Graphical Model

i. Modeling Documents with a Deep Boltzmann Machine (TM-DBM)

In [43], topic modeling with a Deep Boltzmann Machine (DBM) was introduced as appropriate for extracting distributed semantic representations from a large amorphous collection of documents. DBM are usually slow to train therefore fast approximate training method makes it feasible to train model with Contrastive Divergence. An efficient pre-training method was incorporated due to parameter sharing between the hidden and visible Softmax units. In this model, the Over-Replicated Softmax model performs better than standard Replicated Softmax model for features extraction. The Over-Replicated Softmax model provides an efficient way of defining a flexible prior over the latent topic features of Restricted Boltzmann Machines. Fig. 22. shows the Over-Replicated Softmax model.

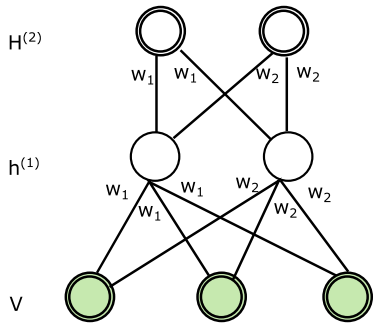

Fig.22. Over-Replicated Softmax model

j. Deep Belief Nets for Topic Modeling (TM-DBN)

In [44], topic modeling with a Deep Belief Nets (DBN) was introduced as generative bag-of-words model and signifies conceptual meanings of documents. DBN has enabled to find a better internal representation of the documents due to the highly non-linear dimensionality reduction because of its deep architecture [45]. The training of model has realized in two phases as pre-training and fine-tuning. In pre-training phase, the approximation of model parameters has achieved. Also, in this phase training was carried out through Gibbs sampling using contrastive divergence as the approximation to the gradient [46]. In fine-tuning phase, the binary output values were computed by adding deterministic Gaussian noise to the output layer which executed the logistic sigmoid activation function gave results close to 0 or 1. Binary latent representation has showed faster results for finding similar documents.

k. Neural Topic Model (NTM)

In [47], NTM was introduced as combination of topic modeling and deep learning techniques. NTM was a neural network topic model consists of two hidden layer for efficiently acquiring the topic-document representation and *n*-gram topic model. The training of the model has been accomplished by applying back-propagation algorithm for adjusting weights and stochastic gradient descent with L_2 norm regularization. The issue of local optima due to back-propagation has been overcome by applying a pre-training procedure using auto-encoder. NTM has shown improved topic representations on document learning tasks such as rating regression, multi-label classification and multi-class classification. Fig. 23. shows the graphical model for NTM.

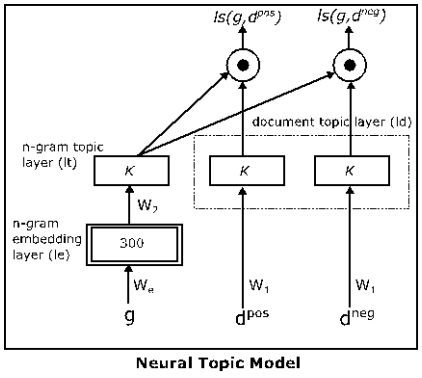

Fig.23. NTM Graphical Model

F. non-LDA based bag of words topic models with supervised learning

a. Supervised Neural Topic Model (sNTM)

In [47], sNTM was introduced as a flexible transformation of the NTM to super-vised tasks by adding a label layer. The training of the model has been accomplished by applying back-propagation algorithm for adjusting weights on the labels of the documents and stochastic gradient descent with ℓ2 norm regularization. The issue of local optima due to back-propagation has been overcome by applying a pre-training procedure using auto-encoder. sNTM has shown improved topic representations on document learning tasks and shown competitive performance on supervised tasks as compared to other supervised topic modeling techniques. Fig. 24. shows the graphical model for sNTM.

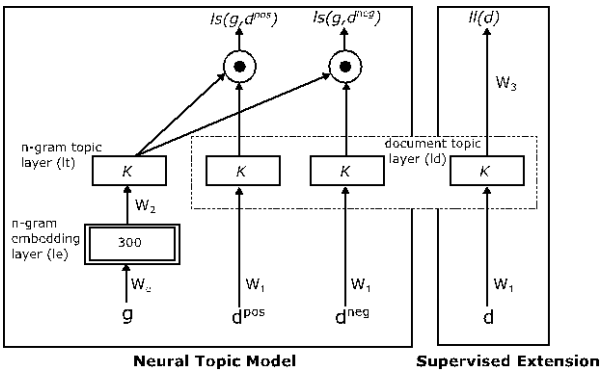

Fig.24. sNTM Graphical Model

G. non-LDA based bag of words topic models with supervised learning

a. Topical N-Grams (TNG)

In [48], TNG was introduced as a generative probabilistic model that tries to alleviate bag-of-words assumption follow in LDA. TNG models N-grams up to arbitrary N instead of LDA which models only unigram. TNG topic model determines topics along with topical phrases. In this model inference has to some extent more complicated than in LDA, but similar approximate inference algorithms are still applicable. The retrieval performance has improved in standard ad-hoc retrieval tasks. TNG have more contented topic representations which enabled the modeling of concepts made of multiple words as compared to earlier probabilistic models but such contention comes at greater computational cost. Fig. 25. shows the graphical model for TNG.

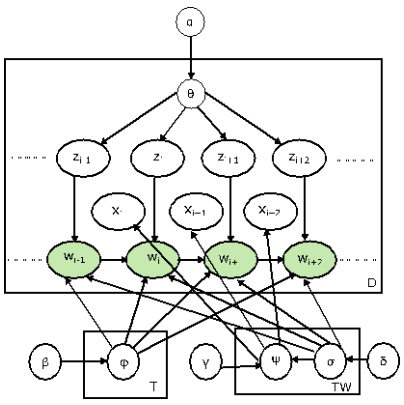

Fig.25. TNG Graphical Model

H. non-LDA based sequence of words topic models with supervised learning

a. Topic segmentation with an aspect hidden Markov model (AHMM)

In [49], AHMM introduced as a novel probabilistic approach for topic classification on unstructured text. Earlier, hidden Markov model (HMM) was used to for probabilistically modeling for sequential data [50]. This model was extended from its predecessor i.e. pLSI into the segmenting HMM to form aspect HMM (AHMM). The Viterbi algorithm [51], a dynamic programming technique, is used to find the most likely hidden sequence of topic states given an observed sequence of word sets. Fig. 26. shows the hidden topic state; *z* as random variable in the trained aspect model.

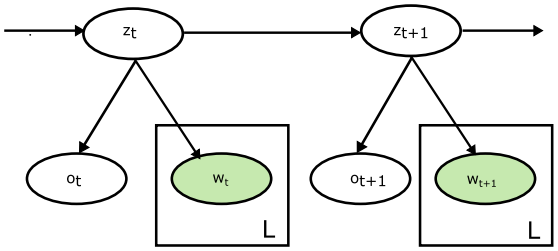

Fig.26. AHMM Graphical Model

b. Supervised Conditional Topic Model (sCdTM)

In [52], sCdTM was introduced as an attempt to improve performance by utilizing the nontrivial input features. The standard maximum-likelihood estimation (MLE) has applied for parameter estimation. sCdTM has used metadata for modeling at word level which allows use of rich feature such as POS tags and ontologies. In this model discriminative max-margin learning method has applied to learn discriminative latent topic models. Fig. 27. shows the graphical model for STM.

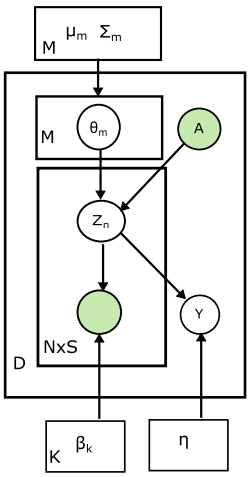

Fig.27. sCdTM Graphical Model

IV. Applications of Topic Modeling

Topic modeling is used for diverse field of study from scientific knowledge discovery to social network analysis to biomedical domain. Here, we are highlighting some of the applications of topic modeling. In [53], LSA model was applied as a review and knowledge extraction to define the taxonomy of research on behavioral operations in supply chain. In [54], pLSA model was applied as an image retrieval to mine the image visual features for each image as a set of visual words from a discrete and finite visual vocabulary. In [55], pLSA was used in automatic question recommendation for recommending the questions to different users. In [56], LDA and ATM model was combines as a role discovery to learn topic distributions based on the communication between senders and receivers. In [57], pairwise-link-LDA model was used to joint modeling of text and citations to generate arbitrary link structure. In [58], LDA model was used for automatic essay grading to solve information retrieval tasks from classification and information filtering to document retrieval. In [59], LDA model was applied as an anti-phishing technique to safeguard the account information, credit card etc. In [60], hierarchical dirichlet process was applied to unstructured nursing notes to discover the topics for mortality prediction and stratify risk for the hospital. In [61], the topic model was applied to drug with analogous side effects are possible to be effective for same disease. Similarity of drugs was found using topic distribution which helps to be replaced the drugs with safer alternatives and in [62], 100 topics was generated from drug labels to identifying the potential adverse effect of drug. In [63], topic modeling technique was applied to uncover the functional groups in each microbiome sample and in [64], latent topics has been identified in wearable wireless sensor devices to track the high-level from low-level physical activities. In [65], topic modeling technique was applied to stock market activity to extract the information that influences the stock market. In [66], topic modeling technique was applied to predicting the popular twitter messages to classifying the users and messages into topical categories. In [67], proposed a method for topic identification in web documents using web design features. In [68], topic modeling was applied to measuring software maintainability.

In summary, topic modeling techniques would be applied in areas like trend analysis of scientific literature, medical, stock market, image retrieval, social networking, and anti-phishing and wireless sensor. These techniques can be further applied to other areas where there is a need to identify the hidden thematic structure in data.

V. Conclusion

A survey on most prominent probabilistic topic modeling techniques is presented and a novel classification is proposed in order to highlight the emergent number of research to discover the latent topics in text corpora. The main motivation to this survey rests in finding a new prospective of research direction. In Table 1, we have summarized the topic modeling techniques from LSA to Deep Learning based on proposed classification along with the method used in each technique for extracting topics with their ability.

Perhaps, the most distinguished contribution of this survey lies in observing the LSS class where none of the models are found. This observation may turn out to be most valuable to the research community for doing further research.

Table 1. Summary of Topic Modeling Techniques

SNo.	Topic Model	C1: L/ NL	C2: B/S	C3: U/S	Method Used	Ability
1	LSA [Deerwester, 1990]	NL	B	U	Singular Value Decomposition	LSA can get from the topic if there are any synonym words.
2	pLSI [Hofmann, 1999]	NL	B	U	Tempered EM	It can generate each word from a single topic; even though various words in one document may be generated from different topics. PLSA handles polysemy.
3	AHMM [Blei, 2001]	NL	S	S	Viterbi and EM	Topical dependency between words.
4	LDA [Blei, 2003]	L	B	U	Variational EM	LDA cannot make the representation of relationships among topics.
5	HLDA [Blei, 2004]	L	B	U	Gibbs sampling	topic hierarchy, number of topics not fixed.
6	ATM [Zvi, 2004], [Zvi, 2010]	L	B	U	Gibbs sampling and MCMC	Topic authorship
7	HLSI [Huang, 2005]	NL	B	U	Hierarchy similarity graph	It preserves the intrinsic structure of the original categorization
8	DTM [Blei, 2006]	L	B	U	Gaussian models, Variational Kalman Filtering and Wavelet Regression	Time evolution of topics
9	CorrTM [Blei, 2006]	L	B	U	Logistic normal distribution, Variational EM	Using of logistic normal distribution to create relations among topics. Allows the occurrences of words in other topics and topic graphs. topic correlations as matrix
10	BLTM [Wallach, 2006]	L	B	U	Gibbs EM algorithm	Extends an unigram model into bigrams.
11	PAM [Li, 2006]	NL	B	U	Gibbs sampling	Topic correlations as DAG
12	BPTM [Nie, 2007]	NL	B	U	EM algorithm	Bigrams topic model that unevenly assigns latent topics to context words
13	TNG [Wang, 2007]	NL	S	U	Gibbs sampling	phrases and n-grams
14	MG-LDA [Titov, 2008]	L	B	U	Collapsed Gibbs sampling	It produces topics based on the aspects of an object rather than global properties of objects
15	SLDA [Blei, 2010]	L	B	S	GLM EM algorithm	Supervised learning of topics
16	STM [Boyd-Graber, 2010]	NL	B	U	Collapsed Gibbs sampling	Topics are generated by combining document specific topic weights and parse-tree specific syntactic transitions
17	sCdTM [Zhu, 2010]	NL	S	S	Variational EM	arbitrary word level features, supervised
18	CTM [Steyvers, 2011]	L	B	U	Collapsed Gibbs sampling	Arbitrary word level features
19	seqLDA [Du, 2011]	L	S	U	Hierarchical two parameter Poisson Dirichlet process and Collapsed Gibbs sampling	It generates sequential topical structure based on multiple segments
20	DMR [Mimno, 2012]	L	B	S	Stochastic EM and Gibbs sampling	Arbitrary document metadata
21	KTM [Henning, 2012]	NL	B	U	Gaussian process and Laplace Bridge	It generates topics based on the temporal, hierarchical, spatial and social structure between documents
22	RLSI [Wang, 2013]	NL	B	U	Distributed computing using MapReduce	It uses l1 and l2-norm regularization and optimized in parallel using MapReduce
23	TM-DBM [Srivastava, 2013]	NL	B	U	MCMC based stochastic approximation	It provides a feature extraction method for classification and retrieval.
24	TM-DBN [Maaloe, 2015]	NL	B	U	Gibbs sampling using contrastive divergence	It provides a better internal depiction of the documents in an output space of lower dimensionality.
25	NTM [Cao, 2015]	NL	B	U	Back propagation, Stochastic gradient descent with L2 norm regularization	Unigram topics are allowed and uniform framework is designed for representation of words and documents.
26	sNTM [Cao, 2015]	NL	B	U	Back propagation, Stochastic gradient descent with L2 norm regularization	Supervised learning of topics
27	SCNTM [Lim & Buntine, 2016]	L	B	S	GEM and HPYP distribution, MCMC and collapsed Gibbs sampler	Bibliographic analysis of authors, topics and documents

REFERENCES

[1] Daud, Ali, Juanzi Li, Lizhu Zhou, and Faqir Muhammad. "Knowledge discovery through directed probabilistic topic models: a survey." Frontiers of computer science in China 4, no. 2 (2010): 280-301.

[2] Blei, David M. "Probabilistic topic models." Communications of the ACM 55, no. 4 (2012): 77-84.

[3] Steyvers, Mark, and Tom Griffiths. "Probabilistic topic models." Handbook of latent semantic analysis 427, no. 7 (2007): 424-440.

[4] Jelisavcic, V., Furlan, B., Protic, J., & Milutinovic, V. M., "Topic Models and Advanced Algorithms for Profiling of Knowledge in Scientific Papers", 35th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO'2012), 1030–1035.

[5] Blei, David M., Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of machine Learning research 3, no. Jan (2003): 993-1022.

[6] Griffiths, D. M. B. T. L., and M. I. J. J. B. Tenenbaum. "Hierarchical topic models and the nested chinese restaurant process." Advances in neural information processing systems 16 (2004): 17.

[7] D. Aldous. Exchangeability and related topics. In ´Ecole d'´ e de probabilit´ et´ es de Saint-Flour, XIII—1983, pages 1–198, (1985) Springer, Berlin.

[8] Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. "The author-topic model for authors and documents." In Proceedings of the 20th conference on Uncertainty in artificial intelligence, pp. 487-494. AUAI Press, 2004.

[9] Rosen-Zvi, Michal, Chaitanya Chemudugunta, Thomas Griffiths, Padhraic Smyth, and Mark Steyvers. "Learning author-topic models from text corpora." ACM Transactions on Information Systems (TOIS) 28, no. 1 (2010): 4.

[10] Blei, David M., and John D. Lafferty. "Dynamic topic models." In Proceedings of the 23rd international conference on Machine learning, pp. 113-120. ACM, 2006.

[11] Blei, David, and John Lafferty. "Correlated topic models." Advances in neural information processing systems 18 (2006): 147.

[12] J. Aitchison, "The statistical analysis of compositional data", Journal of the Royal Statistical Society, Series B, 44(2):139–177, 1982.

[13] Bishop, Christopher M., David Spiegelhalter, and John Winn. "VIBES: A variational inference engine for Bayesian networks." In NIPS, vol. 15, pp. 777-784. 2002.

[14] Wallach, Hanna M. "Topic modeling: beyond bag-of-words." In Proceedings of the 23rd international conference on Machine learning, pp. 977-984. ACM, 2006.

[15] MacKay, David JC, and Linda C. Bauman Peto. "A hierarchical Dirichlet language model." Natural language engineering 1, no. 03 (1995): 289-308.

[16] Titov, Ivan, and Ryan McDonald. "Modeling online reviews with multi-grain topic models." In Proceedings of the 17th international conference on World Wide Web, pp. 111-120. ACM, 2008.

[17] Steyvers, Mark, Padhraic Smyth, and Chaitanya Chemuduganta. "Combining background knowledge and learned topics." Topics in Cognitive Science 3, no. 1 (2011): 18-47.

[18] Mcauliffe, Jon D., and David M. Blei. "Supervised topic models." In Advances in neural information processing systems, pp. 121-128. 2008.

[19] Mimno, David, and Andrew McCallum. "Topic models conditioned on arbitrary features with dirichlet-multinomial regression." arXiv preprint arXiv:1206.3278 (2012).

[20] Lim, Kar Wai, and Wray Buntine. "Bibliographic analysis on research publications using authors, categorical labels and the citation network." Machine Learning 103, no. 2 (2016): 185-213.

[21] Pitman, Jim. "Some developments of the Blackwell-MacQueen urn scheme." Lecture Notes-Monograph Series (1996): 245-267.

[22] Teh, Yee Whye. "A hierarchical Bayesian language model based on Pitman-Yor processes." In Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics, pp. 985-992. Association for Computational Linguistics, 2006.

[23] Du, Lan, Wray Buntine, Huidong Jin, and Changyou Chen. "Sequential latent Dirichlet allocation." Knowledge and information systems 31, no. 3 (2012): 475-503.

[24] Ishwaran, Hemant, and Lancelot F. James. "Gibbs sampling methods for stick-breaking priors." Journal of the American Statistical Association 96, no. 453 (2001): 161-173.

[25] Pitman, Jim, and Marc Yor. "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." The Annals of Probability (1997): 855-900.

[26] Deerwester, Scott, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. "Indexing by latent semantic analysis." Journal of the American society for information science 41, no. 6 (1990): 391.

[27] Cios, K. J., Pedrycz, W., Swiniarski, R. W., & KurganL, A. L., "Data mining: A knowledge discovery approach", New York, NY: Springer, (2007).

[28] Dumais, Susan T. "Latent semantic analysis." Annual review of information science and technology 38, no. 1 (2004): 188-230.

[29] Dumais, Susan T. "LSA and information retrieval: Getting back to basics." Handbook of latent semantic analysis (2007): 293-321.

[30] Han, J., & Kamber, M., "Data mining: Concepts and techniques (2nd ed.)", San 695 Francisco, CA: Morgan Kaufmann Publishers (Elsevier), (2006).

[31] Manning, C. D., Raghavan, P., & Schütze, H., "An introduction to information retrieval", New York, NY: Cambridge University Press (2009).

[32] Landauer, Thomas K. "LSA as a theory of meaning." Handbook of latent semantic analysis (2007): 3-34.

[33] Hofmann, Thomas. "Probabilistic latent semantic indexing." In Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, pp. 50-57. ACM, 1999.

[34] Hofmann, Thomas, Jan Puzicha, and Michael I. Jordan. "Learning from dyadic data." Advances in neural information processing systems (1999): 466-472.

[35] Saul, Lawrence, and Fernando Pereira. "Aggregate and mixed-order Markov models for statistical language processing." arXiv preprint cmp-lg/9706007 (1997).

[36] Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." Journal of the royal statistical society. Series B (methodological) (1977): 1-38.

[37] Huang, Yi, Kai Yu, Matthias Schubert, Shipeng Yu, Volker Tresp, and Hans-Peter Kriegel. "Hierarchy-regularized latent semantic indexing." In Data Mining, Fifth IEEE International Conference on, pp. 8-pp. IEEE, 2005.

[38] Li, Wei, and Andrew McCallum. "Pachinko allocation: DAG-structured mixture models of topic correlations." In Proceedings of the 23rd international conference on Machine learning, pp. 577-584. ACM, 2006.
[39] Nie, Jiazhong, Runxin Li, Dingsheng Luo, and Xihong Wu. "Refine bigram PLSA model by assigning latent topics unevenly." In Automatic Speech Recognition & Understanding, 2007. ASRU. IEEE Workshop on, pp. 141-146. IEEE, 2007.
[40] Boyd-Graber, Jordan L., and David M. Blei. "Syntactic topic models." In Advances in neural information processing systems, pp. 185-192. 2009.
[41] Hennig, Philipp, David Stern, Ralf Herbrich, and Thore Graepel. "Kernel topic models." In Artificial Intelligence and Statistics, pp. 511-519. 2012.
[42] Wang, Quan, Jun Xu, Hang Li, and Nick Craswell. "Regularized latent semantic indexing." In Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp. 685-694. ACM, 2011.
[43] Srivastava, Nitish, Ruslan R. Salakhutdinov, and Geoffrey E. Hinton. "Modeling documents with deep boltzmann machines." arXiv preprint arXiv:1309.6865 (2013).
[44] Maaloe, Lars, Morten Arngren, and Ole Winther. "Deep belief nets for topic modeling." arXiv preprint arXiv:1501.04325 (2015).
[45] Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313, no. 5786 (2006): 504-507.
[46] Hinton, Geoffrey E. "Training products of experts by minimizing contrastive divergence." Neural computation 14, no. 8 (2002): 1771-1800.
[47] Cao, Ziqiang, Sujian Li, Yang Liu, Wenjie Li, and Heng Ji. "A Novel Neural Topic Model and Its Supervised Extension." In AAAI, pp. 2210-2216. 2015.
[48] Wang, Xuerui, Andrew McCallum, and Xing Wei. "Topical n-grams: Phrase and topic discovery, with an application to information retrieval." In Data Mining, 2007. ICDM 2007. Seventh IEEE International Conference on, pp. 697-702. IEEE, 2007.
[49] Blei, David M., and Pedro J. Moreno. "Topic segmentation with an aspect hidden Markov model." In Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 343-348. ACM, 2001.
[50] P. van Mulbregt, I. Carp, L. Gillick, S. Lowe, and J. Yamron, “Text segmentation and topic tracking on broadcast news via a hidden markov model approach”, 1998.
[51] Viterbi, Andrew. "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm." IEEE transactions on Information Theory 13, no. 2 (1967): 260-269.
[52] Zhu, J., Xing, E.P., “Conditional topic random fields”, Proc. 27th Int. Conf. Mach. Learn. 2010, 1239–1246.
[53] Kundu, Anirban, Vipul Jain, Sameer Kumar, and Charu Chandra. "A journey from normative to behavioral operations in supply chain management: A review using Latent Semantic Analysis." Expert Systems with Applications 42, no. 2 (2015): 796-809.
[54] Romberg, Stefan, Eva Horster, and Rainer Lienhart. "Multimodal pLSA on visual features and tags." In Multimedia and Expo, 2009. ICME 2009. IEEE International Conference on, pp. 414-417. IEEE, 2009.
[55] Wu, Hu, Yongji Wang, and Xiang Cheng. "Incremental probabilistic latent semantic analysis for automatic question recommendation." In Proceedings of the 2008 ACM conference on Recommender systems, pp. 99-106. ACM, 2008.
[56] McCallum, Andrew, Xuerui Wang, and Andrés Corrada-Emmanuel. "Topic and role discovery in social networks with experiments on enron and academic email." Journal of Artificial Intelligence Research 30 (2007): 249-272.
[57] Bao, Shenghua, Shengliang Xu, Li Zhang, Rong Yan, Zhong Su, Dingyi Han, and Yong Yu. "Joint emotion-topic modeling for social affective text mining." In Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, pp. 699-704. IEEE, 2009.
[58] Kakkonen, Tuomo, Niko Myller, and Erkki Sutinen. "Applying latent Dirichlet allocation to automatic essay grading." In Advances in Natural Language Processing, pp. 110-120. Springer Berlin Heidelberg, 2006.
[59] Bergholz, Andre, Jeong Ho Chang, Gerhard Paass, Frank Reichartz, and Siehyun Strobel. "Improved Phishing Detection using Model-Based Features." In CEAS. 2008.
[60] Lehman, Li-Wei H., Mohammed Saeed, William J. Long, Joon Lee, and Roger G. Mark. "Risk stratification of ICU patients using topic models inferred from unstructured progress notes." In AMIA. 2012.
[61] Bisgin, Halil, Zhichao Liu, Reagan Kelly, Hong Fang, Xiaowei Xu, and Weida Tong. "Investigating drug repositioning opportunities in FDA drug labels through topic modeling." BMC bioinformatics 13, no. 15 (2012): S6.
[62] Bisgin, Halil, Zhichao Liu, Hong Fang, Xiaowei Xu, and Weida Tong. "Mining FDA drug labels using an unsupervised learning technique-topic modeling." BMC bioinformatics 12, no. 10 (2011): S11.
[63] Chen, Xin, TingTing He, Xiaohua Hu, Yanhong Zhou, Yuan An, and Xindong Wu. "Estimating functional groups in human gut microbiome with probabilistic topic models." IEEE transactions on nanobioscience 11, no. 3 (2012): 203-215.
[64] Kim, Samuel, Ming Li, Sangwon Lee, Urbashi Mitra, Adar Emken, Donna Spruijt-Metz, Murali Annavaram, and Shrikanth Narayanan. "Modeling high-level descriptions of real-life physical activities using latent topic modeling of multimodal sensor signals." In Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE, pp. 6033-6036. IEEE, 2011.
[65] Hisano, Ryohei, Didier Sornette, Takayuki Mizuno, Takaaki Ohnishi, and Tsutomu Watanabe. "High quality topic extraction from business news explains abnormal financial market volatility." PloS one 8, no. 6 (2013): e64846.
[66] Hong, Liangjie, and Brian D. Davison. "Empirical study of topic modeling in twitter." In Proceedings of the first workshop on social media analytics, pp. 80-88. ACM, 2010.
[67] Kazem Taghandiki, Ahmad Zaeri, Amirreza Shirani, “A Supervised Approach for Automatic Web Documents Topic Extraction Using Well-Known Web Design Features", International Journal of Modern Education and Computer Science(IJMECS), Vol.8, No.11, pp.20-27, 2016.DOI: 10.5815/ijmecs.2016.11.03
[68] Mohammad Zavvar, Farhad Ramezani, “Measuring of Software Maintainability Using Adaptive Fuzzy Neural Network", IJMECS, vol.7, no.10, pp.27-32, 2015.DOI: 10.5815/ijmecs.2015.10.04

Authors' Profiles

Deepak Sharma has received his B.E. (in Computer Engineering) and M.Tech. (in Information Technology) from Bharati Vidyapeeth College of Engineering, Bharati Vidyapeeth University, Pune respectively. Currently, he is pursuing Ph.D. in Topic Modeling and Trend Analysis from Department of Computer Engineering, Netaji Subhas Institute of Technology, New Delhi, India. His research interest data mining, natural language processing, text mining, topic modeling.

Bijendra Kumar did his Bachelor of Engineering from H.B.T.I. Kanpur, India. He has done his Ph.D. from Delhi University, Delhi, India in 2011. Presently he is working as a Professor in Computer Engineering Division, Netaji Subhas Institute of Technology, Delhi, India. His areas of research interests are text mining video applications, watermarking, design of algorithms, cloud computing.

Satish Chand did his M.Sc. in Mathematics from Indian Institute of Technology, Kanpur, India and M.Tech. in Computer Science from Indian Institute of Technology, Kharagpur, India and Ph.D. from Jawaharlal Nehru University, New Delhi, India. Presently he is working as a Professor in School of Computer & Systems Sciences, Jawaharlal Nehru University, New Delhi, India. Areas of his research interest are text mining, trend analysis, multimedia broadcasting, networking, video-on-Demand, cryptography, and image processing.

How to cite this paper: Deepak Sharma, Bijendra Kumar, Satish Chand,"A Survey on Journey of Topic Modeling Techniques from SVD to Deep Learning", International Journal of Modern Education and Computer Science(IJMECS), Vol.9, No.7, pp.50-62, 2017.DOI: 10.5815/ijmecs.2017.07.06

