2016 IEEE Second International Conference on Multimedia Big Data

Contextual-LDA: A Context Coherent Latent Topic Model For Mining Large Corpora

Ding Peng
School of Software and
Research Institute of Information Technology
Tsinghua University
Beijing, China
Email: dp13@mails.tsinghua.edu.cn

Dai Guilan, Zhang Yong
Department of Computer Science and Technology and
Research Institute of Information Technology
Tsinghua University
Beijing, China
Email: {daigl, zhangyong05}@mail.tsinghua.edu.cn

***Abstract*—Statistical topic models represented by Latent Dirichlet Allocation (LDA) and its variants are ubiquitously applied to understanding large corpora. Meanwhile, topic models based on bag-of-words (Bow) rarely adopt contextual information, which encompasses enormous amount of serviceable knowledge in a document, into the probabilistic framework. This shortcoming of LDA leads to its failing to learn contextual information in sentences and paragraphs. We present a contextual coherent topic model for text learning namely Contextual Latent Dirichlet Allocation (Contextual-LDA) to include the contextual knowledge without increasing the perplexity of the algorithm very much. In our model, a document is segmented into finely-divided word sequences, each corresponded with one distinct latent topic to capture local context, while the global context is obtained by the location a segment appears in the document. We learn parameters using Gibbs sampling analogous to traditional LDA. Our model takes advantage of statistical strength of BoW through extending LDA without ignoring knowledge contained in the original context of documents. We also demonstrate it in supervised scenario. While comparing to LDA model, experiment results on BBC corpus in both unsupervised and supervised settings reveal our method is finely adapted for text mining.**

***Index Terms*—Topic; Topic Model; Text Mining; Latent Dirichlet Allocation**

I. Introduction

The aim of topic modeling is to discover themes those pervade a collection of documents. This task is inspired by problems of information retrieval and filtering, text summarization and segmentation, natural language processing, and knowledge discovery. A convincing topic model can help gain a precise understanding of the key information from a vast corpus.

A lot of research has been done on topic modeling. Probabilistic topic models represented by probabilistic latent semantic analysis (pLSA) [1] and latent Dirichlet allocation (LDA) [2] [3] and their variants are widely adapted in practice. These unsupervised learning tools make it possible to analyse and summarize a corpus without a priori such as defined categories. Both pLSA and LDA are designed to unsupervisedly learn latent topics in a collection of documents. A statistical topic model regards each document in the corpus as a mixture of same K latent topics in different proportion, and each topic as a distribution over a fixed vocabulary. In addition, a statistical topic model derives a low-dimensional representation of the corpus for further use.

A statistical topic model is described by its fictitious generative process of corpus. pLSA generates the words of a document in a two-stage process using two latent multinomial distribution: 1) randomly choose a latent topic from the per-document topic distribution, 2) randomly choose a word from the chosen-topic word distribution. Both document-topic and topic-term distribution are latent variables in need of estimation to maximize the likelihood of co-occurrence in the corpus, which is usually solved using EM algorithm. LDA model extends pLSA by assuming a uniform Dirichlet priori of topic distributions under a Bayesian framework. To be specific, per-document topic distribution and per-topic word distribution are drawn from two Dirichlet distributions, each governed by a concentration parameter. A symmetric priori over per-topic word distribution makes a prior statement about the sparse extension of topic distribution over words according to the proof in [4], for the sake of topics being distinct and specialized in demand.

However, a major drawback of statical latent topic models is the assumption that each word in the document is generated independently given the latent per-topic word distribution due to the basis of bag-of-words (BoW) model. Ignoring the context is inapposite while the sequence of words contains much more information than merely a word co-occurrence.

For instance, a news report which describes poverty in Somalia and discusses humanitarian aid from France and Saudi Arab could contain several latent topics including: 1) poverty, child, starve, ...; 2) warlord, stagnancy, fragile, ...; 3) pirate, terror, fundamentalism, ...; 4) international, geographical, politic, ...; 5) humanity, concern, money, ... without using large space of text. In traditional topic models, learned latent topics in such scenario are more likely to be coupling, and derive to ambiguous understanding of a large corpus, due to lack of context information. A spatially divided method is able to prevent this since a segment will always describe single topic.

LDA model is also popularly applied to computer vision, where visual words usually fail to capture information over a bigger scope. In this field, researchers pay more attention to utilizing visual context to strengthen efficacy of topic models.

978-1-5090-2179-6/16 $31.00 © 2016 IEEE
DOI 10.1109/BigMM.2016.72

Proposed methods in [5], [6], [7], and [8] model the visual words along with the location or appearance in an image at same time. In this paper, we invite an intelligent and creative frame of mind around 'spatial' into our topic model from these methods on computer vision field.

This paper presents an algorithm, Contextual-LDA, that allows probabilistic topic models using contextual information. We apply a strategy of over-segmentation and clustering to take advantage of both global and local contextual information. In Contextual-LDA, Only one latent topic is assigned to the word tokens within each segment, ensuring the context coherency of the model. We assume that segments describe same topic in a document are more likely gathered. The model regards each document as a two-level structure, the segment level nodes to capture global context over document, and the word level ones to capture local context within a segment.

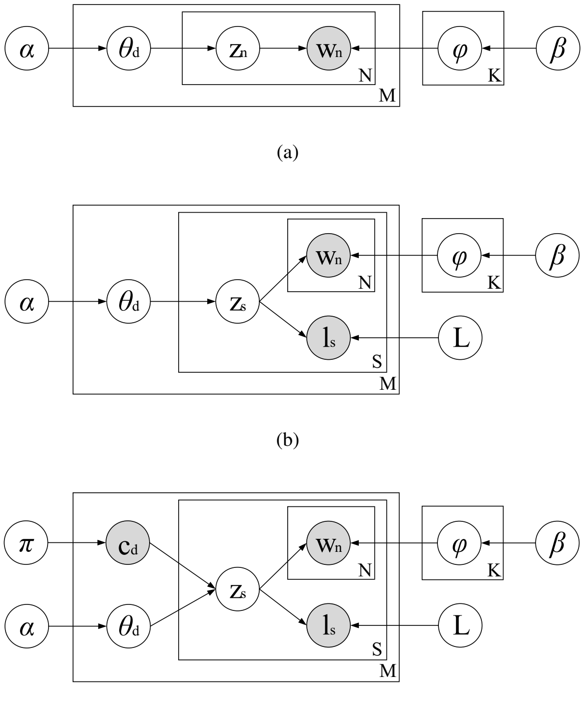
Moreover, there are lots of proposed methods about supervised topic model based on pLSA and LDA, mostly by mixing into the topic's generation step with a response variable associated with each document. Supervised LDA [9] applies response variables from a normal linear model associated with each document into LDA. Labeled LDA [10] deals with multiple labeled corpus by associating each label with one topic in direct correspondence. We also give the supervised methodology based on our algorithm.

The remainder of this paper is organized as follows. We first presents related topic model techniques in Section II. In Section III, the generative model of our Contextual-LDA and algorithm to learn its parameters are presented after a brief overview. Supervised version of our method also goes in Section II. Section IV shows the experimental results of Contextual-LDA under both unsupervised and supervised settings.

II. Related Work

In this section, we introduce how do proposed topic models take contextual or relevant knowledge under consideration. Related approaches powered by tasks in fields beyond text mining are also included.

Correlated topic model (CTM) [11] draws a real valued random vector from a Gaussian distribution instead of a Dirichlet to fix the inability of LDA in modeling topic correlation, though it is not able to capture context in word level. Proposed models not based on LDA sparkle exciting inspirations, yet are still lack of theoretical basis or practice. Bigram topic model (BTM) [12] combines LDA and bigram language model to integrate bigram-based and topic-based approaches. BTM generates word by the definition of $P(w_t = i|w_{t-1} = j, z_t = k)$ to benefit from knowledge of word order. Hidden topic Markov models (HMTM) [13] incorporate HMM into the LDA model to capture local and global context. Admixture of Poisson Markov random fields (APM) [14] takes dependencies between words into consideration under a similar framework with LDA.



(a)

(b)

(c)

Fig. 1. Generative process for unsupervised and supervised Contextual-LDA, along with comparison with LDA. (a) Latent Dirichlet Allocation (LDA). (b) Contextual-LDA. (c) Supervised Contextual-LDA.

Spatial-LDA [5] encodes visual words from clustering groups of image regions to learn spatial structure within a document and co-occurring among image patches. In Spatial-LTM [6], an image was divided into patches of consistent and similar appearance, each allocated single latent topic to ensure coherency of an image region. Spatial DiscLDA (S-DiscLDA) [7], and context aware topic model (CA-TM) [8] extends DsicLDA [15] to model the visual words and their locations in an image at same time. In essence, all these approaches introduce a two-level structure to describe a document or image to capture visual appearance of each patch and influence among neighbors. Within such structures, the definition of spatial or contextual features avoids imperfection of the BoW model.

III. The Contextual-LDA Model

In this section, we present the document generative model of Contextual-LDA and compare it with the traditional LDA model.

To define model the context coherent nature of a document, we enforce words in a restrictive and consecutive segment sharing a same latent topic. A such segment is supposed to be a meaningful unit, such as a sentence/paragraphs, or a sequence of sentences/paragraphs that describe the same topic. It is worth to mentioning that in the extreme case when each document divided into only one segment, it tends to be identical assumption of traditional LDA.

The first step of our algorithm is to over-divide a document into such segments. We choose to apply a domain independent text segmentation approach proposed in [16], which process divisive clustering on a ranking matrix built on cosine similarity measure. The algorithm provides precious result over sparse corpus and is of linear computational complexity. We enforce a loose strategy that accepts more potential boundaries than a standard text segmentation problem, so as to ensure each segment has one rigid distinct topic. We have no intention of bounding particular segmentation algorithm within our model. Indeed, any approach that could generate a rational over-segmentation result could meet our requirements.

A. The Generative Model

After the over-segmentation step, we represent a document in a two-level structure, segment-level and word-level respectively.

Our Contextual-LDA model defines three corpus-level parameters α, β, and L. The former two are Dirichlet priori inherited from LDA, and the latter is added to constrain the segment-level context, while the word-level context is in consideration under the assumption of words in single segment shareing the same topic.

The Contextual-LDA model generates a D-document corpus hypothetically as follows: Firstly, for each document d, we draw a multinomial distribution θ_d over K latent topics from the Dirichlet priori governed by α. Given θ_d, for each segment s in the document, we select a topic z_s. Given segment s and its corresponding z_s, we choose the position l_s of s in document d from the Gaussian distribution denoted by L, which refers a parameter set $\{\mu_k, \sigma_k\}_1^K$. Finally, for each word w_n in the given segment, we pick a token from the word proportion φ for the selected topic z_r, which is drawn from the Dirichlet priori governed by β. The graphical representation of generative process for Contextual-LDA is shown in Figure 1b.

There are two significant differences between LDA and our Contextual-LDA model. Firstly, we assign latent topics to a sequential words, i.e. a segment instead of a single word to encode local context. Secondly, we adopt a topic-relevant distribution L to encode global context.

The joint distribution of $\{z_s, l_s, w_n\}$ given corpus d is written as

$$\begin{aligned} p(z_s, l_s, &w_n|\Theta, \Psi, L) \\ &= \prod_{s=1}^{S} \left(p(z_s|\Theta, d) p(l_s|L, z_s) \prod_{n=1}^{N} p(w_n|\Psi, z_s) \right) \end{aligned}$$

in which Θ, Ψ are multinomial distribution sets drawn over per-document topic priori and per-topic word priori respectively, i.e. α and β. These two sets produce $p(z_s|\Theta, d)$ represented as θ_d in Figure 1b that describes the proportion of topics in document d, and $p(w_n|\Psi, z_s)$ that describes the proportion of words in topic z_s. $p(l_s|L, z_s)$ denotes the Gaussian distribution in L that corresponded with topic z_s. Given l_s and w_n, only observed variables apparently, we find the maximum likelihood estimation for Θ, Ψ, and L as

$$\{\Theta^\star, \Psi^\star, L^\star\} = \underset{\Theta, \Psi, L}{\operatorname{argmax}} \, p(\boldsymbol{l}, \boldsymbol{w}|\Theta, \Psi, L)$$

TABLE I
NOTATIONS

Variables	Notations
M	Number of documents
S_d	Number of finely-divided segments in document d
N_s	Number of words in segments s
V	Number of words in vocabulary
K	Number of latent topics
C	Number of categories in supervised senario
α, β	Parameters of per-document topic and per-topic word Drichlet priori
Θ, Ψ	Per-document topic and per-topic word Drichlet priori, as in $M \times K$ and $K \times V$ matrix
π	Parameters of per-category word Dirichlet priori
L	Parameters of per-topic location Gaussian priori
z_s	The latent topic of s associated to segments
l_s	The centesimal position of s in document d
w_n	The n-th word in segment s
c_d	The category label for document d

B. Parameters Learning

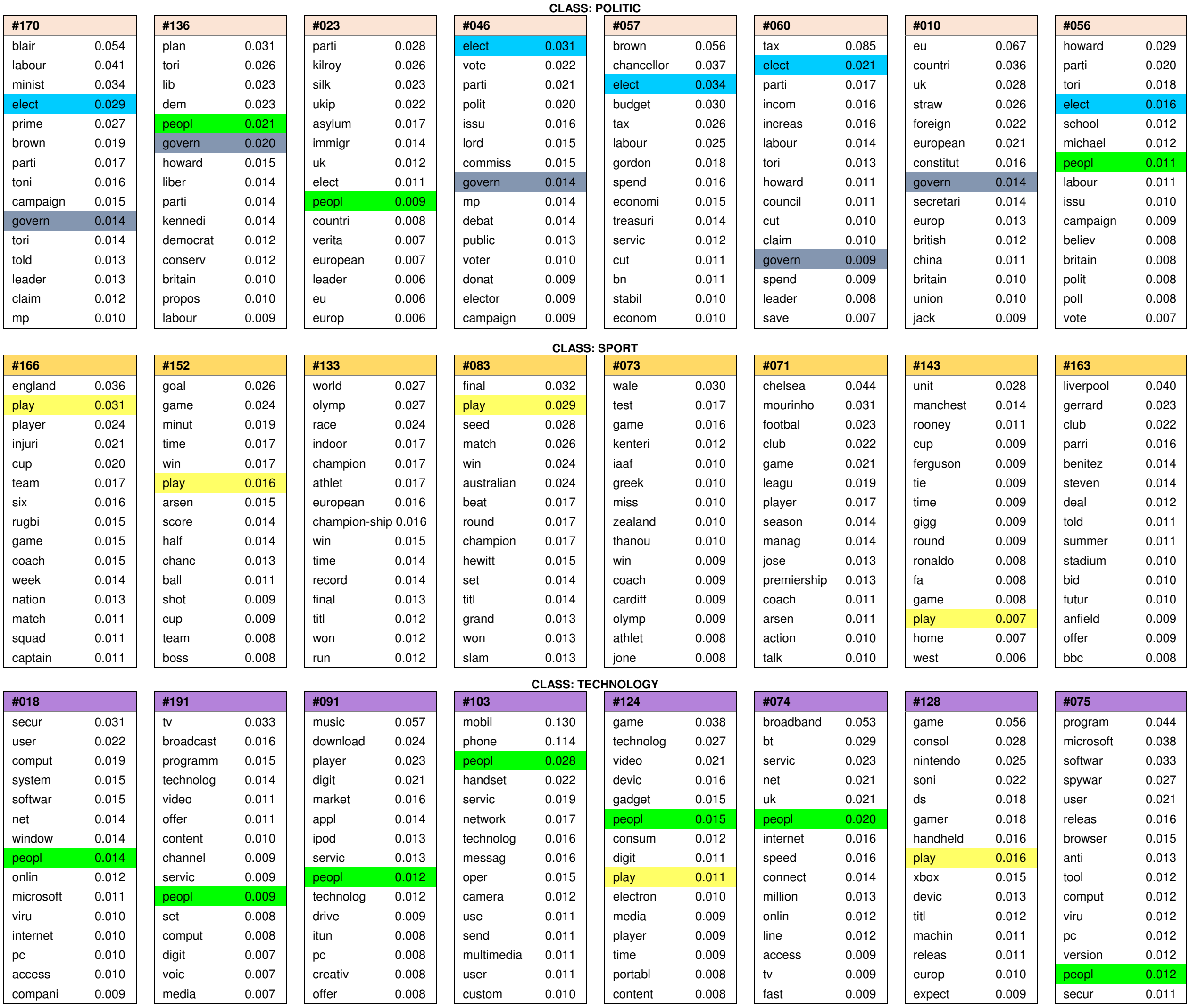
We resort Gibbs sampling, a Markov-Chain Monte Carlo algorithm, to estimate the formulated variables in Section III-A over the posteriori distribution. Given a corpus represented by $\{\boldsymbol{w_d}, \boldsymbol{l_d}\}_{d=1}^M$, where $\boldsymbol{w_d}$ stands for a sequence of segments as in bag-of-word vectors in d, and $\boldsymbol{l_d}$ stands for a sequence of in-document position of corresponding segments in the former. We have the sampling pattern as

$$\begin{aligned} &p(z_s = k|\boldsymbol{z_{-s}}, \boldsymbol{w_d}, \boldsymbol{l_d}, L) \\ &= p(l_s|z_s, L) \cdot \frac{p(\boldsymbol{w_d}|\boldsymbol{z_d})}{p(\boldsymbol{w_{-s}}|\boldsymbol{z_{-s}})} \cdot \frac{p(\boldsymbol{z_d})}{p(-z_s)} \\ &\propto p(l_s|z_s, L) \cdot \frac{\sum_{n=1}^{N_s} \left(\beta_{w_n,k} + q_{-w_n,k}^{(w)} \right)}{\sum_{n=1}^{N_d} \left(\beta_{w_n,k} + q_{-s,k}^{(w)} \right)} \cdot (\alpha_k + q_{-s,d}^{(k)}) \end{aligned}$$

in which $q_{-s,d}^{(k)}$ denotes number of times that latent topic $z_s = k$ assigned to segments in document d expect for s, while $q_{-s,k}^{(w)}$ denotes number of words associated with topic $z_s = k$ expect for those in s, and $q_{-w_n,k}^{(w)}$ denotes number of words associated with topic $z_s = k$ expect for those with value w_n. N_s and N_d stand for the amount of words in s and d respectively.

Given the samples from the posteriori we can estimate Ψ and Θ as in traditional LDA. We give only estimation for the former on account of symmetry of LDA framkework. Since $p(\varphi_k|\boldsymbol{z}, \boldsymbol{w}) \sim Dir(\{\beta_{v,k} + q(s,k)\}_{s=1}^S)$, we can estimate Ψ and Θ as

$$\begin{cases} \varphi_k(w) = \dfrac{q(w,k) + \beta}{q(k) + S\beta} \\ \vartheta_m(k) = \dfrac{q(k,m) + \alpha}{q(m) + S\alpha} \end{cases}$$



CLASS: POLITIC

#170		#136		#023		#046		#057		#060		#010		#056	
blair	0.054	plan	0.031	parti	0.028	elect	0.031	brown	0.056	tax	0.085	eu	0.067	howard	0.029
labour	0.041	tori	0.026	kilroy	0.026	vote	0.022	chancellor	0.037	elect	0.021	countri	0.036	parti	0.020
minist	0.034	lib	0.023	silk	0.023	parti	0.021	elect	0.034	parti	0.017	uk	0.028	tori	0.018
elect	0.029	dem	0.023	ukip	0.022	polit	0.020	budget	0.030	incom	0.016	straw	0.026	elect	0.016
prime	0.027	peopl	0.021	asylum	0.017	issu	0.016	tax	0.026	increas	0.016	foreign	0.022	school	0.012
brown	0.019	govern	0.020	immigr	0.014	lord	0.015	labour	0.025	labour	0.014	european	0.021	michael	0.012
parti	0.017	howard	0.015	uk	0.012	commiss	0.015	gordon	0.018	tori	0.013	constitut	0.016	peopl	0.011
toni	0.016	liber	0.014	elect	0.011	govern	0.014	spend	0.016	howard	0.011	govern	0.014	labour	0.011
campaign	0.015	parti	0.014	peopl	0.009	mp	0.014	economi	0.015	council	0.011	secretari	0.014	issu	0.010
govern	0.014	kennedi	0.014	countri	0.008	debat	0.014	treasuri	0.014	cut	0.010	europ	0.013	campaign	0.009
tori	0.014	democrat	0.012	verita	0.007	public	0.013	servic	0.012	claim	0.010	british	0.012	believ	0.008
told	0.013	conserv	0.012	european	0.007	voter	0.010	cut	0.011	govern	0.009	china	0.011	britain	0.008
leader	0.013	britain	0.010	leader	0.006	donat	0.009	bn	0.011	spend	0.009	britain	0.010	polit	0.008
claim	0.012	propos	0.010	eu	0.006	elector	0.009	stabil	0.010	leader	0.008	union	0.010	poll	0.008
mp	0.010	labour	0.009	europ	0.006	campaign	0.009	econom	0.010	save	0.007	jack	0.009	vote	0.007

CLASS: SPORT

#166		#152		#133		#083		#073		#071		#143		#163	
england	0.036	goal	0.026	world	0.027	final	0.032	wale	0.030	chelsea	0.044	unit	0.028	liverpool	0.040
play	0.031	game	0.024	olymp	0.027	play	0.029	test	0.017	mourinho	0.031	manchest	0.014	gerrard	0.023
player	0.024	minut	0.019	race	0.024	seed	0.028	game	0.016	footbal	0.023	rooney	0.011	club	0.022
injuri	0.021	time	0.017	indoor	0.017	match	0.026	kenteri	0.012	club	0.022	cup	0.009	parri	0.016
cup	0.020	win	0.017	champion	0.017	win	0.024	iaaf	0.010	game	0.021	ferguson	0.009	benitez	0.014
team	0.017	play	0.016	athlet	0.017	australian	0.024	greek	0.010	leagu	0.019	tie	0.009	steven	0.014
six	0.016	arsen	0.015	european	0.016	beat	0.017	miss	0.010	player	0.017	time	0.009	deal	0.012
rugbi	0.015	score	0.014	champion-ship	0.016	round	0.017	zealand	0.010	season	0.014	gigg	0.009	told	0.011
game	0.015	half	0.014	win	0.015	champion	0.017	thanou	0.010	manag	0.014	round	0.009	summer	0.011
coach	0.015	chanc	0.013	time	0.014	hewitt	0.015	win	0.009	jose	0.013	ronaldo	0.008	stadium	0.010
week	0.014	ball	0.011	record	0.014	set	0.014	coach	0.009	premiership	0.013	fa	0.008	bid	0.010
nation	0.013	shot	0.009	final	0.013	titl	0.014	cardiff	0.009	coach	0.011	game	0.008	futur	0.010
match	0.011	cup	0.009	titl	0.012	grand	0.013	olymp	0.009	arsen	0.011	play	0.007	anfield	0.009
squad	0.011	team	0.008	won	0.012	won	0.013	athlet	0.008	action	0.010	home	0.007	offer	0.009
captain	0.011	boss	0.008	run	0.012	slam	0.013	jone	0.008	talk	0.010	west	0.006	bbc	0.008

CLASS: TECHNOLOGY

#018		#191		#091		#103		#124		#074		#128		#075	
secur	0.031	tv	0.033	music	0.057	mobil	0.130	game	0.038	broadband	0.053	game	0.056	program	0.044
user	0.022	broadcast	0.016	download	0.024	phone	0.114	technolog	0.027	bt	0.029	consol	0.028	microsoft	0.038
comput	0.019	programm	0.015	player	0.023	peopl	0.028	video	0.021	servic	0.023	nintendo	0.025	softwar	0.033
system	0.015	technolog	0.014	digit	0.021	handset	0.022	devic	0.016	net	0.021	soni	0.022	spywar	0.027
softwar	0.015	video	0.011	market	0.016	servic	0.019	gadget	0.015	uk	0.021	ds	0.018	user	0.021
net	0.014	offer	0.011	appl	0.014	network	0.017	peopl	0.015	peopl	0.020	gamer	0.018	releas	0.016
window	0.014	content	0.010	ipod	0.013	technolog	0.016	consum	0.012	internet	0.016	handheld	0.016	browser	0.015
peopl	0.014	channel	0.009	servic	0.013	messag	0.016	digit	0.011	speed	0.016	play	0.016	anti	0.013
onlin	0.012	servic	0.009	peopl	0.012	oper	0.015	play	0.011	connect	0.014	xbox	0.015	tool	0.012
microsoft	0.011	peopl	0.009	technolog	0.012	camera	0.012	electron	0.010	million	0.013	devic	0.013	comput	0.012
viru	0.010	set	0.008	drive	0.009	use	0.011	media	0.009	onlin	0.012	titl	0.012	viru	0.012
internet	0.010	comput	0.008	itun	0.008	send	0.011	player	0.009	line	0.012	machin	0.011	pc	0.012
pc	0.010	digit	0.007	pc	0.008	multimedia	0.011	time	0.009	access	0.009	releas	0.011	version	0.012
access	0.010	voic	0.007	creativ	0.008	user	0.011	portabl	0.008	tv	0.009	europ	0.010	peopl	0.012
compani	0.009	media	0.007	offer	0.008	custom	0.010	content	0.008	fast	0.009	expect	0.009	secur	0.011

Fig. 2. Top topics learned organized by their target categories through Contextual LDA. Top 15 terms and their probabilities for each topic are illustrated. While 'play' and 'peopl' appear in 2 different categories of top topics, 'govern' and 'elect' appear only in top topics of subset labelled *politic*. Notice all terms are stemmed while pre-processing the corpus.

C. Supervised Model

There are two ways to learn categories of documents. The straightforward method is combining traditional classifier with computed latent variables from Contextual-LDA. Our method provides a per-document topic multinomial distribution Θ governed by β, which can be regarded as a dimensionality reduction process as same as every each statistical topic models. Given a proper category labeled corpus, one can adopt Θ as selected features to train any modern classifier that supports standard input.

The rest of this section demonstrates another method to learn labels or categories by adding a c_d node into topic z_s generation steps. Generative model for supervised Contextual-LDA is represented in Figure 1c. The difference between generative process in unsupervised and supervised Contextual-LDA is that a distribution c selected from a Dirichlet priori governed by parameter π is also taken into account while drawing a latent topic from θ_d. Each category label c is assigned to a distribution of K latent topics, in other words, each row in a $C \times K$ matrix stands for a Dirichlet distribution θ corresponded to category c, in which C denotes number of categories upon the corpus.

IV. Experiments

We evaluate our Contextual-LDA model on the BBC news dataset [17]. The BBC dataset consists of 265,351 meaningful terms in 2,225 documents from the BBC news website corresponding to stories in five topical areas between 2004 and 2005. All reports are categorized into five labeled class including *business, entertainment, politics, sports* and *technology*.

In pre-processing of our experiments, all terms are stemmed using the nltk[1] library.

A. Latent Topic Allocation

We locate latent topics to the entire corpus ignoring labels for each document. Representative learned topic samples are shown in Figure 2. The topic word distribution covers five labels annotated in the corpus. Top 8 topics pervade in each subset with categories of *politics, sports* and *technology* are selected to be illustrated. These topics are representative features for each label while multiple occurrences of top topics in various categories do not exist. Terms with higher frequency reveals an interesting tricks. While ‘play’ and ‘peopl’ appear in 2 different categories of top topics, ‘govern’ and ‘elect’ appear only in top topics of subset labelled *politic*. This phenomenon can be evidence of the robustness of our model.

Figure 3a illustrates per-word held out log likelihood comparison for Contextual-LDA and LDA, given various numbers of topics, and Figure 3b shows perplexity comparison for the two models. These empirical results indicates that our approach fits the corpus better than LDA.

B. Text Classification

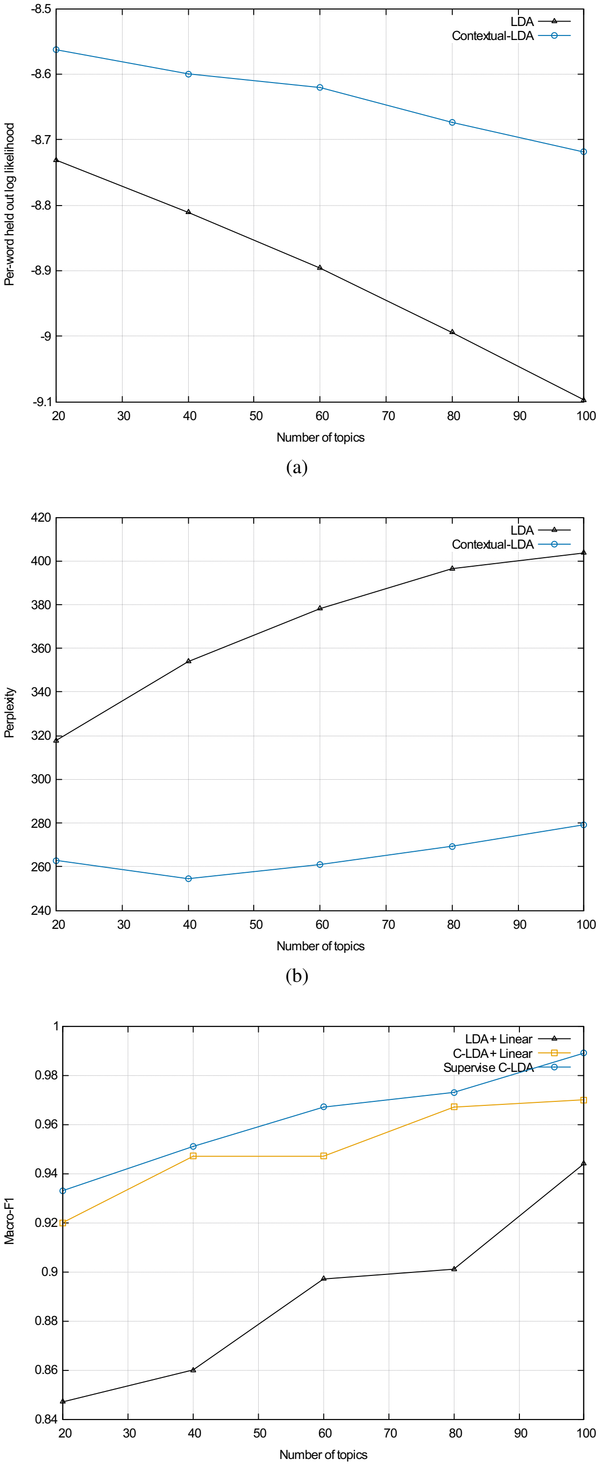
We cast text classification tasks upon the BBC dataset through LDA and Contextual-LDA. Instead of using Supervised LDA or labelled LDA, we adopt linear classifier approach on LDA latent parameters. Θ represents a set of K-dimensional features for M documents. The comparison of classification results exposes that our approach provides better features. Figure 3c shows the classification results we obtain through both LDA and our method. We adopt MacroF1 to measure the accuracy of these two algorithms under various numbers of topics.

The influence of K is non-ignorable. The experiment represents that the Supervised Contextual LDA is highly robust on various range of K. The results indicate that our method outperforms linear classifier on LDA dimensionality reduction. Boosting our approach with a classic classifier such as Naive Bayes or SVM provides also significant contribution.

V. Conclusion

The Contextual-LDA proposed in this paper is an effective and practical model for detecting key information from a corpus. Contextual-LDA gains the merit of classic probabilistic model and benefits from valuable contextual information. When compared to traditional LDA model, experimental results shows that latent topics learned by Contextual-LDA are more convincing and less perplexing. Our approach also provides flying scores under supervised settings.

Possible future directions for this work include limiting the generative model of documents with more certified principle and accelerating the process of Gibbs sampling. As a result of over-segment stage enlarging the set size of latent variables to estimate, the Contextual-LDA model is more likely slower



(a)

(b)

(c)

Fig. 3. (a) Per-word held out log likelihood comparison for Contextual-LDA and LDA. (b) Perplexity comparison for Contexutual-LDA and LDA. (c) Supervised Contextual LDA Performance on BBC dataset for predicting 5 tags compared with LDA + linear classifier.

[1] http://www.nltk.org/

than a traditional LDA method. It would be of more importance to improve efficiency and running speed of parameters learning for larger corpora.

References

[1] T. Hofmann, "Probabilistic latent semantic indexing," in *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 1999, pp. 50–57.

[2] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.

[3] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012.

[4] H. M. Wallach, D. Minmo, and A. McCallum, "Rethinking lda: Why priors matter," 2009.

[5] X. Wang and E. Grimson, "Spatial latent dirichlet allocation," in *Advances in Neural Information Processing Systems*, 2008, pp. 1577–1584.

[6] L. Cao and L. Fei-Fei, "Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.

[7] Z. Niu, G. Hua, X. Gao, and Q. Tian, "Spatial-disclda for visual recognition," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 1769–1776.

[8] ——, "Context aware topic model for scene recognition," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2743–2750.

[9] J. D. Mcauliffe and D. M. Blei, "Supervised topic models," in *Advances in neural information processing systems*, 2008, pp. 121–128.

[10] D. Ramage, D. Hall, R. Nallapati, and C. D. Manning, "Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora," in *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, 2009, pp. 248–256.

[11] D. M. Blei and J. D. Lafferty, "A correlated topic model of science," *The Annals of Applied Statistics*, pp. 17–35, 2007.

[12] H. M. Wallach, "Topic modeling: beyond bag-of-words," in *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006, pp. 977–984.

[13] A. Gruber, Y. Weiss, and M. Rosen-Zvi, "Hidden topic markov models," in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 163–170.

[14] D. Inouye, P. Ravikumar, and I. Dhillon, "Admixture of poisson mrfs: A topic model with word dependencies," in *Proceedings of The 31st International Conference on Machine Learning*, 2014, pp. 683–691.

[15] S. Lacoste-Julien, F. Sha, and M. I. Jordan, "Disclda: Discriminative learning for dimensionality reduction and classification," in *Advances in neural information processing systems*, 2009, pp. 897–904.

[16] F. Y. Choi, "Advances in domain independent linear text segmentation," in *Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference*. Association for Computational Linguistics, 2000, pp. 26–33.

[17] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proc. 23rd International Conference on Machine learning (ICML'06)*. ACM Press, 2006, pp. 377–384.