Activating Topic Models from a Cognitive Perspective

Jie Bai[1, 2] Linjing Li[1] Daniel Zeng[1, 2, 3]

[1] The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China

[2] School of Computer and Control Engineering, University of Chinese Academy of Sciences, China

[3] Department of Management of Information Systems, The University of Arizona, USA

{baijie2013, linjing.li, dajun.zeng}@ia.ac.cn

***Abstract*—Topic modeling is a popular text mining technique for extracting latent semantics from text. It can be widely applied in intelligence analyzing, anti-terrorist, and various other security related tasks. Most existing topic models only focus exclusively on the text literally, and disregard rich contextual, cultural, and language background, hindering the understanding and discovering of the key clues implied in the text. Based on cognitive psychology theories, we justify the classical psychological activation theory named Adaptive Control of Thought from the perspective of information theory. Then, we propose a fast and loosely-coupled activation presentation of text for topic models. Our method mimics the aspect of human cognitive procedure when facing the activation of new concepts based on word correlations and word frequencies. Experimental results on multiple tasks show that our activation presentation models can significantly improve the performance of the topic models with linear time consumption.**

Keywords—topic model; text presentation; activation theory; cognitive psychology

I. Introduction

Topic modeling, represented by Latent Dirichlet Allocation (LDA), is a set of popular techniques for latent semantics analysis and modeling. In a typical topic model, each latent topic is defined as a distribution over words. Hence documents can be transformed to distributions over topics, which provides a macroscopic presentation for the documents as well as a distinct feature space for various text mining tasks. Topic modeling has been widely adopted in various text mining applications and can be applied in many security related tasks, such as intelligence analyzing and anti-terrorist.

Most topic modeling techniques are focusing on learning documents literally and ignore the contextual, cultural, and language background, which are far from accomplishing semantic text mining. As a result, some key clues implied in the text data is hard to understood and discovered. For example, the word "LOL" (the abbreviation of "Laugh Out Loud") usually appears in tweets which is an important signal of surprised or amused sentiments, but it can hardly be caught in text mining works without a specific knowledge base. The lack of linkage between explicit text and implicit concepts leads to a critical bottleneck hindering semantic analysis. In this paper, we attempt to apply the ideas from cognitive theory to enhance the performance of topic modeling and enable a stronger form of latent semantic analysis.

Some cognitive psychological theories maintain that human reading is a process of interactions between the text seen by the reader and the background knowledge kept in the reader's mind. Specifically, the internal knowledge that the reader already has forms the *long-term memory*, according to the psychological terminology, and the external words being read are the *stimulants*. The interaction is a process of activating concepts from long-term memory, based on the given stimulants, which forms the *short-term working memory* ultimately. Then the cognition process of reading involves how difference pieces of internal information in the long-term memory be retrieved based on the external stimulants [1]. Various theories has been proposed to describe the process, among which the Adaptive Control of Thought (ACT) has been widely-accepted [2]. ACT claims that a specific concept can be jointly activated by a set of stimulants with the associations between the stimulants and the concept.

Inspired by the ACT theory, we propose a novel activation presentation to learn implicit information of the given text. An activation presentation model can be treated as a loosely-coupled component of the topic models, which provides a flexible solution for improving the effectiveness of topic modeling. Intuitively, a single document may have some implicit information, but a large collection of documents usually contains massive explicit information which may help understanding of the implicit information for the document. We start from demonstrating the activation theory from the information theory point of view and deducing the key factors to enable activation of implicit information. Our basic intuition is that the more implicit information we exploit from the text, the larger reduction of document perplexity we shall achieve. Based on the inferences under different assumptions, we propose a set of activation presentation models accordingly. The experimental results partly prove our intuition and show that our models bring significant improvements on the effectiveness of topic modeling.

The rest of this paper is organized as follows. Section 2 introduces the related works. In Section 3, we present the related activation theory and our justification on it from the point of view of information theory. Section 4 provides details on our models. Experiments and findings are presented in Section 5 and Section 6 concludes the paper.

978-1-5090-3865-7/16/$31.00 ©2016 IEEE

II. RELAED WORKS

The works investigated in this section are mainly from two research fields, which are 1) topic modeling techniques and 2) cognitive activation theories and their applications in text analysis.

A. Topic modeling

Topic modeling has been extensively studied for long. Latent Semantic Indexing (LSI) [3] was proposed to analyze the high-order term-document associations by taking advantage of Singular-Value Decomposition, which was demonstrat that the model reveals the "topic" of the document [4]. Probabilistic LSI (pLSI) was developed to generate LSI models statistically [5]. Later on, Latent Dirichlet allocation (LDA) [6] generalized pLSI by adding two Dirichlet priors, which is the most common topic model currently in use. Since proposed, LDA has been extensively developed in many aspects, such as Dynamic Topic Model [7] and Supervised LDA [8].

However, the works mentioned above are limited in analyzing literal documents only, and the semantics embedded in word associations are out of consideration. To address this issue, some approaches has been proposed to explore how to improve the effectiveness of the topic modeling with the help of word associations. Andrzejewski et al. applied Dirichlet Forest prior to model Must-Links and Cannot-Links between words, which confine the probabilities that similar words and dissimilar words appear in the same topic [9]. Newman et al. imposed a quadratic regularizer to incorporate word correlations in the topic-word multinomial [10]. Xie et al. embedded the Markov Random Field into LDA latent topic layer, which guides the topic-word distributions by word correlations adaptively [11]. Different from these integrative topic modeling approaches, we propose a loosely-coupled augment presentation for topic models by taking advantage of cognitive psychological theories.

B. Cognitive Activation Theories and Their Applications

The idea of memory activation or recall mechanism, was born in early 1970's when the theory of human cognition was investigated. A representative model is Human Associative Memory (HAM) theory [12], which assumes that one searches the items in long-term memory and evaluates the contextual associations of the items to determine which should be recalled. ACT theory [2] was proposed to include a production rule system of procedural memory to complement HAM's declarative memory.

How memory activation theory can be applied to understand natural languages is not well-studied. Baddeley proposed to use HAM to relate the mechanisms found in short-term working memory to sentence comprehension [13]. Cambria et al. employed dual-process (unconscious and conscious) models into sentiment analysis, and proposed a two-level affective reasoning framework [14]. Zhang et al. built a three-level text presentation framework TRMBK [15] based on the cognitive architecture proposed by Kinstch and Dijk [16]. Compared with these previous researches, we focus on applying memory activation theory to text presentation serving topic modeling tasks.

III. ACTIVATION THEORY AND ITS INSIGHTS

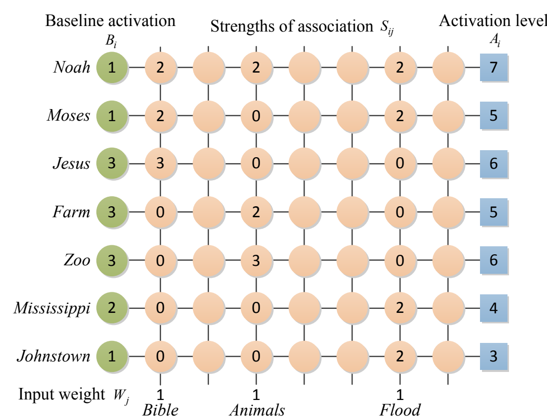
We begin with the introduction of ACT theory. For the sake of convenience, all the notations involved in this paper are summarized in Table 1.

ACT theory qualitatively explain how activation works in human brain. For any potential target word y_i, the activation level A_i can be calculated in a concept network according to the following equation.

$$A_i = B_i + \sum_j W_j S_{ij} \tag{1}$$

Where B_i reflects the activation baseline of the potential target y_i, W_j reflects the weight given to stimulant w_j, and S_{ij} reflects the strength of associations between stimulant w_j and the potential target y_i. The ACT calculation procedure is illustrated in Figure 2. The sample words shown on the figure are from the illustrative example of "Bible, Animals, and Flood" [1]. From the example illustrated by Figure 2 we can explain how the word "Noah" is jointly activated by words "Bible, animals, flood". Although for "Noah" the Baseline activation is relatively low compared with the other words, it has high associations with "Bible", "Animals" and "Flood" simultaneously. Thus the activation level of "Noah" is higher than the other words.

Fig. 1. Example of ACT calculation procedure



Inspired by the ACT theory, we assume that one can activate some implicit information for a single document with the whole collection of documents as the knowledge base. However, there are still some confusion before we apply the ACT theory to topic modeling. On the one hand, we need to technically explain why it can help topic modeling; on the other hand, the qualitative parameters in (1), i.e., baseline activation B_i, input weight W_j and association S_{ij} need to be redefined quantitatively. In the rest of this section, we shall propose a formal justification of ACT in the framework of information theory, which also inspires us the clue of how to choose the parameters involved.

Intuitively, adding a set of implicit words to a document will help improving the performance of text analysis. To quantitatively measure the performance we employ perplexity, which is derived from entropy and commonly used to evaluate probability language model. The perplexity of a given corpus D is defined as

TABLE I. NOTATION SUMMARY

Symbol	Definition	Meaning
D	$D=\{d_1,d_2,...,d_M\}$	Corpus, constructed by collection of documents
$\boldsymbol{V}$	$\boldsymbol{V}=\{w_1,w_2,...,w_V\}$	Vocabulary of the corpus, constructed by a word collection
d	$d=\{w_1,w_2,...,w_{N_d}\}_{seq}$	Document, constructed by a word sequence
$\boldsymbol{w}$	$\boldsymbol{w}=\{w_1,w_2,...,w_{N_w}\}$	Stimulant words set
$\boldsymbol{y}$	$\boldsymbol{y}=\{y_1,y_2,...,y_{N_y}\}$	Implicit words set given d, where $y_i \in \boldsymbol{V}, i=1,2,...,N_y$
$\mathbf{v}_{w_i}$	$\mathbf{v}_{w_i}=\{v^1_{w_i},v^2_{w_i},...,v^L_{w_i}\}$	Word embedding of the specific word w_i
M		The total number of documents that D has
N_d (N_D, $N_{\boldsymbol{w}}$, $N_{\boldsymbol{y}}$)		The total number of words that d (D, $\boldsymbol{w}$, $\boldsymbol{y}$) has
N^y_D		The frequency that word y appears D
V		Vocabulary size
L		The dimensionality of the word embeddings
A^d_y		The activation level of a specific word y, given d
$S_{i,j}$ / $S_{y,\boldsymbol{w}}$		The association of the specific words w_i and w_j /the word y and the word collection $\boldsymbol{w}$
α	$\alpha=N_{\boldsymbol{y}}/N_{\boldsymbol{w}}$	The activation rate α decides how many implicit words are activated from per stimulant word
β	$\beta\in(0,1)$	The hyper parameter that tradeoff between the two factors involved in the activation function

$$Perp(D)=\exp\{-\frac{\sum_{j=1}^{N_D}\log p(w_j)}{N_D}\} \tag{2}$$

Following the same logic, we can also define the perplexity for a single document d as

$$Perp(d)=\exp\{-\frac{\sum_{j=1}^{N_d}\log p(w_j)}{N_d}\} \tag{3}$$

As for a specific document d, our goal is to activate appropriate implicit words set $\boldsymbol{y}$ given the stimulant words set $\boldsymbol{w}$, so as to reduce the document perplexity, i.e., $Perp(\boldsymbol{w},\boldsymbol{y})<Perp(\boldsymbol{w})$. This can be formulated with (3) as:

$$\exp\{-\frac{\sum_{j=1}^{N_d}\log p(w_j)+\sum_{i=1}^{N_y}\log p(y_i\mid\boldsymbol{w})}{N_d+N_{\boldsymbol{y}}}\}<\exp\{-\frac{\sum_{j=1}^{N_d}\log p(w_j)}{N_d}\} \tag{4}$$

It can be proved that (4) is equivalent to:

$$\frac{\sum_{i=1}^{N_y}\log p(y_i\mid\boldsymbol{w})}{N_{\boldsymbol{y}}}>\frac{\sum_{j=1}^{N_d}\log p(w_j)}{N_d} \tag{5}$$

Then we analyze (5) under different assumptions.

Assumption 1: $\boldsymbol{w}$ and $\boldsymbol{y}$ are independent identically distributed respectively. i.e.:

$$p(w_1)=p(w_2)=...=p(w_{N_d})=p(y_1)=p(y_2)=...=p(y_{N_y}),$$
$$p(y_1\mid\boldsymbol{w})=p(y_2\mid\boldsymbol{w})=...=p(y_{N_y}\mid\boldsymbol{w})$$

Then (4) can be transformed as follows.

$$\log(\frac{p(y_i,\boldsymbol{w})}{p(\boldsymbol{w})p(y_i)})>0 \tag{6}$$

The left hand of (6) is similar to the definition of Pointwise Mutual Information (PMI)[17], which means that the association between y_i and $\boldsymbol{w}$ should be strong enough to make the activation effective.

Assumption 2: $\boldsymbol{w}$ and $\boldsymbol{y}$ are independent, but not identically distributed, which means that the marginal probability of each word is unequal.

By adopting the activation rate $\alpha=N_{\boldsymbol{y}}/N_d$, (5) can further be transformed as:

$$\sum_{i=1}^{N_y}\log(\frac{p(y_i,\boldsymbol{w})}{p(\boldsymbol{w})p(y_i)})+\sum_{i=1}^{N_y}\log(p(y_i))-\alpha\sum_{j=1}^{N_d}\log p(w_j)>0 \tag{7}$$

Inequation (7) shows that we should search for the implicit words $\boldsymbol{y}=\{y_1,...,y_{N_y}\}$ such that the associations of any y_i and $\boldsymbol{w}$ should be strong enough and the marginal probability of y_i should be high enough, given $\boldsymbol{w}=\{w_1,...,w_{N_d}\}$ and $N_{\boldsymbol{y}}=\alpha N_d$. This conclusion is highly identical with the ACT theory if we explain the baseline activation index B_i as the marginal probability term $\log(p(y_i))$, and the word association index $\sum_j W_j S_{ij}$ in (1) as the PMI term $\log(\frac{p(y_i,\boldsymbol{w})}{p(\boldsymbol{w})p(y_i)})$.

The justification above can serve as an explanation of ACT theory from the information theory point of view, and it also explains how ACT theory can help with topic modeling. On the one hand, we have got a technically sound approach to activate implicit information for given documents; on the other hand, it has been proved that the activated implicit information will bring reduce to the per-word perplexity.

IV. COGNITIVE PRESENTATION MODELS

A. *Association Computing Based on Word Embeddings*

Based on the above deduction, a set of activation presentation models can be built for topic modeling to enable stronger form of latent semantic analysis. However, it should be noticed that the PMI factor $\log(\frac{p(y_i,\boldsymbol{w})}{p(\boldsymbol{w})p(y_i)})$ is actually unreachable, since $p(y_i,\boldsymbol{w})$ is usually 0 in practical document collections where $\boldsymbol{w}$ as a whole may not appear repeatedly. So it is necessary to replace the PMI factor with another appropriate metric.

It is widely accept that PMI was proposed as a measure of word association, while there are also other word association measures, like cosine distance. As a matter of fact, there are internal correlations exist among different measures. Taking a look at $\log(\frac{p(y_i,\boldsymbol{w})}{p(\boldsymbol{w})p(y_i)})$ from the perspective of "cosine distance". If we present a specific word as a vector to indicate whether it appears in each document respectively, the cosine distance between any specific w_i and w_j can be formulated in (8), which is highly correlated with PMI:

$$d_{\cos}(w_i, w_j) = \frac{w_i w_j}{\| w_i \| \| w_j \|} = \frac{p(w_i, w_j)M}{\sqrt{p(w_i)p(w_j)M^2}}$$

$$= \frac{p(w_i, w_j)}{p(w_i)p(w_j)} \sqrt{p(w_i)p(w_j)} \quad (8)$$

Based on the similarity of structure insights between PMI and cosine distance, it is reasonable to replace the PMI factor with the cosine distances of some fine designed word vectors. We choose the word embeddings trained by skip-gram model [18]. As described in Table 1, a word w_i can be denoted by word embedding $\mathbf{v}_{w_i} = \{v_{w_i}^1, v_{w_i}^2, ..., v_{w_i}^L\}$. Then $S_{i,j}$ equals to the cosine distance of $\mathbf{v}_{w_i}$ and $\mathbf{v}_{w_j}$:

$$S_{i,j} = d_{\cos}(\mathbf{v}_{w_i}, \mathbf{v}_{w_j}) = \frac{\mathbf{v}_{w_i}\mathbf{v}_{w_j}}{\|\mathbf{v}_{w_i}\|\|\mathbf{v}_{w_j}\|} \quad (9)$$

As for any word y and a specific word collection $\boldsymbol{w}$, we approximate their associations as the cosine distance of y and the "center" of $\boldsymbol{w}$, i.e.:

$$S_{y,\boldsymbol{w}} = \frac{\mathbf{v}_y \mathbf{v}_{\bar{w}}}{\|\mathbf{v}_y\|\|\mathbf{v}_{\bar{w}}\|}, \mathbf{v}_{\bar{w}} = \frac{\sum_{w \in \boldsymbol{w}} \mathbf{v}_w}{N_{\boldsymbol{w}}} \quad (10)$$

It should be noticed that the calculation of $S_{y,\boldsymbol{w}}$ is the bottleneck as for the whole activation computing process. Thus transforming $\boldsymbol{w}$ into $\bar{w}$ reduce the time cost greatly.

B. Activing Process for Topic Modeling

Three models are built on the progressive assumptions in Section 3. Specifically, we define three activation functions in accordance with the deduction above to evaluate the activation level for every word in $\boldsymbol{V}$, given the stimulant words set $\boldsymbol{w}$. For each document d in corpus D, the number of activated words $N_{\boldsymbol{y}}$ is denoted by αN_d, then the implicit word set $\boldsymbol{y}$ is constructed by the activated words, which are selected from $\boldsymbol{V}$ according to their activation level. The activated words in $\boldsymbol{y}$ reveals the implicit information of the document d, and they can be added to the original Bag-of-Words presentation of the documents as the training corpus of the topic models.

The activation presentation models are given as follows:

1) Global Equal Base Model

This model is built under Assumption 1, i.e., all of the words are independent identically distributed. Then the activation level of a specific word y is decided by its association with the words in the given document d. According to (10), this can be formulated as:

$$A_y^d = S_{y,d} = \frac{\mathbf{v}_y \mathbf{v}_{\bar{w}}}{\|\mathbf{v}_y\|\|\mathbf{v}_{\bar{w}}\|}, \mathbf{v}_{\bar{w}} = \frac{\sum_{w \in d} \mathbf{v}_w}{N_d} \quad (11)$$

2) Global Activation Model

This model is built under Assumption 2, i.e., the activation level of a specific word y is decided by its occurrence probability as well as its association with the words in the given document d. This can be formulated as:

$$A_y^d = \beta p(y) + (1-\beta) S_{y,d} = \beta \frac{N_D^y}{N_D} + (1-\beta) \frac{\mathbf{v}_y \mathbf{v}_{\bar{w}}}{\|\mathbf{v}_y\|\|\mathbf{v}_{\bar{w}}\|}, \mathbf{v}_{\bar{w}} = \frac{\sum_{w \in d} \mathbf{v}_w}{N_d} \quad (12)$$

3) Local Activation Model

Different from the models above, Local Activaion is built based on an additional assumption that the implicit words are activated by some subsequences instead of the whole document. So we build the model to extract the subsequences from a particular document d and activate words based on these subsequences. Specifically, the activation level of the word y given a subsequence $\boldsymbol{w}_t$ can be formulated as:

$$A_y^{\boldsymbol{w}_t} = \beta p(y) + (1-\beta) S_{y,\boldsymbol{w}_t} = \beta \frac{N_D^y}{N_D} + (1-\beta) \frac{\mathbf{v}_y \mathbf{v}_{\bar{w}}}{\|\mathbf{v}_y\|\|\mathbf{v}_{\bar{w}}\|}, \mathbf{v}_{\bar{w}} = \frac{\sum_{w \in \boldsymbol{w}_t} \mathbf{v}_w}{N_{\boldsymbol{w}_t}} \quad (13)$$

The workflow of Local Activate for d is described in Table 2.

TABLE II. LOCAL ACTIVATE ALGORITHM

Input: Vocabulary V, the document d, the activation rate α
Output: The implicit word set $\boldsymbol{y}$
Local Activate: Initialize $\boldsymbol{y} = \varnothing$, $N_{\boldsymbol{y}} = \alpha N_d$ while # of words in $\boldsymbol{y} < N_{\boldsymbol{y}}$, do Extract a subsequence $\boldsymbol{w}_t$ randomly from d, $N_{\boldsymbol{w}_t} \sim U[0, N_d]$ $N_t = \alpha N_{\boldsymbol{w}_t}$ $\boldsymbol{y}_t \leftarrow$ activate N_t words according to (13) while # of words in $\boldsymbol{y} < N_{\boldsymbol{y}}$, do Pop a word form $\boldsymbol{y}_t$ and append it to $\boldsymbol{y}$ return $\boldsymbol{y}$

V. EXPERIMENTS

Various experiments have been conducted to evaluate the effectiveness of the proposed activation presentation models. This section summarizes the experimental setup and report the main findings.

A. Setup

We chose "20 Newsgroups"[1] as the main dataset for our experiments. This dataset is a collection of approximately 20,000 documents partitioned across 20 different newsgroups, in which 11314 documents are in the train set and 7532 documents are in the test set.

We trained word embeddings with Skip-gram model[2] ahead before the model building. The training corpus was the whole "20 Newsgroups" dataset, including the training set and the test set. The hidden layer size was set to be 50, i.e., the dimensionality of the word embeddings is 50. The models as discussed in Section 3 were built for LDA[3] and were denoted as Global EB, Global Actv, and Local Actv separately. The results of LDA tested on the raw data were adopted as one of the baselines (denoted as LDA). Besides, we selected words in the vocabulary randomly for each document as the control approach, which was another baseline denoted as Random. All of the topic models trained was set as 50 topics. During the experiment, the hyper parameter α and β was learned through cross validation if there is no additional explanation.

[1] http://qwone.com/~jason/20Newsgroups/
[2] https://code.google.com/p/word2vec/
[3] https://radimrehurek.com/gensim/index.html

The experiment includes two parts, which evaluated the model quality and the performance on the downstream tasks separately. Each part may contains several subtasks, trying to evaluate the models comprehensively.

B. Model Quality

Following the literature [6], the perplexity of a test set D_{test} is formulated under topic models as:

$$Perp(D_{test}) = \exp\{-\frac{\sum_{j=1}^{M}\sum_{i=1}^{N_j}\sum_{k=1}^{K}\log p(w_i \mid z_k)p(z_k \mid d_j)}{\sum_{j=1}^{M} N_j}\} \quad (14)$$

To explore the extent of the activation and make internal comparisons on our proposed models, we set the activation rate α ranging from 0.2 to 5.0. The experiment results are shown in Figure 4.

The curves in Figure 4 show remarkable improvements that our proposed models enabled as to the effectiveness of topic modeling. The perplexity decreases significantly and then stabilizes as activation rate grows. The curve corresponding to the Random approach partly shows that this improvement has no relation to the number of words, which is also consistent with the analysis in section 3.

To examine the topic distribution details qualitatively, we compared the topics trained with the different activating models. The results are shown in Table 3, where 6 representative topics among 50 are listed and each topic is presented by its top 8 words. The words highlighted with the bold font are noise or those that are inconsistent with the topic more or less. It can be seen from Table 3 that the generated topics from our proposed models are much more reasonable than that from Origin or Random, according to the human intuitions. This may because of our proposed models adds more related background information for each single document, and this leads to a better performance of the topic modeling, which relies on word co-occurrence heavily in model training.

Fig. 2. Comparisons of Test Perplexity

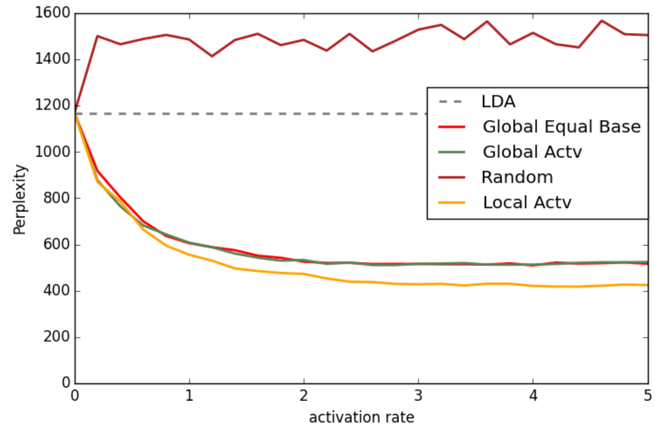


TABLE III. TOPIC DISTRIBUTIONS

Topics	LDA	Global EB	Global Actv	Local Actv	Random
space	space program nasa satellite **data jpeg** propulsion **faq**	space shuttle satellite lunar propulsion flight earth launch	space nasa telescope shuttle satellite earth lunar solar	space nasa lunar satellite shuttle flight propulsion telescope	space nasa satellite shuttle lunar earth **mission henry**
religion	god jesus people **don book** bible christian **life**	god jesus bible romans questioning christ people prophecy	jesus christ church god lord christian romans christians	god jesus romans questioning Christ bible prophecy psalm	god **life** christian **don** faith people **bullock writes**
windows	file windows dos software system program serial mail	file window program set windows image jpeg display	dos windows setup file slipper **finder** mscdex software	windows file slipper dos logitech ncr **ethersl** program	dos file windows system software program mail files
politics	people **don** president school **ancient time** clinton **writes**	people president **don** money clinton stephanopoulos care **applause**	president clinton money tax health care government plan	clinton president **applause** stephanopoulos protestors hostage legislation schumer	president people clinton **time don** activities **article** public
mideast	jews jewish israeli israel arab war bodies **law**	jews israel israeli arab jewish war leanings nationalism	jews israel war people israeli **soldiers** arab jewish	israel israeli jews invasions arab nationalism zionists prisoners	jews israel israeli jewish people arab **rights** human
guns	gun weapons guns weapon control **writes calgary article**	gun weapons fbi guns infantry police control overturn	gun weapons people law militia control government zealots	gun weapons **bullock** semiautomatic infantry overturn **schumer** police	gun weapons control police guns tanks crime firearms

C. Downstream Tasks

To test the models in practical application scenes, we conducted multiple experiments including document retrieval, document clustering and document classification.

- For document retrieval, we trained topic models with the training set, and generated document vector presentations with the trained topic models for the training set and the test set separately. Retrieval was setup as the test set acted as the queries and the training set acted as the database. The curves that the retrieval precision varies with the recall are shown in Figure 6.
- For document clustering, the training and test set were merged to perform the topic modeling and K-means clustering (K=20). The results shown in Table 4 compares different models through homogeneity, completeness and V_measure.
- For document classification, we trained topic models with the training set like that in document retrieval. LIBSVM was used to perform the 20 class classification task. The experimental results are shown in Table 5 which compares average precision, average recall and average F1 score.

From the experimental results we can see that different model outperforms the others in different tasks. Although Local Actv has lower test perplexity compared with the other two activation models, it doesn't performs better in practical applications. As for the parameter learned, α usually ranges from 0.5 to 1.0, while β usually achieves best performance at around 0.8.

In general, the experimental results shows that our proposed activation presentation models outperform Origin and Random in not only test perplexity and the reasonability of the topic distributions, but also in various downstream document analysis tasks.

Fig. 3. Document retrieval results

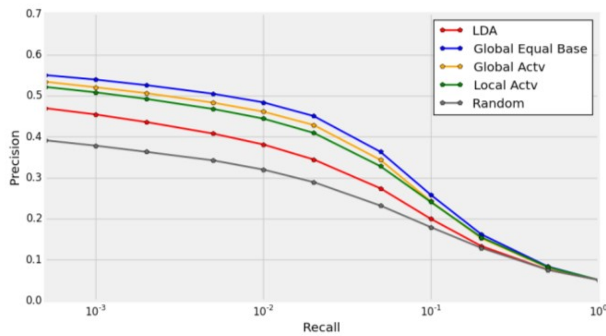


TABLE IV. DOCUMENT CLUSTERING RESULTS

	LDA	Global EB	Global Actv	Local Actv	Random
Homogeneity	0.339	0.481	0.486	0.448	0.298
Completeness	0.373	0.505	0.508	0.472	0.362
V_measure	0.355	0.493	0.497	0.459	0.327

TABLE V. DOCUMENT CLASSIFICATION RESULTS

	LDA	Global EB	Global Actv	Local Actv	Random
Precision	0.561	0.627	0.626	0.611	0.524
Recall	0.561	0.617	0.613	0.601	0.509
F1 Score	0.555	0.613	0.612	0.596	0.503

VI. CONCLUSION

In this paper, we deduced the key factors and conditions to enable the activation of implicit information from the given text inspired by cognitive psychology, and proposed a fast and loosely-coupled activation presentation of text for topic modeling. The experiments show that the activation models we proposed can provide significant performance improvements in various application tasks, with linear time consumption. The contributions of our work are mainly three-fold. First, we applied the ACT theory to improve the effectiveness of topic modeling. To the best of our knowledge, it is the first time that a cognitive psychological theory be used in this field. Second, we conducted a formal justification to reveal the insight of ACT theory in the language of information theory. Third, we proposed a set of novel activation presentation models inspired by the ACT theory.

In our ongoing work, we are exploring how to embed the activation process into neural network topic modeling techniques, which can be used to learn the word presentations and the topics simultaneously. We are also trying to apply the proposed activation models to other circumstances since the output of the activation is appropriate for other Bag-of-Words models as well. Additional computational and empirical testing of the proposed approach is under way as well.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China under Grants 71202169, 71402177, 71602184, 71621002 and U1435221, as well as the Early Career Development Award of SKLMCCS.

REFERENCES

[1] J. R. Anderson: "Cognitive Psychology and its Implications", WH Freeman/Times Books/Henry Holt & Co, 1990

[2] J. R. Anderson: "Language, Memory, and Thought", Psychology Press, 1976

[3] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman: "Indexing by Latent Semantic Analysis", Journal of the American Society for Information Science, 1990, vol. 41, pp. 391

[4] C. H. Papadimitriou, H. Tamaki, P. Raghavan, and S. Vempala: "Latent Semantic Indexing: A Probabilistic Analysis". Proceedings of the Seventeenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, Seattle, June 1998

[5] T. Hofmann: "Probabilistic Latent Semantic Indexing". Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, August 1999

[6] D. M. Blei, A. Y. Ng, and M. Jordan: "Latent Dirichlet Allocation", the Journal of Machine Learning Research, 2003, vol. 3, pp. 993-1022

[7] D. M. Blei and J. D. Lafferty: "Dynamic topic models". Proceedings of the 23rd international conference on Machine learning, Pittsburgh, 2006

[8] J. D. Mcauliffe and D. M. Blei: "Supervised topic models". Twenty-Second Annual Conference on Neural Information Processing Systems, Vancouver, December 2008

[9] D. Andrzejewski, X. Zhu, and M. Craven: "Incorporating domain knowledge into topic modeling via Dirichlet forest priors". Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, June 2009

[10] D. Newman, E. V. Bonilla, and W. Buntine: "Improving topic coherence with regularized topic models". Twenty-Fifth Annual Conference on Neural Information Processing Systems, Granada, December 2011

[11] P. Xie, D. Yang, and E. P. Xing: "Incorporating Word Correlation Knowledge into Topic Modeling". Conference of the North American Chapter of the Association for Computational Linguistics, Denver, 2015

[12] J. R. Anderson and G. H. Bower: "Human Associative Memory", Psychology press, 2014

[13] A. Baddeley: "Working memory: looking back and looking forward", Nature reviews neuroscience, 2003, vol. 4, pp. 829-839

[14] E. Cambria, D. Olsher, and K. Kwok: "Sentic Activation: A Two-Level Affective Common Sense Reasoning Framework". Twenty-Sixth AAAI Conference on Artificial Intelligence, Toronto, July 2012

[15] J. Zhang, Q. Li, X. Luo, and X. Wei: "A Multi-level Text Representation Model Within Background Knowledge Based on Human Cognitive Process". 12th IEEE International Conference on Cognitive Informatics & Cognitive Computing (ICCI* CC) Palo Alto, August 2013

[16] W. Kintsch and T. A. Van Dijk: "Toward a Model of Text Comprehension and Production", Psychological Review, 1978, vol. 85, pp. 363

[17] K. W. Church and P. Hanks: "Word Association Norms, Mutual Information, and Lexicography", Computational Linguistics, 1990, vol. 16, pp. 22-29

[18] T. Mikolov, W.-t. Yih, and G. Zweig: "Linguistic Regularities in Continuous Space Word Representations". The 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Atlanta, June 2013