International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:11, No:2, 2017

Incremental Learning of Independent Topic Analysis

Takahiro Nishigaki, Katsumi Nitta, Takashi Onoda

***Abstract*—**In this paper, we present a method of applying Independent Topic Analysis (ITA) to increasing the number of document data. The number of document data has been increasing since the spread of the Internet. ITA was presented as one method to analyze the document data. ITA is a method for extracting the independent topics from the document data by using the Independent Component Analysis (ICA). ICA is a technique in the signal processing; however, it is difficult to apply the ITA to increasing number of document data. Because ITA must use the all document data so temporal and spatial cost is very high. Therefore, we present Incremental ITA which extracts the independent topics from increasing number of document data. Incremental ITA is a method of updating the independent topics when the document data is added after extracted the independent topics from a just previous the data. In addition, Incremental ITA updates the independent topics when the document data is added. And we show the result applied Incremental ITA to benchmark datasets.

***Keywords*—**Text mining, topic extraction, independent, incremental, independent component analysis.

I. Introduction

THERE are a lot of studies of topic extraction from a large amount of document data. In this paper, we focus on topic extraction which is one of the challenges of text mining. The topic is the information represented by a co-occurrence of words between the large numbers of documents. As a method of topic extraction, there are a lot of studies of topic model such as PLSA (Probabilistic Latent Semantic Analysis) [15] proposed by Hofmann et al. and LDA (Latent Dirichlet Allocation) [8] proposed by Blei et al. The topic model is a method of extracting the topic focusing on probabilistic generative model [9]. The topic model generates the model between the document, word and topic. The topic is latent variable in the topic model. Moreover, the topic model represents the bias of the topic with the document by defining the distribution of the topic for each document. In many of the topic model, a document is represented as the bag of its words (bag-of-words), disregarding grammar and even word order but keeping multiplicity. In the bag-of-words, it commonly used the frequency of the words or tf-idf [22]. It is possible to represent the probability model of the relationship between the document, the word and the topic by using the topic model. However, the PLSA and the LDA is not focus on the relation between topics as correlation relationship or independence. On the other hand, LSI (Latent Semantic Indexing) [12] and ITA (Independent Topic Analysis) [25] focus on the relation between topics. LSI is able to extract the topic as large variance of the word or the document. It is possible to extract the topic without the correlation relationship between each topic by using LSI. The topic without the correlation relationship between each topic shows that there is no linear relationship such as a topic increase as well as another topic increase. ITA is able to extract the highly independent topic by using ICA (Independent Component Analysis) [17]. ICA is a computational method for separating a multivariate signal into additive subcomponents in signal processing. There are the service by using ITA: IT-DMS (Independent Topic-based Document Management System) [26], expanded IT-DMS [29]. The highly independent topic shows the mutual information [10] between each topic is small topic. It is easy to make document summarization with a large of information by extracting highly independent topic. The independent topic contains the topic without the correlation relationship. In other words, the independent topic is not equal to the topic without the correlation relationship. Note that when the topic in the document is normal distribution, the independent topic is equal to the topic without the correlation relationship. However, the topic in actual document data is not necessarily the normal distribution. Thus, it is necessary to extract the independent topic for extracting the topic which is small mutual information.

In this paper, we describe the ITA which is topic extraction method focusing on the independence of the topics. However, ITA extracts the independent topics using all document data. Therefore, ITA apply to the increasing number of document data is a hard because temporal and spatial cost is large. So it is difficult that ITA extract the independent topics from increasing the number of document data. In this study, we propose a method to resolve the problem. We refer to the proposed method as Incremental ITA. Incremental ITA can extract the independent topics from increasing the number of document data.

In the following sections, we introduce the related works of topic extraction with user constraints and constrained clustering in Section II, and we introduce the ITA (Independent Topic Analysis) in Section III. In Section IV, we describe the presented concept and ITA with the user constraints. In Sections V and VI, we show the experimental setup and experimental result. Finally in Section VII, we describe the conclusion and future works.

II. Related Works

In this section, we introduce the related work of the topic extraction from the increasing data. Furthermore, we introduce the methods of applying the increasing data to PCA (Principal Component Analysis) or ICA (Independent Component Analysis).

T. Nishigaki is with the Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, Yokohama, Japan (e-mail: nishigaki@ntt.dis.titech.ac.jp).

K. Nitta is Department of Computer Science, Tokyo Institute of Technology, Yokohama, Japan (e-mail: nitta@dis.titech.ac.jp).

T. Onoda is Department of Industrial and Systems Engineering, Aoyama Gakuin University, Kanagawa, Japan, e-mail: onoda@ise.aoyama.ac.jp.

As adapting the increasing data to topic model, several methods have been proposed [5]. For example, Neal et al. proposed Online vMF [20]. The vMF is a mixure model that uses von Mises-Fisher distribution as the components. In the mixture of von Mises-Fisher distributions model, a document is represented as an unit vector that is simply the L2 normalized version of the tf-idf vector corresponding to the document. Thus, all documents lie on the surface of the unit hypersphere [5]. Banerjee et al. focus on the spherical kmeans algorithm, which is a popular special case of the general vMF model, and proposed a version that is fully online. If the vMF model was given the new document, the parameters of the model need to be updated based on the new document. While there are several choices of doing such an update, the online vMF chooses a simple approach of only updating the parameters of the mixture component to which the current document got assigned to.

As the other example, there is an EDCM to an extension of DCM [5]. The DCM (Dirichlet Compound Multinomial) model is not an exponential family distribution, so the simple recursive update is not appropriate for the mixture of DCM model [13], [19]. In fact, the EDCM model is an exponential family approximation of the DCM. EDCM is actually not an exponential family model since the cumulant function has not been determined exactly [4], [3]. EDCM was resorted to a more explicit windowed update. If the existing mixture of EDCM models were given the new document, compute the probability of assigning the document to the most likely components from the existing model components. After assigning the document to the most likely component, EDCM update the component parameters. The parameter of the EDCM components are updated as a moving average of the new estimated parameters and the existing parameter values over the sliding window. Moreover, Song et al. proposed incremental LDA [28]. In the incremental LDA, batch LDA is initially run on a small window of the incoming data stream and the LDA parameters are initialized using the MAP estimates. This parameter updates with every new incoming document. And the estimated topic assignments depend only on the accumulated counts and the word in the current document. Incremental PLSA [11] and Online PLSA [6] was an expanded the PLSA. Consequently, there has been a lot of study about the topic model with increasing data. On the other hand, several incremental principal component analysis (IPCA) algorithm has been proposed [14], [21], [23]. Weng et al. proposed a fast IPCA algorithm, called candid covariance-free IPCA (CCIPCA) [30]. Principal component analysis is a well-known technique [27]. CCIPCA based on the amnesic average technique which is also used to dynamically determine the retaining rate of the old and new data, instead of a fixed learning rate. In addition, an extension of the ICA, such as online ICA [24] and recursive ICA [1] has been proposed. These methods were a method of both using the information maximization (Infomax) [7], [2] algorithm. However, we focus on the independent topic analysis which used the kurtosis estimation modifications. Futhermore, there is a little study on how to deal with increasing data to ITA. We propose a method of extracting the independent topic from document data which is increasing. In this paper, we propose the Incremental Independent Topic Analysis (Incremental ITA). This method is able to extract the independent topics form increasing document data.

III. ITA: Independent Topic Analysis

In this section, we introduce the Independent Topic Analysis (ITA) [26]. This method extract to topics from the document data by Independent Component Analysis (ICA) [17]. In the followings that a small letter expresses scalar, a bold small letter expresses vector, and a bold capital letter expresses matrices. As common variables, $t \in \{1, \ldots, k\}$ express topic variables, $d \in \{1, \ldots, n\}$ express document variables, and $w \in \{1, \ldots, m\}$ express word variables.

Firstly, we describe the concepts of ITA. Matrices $\mathbf{V}$ have m rows and k columns, which called "importance of the word w in the topic t". And vector $\mathbf{v}_t$ represent the t-th columns of vector of matrices $\mathbf{V}$. The vector $\mathbf{v}_t$ is $(v_{1,t}, \ldots, v_{m,t})^{\mathrm{T}}$. Vector $\mathbf{v}_w^{\mathrm{T}}$ represent the transposition of the w-th rows of vector of matrices $\mathbf{V}$. The vector $\mathbf{v}_w$ is $(v_{w,1}, \ldots, v_{w,k})$. Matrices $\mathbf{U}$ have n rows and k columns, which called "importance of the document d in the topic t". And vector $\mathbf{u}_t$ represent the t-th columns of vector of matrices $\mathbf{U}$. The vector $\mathbf{u}_t$ is $(u_{1,t}, \ldots, u_{n,t})^{\mathrm{T}}$. Vector $\mathbf{u}_d^{\mathrm{T}}$ represent the transposition of the d-th rows of vector of matrices $\mathbf{U}$. The vector $\mathbf{u}_d^{\mathrm{T}}$ is $(u_{d,1}, \ldots, u_{d,k})$. In the same way, matrices $\mathbf{A}$ have n rows and m columns, which called "frequency of word w in a document d". And $\mathbf{a}_w$ represent the w-th columns of vector of matrices $\mathbf{A}$. The vector $\mathbf{a}_w$ is $(a_{1,w}, \ldots, a_{n,w})^{\mathrm{T}}$. $\mathbf{a}_d^{\mathrm{T}}$ represent the transposition of the d-th rows of vector of matrices $\mathbf{A}$. The vector $\mathbf{a}_d^{\mathrm{T}}$ is $(a_{d,1}, \ldots, a_{d,m})$. So, we define the matrices $\mathbf{V}$, $\mathbf{U}$, and $\mathbf{A}$ as:

$$\mathbf{V} = \begin{pmatrix} v_{1,1} & \cdots & v_{1,k} \\ \vdots & \ddots & \vdots \\ v_{m,1} & \cdots & v_{m,k} \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_{1,1} & \cdots & u_{1,k} \\ \vdots & \ddots & \vdots \\ u_{n,1} & \cdots & u_{n,k} \end{pmatrix}$$

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & \cdots & a_{1,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,m} \end{pmatrix}$$

We use the kurtosis of fourth moment of the standard score as the measure of topic of independence. We define the measure of topic of independence as:

$$\sum_{w}^{m} \left(v_{w,t}^4 P(w) \right) - 3 \left(\sum_{w}^{m} v_{w,t}^2 P(w) \right)^2$$

where $v_{w,t}$ is component of w-th rows and t-th columns of the matrix $\mathbf{V}$. And $P(w)$ is defined as:

$$P(w) \equiv \frac{\sum_{d}^{n} a_{d,w}}{\sum_{d,w}^{n,m} a_{d,w}}$$

where $a_{d,w}$ is component of d-th rows and w-th columns of the matrix $\mathbf{A}$. When this measure is large, many components of matrix $\mathbf{V}$ and $\mathbf{U}$ become 0 value. So it is possible to express the topic only in a small number of words and documents.

Secondly, we describe the algorithm of ITA. ITA is formulated as an optimization problem as:

$$\underset{\mathbf{R}}{\text{maximize}} \quad \left| \sum_{t}^{k} \left\{ \sum_{w}^{m} \left((\mathbf{VR}).^4 P(w) \right) - 3 \left(\sum_{w}^{m} (\mathbf{VR}).^2 P(w) \right)^2 \right\} \right|$$

$$\text{subject to} \quad \mathbf{R}^{\mathrm{T}}\mathbf{R} = 1, \quad \|\mathbf{R}\| = 1$$

where $(\mathbf{VR}).^4$ is fourth power of each component of the matrix $\mathbf{VR}$. In follow, we show algorithm of ITA. In this method, the number of topics k is random variable.

1) Get the matrix $\mathbf{A}$. And using by [16], we normalize the $\mathbf{A}$ to make $\tilde{\mathbf{A}}$.
2) Performs a singular value decomposition of the matrix $\tilde{\mathbf{A}}$, such that $\hat{\mathbf{U}}^{\mathrm{T}}\tilde{\mathbf{A}}\hat{\mathbf{V}} = \hat{\mathbf{S}}$. Where $\hat{\mathbf{S}}$ is a diagonal matrix of singular values.
3) Extrac the matrix $\mathbf{U}$, $\mathbf{S}$ and $\mathbf{V}$ from the matrix $\hat{\mathbf{U}}$, $\hat{\mathbf{S}}$ and $\hat{\mathbf{V}}$. Extracted by k components in descending order of the value of the matrix $\hat{\mathbf{S}}$.
4) The matrix $\mathbf{X}$ of the topic in the k-dimensional space is defined as follows.
$$\mathbf{X} = \mathbf{S}^{-1/2}\mathbf{U}^{\mathrm{T}}\tilde{\mathbf{A}}$$
5) Independence maximization between each topic: calculate the rotation matrix $\mathbf{R}$ of the maximum independence based on FPIC as:
   a) Initialize the $\mathbf{R}$ to $k \times k$ zero matrix.
   $$\mathbf{R} = \mathbf{0}$$
   b) Substitute t-th column vector of the $\mathbf{R}$ to t-th column vector $\mathbf{e}_t$ of identity matrix $\mathbf{I} = (\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k)$
   $$\mathbf{r}_t = \mathbf{e}_t$$
   c) Initialize the $\mathbf{r}^{(old)}$ to $k \times 1$ zero vector.
   $$\mathbf{r}^{(old)} = (0, 0, \ldots, 0)^{\mathrm{T}}$$
   d) Update the $\mathbf{r}_t$ as:
   $$\mathbf{r}^{(old)} = \mathbf{r}_t, \quad \mathbf{r}_t = \mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{r}_t).^3 - 3\mathbf{r}_t$$
   $$\mathbf{r}_t = \mathbf{r}_t - \mathbf{R}\mathbf{R}^{\mathrm{T}}\mathbf{r}_t, \quad \mathbf{r}_t = \mathbf{r}_t / \|\mathbf{r}_t\|$$
   where $(\mathbf{X}^{\mathrm{T}}\mathbf{r}_t).^3$ is cube of components of the $\mathbf{X}^{\mathrm{T}}\mathbf{r}_t$.
   e) If $\mathbf{r}$ is convergence under the same conditions as FPICA [16], go to Step (5f). Otherwise go to Step (5d).
   f) If $t < k$, increasing one t and go to Step (5b). If $t = k$, go to Step (6).
6) Calculate the ${}^*\mathbf{V}$ and the ${}^*\mathbf{U}$ as follows.
$${}^*\mathbf{V} = \mathbf{VR}, \quad {}^*\mathbf{U} = \mathbf{UR}$$

Thus, it is possible to extract highly independent topics, such as shown in Fig. 1. We apply ITA to Los Angeles Times (LA Times) data as benchmark data. In Table I, it is shown that the important words of extracted 6 topics by applying ITA to LA Times. The important words which are the word with the large value of $v_{w,t}$. In Table I, we think that the topic 1 indicates "Soccer", the topic 2 indicates "Foreign", the topic 3 indicates "Entertainment", the topic 4 indicates "President", the topic 5 indicates "Finance" and the topic 6 indicates "Los Angeles". However, it is difficult that ITA extract the independent topics from the increasing number of document data. Because ITA extracts the independent topics using all current document data as Fig. 2. In Fig. 2, firstly we apply the ITA to the current document data A. Secondly we re-apply the ITA to the all current document data A+B. When the adding the document data, we re-apply the ITA to the all current document data. So applying ITA to the increasing number of document data is a hard because temporal and spatial cost is large. Therefore, we propose a method of applying ITA when the number of document data is increasing. In Section IV, we describe the method of applying ITA to increasing the number of document data.

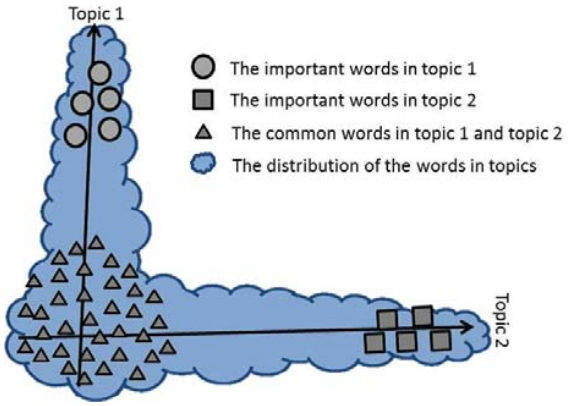

Fig. 1 The image of ITA

TABLE I
THE IMPORTANT WORDS OF THE EXTRACTED 6 TOPICS BY APPLYING ITA TO LA TIMES

No. topics	important words of each topic				
	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	scor	game	lead	quarter	rebound
2	soviet	afghanistan	israel	foreign	militari
3	aleen	macmin	art	entertain	report
4	bush	polic	budget	senat	tower
5	million	earn	bank	quarter	billion
6	polic	counti	offic	orang	citi

IV. INCREMENTAL ITA: INCREMENTAL INDEPENDENT TOPIC ANALYSIS

In this section, we describe the method of applying ITA to increasing the number of document data. We refer to the proposed method as Incremental ITA.

We explain the concepts of Incremental ITA. First step, Incremental ITA extracts the independent topics from a current document data. Next step, Incremental ITA updates the extracted independent topics every time data is added.

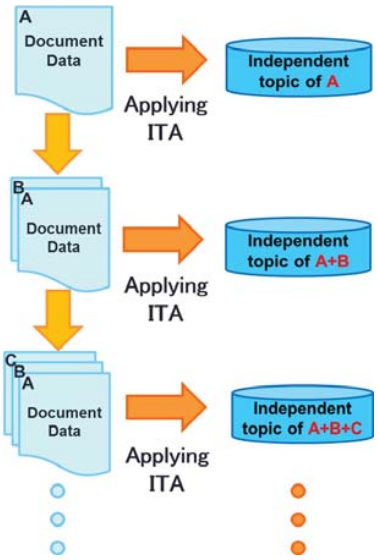

Fig. 2 The issue of ITA

There two steps are repeated.

We describe the algorithm of Incremental ITA as:

1) Extract the k independent topics from a current document data using the ITA.
2) Select the k independent topics.
3) Select the k data which have the maximum of absolute value of $\mathbf{U}$ in each topics.
4) Combine a new document data and the data of the Step 2, 3.
   - A number of new document data is fewer than the data of Step 1.
5) Update the k independent topics from the data of Step 4 using the ITA.
   - The initial matrix of $\mathbf{R}$ is to use the matrix calculated in Step 1.
6) Extracted the new independent topics after adding the data.

When the data is added, it is possible to extract the new independent topic by repeating this step 6 from step 2. We show the image of IITA in Fig. 3.

In Fig. 3, firstly we apply the ITA to the document data. Secondly, we apply the Incremental ITA to the independent topic and adding document data. When the adding the document data, we re-apply the Incremental ITA to the independent topic and adding document data.

V. Experimental Set up and Evaluation Metrics

In this section, we apply Incremental ITA to the benchmark dataset as follows. We used three benchmark dataset as:

- Los Angeles Times (LA Times) [31], [32]: the number of document $\times$ words is 6279×31472
- DAILY KOS blog (KOS blog) [18]: the number of document $\times$ words is 3430×6906

In incremental ITA, the number of data of the first step to use the 50%, 70% and 90% of the total documents. The

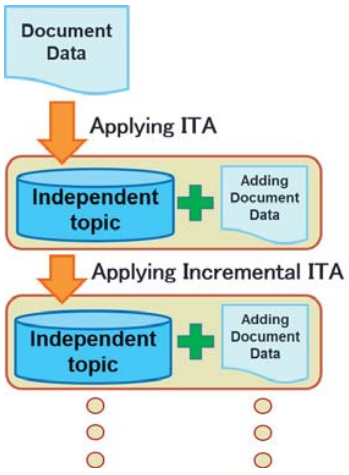

Fig. 3 The image of IITA

TABLE II
The Number of Document Data to Be Used in the First Step

percentage	50%	70%	90%
LA Times	3139	4395	5651
KOS blog	1715	2401	3087

number of data to add was 100. Table II shows the number of document data to be used in the first step. And the number of topics is fixed in 6.

The present experiments investigated whether: the independent topic by Incremental ITA was the same as the topic when using all document data. We used the absolute value of the cosine between the two topics for evaluation as:

$$|Cos_{ij}| = \left| \frac{\mathbf{v}_i^{\mathrm{T}} \cdot \mathbf{v}_j}{\sqrt{(\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i)(\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j)}} \right|$$

where $\mathbf{v}_i$ and $\mathbf{v}_j$ are the vector of importance of the each word in the topic i and topic j. When this value will be large, the topic i and topic j have the similar contents.

In the experiments, we calculate the cosine of the independent topic by Incremental ITA and the topic when using all document data. It is evaluated by the sum of this absolute value. If this value was large, these topics could show the similar topics. Furthermore, the maximum value of this value is 6 since the number of topics is fixed in 6. Thus, when this value is 6, the independent topic by Incremental ITA and the topic when using all document data are the same topic. We aim that this value is a large after repeating the incremental step.

VI. Experimental Results and Discussion

In this section, we show the experimental results of applying Incremental ITA to the benchmark data, and we discuss it. In the experiment, we firstly choose the 50% , 70% and 90% of the benchmark data at random. Secondly, we extract the independent topic from the data. Thirdly, we update the independent topic after adding the 100 data.

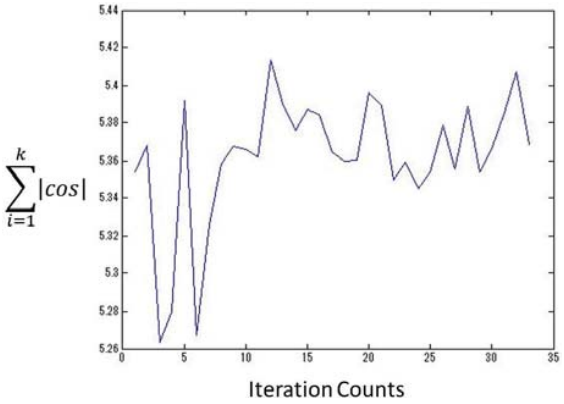

Fig. 4 Applying Incremantal ITA (50%) to LA Times; The absolute value of cosine and the number of times adding the data

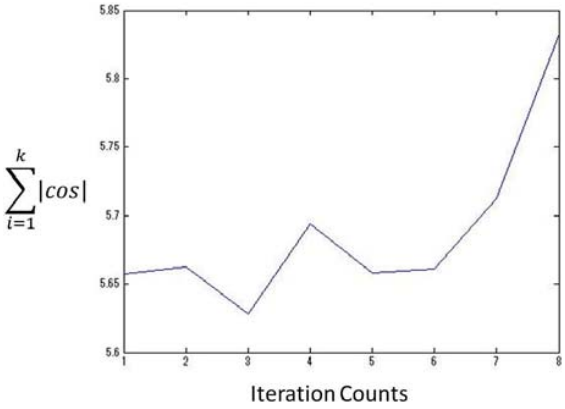

Fig. 6 Applying Incremantal ITA (90%) to LA Times; The absolute value of cosine and the number of times adding the data

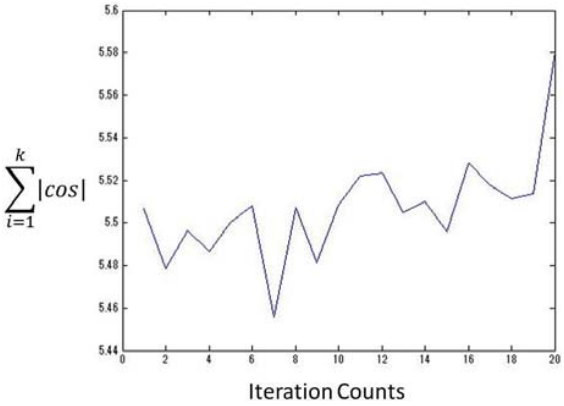

Fig. 5 Applying Incremantal ITA (70%) to LA Times; The absolute value of cosine and the number of times adding the data

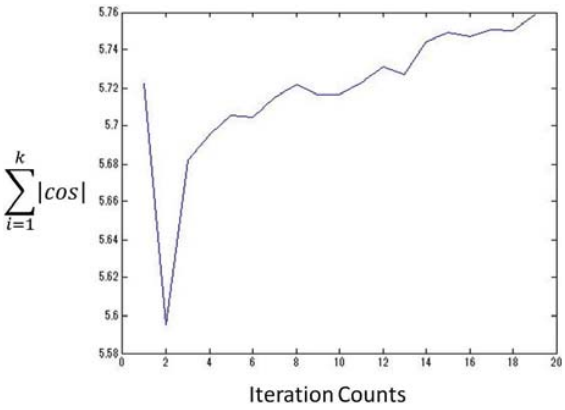

Fig. 7 Applying Incremantal ITA (50%) to KOS Blog; The absolute value of cosine and the number of times adding the data

We compare the extracted independent topics by Incremental ITA and the extracted independent topics by ITA from all data.

First, we show the result of Incremental ITA apply to LA Times. In Figs. 4-6, we show the absolute value of the total of cosine between two methods. This values indicates an average of 5 times. In these figures, they are shown the result of applying the Incremental ITA to the LA Times. In these figures, the horizontal axis shows the number of times adding the data and the vertical axis shows the absolute value of the cosine of an independent topic extracted by ITA. In these figure, they show that all cases have become a large value after adding the data. In particular, they show that it will be the large value in case of 70% and 90%. On the other hand, in case of 50%, the values are unstable because that depends on the initial data. LA Times has some of the independent topics, but the number of document in that topic is not equality. Therefore, its result very depends on the first document to be selected. We consider that the Incremental ITA is directed to the data which have approximately the same number of document in each topic.

Secondly, we show the result of Incremental ITA apply to KOS bolg. Figs. 7-9 show the result of applying the Incremental ITA to the KOS Blog. This value indicates an average of 5 times. We compare the extracted independent topics by Incremental ITA and the extracted independent topics by ITA from all data. These figures show the result of applying the Incremental ITA to the KOS bolg. These figures show that some cases have become a large value after adding the data. In particular, they show that it will be the large value in case of 50% and 70%. On the other hand, in case of 90%, the values are unstable because the influence of added data immediately before the Incremental ITA step. KOS blog is small scale of the data than the LA Times. When Incremental ITA uses the 90% of the KOS blog, it extracted the same topics used by all data. And when adding the data, extracted the independent topic is largely affected by the added data immediately. Therefore, Incremental ITA results depend on added data immediately before update step of Incremental ITA algorithm. We consider that Incremental ITA is directed to a large number of data.

Next we show a result of the important words of LA Times. We represent the details of the results of the Incremental ITA which applied to LA Times. In Table III, it is shown that the important words applying Incremental ITA of 1st time used 90%. When comparing Tables III and I, we think that the topic 1 and the topic 3 are exactly the same topic. And we think that the topic 2, the topic 5 and the topic 6 are almost the same topic. Moreover, we can understand that the

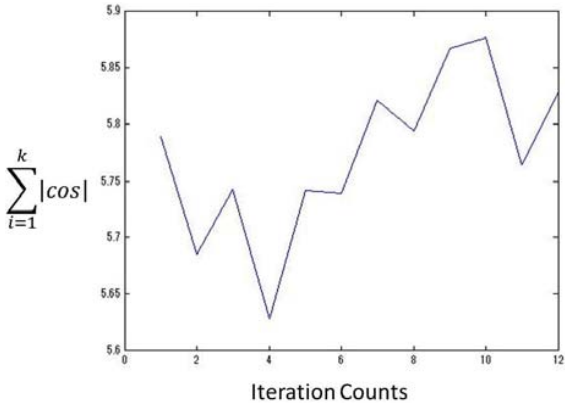

Fig. 8 Applying Incremantal ITA (70%) to KOS Blog; The absolute value of cosine and the number of times adding the data

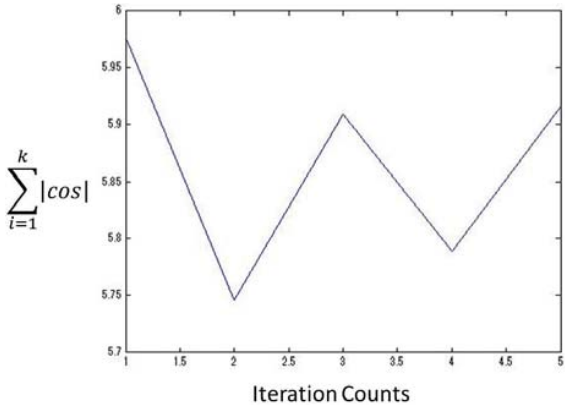

Fig. 9 Applying Incremantal ITA (90%) to KOS Blog; The absolute value of cosine and the number of times adding the data

topic 4 include "game" and "team" are the word representing the "Soccer". However, the topic 4 indicates "President" in Table I. Moreover, in Table IV, it is shown the important words applying Incremental ITA of 8th time used 90%. Firstly, we compare Table IV to Table I. We can understand that Incremental ITA can extract the independent topics which are almost the same as the topic of Table I. Secondly we compare Table IV to Table III. In Table III, "game" and "team" are contained in the topic 4 as "President" topic. On the other hand, "game" and "team" are not included in the topic 4 of Table IV. In Table IV, we can understand that the topic 4 indicates "President". This result is almost the same as Table I. We show that the extracted independent topics by Incremental ITA after adding the data approach to the independent topics by ITA. In these experimental results, extracted the independent topics by Incremental ITA are almost the same the independent topics by ITA, they show that Incremental ITA is directed to a large number of data and directed to the data which have much the same number of document in each topic.

VII. Conclusion

In this paper, we presented Incremental IT). This method was to solve one of the problems of the ITA. The problem

TABLE III
THE IMPORTANT WORDS OF THE EXTRACTED 6 TOPICS BY APPLYING INCREMENTAL ITA OF 1ST TIME USED 90% TO LA TIMES

No.	important words of each topic				
topics	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	scor	game	lead	quarter	rebound
2	soviet	afghanistan	israel	foreign	govern
3	aleen	macmin	art	entertain	report
4	bush	budget	game	counti	team
5	million	earn	quarter	compani	rose
6	polic	bush	counti	car	arrest

TABLE IV
THE IMPORTANT WORDS OF THE EXTRACTED 6 TOPICS BY APPLYING INCREMENTAL ITA OF 8TH TIMES USED 90% TO LA TIMES

No.	important words of each topic				
topics	$w=1$	$w=2$	$w=3$	$w=4$	$w=5$
1	scor	game	lead	quarter	rebound
2	soviet	afghanistan	israel	foreign	afghan
3	aleen	macmin	art	entertain	report
4	bush	polic	budget	reagan	presid
5	million	earn	quarter	bank	compani
6	polic	counti	orang	offic	citi

is that it is difficult to extract the independent topic from increasing number of document data by ITA. The Incremental ITA is able to the increasing number of document data. To evaluate the performance of the proposed method as Incremental ITA, we implemented and tested them in Matlab. We applied IITA to benchmark data. These experimental results show the new topics extracted by Incremental ITA is able to the increasing number of document data. Moreover, Incremental ITA is directed to a large number of data which have much the same number of document in each topic.

For our future works, we plan the following. First, it is necessary to apply other large benchmark data. Second, we want to theoretical proof of the convergence of the topic extracted by Incremental ITA.

Acknowledgment

This work was supported by the Japan Science and Technology (JST) agency under the EMS-CREST program.

References

[1] Akhtar, M. T., Jung, T.-P., Makeig, S., and Cauwenberghs, G., 2012. Recursive independent component analysis for online blind source separation, *IEEE International Symposium on Circuits and Systems*.

[2] Amari, S., Cichocki, A., and Yang, H.-H., 1996. A new learning algorithm for blind signal separation, *In Advances in Neural Information Processing Systems, D.S. Touretzky, M.C. Mozer, and M.E. Hasselmo, Editor*, Vol.8, pp.757–763, The MIT Press.

[3] Azoury, K. S., and Warmuth, M. K., 2001. Relative loss bounds for on-line density estimation with the exponential family of distribution, *Machine Learning*, Vol.42, No.3, pp.211–246.

[4] Banerjee, A., Merugu, S., Dhillon, I., and Ghosh, J., 2005. Clustering with Bregman divergences, *Journal of Machine Learning Research*, Vol. 6, pp.1705–1749.

[5] Banerjee, A., and Basu, S. 2007. Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning, *SIAM International Conference on Data Mining*.

[6] Bassiou, N. K., and Kotropoulos, C. L., 2014. Online PLSA: Batch Updating Techniques Including Out-of-Vocabulary Words, *IEEE Transaction on Neural Networks and Learning Systems*, Vol.25, Issue.11, pp.1953–1966.
[7] Bell, A. J., and Sejnowski, T. J., 1997. An information-maximization approach to blind separation and blind dexonvolution, *Neural Computation*, Vol.7, pp.1129–1159.
[8] Blei, D. M., Ng, A. Y., and Jordan, M. I. 2003. Latent dirichlet allocation, *The Journal of Machine Learning Research*, Vol. 3, pp. 993–1022.
[9] Blei, D. M. 2012. Probabilistic topic models, *Commun. ACM*, Vol. 55, No. 4, pp. 77–84.
[10] Brown, G., Pocock, A., Zhao, M-J., and Luján, M. 2012. Conditional likelihood maximisation: A unifying framework for information theoretic feature, *Journal of Machine Learning Research (JMLR)*, Vol. 13, pp. 27–66.
[11] Chien, J.-T., and Wu, M.-S., 2007. Adaptive Bayesian Latent Semantic Analysis, *IEEE Transactions on Audio, Speech and Language Processing*, Vol.16, Issue.1, pp.198–207.
[12] Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. 1990. Indexing by latent semantic analysis, *Journal of the American Society of Information Science*, Vol. 41, No. 6, pp. 391–407.
[13] Elkan, C., 2006. Clustering documents with an exponential-family approximation of the Dirichlet compund multinomial distribution, *In Proceedings of the 23rd International Conference on Machine Learning*.
[14] Hertz, J., Krogh, A., and Palmer, R. G., 1991. Introduction To the Theory of Neural Computation, *Addison-Wesley, Reading, MA*.
[15] Hofmann, T. 1999. Probabilistic latent semantic analysis, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI'99)*, pp. 289–29, Morgan Kaufmann Publishers Inc..
[16] Hyvärinen A. 1999. Fast and robust fixed-point algorithms for independent component analysis, *IEEE Trans. on Neural Networks*, Vol. 10, No. 3.
[17] Hyvärinen, A., Karhunen, J. and Oja, E. 2001. Independent component analysis, John Wiley & Sons.
[18] Lichman, M. 2013. UCI machine learning repository, http://archive.ics.uci.edu/ml , Accessed on 11/11/2016.
[19] Madsen, R., Kauchak, D., and Elkan, C., 2005. Modeling word burstiness using the Dirichlet distribution, *In Proceedings of the 22nd International Conference on Machine Learning*.
[20] Neal, R.M., and Hinton, G.E., 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. In M.I. Jordan, editor, *Learning in Graphical Model*, pp.355–368, MIT Press.
[21] Oja, E., and Karhunen, J., 1985. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix, *Journal of Mathematical Analysis and Application*, Vol.106, pp.69–84.
[22] Salton, G., Fox, E. A., Wu, H. 1983. Extended boolean information retrieval, *Commun. ACM*, Vol. 26, No. 11, pp. 1022–1036.
[23] Sanger, T. D., 1989. Optimal unsupervised learning in a single-layer linear feedforward neural network, *IEEE Transaction on Neural Networks*, Vol.2, pp. 459–473.
[24] Schraudolph, N. N., and Giannakopoulos, X., 1999. Online Independent Component Analysis With Local Learning Rate Adaptation, *NIPS*, pp. 789–795
[25] Shinohara, Y. 1999. Independent Topic Analysis : Extraction of Characteristic Topics by maximization of Independence, Technical report of IEICE.
[26] Shinohara, Y. 2000. Development of Browsing Assistance System for finding Primary Topics and Tracking their Changes in a Document Database, CRIEPI Research Report.
[27] Sirovich, I., and Kirby, M., 1987. Low-Dimensional procedure for the caracterization of human faces, *Journal of Optical Society of America A*, Vol.4, No.3, pp.519–524.
[28] Song, X., Lin, C.-Y., Tseng, B. L., Sun, M.-T., 2005. Modeling and predicting personal information dissemination behavior, *In Proceedings of the 11th ACM SIGKKD International Conference on Knowledge Discovery and Data Mining*.
[29] Tanaka, M, Shinohara, Y. 2003. Topic-Based Dynamic Document Management System for discovering Important and New Topics, CRIEPI Research Report.
[30] Weng, J., Zhang, Y, and Hwang, W.-S., 2003. Candid Covariance-free Incremental Principal Compoenent Aalysis, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.25, No.8, pp.1034–1040.
[31] Zhao, Y. and Karypis, G. 2002. Evaluation of hierarchical clustering algorithms for document datasets, *Conference of Information and Knowledge Management (CIKM)*, pp. 515–524, ACM.
[32] Zhong, S., and Ghosh, J. 2003. A comparative study of generative models for document clustering, *Data Mining Workshop on Clustering High Dimensional Data and Its Applications*.

Takahiro Nishigaki Takahiro Nishigaki was born in Japan 1987. He is a PhD student in Department of Computational Intelligence and Systems Science at Tokyo Institute of Technology. He recieved the B. E. Degree from Kyoto Institute of Technology in 2011. He recieved the M. E. Degree from Tokyo Institute of Technology in 2013. He is a student member of the Japanese Society of Artficial Intelligence (JSAI).

Katsumi Nitta Katsumi Nitta received his B.E., M.E. and Dr.Eng. from Tokyo Institute of Technoloy in1975, 1977 and 1980, respectively. In 1980, he joined Electrotechnical Laboratory as a computer scientist, and in 1989, he joined Institute of New Generation Computer Technology. Since 1996, he has been a Professor of the Interdiscipinaly Graduate School of Science and Engineering, Tokyo Institute of Technology. His research interest includes the legal informatics and man-machine interaction. He is a member of the Information Processing Society of Japan (IPSJ), the Japanese Society for Artificial Intelligence (JSAI), and the Institute of Electronics, Information and Communication Engineers (IEICE).

Takashi Onoda Takashi Onoda was born in Japan in 1962. He received the B. S. Degree from International Christian University in 1986. He received the M. S. Degree from Tokyo Institute of Technology in 1988. He received the Dr. Eng. Degree from University of Tokyo in 2000. He was a Visiting Researcher in GMD FIRST (Fraunfoher FIRST) from 1997 to 1998. Since 2007, he has also been an Visiting Professor at Department of Computational Intelligence and System Science, Tokyo Institute of Technology, Japan. He is a member of JSAI (Japanese Society of Artficial Intelligence). He researched in Central Research Institute of Electric Power Industry from 1988 to 2015. He has been a Professor with Aoyama Gakuin University since 2016.