Visual Materials to Teach Gibbs Sampler

Yukari Shirota, Takako Hashimoto, and Basabi Chakraborty

Abstract—**Bayesian model of inference is widely used in various application fields such as data engineering or text processing. Using Bayes' theorem, we can obtain the posterior distribution function. When we conduct sampling using Markov chain Monte Carlo (MCMC), the most prominent MCMC algorithms are the Metropolis-Hastings and the Gibbs sampler, the latter being particularly useful in Bayesian analysis. This paper presents the visual teaching material for studying Gibbs sampler algorithm. Interaction with this material is supposed to enable students to deeply understand the mathematical process behind Gibbs sampling and encourages them to comprehend the mathematical expressions in the textbooks.**

Index Terms—**Bayesian inference, MCMC, simple topic model, Gibbs sampler, visualization.**

I. Introduction

Bayesian inference is widely used in various application fields such as data engineering and econophysics. Using Bayes' theorem, we can obtain the posterior distribution function. While evaluating expectations of the functions, a serious problem is the limitations in spaces of high dimensionality [1]. Thus we need to apply sampling using Markov chain Monte Carlo (MCMC) algorithms. The most well-known MCMC algorithms are the Metropolis-Hastings and the Gibbs sampler, the latter being particularly useful in Bayesian analysis [2].

Gibbs sampler programs have been offered in various programming languages such as R and Python. Therefore students and researchers can easily apply the Bayesian inference in their problems and obtain results. The problem is, however, that many of them do not understand the mathematical process behind the algorithm. We believe that the users should understand the mathematical process of the algorithm in order to have better grasp on its use. Otherwise they are unable to analyze the results correctly leading to their inability of making a good decision regarding the convergence of the simulation and finally they cannot make an extension of the algorithm according to their needs.

We have been using Gibbs sampler in text mining researches [3]-[7]. We let students use our developed text mining tools which were developed based on the topic model [8]. Being undergraduate students of economics /

Y. Shirota is with Faculty of Economics, GakushuinUniversityy, 1-5-1 Mejiro, Toshima-ku Tokyo, 171-8588 Japan (e-mail: yukari.shirota-atmark-gakushuin.ac.jp).

T. Hashimoto is with Chiba University of Commerce, 1-3-1, Konodai Ichikawa City, Chiba, 272-8512, Japan (e-mail: takako-atmark-cuc.ac.jp).

B. Chakraborty is with the Department of Software and Information Science, Iwate Prefectural University, 152-52 Sugo, Takizawa, Iwate 020-0693, Japan (e-mail: basabi-atmark-iwate-pu.ac.jp).

ministration departments, it is too difficult for them to read and understand the mathematical process behind the theory of topic model and Gibbs sampler described in the textbooks. Therefore we have developed visual teaching materials to make them understand the Gibbs sampler. By the visual materials, the number of students who grasp the concept of Gibbs sampler has been increased.

In the next section, we shall explain in brief, the simple topic model and Gibbs sampler because we have adopted the simple topic model instead of the general topic model. In Section III, we would like to present our visual teaching materials. Then we shall describe the existing teaching visual materials on Gibbs sampling. In Section IV, we would like to conclude the paper.

II. Simple Topic Model and Gibbs Sampler

In the section, we shall explain in brief, the simple topic model and Gibbs sampler. Let us suppose that our target application is topic extraction. In topic extraction, the topic model has been used as the model and as its implementation Latent Dirichlet Allocation (LDA) model has been widely used. Many introductory texts of LDA exist [9]-[11].

A. Simple Topic Model

In the visualization of Gibbs sampler, we adopted, instead of the topic model, a simple topic model (or mixture of unigrams). A simple topic model is a simplified version of the topic model with the limitation that a document has only one topic. Figure 1 illustrates three graphical models to note the differences between them [8], [12], [13]. They are unigram model, simple topic model, and topic model. For the details of a graphical model, the introductory textbooks such as [1] is referred.

The unigram model is much simpler version of the topic model in which the whole document has just one topic. The parameters α, β in Fig. 1 are hyper-parameters of the Dirichlet distributions. In a simple topic model, one topic has its word distribution. The simple topic model allows a document to have just one topic.

We have selected a simple topic model as our target model because the simple topic model shares basic concept of the topic models and we think that understanding the simple topic model will lead one smoothly to understand the other topic models. Another reason of the selection was that the other topic model was too complicated to be visualized on the screen.

B. Gibbs Sampler

In MCMC, the sample sequence is constructed so that, successive samples are generated from distributions that probably get closer and closer to the given desired posterior [14] distribution. In Gibbs sampler, we iterate over each of the unsolved variables, sampling a new value for each

variable given our current sample for **all other variables** [14].

When the topic extraction of documents is executed, a topic identification of each document is in turn decided from other $(n-1)$ documents' current status. We visualize the meaning of this as shown in Fig. 2.The details of the illustration will be explained in the next section.

In MCMC, after enough time repetition of the substitutions, the target distribution $P(x)$ becomes the invariant distribution of the Markov chains. This means that when we generate a sample x from the distribution P, substituting x by x', then the distribution of x' again gets to become P as follows [15].

$$\sum_{x_i}\{P(x_i'|x_1,\cdots,x_{i-1},\quad x_{i+1},\cdots,x_N) \times$$

$$\underline{P(x_1,\cdots,x_{i-1},x_i,x_{i+1},\cdots,x_N)}\}$$

$$= P(x_i'|x_1,\cdots,x_{i-1},\quad x_{i+1},\cdots,x_N)$$

$$\times P(x_1,\cdots,x_{i-1},\quad x_{i+1},\cdots,x_N)$$

$$= \underline{P(x_1,\cdots,x_{i-1},x_i',x_{i+1},\cdots,x_N)}$$

The Gibbs sampler update probability is the following:

$$P(x_i'|x_1,\cdots,x_{i-1},\quad x_{i+1},\cdots,x_N)$$

In the Gibbs sampler, we must need to know whether the distribution converges to the required invariant distribution or not. For that, we think that the visualization of the sampling sequences will enable us to conceive the level of the convergence.
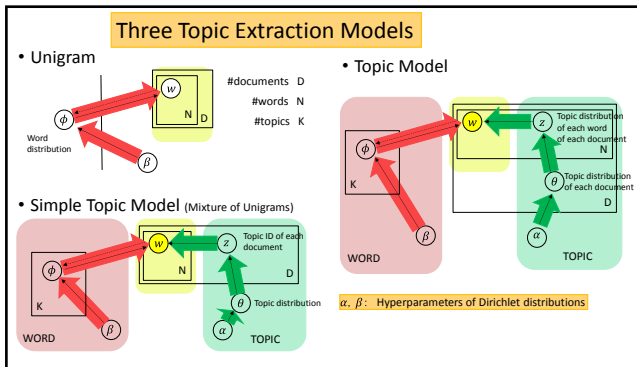

Fig. 1. Three graphical models for topic extraction which are unigram model, simple topic model (mixture of unigrams model) and topic model.

III. GIBBS SAMPLER VISUAL MATERIALS

In the section, we present the visual teaching material for Gibbs sampling (see Fig. 3). In this example, the simple topic model is used for its simplicity. The number of documents is five, the number of topics is seven. The document is expressed by a bag of words which consists of six words.

The graphics in Fig. 3 is made in Mathematica by Wolfram and we transformed the Mathematica programs to the Wolfram CDF version and have published them on web so that everyone can access it. The Wolfram CDF player is free software and by installing the player, everyone can conduct interactive operations with that program on web browsers.

With the teaching materials, the user can interactively operate the sampling by using the top located slider. The

reverse motion is also available. In the initial state, there is no topic. Then gradually, topics are organized and its word distributions are also developed. The user can see the changes on screen.
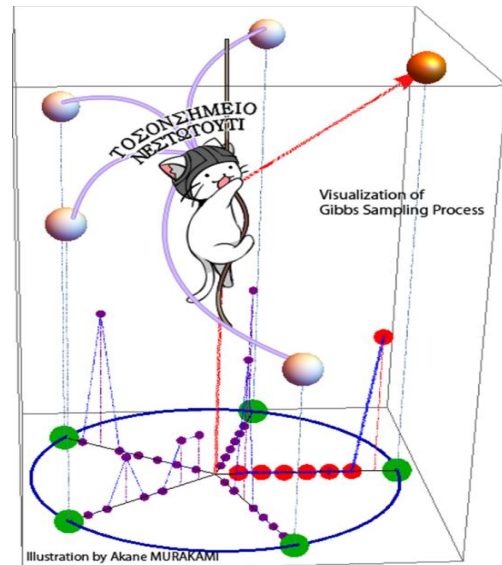

Fig. 2. Concept image of Gibbs sampler of which the background image was developed by Mathematica.

In one turn of Gibbs sampler, at first, the target document that is denoted by an arrow is isolated from others. Then seeing the word distribution of the target document, the program considers which topic word distribution is the most similar to the target document's word distribution. Then the topic ID of the target document has been determined and the target document is set to the topic ID group member. That is the classification process. Then, the next turn will begin on the next target document.

In Fig. 2, there are five balls on the cylinder edge. The height of the ball indicates its topic identification number. On the bottom plane, there is a circle and the five radius lines are illustrated. On the radius line, the topic distribution probability of each document is shown.

In Fig. 2, the target document group number has become seven because the probability of the seventh topic is higher than others; the other six probability levels are zero.

Next we show the image of our teaching material in Fig. 3. In Fig. 3, the target document group number has become three. In Fig. 3, on the bottom, there are three graphics.

The left graphics shows the number of each group members. In the centre graphics, the topic distribution of each document is shown. The right graphics shows the word distribution of each topic. As shown here, each topic has its word distribution in a simple topic model. By using this, the programme will be able to make a decision to which topic the target document belongs to.

Advantages of this visualization are 1) we can see the most similar topic group of the target document, 2) we can see a set of words of the target document are immigrated to the topic group, and 3) we might see the convergence of the Markov chains.

IV. RELATED WORK

In the section, we shall present a survey of the existing

researches on visualization of Gibbs sampling.

The most famous one would be the illustration by David JC MacKay [16]. The illustration shows alternate updates of just two variable density function $P(x_1, x_2)$. In the famous text book by Bishop, the same kind of illustration on $P(x_1, x_2)$ which is a correlated Gaussian is also presented [1], [17]. Similar graphics of these two variable density functions have been published; for example one in the Mathematica online manual page titled "how to / Perform a Monte Carlo Simulation."
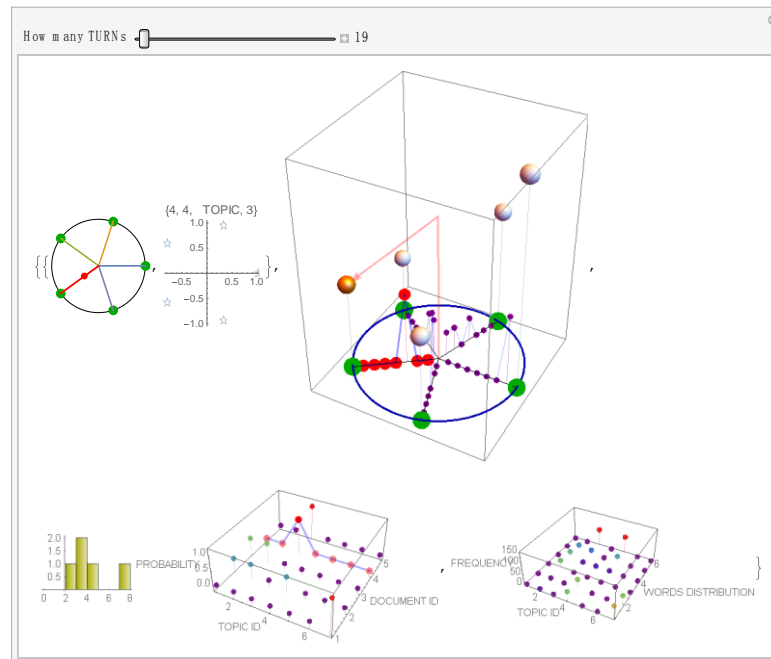

Fig. 3. Visual teaching material for Gibbs sampler.

However, as a practical use, we need multivariable cases of more than three variables, because the two variable density function case does not express enough about the intrinsic feature of Gibbs sampling.

In the field of the topic model visualization, there are many researches on visualizations [18]. However, they focus on relationship between a topic and the word distribution. There is no research on visualization to explain Gibbs sampler on the topic model, especially focusing on the explanation of the mathematical process.

V. CONCLUSIONS

This paper presents the visual teaching materials to study Gibbs sampler using a simple topic model. The Gibbs sampler is widely used in Bayesian inferences, especially together with the topic model. We think that the researchers who use the topic model and Gibbs sampler should understand the mathematical process behind this. If they cannot understand the mathematics, even if they can use the tools, they cannot extend the model or perform correct analysis of the result. Then visualization teaching materials are helpful and they encourage users to comprehend the mathematical expression.

Many graduate students in engineering with interest in Gibbs sampling commented that the visualization was effective. We will continuously develop various visual teaching materials on statistical techniques.

ACKNOWLEDGMENT

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[2] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method*, second edition, John Wiley & Sons, 2009.

[3] T. Hashimoto, T. Kuboyama, and Y. Shirota, "Graph-based consumer behavior analysis from buzz marketing sites," in *Proc. 21st European Japanese Conference on Information Modelling and Knowledge Bases*, Estonia, June 6-10, 2011.

[4] Y. Shirota, T. Kuboyama, T. Hashimoto, S. Aramvith, and T. Chauksuvanit, *Study of Thailand People Reaction on SNS for the East Japan Great Earthquake — Comparion with Japanese People Reaction,* p. 62, Research Institute for Oriental Cultures Gakushuin University, 2015.

[5] Y. Shirota, T. Hashimoto, and S. Tamaki, "Monetary policy topic extraction by using lda — Japanese monetary policy of the second abe cabinet term," in *Proc. IIAI International Congress on Advanced Applied Informatics 2015*, 12-16 July, 2015, Okayama, Japan, pp. 8-13, 2015.

[6] Y. Shirota, T. Hashimoto, and T. Sakura, "Topic extraction analysis for monetary policy minutes of Japan in 2014," *Advances in Data Mining: Applications and Theoretical Aspects*, pp. 141-152, Springer International Publishing, 2015.

[7] T. Hashimoto and Y. Shirota, "Framework of an advisory message board for women victims of the east Japan earthquake disaster," in *Prof. of JADH2013 (Japanese Association for Digital Humanities)*, Sept 19-21, Kyoto, pp. 31-32, 2013.

[8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.

[9] P. D. Hoff, *A First Course in Bayesian Statistical Methods*, Springer, 2010.

[10] M. D. Lee and E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course*, Cambridge University Press, 2014.

[11] J. Kruschke, *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*, Academic Press, 2014.

[12] D. Blei and J. Lafferty, "Dynamic topic models," in *Proc. the 23rd International Conference on Machine Learning*, 2006.

[13] D. M. Blei, "Probabilistic topic models," *Commun. ACM,* vol. 55, no. 4, pp. 77--84, 2012.

[14] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*, The MIT Press, 2009.

[15] Y. Iba, "Bayesian statistics and statistical physics," *Iwanami-Shoten*, 2003.
[16] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, 2003.
[17] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, third edition, Chapman & Hall/CRC, 2004.
[18] A. J.-B. Chaney and D. M. Blei, "Visualizing topic models," in *Proc. the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland*, June 4-7, 2003.

Yukari Shirota is the faculty of Economics, Gakushuin University. Her research fields are visualization of data on the web data visualization, social media analysis, and visual education methods for business mathematics. For over 17 years, she has developed visual teaching materials for business mathematics and statistics. In VINCI 2015, the tutorial titled "Visually do statistics for business persons: Visual materials from regression to black-sholes model" she talked. Various visual teaching material sites have been published on the web which are freely available as follows: http://www.gakushuin.ac.jp/univ/eco/english/teacher/sirota.html etc.

Takako Hashimoto received a Ph.D. in computer science, specialization in multimedia information processing, from the Graduate School of Systems and Information Engineering of University of Tsukuba in 2005. From April of 2009, she was involved in Chiba University of Commerce as associate professor. In 2015, she has become professor of Chiba University of Commerce and stayed at University of California, Los Angeles as a visiting researcher. She has focused on the data mining research and the social media analysis, especially topic extraction from millions of tweets related to the East Japan Great Earthquake. She is developing the high performance feature selection technique for big data. She's also conducting global researches for developing the social media analysis platform in multi-language/cultural environment.

Basabi Chakraborty received B.Tech, M.Tech and Ph. D degrees in radio physics and electronics from Calcutta University, India and worked in Indian Statistical Institute, Calcutta, India until 1990. She received another Ph. D in information science from Tohoku University, Japan in 1996. Currently she is a full professor in Software and Information Science Department of Iwate Prefectural University, Japan. Her main research interests are in the area of pattern recognition, machine learning, soft computing techniques, data mining and online social media mining. She is a senior member of IEEE, member of ACM, INNS and Japanese Society of Artificial Intelligence.