Information Engineering Express
International Institute of Applied Informatics
2017, Vol.3, No.1, P.1 – 8

Visualization of Deductive Reasoning for Joint Distribution Probability in Simple Topic Model

Yukari Shirota *, Takako Hashimoto†, Basabi Chakraborty ‡

Abstract

Bayesian inference is widely used in various application field such as data engineering. When we derive the posterior, we have to combine many theorems or rules such as the Bayes' theorem. The derivation of the posterior expression is quite difficult, even if we use the probabilistic graphical model. So we propose a deductive reasoning based approach for that. The concrete deductive diagram for a simple topic model is presented in the paper. The deductive reasoning diagram clarifies which theorems and how they are used in the deduction. In addition, the three conditional independence pattern rules which are used frequently in the posterior derivation are explained visually

Keywords: Bayesian inference, MCMC, simple topic model, Gibbs sampler, deductive reasoning, visualization, conditional independence

1 Introduction

Bayesian inference is widely used in various application fields [1-3]and we have developed a tool to visualize the Markov chain Monte Carlo (MCMC) process and used the tool as the teaching material [4]. As we have been conducting researches on topic extraction by using Latent Dirichlet Allocation model or so [5-9], we have thought that it is quite difficult problem to teach the Bayesian theorems to our students. They are unfamiliar to Markov chain and the three conditional independence rules [10] so they cannot derive the posterior distribution probability. We believe that we must teach the math processes and make them understand them before their using the convenient tools or programs. This is because if they do not understand the math process, (1) they cannot analyze correctly the results, and (2) they cannot make the advanced model. For example, they should understand the distinction between a topic model [11] and a dynamic topic model or so [12, 13].

From long experience of teaching mathematics to students, we believe that the essence of teaching math is to foster their reasoning skills and we have developed the e-learning systems

* Gakushuin University, Tokyo, Japan
† Chiba University of Commerce, Chiba, Japan
‡ Iwate Prefectural University, Iwate, Japan

[14-19]. In general, there are various kinds of reasoning such as deduction, abduction, and induction[20]. Among them, the deductive reasoning is mostly used in Bayesian inference. Therefore we adopt a deductive approach for the teaching. In the deduction process, first (1) given data and conditions and then (2) unknowns are given[21, 22]. Finding the missing link between the given data and the unknowns is solving the problem. The hard and laborious work there is to contrive the way of selection and using theorems. Then the visualization of the whole deduction process is very helpful to foster the deductive skills. We had researched the visualization of the deductive reasoning processes and evaluated the effectiveness in business mathematics [22-25]

In the paper, we shall describe our teaching materials for the simple topic model. By the teaching materials, many students could understand the derivation of the posterior. In the next section, we shall shortly explain the simple topic model and the Gibbs sampler. In Section 3, we present our deductive approach and teaching materials. In Section4, we shall conclude the paper.

2 Simple Topic Model and Gibbs Sampler

In the section, we shall explain shortly the simple topic model and Gibbs sampler. A simple topic model is a simply modified version of the topic model with the limitation that a document has only one topic [26]. We have selected a simple topic model because the simple topic model shares a basic model concept as the topic model and to understand the simple topic model enables us to understand smoothly the topic model.

Gibbs sampler is one of Markov chain Monte Carlo (MCMC) algorithms. Although the first sample may be generated from the prior, successive samples are generated from distributions that probably get closer and closer to the desired posterior [27]. 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 [27]. In case of the topic extraction of documents, a topic identification of each documents is in turn decided from other (n – 1) documents current status. 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', by the Gibbs update operation $P(x_i'|x_1, \cdots, x_{i-1}, \quad x_{i+1}, \cdots, x_N)$ the distribution of x' again becomes P.

Let us simply explain the Gibbs update operation which slightly modifies the topic distribution of the documents. We illustrates the Gibbs sampling in Figure 1 where the number of documents is five and the number of topics is seven. There the cat is the figurative existence of the Gibbs sampling system. The documents are placed around on the circle and its height shows the topic ID. On the radius from the center of the circle to the bottom of each document, the probability distribution function on topics is illustrated. In Figure 1, the system (cat) pointed the document and the document probability density function shows that the topic 7 probability is higher than others. As shown Figure 1, the probability density function shows the value set (0, 0, 0, 0, 0, 0, 1) where we count the ID numbers from the circle center to the circle line like 1, 2, 3, 4, 5, 6, and 7. First, the system lets the target document get out of the circle. The word distribution of the document is set back from the calculation. The system calculates the probability density of the topic using (n-1) document word information. This is illustrated by the cat wearing a helmet which has connections to the (n-1) documents, in this case, four documents.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

After the document topic is calculated and the document height is determined, the cat moves to the next document. And the cat goes round the circle and in turn and in turn, a document topic ID is calculated and the document topic distribution is getting to the equilibrium status. The teaching material of the Gibbs sampling is written in Wolfram Mathematica and the CDF material version has been published on the site http://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/. The Wolfram CDF player is free software. Therefore, everyone can moves the Gibbs sampling program without charge by installing the Wolfram CDF player.

3 Deductive Reasoning Approach

In the section, we describe our deductive reasoning approach to teach the derivation of the posterior in Bayesian inference.

In a deductive reasoning, we are first given the given data and the unknown. Between the given data and the unknown, we will have to find a missing link so that we can connect the given data and the unknown. To connect them, we use the formulas and theorems. The deductive reasoning is the process in which the transformation of expressions is conducted by the selected formula. It is a deductive reasoning process to find the missing link. In other words, to conduct the deductive reasoning is the way of solving a problem so far as the problem is a solution find type, not a proof problem.

We would like to calculate the posterior of a simple topic model. When we handle and solve the model, probabilistic graphical models (abbrev. graphical model) [27] are so useful because they provide a simple way to visualize the structure of the model and because insights into the properties of the model, including conditional independence properties, can be obtained inspection of the graph [10]. And an important concept for probability distributions over multiple variables is that of conditional independence. Sato says that drawing the graphical model and using three patterns of conditional independence [28] enables us to easily expand the joint distribution and addresses the importance of the graphical model and the conditional independence [29]. Bishop explains the conditional independence properties of directed graphs by considering three simple examples each involving graph having just three nodes [10]. If our students could read directly Bishop's textbook, there would be no need to visual and deductive teaching materials. However, they had difficulties to read the textbooks. Therefore some interpretation illustration is needed.

We illustrates the three patterns so that students can easily memorize them (See Figure 2). There variable c is given and fixed, and then we may think c is a block wall. The pattern names such as "Tail-to-Tail", "Head-to-Tail", and "Head-to-Head" were cited from [29] (Sato named them so). In the Tail-Tail and Head-Tail types, given c, variable a and b are independent. In the Head-Head type, given c, variable a and b are not independent. We name the three types as follows for students' memorization:

"Run-away",

"Chased but blocked by the given c", and

"Confrontation".

We shall conduct the deductive reasoning to obtain the posterior. Then the formulas and theorems to be used are (1) Bayesian theorem, (2) joint probability theorem, and (3) conditional independ-ent theorems.

In the existing textbooks, they explain that using the graphical model and the conditional independence precisely. We however think their explanations are not enough. The concrete deductive reasoning process needs to be offered. Therefore we have illustrated that (See Figure 2).

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

That is the posterior derivation of the simple topic model. The concrete math expressions are described in [26]. In Figure 2, there are 14 times expression transformations, in other words, 14 times inferences/reasoning. We think that in each transformation which theory/rule was used there is explained. In advance, the available rules/theories are offered; in this case they are (1) Bayesian theorem, (2) conditional independence rules, (3) joint probability by conditional probability $p(A,B) = p(A|B)p(B)$. Figure 3 shows the probability transformation for the posterior of the d-th document which belongs to topic ID k. The posterior $p\left(z_d = k \middle| W, z_{\backslash d}, \alpha, \beta\right)$ is corresponding to the Gibbs update operation $P\left(x_i' | x_1, \cdots, x_{i-1}, \quad x_{i+1}, \cdots, x_N\right)$ in Gibbs sampler. In Figure 3, the used theory name is shown with a mark such as "Bayes" and "joint" which means the joint probability by conditional probability $p(A,B) = p(A|B)p(B)$. The almost transformation means equal relationship "=" but partly in two teansformations the proportional relationships denoted by "$\propto$" appear. This is because if the expression includes no z_d the denominator expression has no effect on the posterior. The latter part in Figure 3, the conditional independence rule is twice used to simplify the probability expression. There we use the marks of "Head" and "Tail". In addition, the dependency relationship is illustrated by the mark "Chain" such as a relationship betweenα and z_d. In addition, a plate expression in the graphical model for a repetition should be illustrated separately as shown there in order to make students conceive the relationship $W = W_{\backslash d} \cup \{w_d\}$.

4 Conclusions

We discuss deductive and visual teaching methods of derivation of the posterior for a simple model. The derivation of the posterior expression is difficult to understand. Therefore we have made the teaching methods which enable students to easily understand the deductive reasoning process. Although the number of students who need mathematics for kind of the topic model probabilistic inference is limited, studying the math is essential to understanding the model. The deductive reasoning diagram is helping students to understand the derivation in our classes. We will continue to develop helpful visualization teaching materials.

We have discussion about the effectiveness of the deductive reasoning in math education. The Gibbs sampling teaching materials presented here were used in our PRICAI tutorial lecture entitled "Visually See Text Mining Math Processes on LSA, SVD, and Gibbs Sampling"§. Many lecturers have the same problem that is how to teach the Bayesian inference mathematics to their students as ours. And through the discussions, we thought the visualization and the deductive reasoning approach are so effective for students. We shall continuously develop the visual and deductive-based teaching materials .

Aknowledgement

We thank Prof Tetsuji Kubotyama (Computer Centre, Gakushuin University) for his wide range of knowledge about machine learning that helps our research. In addition, we thank Ms. Akane Murakami for her scientific illustrations of the Gibbs sampler. This research was partly supported by funds from Gakushuin University Computer Centre projects from 2015-2016, by funds from the Telecommunications Advancement Foundation research project in 2015 to 2016, and by a grant from the Japanese Society for the Promotion of Science from 2015-2017 (15K00314). We sincerely express our gratitude to the Society for its support.

§ 14th PRICAI 2016 (Pacific Rim International Conference on Artificial Intelligence) tutorial: http://aiat.in.th/pricai2016/front/show/tutoriallist

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

References

[1] P. D. Hoff, *A First Course in Bayesian Statistical Methods*: Springer, 2010.
[2] M. D. Lee, and E.-J. Wagenmakers, *Bayesian Cognitive Modeling: A Practical Course*: Cambridge University Press, 2014.
[3] J. Kruschke, *Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan*: Academic Press, 2014.
[4] Y. Shirota, T. Hashimoto, and B. Chakraborty, "Visual Materials to Teach Gibbs Sampler," *2016 International Conference on Knowledge (ICOK 2016), London, UK, May 7-8, 2016.*, pp. (in printing), 2016.
[5] T. Hashimoto, T. Kuboyama, and Y. Shirota, "Graph-based Consumer Behavior Analysis from Buzz Marketing Sites," *Proc. of 21st European Japanese Conference on Information Modelling and Knowledge Bases, Estonia, June 6-10*, 2011.
[6] 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.^pp. 62: Research Institute for Oriental Cultures Gakushuin University, 2015.
[7] Y. Shirota, T. Hashimoto, and S. Tamaki, "MONETARY POLICY TOPIC EXTRACTION BY USING LDA －JAPANESE MONETARY POLICY OF THE SECOND ABE CABINET TERM－," *Proc. of IIAI International Congress on Advanced Applied Informatics 2015, 12-16 July, 2015, Okayama, Japan*, pp. 8-13, 2015.
[8] 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*, Lecture Notes in Computer Science P. Perner, ed., pp. 141-152: Springer International Publishing, 2015.
[9] T. Hashimoto, and Y. Shirota, "Framework of an Advisory Message Board for Women Victims of the East Japan Earthquake Disaster," *Prof. of JADH2013 (Japanese Association for Digital Humanities), Sept 19-21, Kyoto*, pp. pp. 31-32, 2013.
[10] C. M. Bishop, *Pattern Recognition and Machine Learning*: Springer, 2006.
[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research,* vol. 3, pp. 993-1022, 2003.
[12] D. Blei, and J. Lafferty, "Dynamic topic models," *Proceedings of the 23rd International Conference on Machine Learning*, 2006.
[13] D. M. Blei, "Probabilistic topic models," *Commun. ACM,* vol. 55, no. 4, pp. 77--84, 2013/07/01, 2012.
[14] Y. Shirota, T. Hashimoto, and S. Suzuki, "Knowledge Visualization of Reasoning for Financial Mathematics with Statistical Theorems," *Proc. of the DNIS (Databases in Networked Information Systems) 2014, LNCS 8381, Springer, Heidelberg,*, pp. 132-143, 2014.
[15] Y. Shirota, T. Hashimoto, and P. Stanworth, "Knowledge Visualization of Deductive Reasoning for Word Problems in Mathematical Economics," *Prof. of 8th Workshop pn Databases in Networked Information Systems (DNIS 2013), March 25 to 27, 2013, University of Aizu, Japan* vol. (DNIS 2013 Vol. LNCS 7813). , 2013
[16] Y. Shirota, T. Hashimoto, and T. Kuboyama, "A Concept Model for Solving Bond Mathematics Problems," *Frontiers in Artificial Intelligence and Applications, Edited by Jaak Henno, Yasushi Kiyoki, Takehiro Tokuda, Hannu Jaakkola, Naofumi Yoshida, Frontiers in Artificial Intelligence and Applications, Information Modelling and Knowledge Bases XXIII*, pp. 271-286: IOS Press, 2012.
[17] Y. Shirota, and T. Hashimoto, "Plausible Deductive Reasoning Plan for Business Mathematics Learners - Solution Plan Graph Generator -," *Proc. of 2nd Uncertainty Reasoning and Knowledge Engineering (URKE2012), Jakarta, August 14-15*, pp. 5-8, 2012.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

[18] Y. Shirota, T. Hashimoto, and T. Kuboyama, "A Concept Model for Solving Bond Mathematics Problems," *Proc. of 21st European Japanese Conference on Information Modelling and Knowledge Bases, Estonia , June* pp. 6-10, 2011.

[19] P. Stanworth, and Y. Shirota, "Review of the "Web:How2SolveIt" Website," *Gakushuin Economics Papers,* vol. 50, no. 1, pp. 1-18, 2013

[20] N. C. Betsur, *Reasoning Strategies in Mathematics*: Anmol Publications PVT. LTD., 2006.

[21] G. Polya, *How to Solve It: A New Aspect of Mathematical Method* Princeton Science Library, 2004.

[22] Y. Shirota, Y. Takahasi, N. Tanaka, and M. Morita, "Visually Do Statistics for Business Persons Visual Materials from Regression to Black-Sholes Model (the tutorial session)," *Proc. of VINCI 2015, ACM, 24-26 August, 2015, Tokyo*, pp. 170-173, 2015.

[23] Y. Shirota, and B. Chakraborty, "Visual Explanation of Mathematics in Latent Semantic Analysis," *Proc. of IIAI International Congress on Advanced Applied Informatics 2015, 12-16 July, 2015, Okayama, Japan*, pp. 423-428, 2015.

[24] Y. Shirota, and S. Suzuki, "Visualization of the Central Limit Theorem and 95 Percent Confidence Intervals," *Gakushuin Economics Papers,* vol. Vol.50, no. No. 3, 4, pp. pp., 2014.

[25] Y. Shirota, and T. Hashimoto, "Animation Teaching Materials for Explaining Recurrence Formula to FInd the Bond Price with the Spot Rate," *Journal of Japan Society of Business Mathematics,* vol. 33, no. 1/2, pp. 57-69, 2012.

[26] T. Iwata, *Topic Model*, Kodansha, 2015 (written in Japanese).

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

[28] A. P. Dewid, "Conditional independence for statistical operations," *Annals of Statistics* vol. 8, pp. 598-617, 1980.

[29] I. Sato, *Statistical Latent Semantics Analysis Based on Topic Model*: Corona publishing Co., 2015.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

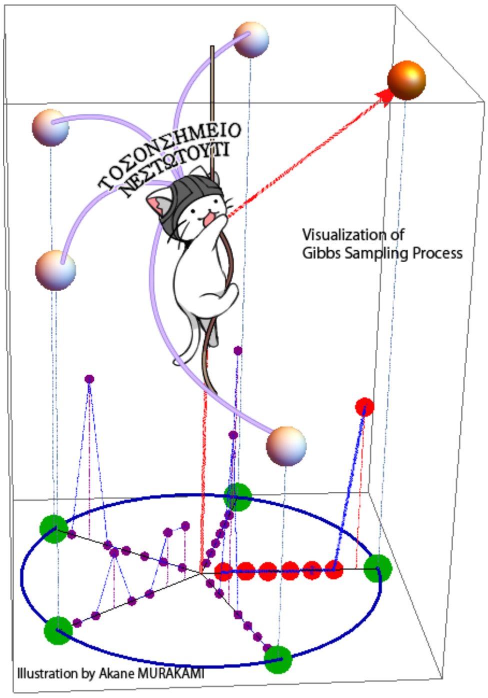

Figure 1: An image illustration of the Gibbs sampling.

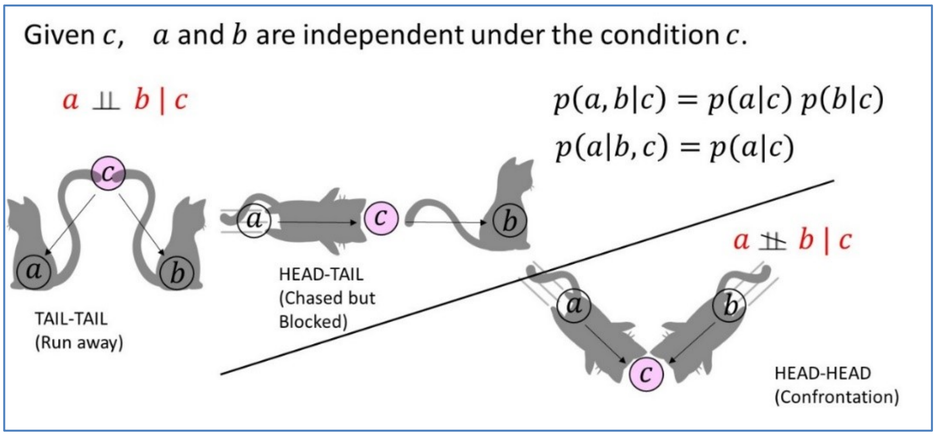

Figure 2: Theree rules on conditional independence in graphical models.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

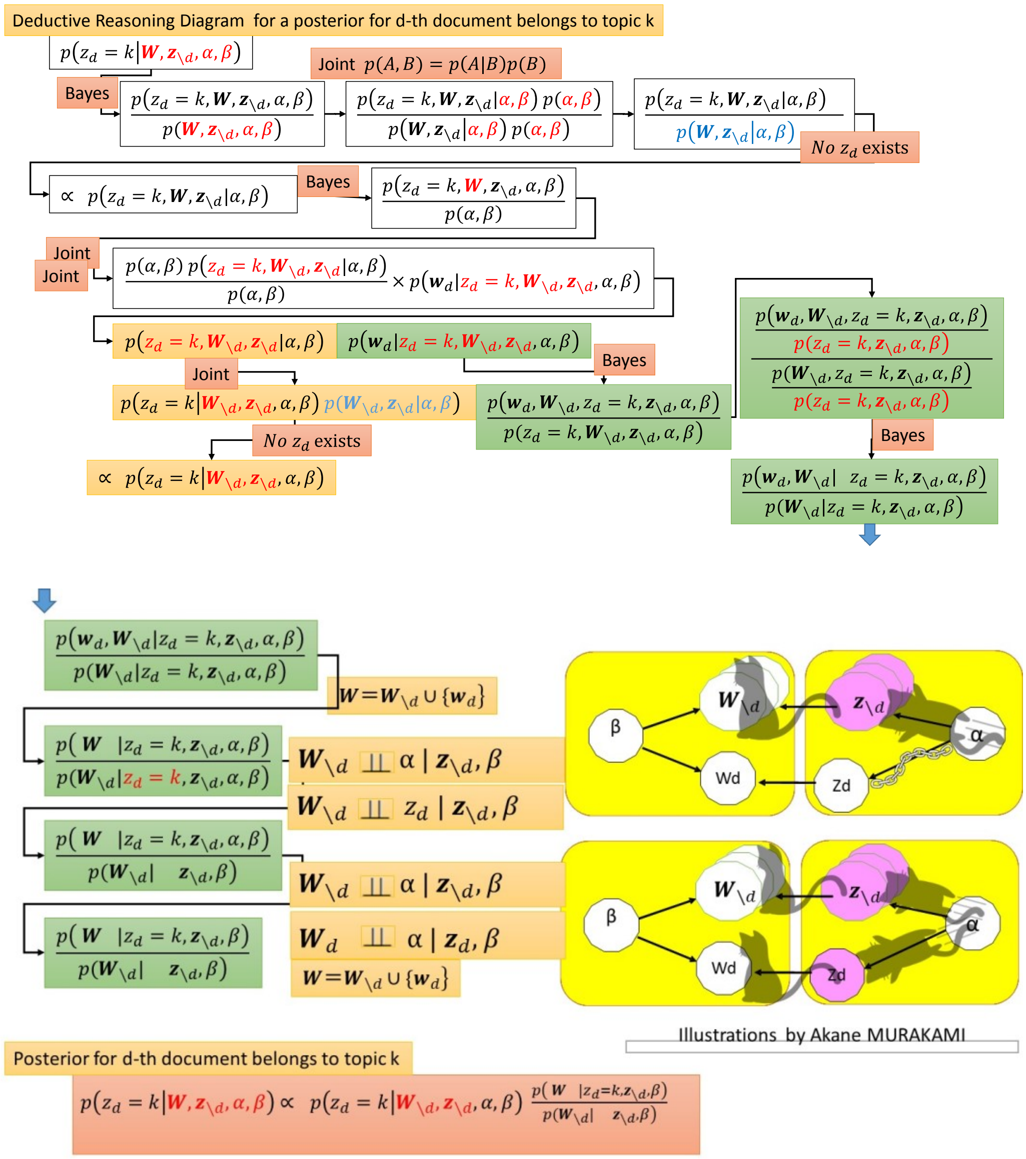

Figure 3: A deductive reasoning diagram for the posterior of d-th document which belongs to topic ID k.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.