

ベイズ推論による大学ブランドイメージ分析

白田 由香利*, 橋本 隆子**, バサビ・チャクラボルティ***

要旨

ベイズ推論は、文書のトピック分析を始め、画像のパターン認識のほか、画像を用いた服装コーディネート推薦、買い物履歴情報からの消費パターン分析など、広い分野で活用されている。本稿では、ベイズ推論の手法を大学ブランドイメージ分析に適用する。我々は、マルコフ連鎖モンテカルロ法 (MCMC) の代表的アルゴリズムであるギブズサンプラーの実装により、シンプルトピックモデルによるベイズ推論のクラスタリングの過程を可視化するツールを開発した。このツールにより、大学ブランドイメージという共通の知識を共有するデータを対象として、クラスタリングのマルコフ連鎖の変動のようすを可視化する。このような手法のポイントは、共有する知識に基づく分析事例を示し、その事例からその背景にある数学的手法を連想させ、理解させる、ことである。このMCMCにおける定常状態の連鎖プロセスにおいては、ある大学が2つのクラスの間を頻繁に行き来するなどの興味深い変動が見られた。こうした可視化により、クラスタリングのツールをただ使って結果を出すだけでなく、その背景にあるギブズサンプラー、シンプルトピックモデル、などの数学プロセスを理解することが可能となる。そうした数学プロセスの理解は、的確な分析に必須であるので、こうした数学教育はベイズ推論において重要と考える。

キーワード：大学ブランド分析，ベイズ推論，マルコフ連鎖モンテカルロ法，可視化，シンプルトピックモデル，ギブズサンプラー，クラスタリング。

I. 始めに

ベイズ推論は、文書のトピック分析を始め、画像のパターン認識のほか、画像を用いた服装コーディネート推薦、買い物履歴情報からの消費パターン分析など、広い分野で活用されている [1-3]。我々はベイズ推論におけるマルコフ連鎖モンテカルロ法 (Markov chain Monte Carlo: MCMC) におけるマルコフ連鎖の可視化ツールを開発した [4,5]。この可視化ツール

*) 学習院大学経済学部

**) 千葉商科大学商経学部

***) 岩手県立大学ソフトウェア情報学部

のモデルはシンプルトピックモデル（混合ユニグラムモデルとも呼ぶ）[6]であり、MCMC アルゴリズムは、ギブズサンプラー（Gibbs sampler）[7,8]を用いている。この可視化ツールを開発した背景として、我々が長年、文書からのトピック抽出の研究をしてきたことがある。手法としては、潜在的ディリクレアロケーションモデル（Latent Dirichlet Allocation model）、及び、潜在的意味分析法（Latent Semantic Analysis, LSA）などを用いてきた[9-13]。ゼミ演習などでのトピック抽出の実験の際、始めからベイズ推論の数学及びアルゴリズムを教えるのは、内容が難しいため、効率的ではない。まず本可視化ツールで、ギブズサンプラーのアルゴリズムを説明することで、概要がつかめ、詳細なアルゴリズム及び数式を読むことが可能となる。本可視化ツールにより、従来のテキストでは分からなかった学生でも、ギブズサンプラーが理解できるようになり、大きな教育効果を上げている[4,5]。

本稿では、本可視化ツールにより大学ブランドイメージのクラスタリングを行う[14]。テキストマイニングの分野では、ベイズ推論による分析は広く普及している。マーケティングの分野では web アンケート及び SNS の日本語分析などが広く行われている[15]。しかし、マーケティングのブランド分析の分野での利用は、テキストマイニングの分野に比較して顕著ではないと思われる。本稿では、ブランド分析への具体的な適用事例を示す。大学ブランドイメージ分析という学生によって興味深いテーマを用いることで、分析手法への理解も深まる。学生は予め大学イメージに関する類似した知識を有している。その知識を使うことで、可視化ツールの示すクラスタリング結果も理解可能となる。このような手法のポイントは、共有する知識に基づく分析事例を示し、その事例からその背景にある数学的手法を連想させ、理解させることである。我々は、既に LSA の分析手法においても類似の教法を使って、教育効果を上げている[16]。

次節では、シンプルトピックモデルとギブズサンプラーについて簡単に説明する。第3節では、リクルート社の行ったアンケート結果に基づきクラスタリングを行った事例を紹介する。第4節は既存研究のサーベイで、第5節はまとめである。

II. シンプルトピックモデルとギブズサンプラー

本節では、簡単に本可視化ツールが使っている、シンプルトピックモデルとギブズサンプラーを説明する。一般的なベイズ推論に関するテキストは[1-3]を参照して頂きたい。

A. シンプルトピックモデル

トピックモデルには、その簡略バージョンとして、ユニグラムモデル、シンプルトピックモデルの2つが有る。その3種類のモデルをグラフィカルモデルで表現し、比較した図を示す(図1参照)[17]。グラフィカルモデルでは、四角いパネルは、類似変数の繰り返しを表す。例えば、シンプルトピックモデルにおいて、変数 z は D （ドキュメント数）と書かれたパネル内に存在するが、これは、 z_1, z_2, \dots, z_D を意味する。シンプルトピックモデルは、モデルとして簡素であり、ひとつの文書はひとつのトピック ID しかもてない。そのトピック ID が変数 z_i の値である。例えば、文書のトピックとして政治、経済、文学の3つがあった場合、一つの文書のトピックはそのどれかに決めなくてはならない。一般的には、ひとつの文書は複数のトピックを持てるようにしたほうがモデルとしての表現力は高くなる。トピックモデルでは、ひとつの文書は複数のトピックを有することが可能であり、その比率も変えられる。図1のトピックモデルのグラ

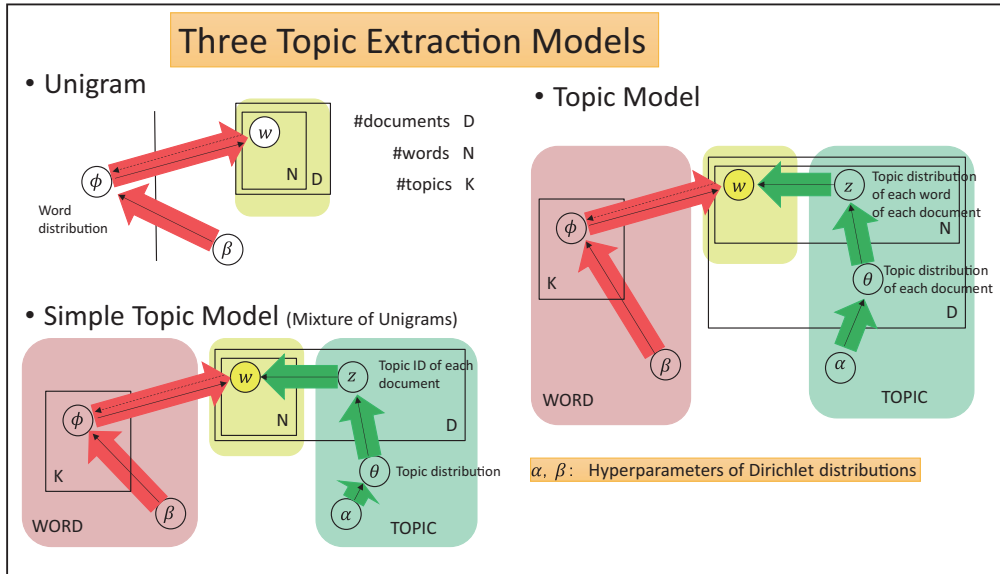


Fig.1. トピックモデル3種類のグラフィカルモデル表現。

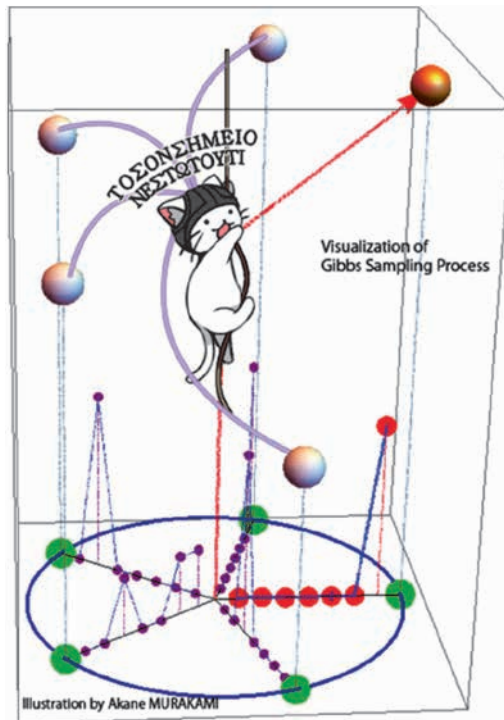


Fig.2. ギブズサンプラーの概念をイラスト表現したもの

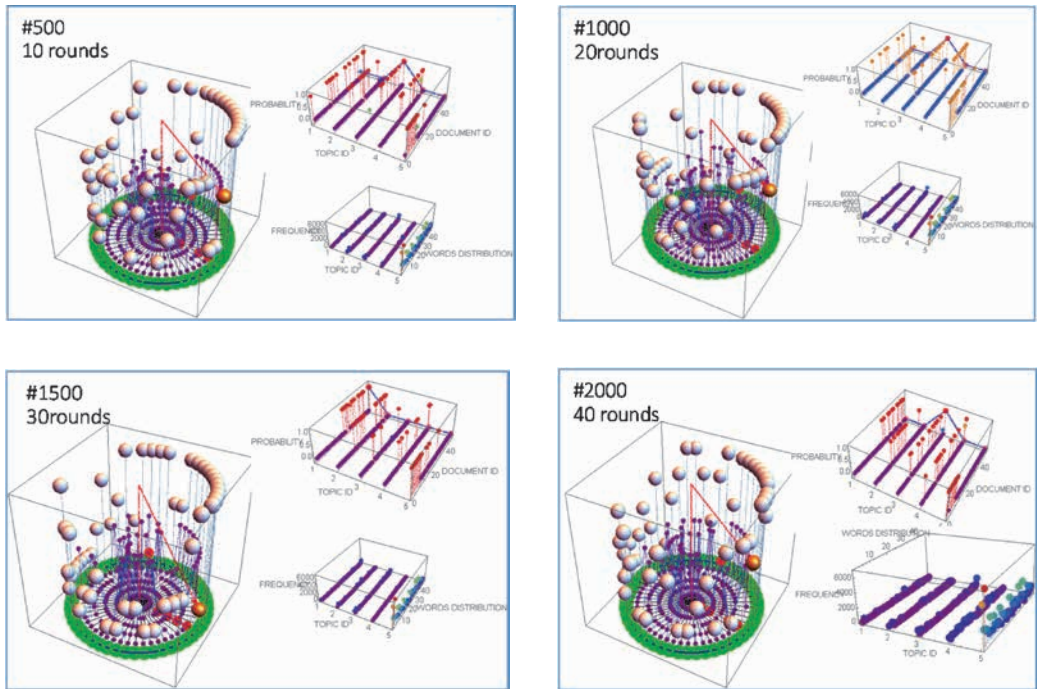


Fig.3. 大学ブランドイメージをギブズサンプラーでクラスタリングしたようす

フィカルモデルでは、トピック分布 θ が D のパネル内に入っていることは、これを意味する。グラフィカルモデルの詳細については [18] が詳しいので参照して頂きたい。

可視化教材を作る場合、トピックモデルでは構造が複雑過ぎて、適さない。シンプルトピックモデルが適していると考える。

B. ギブズサンプラー

マルコフ連鎖とは、次の状態 x' をとる確率が、直前の状態 x にのみ依存するようなルールで動くシステムのことである。MCMCで用いるマルコフ連鎖の遷移確率 π は、メンバーを少しずつ動かし、 π はゼロあるいは殆どゼロとなるようにする。MCMCの代表的アルゴリズムにギブズサンプラーがある [19, 20] [21]。ギブズサンプラーでは、状態を構成する文書が n 個ある場合、 $(n-1)$ 個の文書の状態はそのまま保持し、1 個の文書のみを移動させる。現在所属しているクラスからはずし、新たにどのクラスに属することが確率最大となるかを、与えられた事後確率を用いて計算する。システムは、対象文書が K 個のクラス其々に入った場合の事後確率を計算し、その値が最大となるクラスを選択し、そのメンバーとする。これがギブズサンプラーにおける遷移であり、ギブズ更新操作と呼ばれる。

ギブズサンプラーのアルゴリズムは複雑であるので、その概要をイラストで説明する (図 2 参照)。この図では、我々の作成したギブズサンプラー可視化ツールでバックの図を作成している。この図のクラスタリングでは、文書数は 5、クラス数は 7 としている。真ん中のネコはシステムであり、順番に文書を指して、その文書を新たなクラスに移動させる。その計算の際

に, 他の $(n-1)$ 個の文書の状態はそのまま参照するが, その様子をイラストでは, ネコのヘルメットに, $(n-1)$ 個の文書のクラス及びその単語分布のデータが入っているように描いた。円形に文書が配置され, その高さがクラス ID を示している。システムであるネコは, 回転しながら, 順番に文書のクラスを決めていく。そして, その文書が決められたクラスに遷移する。このプログラムは WEB 公開している。このプログラムは Mathematica で作った。フリーソフトウェアの Wolfram CDF Player をインストールすることで, 自分の PC で動かすことができる (<http://www-cc.gakushuin.ac.jp/~20010570/mathABC/SELECTED/Gibbs/GibbsCat.cdf>)。

遷移が繰り返されるに従い, 文書の球の位置が変動する。状態遷移を十分な回数繰り返した後, メンバーが定常分布になるとその確率分布は変わらなくなる。実際の応用では文書数は有限であり, しかも小さい数であるので, 文書の確率分布は少し変動する。本論文のポイントは, この MCMC における定常状態の変動のようすを可視化により見ることである。

III. 大学ブランドイメージのクラスタリング

本節では, 大学ブランドイメージのクラスタリングを, 本可視化ツールを用いて行った結果を報告する。使用したデータは, リクルート社が2016年3月卒業予定の高校3年生の, 大学に対するイメージなどをアンケートした結果である [22]。本調査は大学関係者間では信頼できるデータとしてよく利用されるデータである。エリアとして関東, 東海, 関西の3つがあるが, 本クラスタリングでは, 関東エリアのデータを用いた。有効回答数は4974件と大きい。調査対象大学数は304である。高校生が個別の大学に対してどのようなイメージをもつかを, 47個のフレーズ項目で表している。うち, 大学の機能的価値として32項目 (例: 教育内容のレベルが高い), 感性的価値15項目 (例: 活気がある感じがする) がある。各項目上位15位までを掲載している。本データでは, ひとつの大学に対して, その項目の%値が示されている。例えば, 「伝統や実績がある」の項目で東京大学が第1位で, 71.4%である。よって, 上位15大学の合計値は100を超える場合が多く, 項目ごとに合計値は異なってくる。

通常, トピック分析は, 文書の単語分布に対して行う。今回の大学ブランドイメージ分析では, 我々は, 大学を文書とみなし, フレーズ項目ごとの得点を単語分布と見なした。分析対象は, 47項目の得点合計の上位50大学とした。モデルはシンプルトピックモデル, 初期値は, 乱数により各大学をクラスに割り振った。トピックモデル及びシンプルトピックモデルを使う際は, 始めにトピック数を指定する必要がある。自動的に適切な数が決まるわけではない。今回の実験では, トピック数 (つまりクラス数) を5とした。これは, 大学数50を考慮し, また, 4から7までトピック数を変化させて結果を見た後に適切と考えられる数として5を選択した。以下では, 40回の遷移を1ラウンドと呼ぶことにする。これは図2のイラストにおいて, システム (ネコ) が1周することに相当する。

可視化ツールでの, この MCMC のようすは図3のようになる。10ラウンド後, 20ラウンド後, 30ラウンド後, 40ラウンド後のようすを示した。20ラウンドから30ラウンド後の間で, 定常状態に入ることが分かる。通常, トピックモデルでは, 25回から100回の繰り返しで十分である, と言われているので, これは納得できる結果である。図2では大学数 (文書数) が5の場合であったのに比較して, 図3では, 大学数が50であるので, 底面の円周は, 50個の球で埋め尽くされている。各球の垂直位置がその大学のクラス ID (トピック ID) を示している。ラウンド

20後以降、クラス ID が変わらない大学群と、異なるクラス ID 間を移動している大学があることが見て取れる（図3参照）。

シンプルトピックモデルでは、トピック毎の単語分布が得られる。本事例では、クラス毎のフレーズ項目分布が得られる。例として、図4にクラス ID 5のフレーズ項目分布を示す。このフレーズ項目分布から、そのクラスの特徴を解釈していく。我々がまとめた各クラスの特徴を表1に示した。クラス ID 5は、他の4クラスに比較して非常にスコアが高い。クラス ID 2は、医大、薬科大、農業工業関連大学などが含まれていた。高校生が専門性が高い大学として評価していることに合致していると考えられる。クラス ID 3は、高校生のイメージが、上品、おしゃれ、伝統的、などが高く、クラスタリングの結果の8大学中、5大学が女子大学であり、他3大学も上品というイメージの大学であった。クラス ID 4は、高校生のイメージが教育内容のレベルが高い、などが高く、クラスタリングの結果も、8大学中殆どが国公立大学であった。クラス ID 5は、スコアが最も高いクラスであり、東大、早稲田、慶応などが含まれていた。総合的に見て、クラスタリングの結果は、東京在住の大学教員2名の常識に照らしても、的確なクラスタリングが行われていると判断できた。

表1 クラスタリングの結果

クラス ID	大学数	特長
1	7	有名・クラブ活動
2	11	専門性・資格取得
3	8	上品・おしゃれ・伝統的
4	8	高等教育・学力・学費
5	13	最難関・有名・伝統・自慢できる

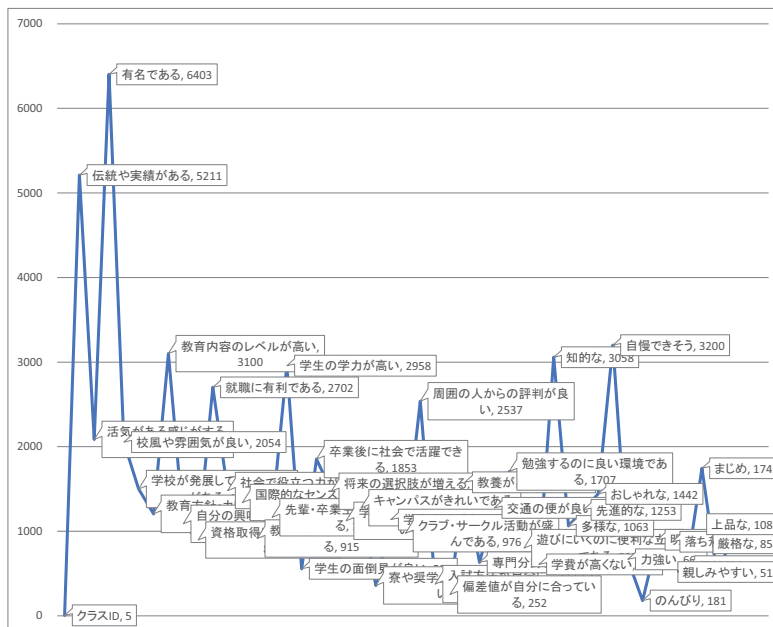


図4 クラス ID 5の大学のフレーズ項目分布

このクラスタリング結果を踏まえて、再度、可視化ツールのようにすを見ていく (図3参照)。10ラウンド後において、既にクラス5の早稲田・慶応などのクラスは確定しており、以後も動かない。円周中で常に高い位置で留まっている球の集団がこれに相当する。図3の各図の右上は、TOPIC ID と document ID の確率分布である。換言すると、クラスと大学の確率分布である。どの大学がどのクラスに属しているかを示している。ラウンド10後の各大学のクラス確率分布 (図では、TOPIC ID) を見ると、高校生のイメージで最著名大学としての評価が高い大学の多くは、クラス5で確率1になっていることが見える。興味深いことに、東大はクラス5とクラス4をMCMC定常状態で行き来している (図5参照)。図5は、横軸に大学および、ラウンド数、垂直方向にクラスIDを取っている。東京大学は、20ラウンド以降、クラス5と4を行き来していることが分かる。

一般に多くの大学では、そのクラス確率分布は、複数のクラスに分散している。これは大学の特色が表1に示したような、ひとつに固定されているのではなく、複数の特色をもつことを意味する。今回の分析はシンプルトピックモデルを使って分析したが、より複雑なトピックモデルを使って、複数のトピックをもつことを許可した場合、複数の特色をもったことを前提としたクラスタリングとなる。しかし、可視化でMCMCの動きを理解する、という教育的視点に立つ場合、このようなシンプルトピックモデルにしたほうが分かり易い。複雑過ぎる構成では、分かりにくくなる可能性が高いからである。

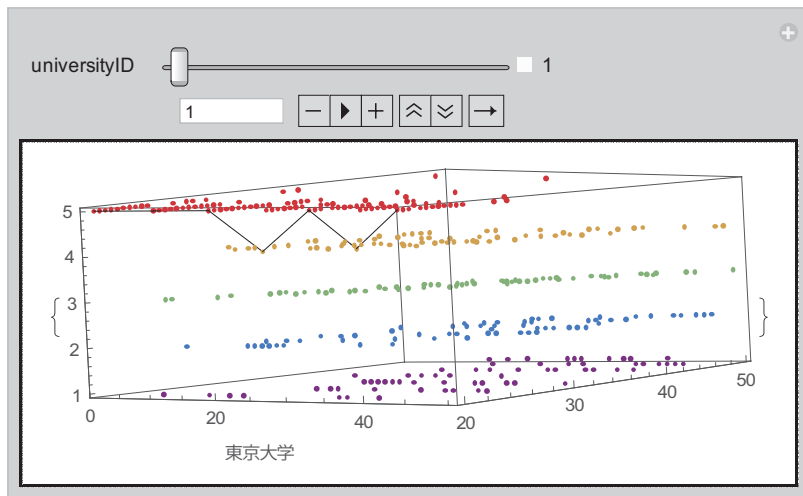


Fig.5. 20ラウンド以降50ラウンドまでの、1ラウンドごとの東京大学のクラスの推移。

図3の各図の右下のグラフは、TOPIC ID と word distribution の分布である。換言すると、クラスIDとそのフレーズ項目分布である。圧倒的にクラス5の分布が大きいことが見て分かる。これを各クラスでグラフ化したものが、図4である。

大学ブランドイメージ・クラスタリングの例を使って、ギブズサンプラー及びシンプルトピックモデルのアルゴリズムを説明する。ギブズサンプラーでは、ある大学 (以下、Aとする)

の新しい状態（クラス ID）を決めるために、まず大学 A の影響をフレーズ項目分布から除く。他の $(n-1)$ 個、ここでは49個の大学のデータのみで構成される、クラス ID のフレーズ項目分布（図 4 に相当）を作る。そして、大学 A のフレーズ分布が最も類似しているフレーズ項目分布を選択する。システムは図 4 の円の半径の上に表示される確率密度分布の最も値が大きいクラス ID を選択する。そして、そのクラスに大学 A を入れる。こうして、新しい状態が決まる。

以下では、この 5 つのクラスへのクラスタリングを基に、MCMC の定常状態で、大学がどのようなクラス間変動をするかを可視化ツールで見ていく。興味深い発見として以下があった。

- (1) クラス 3 と 5 を頻繁に行き来する大学が 2 つあった。ひとつは女子大であった。もうひとつは共学大学であった（図 6 参照）。ラウンド 20 後時点で、クラス 3 に属していた大学のうち、クラス 1 と 3 を頻繁に行き来するものがある。これは、本来、「大学は 2 つ以上のクラス特長をもつ場合がある」ことに対応している。この特質が可視化で見られることは、本可視化ツールの特長であり、それによりトピックモデルの理解が容易になる。トピックモデルとのモデルの違いを理解させるために、まずこの可視化ツールでシンプルトピックモデルでの動きを見せたのち、図 1 のグラフィカルモデルを見せて、トピックモデルでは、大学のフレーズ項目ごとに異なる確率分布をもてることを説明すると、学生も、その結果として、ひとつの大学が複数のクラスがもてる、ということを理解し易いであろう。
- (2) 50 という有限個数の大学数の MCMC 定常状態においては、クラス間の移動がかなり大きい。これはこの应用到に特化したことなのか否かこの実験だけからでは判定不能であるが、予想よりも大きい移動があることを視覚的に確認できた。これは可視化ツールの効用と言える。

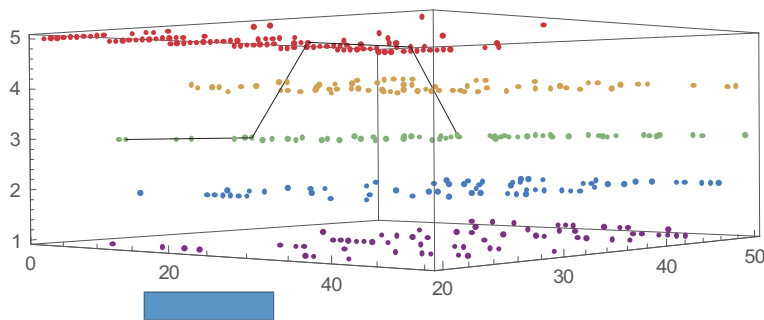


Fig.6. MCMC の定常状態で、クラス3と5を行き来する大学のようす

IV. 関連研究

本節では、ギブズサンプラー及びトピックモデルの可視化に関する関連研究について述べる。

トピックモデルの事後確率分布の導出については [23] が丁寧に説明していて分かり易い。ギブズサンプラーの可視化の最も古く有名な可視化は David JC MacKay によるものである [24]。これらにおいては 2 変数ガウス分布でギブズサンプラーにより MCMC を行う様子が示されている [18, 25]。Mathematica のオンラインマニュアルの項目 “howto/PerformAMonteCarloSimulation” においても、2 変数の確率分布が用いられたギブズサンプラーのプログラムが示されている。しかし、本稿で示したように、実分析では大学数 50 というように、扱う確率変数の数は大きい。ビッグデータの分析では、10 万以上にもなる。ギブズサンプラーの特色は $(n-1)$ 個の情報を基に決める、という点であるが、2 変数関数ではそれを表現しきれていない。

また、トピックモデルの可視化は各種行われているが [26]、それらはトピックと単語分布の関係の可視化を中心にしたもので、ギブズサンプラーのアルゴリズムを可視化で説明しようとしたものではない。我々の可視化ツールは、アルゴリズムやその数学プロセスを説明することに注力したものであり、 $(n-1)$ 個の情報を使うことを強調するために、円周上に大学を並べ、それらを順番に更新していくレイアウトにしたところに、分かり易さの工夫があると考えられる。

V. まとめ

ベイズ推論は、文書のトピック分析を始め、画像のパターン認識のほか、画像を用いた服装コーディネート推薦、買い物履歴情報からの消費パターン分析など、広い分野で活用されている。本稿では、ベイズ推論の手法を大学ブランドイメージ分析に適用した。我々は、マルコフ連鎖モンテカルロ法の代表的アルゴリズムであるギブズサンプラーの実装により、シンプルトピックモデルによるベイズ推論のクラスタリングの過程を可視化するツールを開発した。このツールにより、大学ブランドイメージという共通の知識を共有するデータを対象として、クラスタリングのマルコフ連鎖の変動のようすを可視化した。

応用事例として、リクルート社の高校 3 年生対象のアンケート結果を用いて、関東のスコア上位 50 の大学のクラスタリングを行った。5 つのクラスを作ったところ、5 つの明確な特質をもったクラスが作られた。分析結果に対しては、「高校生の知識が、我々のような大学教員に比較して豊富ではないため、実体と違うイメージをもっているのではないか」という疑問を感じる箇所もあったが、これも高校生のブランドイメージの実態の分析ということで、その分析結果には意味があると考えられる。例えば、理系の偏差値の高い大学のブランドイメージが我々の予想と異なる例があったが、これは調査対象の高校生の理系比率が低いいため、結果として理系大学について興味が低い人が多かったからではないか、と推察した。

総体的に見てクラスタリングの結果は、大学教員の著者 3 人にとっても、納得のいく予想とほぼ一致するものであった。このような、ある程度予想がつく、既知のデータを使う利点は、共有する知識に基づく分析事例を示し、その事例からその背景にある数学的手法を連想させ、理解させる、ことである。この MCMC における定常状態の連鎖プロセスにおいては、ある大学が 2 つのクラスの間を頻繁に行き来するなどの興味深い変動が見られた。例えば、東大京大などが頻繁にクラス 5 と 4 を行き来するようすが見られた。これは東大京大に対する高校生のイメージが、有名校と国立大学、という 2 つの代表的イメージをもつからである。

可視化の利点は、定常状態に入ったか否かの判定を視覚的に行える点である。本事例では、

20から30ラウンドの間で定常状態になったと判断できた。また、定常状態の可視化は、クラスタリングのツールをただ使って結果を出すだけの作業とは違い、その背景にあるギブズサンプラー、シンプルトピックモデル、などの数学プロセスを理解するための教材にもなる。そうした数学プロセスの理解は、MCMCによる分析において的確な分析を行うために必須であるので、こうした数学教育は重要と考える。我々は今後とも、ベイズ推論に係る数学の可視化教材を作っていくつもりである。

謝辞

本研究の一部は2015年度、学習院大学計算機センター特別研究プロジェクトの助成を受けた。また、常日頃、機械学習についてご教示頂いている、同 計算機センター久保山哲二教授に感謝する。また、サイエンス・イラストとして、ネコのイラストを描いてくれた経済学部副手 村上朱音氏に感謝する。

参考文献

- [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] Y. Shirota, T. Hashimoto, and B. Chakraborty, "Deductive Reasoning for Joint Distribution Probability in Simple Topic Model," *Proc. of IIAI International Congress on Advanced Applied Informatics 2016, 12-16 July, 2016, Kumamoto, Japan*, 2016.
- [6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp.993-1022, 2003.
- [7] D. Blei, and J. Lafferty, "Dynamic topic models," *Proceedings of the 23rd International Conference on Machine Learning*, 2006.
- [8] D. M. Blei, "Probabilistic topic models," *Commun. ACM*, vol. 55, no. 4, pp. 77-84, 2013/07/01, 2012.
- [9] 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*.
- [10] Y. Shirota, T. Kuboyama, T. Hashimoto, S. Aramvith, and T. Chauksuvanit, *Study of Thailand People Reaction on SNS for the East Japan Great Earthquake -Comparison with Japanese People Reaction-*, Occasional Papers No. 59: Research Institute for Oriental Cultures Gakushuin University, 2015.
- [11] 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.

- [12] 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.
- [13] 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.
- [14] Y. Shirota, T. Hashimoto, and B. Chakraborty, "Visualization for University Brand Image Clustering With Simple Topic Model," *Proc. of IIAI International Congress on Advanced Applied Informatics 2016, 12-16 July, 2016, Kumamoto, Japan*, 2016.
- [15] 橋本隆子, 久保山哲二, 白田由香利, "ソーシャルメディアを対象としたマーケティング解析 - 時事問題をきっかけとした想定外の消費行動抽出 -, " *学習院大学経済論集, Vol.47, No.4, 2012年1月*, pp.263-280, 2011.
- [16] 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.
- [17] 岩田具治, *トピックモデル: 講談社サイエンティフィク*, 2015.
- [18] C. M. Bishop, *Pattern Recognition and Machine Learning*: Springer, 2006.
- [19] 伊庭幸人, *ベイズ統計と統計物理*: 岩波書店, 2003.
- [20] 伊庭幸人, 種村正美, 大森裕浩, 和合肇, 佐藤整尚, 高橋明彦, *計算統計II マルコフ連鎖モンテカルロ法とその周辺*: 岩波書店, 2005.
- [21] D. Koller, and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*: The MIT Press, 2009.
- [22] リクルート社, "進学ブランド力調査2015," *リクルート カレッジマネジメント*, vol.194 / Sep. - Oct. 2015, pp. 6-44, 2015.
- [23] I. Sato, *Statistical Latent Semantic Analysis Based on Topic Model*: Corona Publishing Co., 2015.
- [24] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*: Cambridge university press, 2003.
- [25] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, third edition ed.: Chapman & Hall / CRC, 2004.
- [26] A. J.-B. Chaney, and D. M. Blei, "Visualizing Topic Models," *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*, 2003.