

ディリクレ過程混合モデルに基づく 離散データの共クラスタリング

桑田修平^{†1} 山田武士^{†2} 上田修功^{†2}

ディリクレ過程混合モデルを用いた共クラスタリング手法を提案する。共クラスタリングとは、ユーザのアイテム購入履歴などのような行列形式で表現可能なデータに対して、行（ユーザ）と列（アイテム）を同時にクラスタリングする問題である。提案法は、ユーザ（もしくはアイテム）クラスごとにアイテム（もしくはユーザ）クラス次元の多項分布を仮定し、互いに同じクラスを選択しあったときに購入行動が生じると仮定したモデルに基づいて共クラスタリングする。提案法は、ユーザ（アイテム）クラス数を事前に設定することなく共クラスタリングができ、特に、購買履歴のような欠損値を含むデータに対してより良いクラスタリング精度を示す。実データを用いた実験により、ディリクレ過程混合モデルに基づく従来手法（無限関係モデル）と比べて、より精度の高い共クラスタリング結果が得られることを示す。

Co-clustering Discrete Data Based on the Dirichlet Process Mixture Model

SHUHEI KUWATA,^{†1} TAKESHI YAMADA^{†2}
and NAONORI UEDA^{†2}

We propose a new co-clustering method based on the Dirichlet process mixture model (DPM). Co-clustering is the problem of simultaneously clustering rows and columns of a data matrix, such as purchase history data of users and catalog items. The proposed method assumes that each user (or item) class has a multinomial distribution over item (or user) classes to select, and a purchase occurs when both selections of user and item classes match. The proposed method can co-cluster users and items without knowing the true numbers of clusters. The experimental results show that the proposed method can provide better co-clustering results compared with Infinite Relational Model (IRM), another previously proposed co-clustering method based on the DPM, especially for data matrices that contain missing data.

1. はじめに

インターネットの普及などにより世界中から容易に商品を購入できるようになった。その結果、消費者（以降、ユーザと呼ぶ）は自分の嗜好により適した商品（以降、アイテムと呼ぶ）を購入することが可能となり、企業にとって、ユーザの嗜好やニーズをより詳細に把握することがますます重要になってきている。これまで、ユーザのニーズを把握する1つのアプローチとして、複数のユーザの商品購入履歴（ユーザがすでに購入したアイテム集合、以降、購買履歴と呼ぶ）をもとに、ユーザを、購買履歴の似通ったグループに分けるクラスタリング手法が利用されている¹⁴⁾。ユーザが本質的にいくつかの範疇に分けられるとは限らないが、現実を簡略化した近似としてユーザクラスごとに特徴を分析することで、ユーザの購買傾向を把握するアプローチである。

これに対し、ユーザだけでなく、アイテムも同時にクラスタリングする「共クラスタリング (Co-clustering)」を利用した手法が提案されている^{4),18)}。ユーザやアイテムの一方をクラスタリングする際に、他方のクラスタリング結果を利用しながらユーザとアイテムを相互にクラスタリングすることで、ユーザクラスとアイテムクラスから定まる“ユーザ・アイテムブロック”が求まる。ユーザのみをクラスタリングするアプローチと比べて、より粒度の細かい購買傾向に基づくクラスタリング結果が得られることになり、得られたユーザ・アイテムブロックごとに購買傾向の特徴を把握することによって、消費行動に関するより有用な情報を得ようとするアプローチである。

共クラスタリングは、購買履歴に限らず遺伝子データやネットワークデータなど様々な領域において適用されており、これまで、

- 非負行列因子分解に基づく手法²⁾,
- グラフの分割に基づく手法^{5),16)},
- 確率モデルに基づく手法^{6),17)},

など多くの手法が提案されている。しかし、これらの従来法は、共クラスタリングを行う際にクラス数をあらかじめ設定しておく必要がある。通常のデータ分析においてはクラス数は未知であることが多いため、これらの手法を用いる際にはクラスタリング基準とは別の、

^{†1} 株式会社 NTT データ技術開発本部
R&D Headquarters, NTT Data Corporation

^{†2} NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories

クラス数を決定するための基準（たとえば，AIC や BIC などのモデル選択基準）が必要となる。

クラス数をあらかじめ設定する必要のあるこれらの従来法に対して，Kemp らにより提案された無限関係モデル（Infinite Relational Model）¹⁰⁾ は，確率モデルに基づいて，クラス数を事前に与えることなく複数の集合を同時にクラスタリングする手法であり，購買履歴を用いたユーザとアイテムの共クラスタリング問題にも適用可能である。このモデルでは，クラス数の決定基準がクラスタリングをする際に用いるモデル学習の基準に含まれており，クラス数は与えられたデータから学習される。しかし，無限関係モデルには欠損値の扱いが得意でないという問題点があり，一般に多くの欠損値が存在すると想定される購買履歴に適用することは適切であるとはいえない。

本論文では，購買履歴のような欠損値を含むデータに適した確率モデルに基づく共クラスタリング手法を提案する。ここで，無限関係モデルと同様に，提案法もクラス数を事前に与える必要がないという特長を有する。具体的には，ユーザのアイテム購入行動に対して，無限関係モデルが「ある確率に従ってアイテムを購入する/購入しないを決定する」モデル化を行っているのに対して，提案法は「ユーザは複数候補の中からある確率に従ってアイテムを選択して購入する」モデル化を行う。つまり，アイテムを購入する事象のみに着目したモデル化を行うことで欠損値に対処する。提案法は，クラス数に対する事前分布としてディリクレ過程（Dirichlet Process）を仮定した多項分布モデルに基づいており，ディリクレ過程混合モデル（Dirichlet Process Mixture Model）¹²⁾ を応用したモデルといえる。映画の評価データを用いた実験の結果，無限関係モデルと比べてよりクラスタリング精度の高い共クラスタリング結果が得られることを示す。

本論文の構成は以下のとおりである。まず，2章で，本研究の問題設定を示す。次に，3章で従来法を説明し，4章で無限関係モデルとその問題点について説明する。続いて，5章で提案法の詳細を述べ，6章で評価実験と考察を行う。最後に7章でまとめと今後の課題を述べる。

2. 問題設定

N 人のユーザと M 個のアイテムからなる， $N \times M$ 行列の購買履歴データを R とする。ここで， R の第 (i, j) 要素 $r_{i,j}$ は，ユーザ i とアイテム j の購買関係，

$$r_{i,j} = \begin{cases} 1, & \text{ユーザ } i \text{ はアイテム } j \text{ を購入済,} \\ 0, & \text{ユーザ } i \text{ はアイテム } j \text{ を未購入,} \end{cases}$$

を表す。本研究で扱う問題は，購買履歴データ R を用いて， N 人のユーザと M 個のアイテムの両方をクラスタリング（共クラスタリング）することであり，ユーザのクラス数 K とアイテムのクラス数 S は未知であるものとする。ここで，ユーザ i ($=1, 2, \dots, N$) の帰属クラスインデックスを $z_i \in \{1, \dots, k, \dots, K\}$ ，アイテム j ($=1, 2, \dots, M$) の帰属クラスインデックスを $w_j \in \{1, \dots, s, \dots, S\}$ とし， $Z = \{z_1, z_2, \dots, z_N\}$ ， $W = \{w_1, w_2, \dots, w_M\}$ とすると，具体的には，購買履歴データ R が与えられた下で，ユーザのクラス割当て Z とアイテムのクラス割当て W を求めることが問題である。ただし，購買履歴データ R には欠損値が存在するものとする。つまり， $r_{i,j} = 1$ が欠損して $r_{i,j} = 0$ となっているデータが存在するものとする。本研究の目的は，欠損値に対処した共クラスタリング手法を提案することである。

ここに，将来購入される可能性の高いアイテム（現在は $r_{i,j} = 0$ であるが将来 $r_{i,j} = 1$ となる可能性が高いアイテム）が，購入される可能性のより高いユーザ・アイテムブロック (k, s) に割り当てられている程度が大きいほど，良い共クラスタリング結果である。

3. 従来研究

共クラスタリングの従来手法として，

- 非負行列因子分解に基づく手法²⁾，
- グラフの分割に基づく手法^{5),16)}，
- 確率モデルに基づく手法^{6),17)}，

などがあげられる。以下，各手法の概要を説明する。

3.1 非負行列因子分解に基づく手法

非負行列因子分解（Non-negative Matrix Factorization）に基づく手法では，以下の基準に基づいて，購買履歴データ R を U, J, V の3つの行列に分解する：

$$\begin{aligned} & \min_{U \neq 0, J \neq 0, V \neq 0} \|R - UJV^T\|^2, \\ & \text{subject to } U^T U = V^T V = I. \end{aligned}$$

U は $N \times K$ ， J は $K \times S$ ， V は $M \times S$ の非負行列をそれぞれ表す。ここで， U はユーザごとのユーザクラス帰属度， V はアイテムごとのアイテムクラス帰属度をそれぞれ表してお

り, U, V を用いることで共クラスタリングを行う。

なお, 関連手法として, 特異値分解を利用して $R=U\Sigma V^T$ と分解する LSI (Latent Semantic Indexing)³⁾ がある。ここで, U は $N \times K$, V は $M \times K$ の行列であり, Σ は $K \times K$ の対角行列である。 U, Σ, V から, フロベニウスノルムの意味で最も良い近似行列が得られ, ユーザとアイテムは共通な K 個のクラスにクラスタリングされる。また, LSI を確率モデルとして拡張した pLSI (probabilistic LSI)⁸⁾ は, 各クラスへの帰属確率をもとにユーザとアイテムを K 個にクラスタリングする。LSI や pLSI はユーザクラスとアイテムクラスが共通であるため, ユーザとアイテムをそれぞれ異なるクラスにクラスタリングする共クラスタリングの方が, より一般的で柔軟であるといえる。

3.2 グラフの分割に基づく手法

グラフの分割に基づく手法では, ユーザとアイテムそれぞれをノードとする 2 部グラフを用いて購買履歴データ R を表現する。ここで, 購入したユーザと購入されたアイテムの間にはエッジ (リンク) が張られている。このグラフに対して, 異なるクラスどうしのノードをつなぐリンク数を最小にするようにグラフを分割する操作 (Minimum Cut) などにより, ノードで表現されたユーザとアイテムを共クラスタリングする。

3.3 確率モデルに基づく手法

確率モデルに基づく手法では, 購買履歴データ R に対して確率分布を仮定する。

文献 6) では, 共クラスタリング前のユーザとアイテムの同時分布から計算される相互情報量をできるだけ保存するように, ユーザとアイテムをクラスタリングする。つまり, ユーザクラス数 K とアイテムクラス数 S を制約として与えた下で, ユーザクラスとアイテムクラス単位でまとめた同時分布から計算される相互情報量を最大化することによって, 共クラスタリングを行う。

また, 文献 17) では, ユーザクラス/アイテムクラスごとに確率分布を仮定した混合モデルに基づいて, 購買履歴データ R から, ユーザ/アイテムそれぞれのクラス帰属確率を学習する。学習の結果得られた各クラスへの帰属確率を用いることにより, ユーザとアイテムを共クラスタリングする。

3.4 従来法の共通点

これらの従来法に共通する点は, クラス数をあらかじめ設定しておく必要がある, という点である。しかし, クラス数は事前には未知である場合が多い。そのため, クラス数を決定する際には, クラス数が異なるクラスタリング結果を求めた後, その中から, たとえば AIC や BIC などのモデル選択基準を用いてクラス数を決定することなどが必要となる。つまり,

クラス数を決定するためには, クラスタリングを行う基準とは別の基準を用意する必要がある, クラスタリングとクラス数の決定が異なる枠組みで扱われているといえる。

4. 無限関係モデル

これらの従来法に対して, 無限関係モデル¹⁰⁾ は, クラス数をあらかじめ設定することなく, ユーザとアイテムを共クラスタリングできる。クラス数は購買履歴データ R から学習される。無限関係モデルは, ノンパラメトリックベイズモデル¹⁵⁾ の 1 つであり, クラス数に対しても確率分布を仮定することで, クラス数の生成をモデルに組み込み, クラス数の決定をモデルパラメータの学習の枠組みで行う。つまり, クラスタリングとクラス数の決定を同じ枠組みで行う。クラス数に対する分布として, ディリクレ過程 (Dirichlet Process: DP) やポアソン-ディリクレ過程 (Poisson-Dirichlet process: PD)¹⁹⁾ などが用いられる。DP や PD はノンパラメトリックな分布であることから, ノンパラメトリックベイズモデルと呼ばれており, 無限関係モデルのほかに, 分析対象データに合わせたモデルが提案されている。たとえば, 単語の共起データなどの頻度データを想定したモデル (Frequency-based IRM)¹¹⁾ や, アイテムに対するユーザの評価値からなる “評点” 履歴データを想定した無限隠れ関係モデル (Infinite Hidden Relational Model)²⁰⁾ などがある。

ここで, クラス数の決定について, 前述したモデル選択基準が利用可能である。しかし, モデル選択の枠組みでは, クラス数に関する尤度に基づいてクラス数が決定されることになり, クラス数に関する事後確率を用いてクラス数を決定する無限関係モデル (ノンパラメトリックベイズモデル) とは異なる。

無限関係モデルは, 種類の異なる複数の集合間の関係を表現する確率モデルであり, 関係の類似性から複数の集合を同時にクラスタリングする。ユーザ集合とアイテム集合の 2 つの集合間の購買関係からなる購買履歴データ R においては, 購買履歴の類似性から, ユーザとアイテムをそれぞれ分割できる。

4.1 購入履歴の生成に対するモデル

ユーザの帰属クラスの割当て Z と, アイテムの帰属クラスの割当て W が与えられた下での, 購買履歴データ R の生成モデル $p(R|Z, W, \Theta)$ を, ベルヌーイ分布を用いて以下のように表現する。

$$p(R|Z, W, \Theta) = \prod_{i=1}^N \prod_{j=1}^M p(r_{i,j} | z_i, w_j, \Theta),$$

$$p(r_{i,j}|z_i, w_j, \Theta) = (\theta_{z_i, w_j})^{r_{i,j}} (1 - \theta_{z_i, w_j})^{1-r_{i,j}}. \quad (1)$$

ここで、 Θ はモデルパラメータを表す．このモデルでは、ユーザの帰属クラス z_i とアイテムの帰属クラス w_j により定まるユーザ・アイテムブロック (z_i, w_j) ごとに購入確率 θ_{z_i, w_j} ($\in \Theta, 0 \leq \theta_{z_i, w_j} \leq 1$) が割り当てられており、ブロックごとの購入確率をもとにベルヌーイ分布に従って履歴 $r_{i,j}$ が生成される．つまり、ユーザクラス k に属するユーザは、アイテムクラス s に属するアイテムを $\theta_{k,s}$ の確率で購入するモデルである．

4.2 クラス数の生成に対するモデル

クラス数をデータから学習する無限関係モデルでは、帰属クラスの割当て Z, W に対して、すべての可能な（一般には無限に存在する）分割に対する分布であるディリクレ過程 (DP) を事前分布 $p(Z), p(W)$ として仮定する（付録 A.1 参照）．DP はノンパラメトリックな分布のため、実際に DP を構成する際には Chinese Restaurant Process (CRP)¹³⁾ や Stick-breaking Process (SBP)¹⁾ などが用いられる．文献 10) では、CRP を用いて DP を構成している．以下、CRP を用いた $p(Z)$ の表現方法を具体的に説明する．ただし、 $p(W)$ についても同様な説明が可能である．

いま、 $i-1$ 人の客（ユーザ）がすでに“中華料理店”のいずれかのテーブル（クラス）に着席して（帰属して）いるとする．このとき、新たに店に入ってきた i 番目の客（ユーザ）がテーブル（クラス） k を選択し、着席する確率 $P(z_i = k)$ は以下のように表現される．

$$P(z_i = k) = \begin{cases} \frac{n_k}{i-1+\alpha}, & n_k > 0, \\ \frac{\alpha}{i-1+\alpha}, & \text{otherwise.} \end{cases} \quad (2)$$

ここで、 n_k はテーブル k にすでに着席している客の数、 $\alpha (> 0)$ は既知のハイパーパラメータ（集中度パラメータと呼ばれる）である．式 (2) から分かるとおり、すでに客が着席しているテーブル（既存クラス）を選ぶだけでなく、誰も着席していない新規テーブル（新規クラス）を選ぶ可能性もあるため、クラス数は可変となる．また、新規テーブル（クラス）を選択する確率は α によって調節され、既存テーブルについては、すでに着席している客が多いほど選択されやすい．式 (2) を $i=1$ から N まで繰り返すことにより以下を得る．

$$p(Z; \alpha) = \frac{\alpha^K \prod_{k=1}^K (n_k - 1)!}{\alpha(\alpha + 1) \cdots (\alpha + N - 1)}.$$

K は最終的に得られたユーザクラス数を表す．式 (2) から、 α の値が小さいほど少ないクラス数が生成されやすく、 α の値が大きいほど大きなクラス数が生成されやすいことが分か

る．なお、 α の値は、分布 $p(\alpha)$ を仮定することにより推定可能であるが¹³⁾、本論文では既知のハイパーパラメータとする．

4.3 無限関係モデルにおける同時分布

以上より、購買履歴データ R の同時分布は、

$$p(R, Z, W, \Theta; \alpha, \beta, \gamma) = p(R|Z, W, \Theta)p(Z; \alpha)p(W; \beta)p(\Theta; \gamma),$$

と表される．ここで、 $p(\Theta; \gamma)$ は購入確率 Θ に対する分布であり、ベルヌーイ分布 $p(R|Z, W, \Theta)$ の共役事前分布であるベータ分布を $p(\Theta; \gamma)$ に仮定する．つまり、上式右辺の各分布は以下のように表される．

$$p(R|Z, W, \Theta) = \prod_{i=1}^N \prod_{j=1}^M (\theta_{z_i, w_j})^{r_{i,j}} (1 - \theta_{z_i, w_j})^{1-r_{i,j}},$$

$$p(Z; \alpha) = \frac{\alpha^K \prod_{k=1}^K (n_k - 1)!}{\alpha(\alpha + 1) \cdots (\alpha + N - 1)},$$

$$p(W; \beta) = \frac{\beta^S \prod_{s=1}^S (m_s - 1)!}{\beta(\beta + 1) \cdots (\beta + M - 1)},$$

$$p(\Theta; \gamma) = \prod_{k=1}^K \prod_{s=1}^S \frac{\Gamma(\gamma_1 + \gamma_0)}{\Gamma(\gamma_1)\Gamma(\gamma_0)} \times (\theta_{k,s})^{\gamma_1-1} (1 - \theta_{k,s})^{\gamma_0-1}.$$

ここで、 $\alpha, \beta, \gamma = \{\gamma_0, \gamma_1\}$ は既知のハイパーパラメータであり、 K, S は学習されたユーザクラス数、アイテムクラス数をそれぞれ表す．また、 n_k はユーザクラス k に帰属するユーザ数、 m_s はアイテムクラス s に帰属するアイテム数をそれぞれ表す：

$$n_k = \sum_{i=1}^N I(z_i = k), \quad m_s = \sum_{j=1}^M I(w_j = s).$$

ただし、関数 I は指数が真のとき 1、そのほかのとき 0 を返すインジケータ関数を表す．

4.4 学習アルゴリズム

購買履歴データ R を用いて上記モデルを学習する．すなわち、ユーザの帰属クラスの割当て Z とアイテムの帰属クラスの割当て W の事後確率 $P(Z, W|R; \alpha, \beta, \gamma)$ を最大化する Z, W を求める．具体的には、

$$P(Z, W|R; \alpha, \beta, \gamma) \propto P(R|Z, W; \gamma)P(Z; \alpha)P(W; \beta),$$

$$= \int_{\Theta} P(R|Z, W, \Theta) p(\Theta; \gamma) d\Theta P(Z; \alpha) P(W; \beta),$$

により計算される事後確率を最大化する．ここで，自然共役な事前分布を仮定しているため上式の積分は解析的に解ける．ギブズサンプリング¹³⁾を用いた具体的な学習アルゴリズムについては文献 10) を参照されたい．

4.5 無限関係モデルの問題点

式 (1) が示すとおり，無限関係モデルは，購買履歴データ R における $r_{i,j}=1$ (購入済) と $r_{i,j}=0$ (未購入) の両方をモデル化している．しかし，多くの購買履歴データの場合， $r_{i,j}=1$ に比べて $r_{i,j}=0$ となる要素数が圧倒的に多く (すなわちデータは疎で)，また，アイテム数が十分多い場合，ユーザは特定のアイテムを現在は未購入であっても存在を知らないだけで，将来的には購入する ($r_{i,j}=0$ の箇所が $r_{i,j}=1$ となる) 可能性がある．すなわち，これらを欠損値として扱う必要があり，購買履歴データ R は数多くの欠損値を含むデータであるといえる．つまり， $r_{i,j}=0$ すべてを「未購入」として扱うことは適切ではない．したがって，購入する/購入しないをモデル化した無限関係モデルを，購買履歴データに適用するうえでは限界があるといえる．欠損値を含んだデータに適したモデル化が必要であることが，提案法の導入目的である．

5. 提案法

無限関係モデルは「ユーザは，アイテムごとに購入する/購入しないを確率的に決定する」という仮定に基づくモデルであるが，提案法は「ユーザは，複数のアイテムの中から購入するアイテムを確率的に選択し購入する」という仮定に基づいてモデル化を行う．このような仮定に基づく提案法は，モデルを学習する際に，購入する事象 ($r_{i,j}=1$) のみを考慮し，購入しない事象 ($r_{i,j}=0$) を考慮しない．つまり， $r_{i,j}=0$ に該当するデータを生成モデルに含めないことで，欠損値に対処する．ここで，ユーザの帰属クラスの割当て Z とアイテムの帰属クラスの割当て W の生成については，無限関係モデルと同様に DP を仮定する．

5.1 購買履歴の生成に対するモデル

提案モデルは，購入する事象 ($r_{i,j}=1$) のみのデータ集合，つまり，購買履歴集合 $\mathcal{R} = \{r_{i,j} | r_{i,j} \in R, r_{i,j} = 1\}$ を観測値とする確率モデルである．つまり，ユーザの帰属クラスの割当て Z と，アイテムの帰属クラスの割当て W が与えられた下での，購買履歴集合 \mathcal{R} の出現確率を多項分布を用いて以下のように表現する．

$$p(\mathcal{R}|Z, W, \Theta, \Phi) = \prod_{r_{i,j} \in \mathcal{R}} p(r_{i,j} | z_i, w_j, \Theta, \Phi),$$

$$p(r_{i,j} | z_i, w_j, \Theta, \Phi) = \theta_{z_i, w_j} \phi_{w_j, z_i}.$$

ここで， $\theta_k \in \Theta$ ， $\phi_s \in \Phi$ はモデルパラメータを表し， $\theta_k = \{\theta_{k,1}, \theta_{k,2}, \dots\}$ ($0 \leq \theta_{k,s} \leq 1$ ， $\sum_s \theta_{k,s} = 1$ を満たす) は，ユーザクラス k のアイテムクラス選択確率 (アイテムクラス次元の多項分布のパラメータ) を表し， $\phi_s = \{\phi_{s,1}, \phi_{s,2}, \dots\}$ ($0 \leq \phi_{s,k} \leq 1$ ， $\sum_k \phi_{s,k} = 1$ を満たす) は，アイテムクラス s のユーザクラス選択確率 (ユーザクラス次元の多項分布のパラメータ) を表す．モデルパラメータ Θ ， Φ の満たすべき条件から，以下の式が満たされる：

$$\sum_{s'=1}^S \sum_{k'=1}^K \theta_{k,s'} \phi_{s,k'} = 1.$$

つまり，選択候補 (ユーザクラス \times アイテムクラス) に関する和が 1 となるモデルである．上記提案モデルは，以下の 3 つのステップを経て，履歴 $r_{i,j}=1$ が生成されると仮定するモデルである．

1. ユーザ i はアイテムクラス選択確率 θ_{z_i} に従って，アイテムクラス s' を選択，
2. アイテム j はユーザクラス選択確率 ϕ_{w_j} に従って，ユーザクラス r' を選択，
3. $s'=w_j$ ，かつ， $r'=z_i$ であるとき，ユーザ i はアイテム j を購入する ($r_{i,j}=1$ となる)．

前述のとおり，無限関係モデルでは，ユーザがアイテムを購入する事象と購入しない事象の両方をモデル化しているのに対して，提案モデルでは，ユーザがアイテムを購入する事象 (ユーザとアイテムが互いに選択しあう事象) のみをモデル化する．これにより，購買履歴データにおける欠損部分，つまり，今は購入していないが将来購入されるアイテムに対処する．

提案法は，多項分布の次元数をユーザクラス，アイテムクラスとしたディリクレ過程混合モデル¹²⁾ であるといえ，無限関係モデルが，ユーザアイテムブロックにおけるベルヌーイ分布に基づきユーザとアイテムを共クラスタリングするのに対して，提案モデルは，ユーザクラス/アイテムクラス上での多項分布に基づきユーザとアイテムを共クラスタリングする．

5.2 提案法における同時分布

購買履歴集合 \mathcal{R} の同時分布は，

$$\begin{aligned} p(\mathcal{R}, Z, W, \Theta, \Phi; \alpha, \beta, \gamma, \eta) \\ = p(\mathcal{R}|Z, W, \Theta, \Phi) p(Z; \alpha) p(W; \beta) p(\Theta; \gamma) p(\Phi; \eta), \end{aligned}$$

となる．ここで， $p(\Theta; \gamma)$ ， $p(\Phi; \eta)$ に対して，多項分布 $p(\mathcal{R}|Z, W; \Theta, \Phi)$ の共役事前分布である（対称）ディリクレ分布を仮定する．つまり，上式右辺の各分布は以下のように表される．

$$\begin{aligned} p(\mathcal{R}|Z, W, \Theta, \Phi) &= \prod_{r_{i,j} \in \mathcal{R}} \theta_{z_i, w_j} \phi_{w_j, z_i}, \\ &= \prod_{s=1}^S \prod_{k=1}^K (\theta_{k,s} \phi_{s,k})^{l(k,s)}, \end{aligned}$$

$$p(\Theta; \gamma) = \prod_{k=1}^K \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \prod_{s'=1}^S (\theta_{k,s'})^{\gamma-1},$$

$$p(\Phi; \eta) = \prod_{s=1}^S \frac{\Gamma(K\eta)}{\Gamma(\eta)^K} \prod_{k'=1}^K (\phi_{s,k'})^{\eta-1},$$

$$p(Z; \alpha) = \frac{\alpha^K \prod_{k=1}^K (n_k - 1)!}{\alpha(\alpha+1) \cdots (\alpha+N-1)},$$

$$p(W; \beta) = \frac{\beta^S \prod_{s=1}^S (m_s - 1)!}{\beta(\beta+1) \cdots (\beta+M-1)}.$$

$l(k, s)$ は，ユーザクラス k に所属するユーザがアイテムクラス s に所属するアイテムを購入した回数，

$$l(k, s) = \sum_{i=1}^N \sum_{j=1}^M \mathbf{I}(z_i = k) \mathbf{I}(w_j = s) r_{i,j}. \quad (3)$$

を表し， $\alpha, \beta, \gamma, \eta$ は既知のハイパーパラメータを表す．

提案法では以下の事後確率 $P(Z, W|\mathcal{R}; \alpha, \beta, \gamma, \eta)$ を最大化する Z, W を求める（導出は付録 A.2 参照）．

$$\begin{aligned} &P(Z, W|\mathcal{R}; \alpha, \beta, \gamma, \eta) \\ &\propto \prod_{k=1}^K \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \frac{\prod_{s'=1}^S \Gamma(\gamma + l(k, s'))}{\Gamma(S\gamma + \sum_{s'=1}^S l(k, s'))} \times \prod_{s=1}^S \frac{\Gamma(K\eta)}{\Gamma(\eta)^K} \frac{\prod_{k'=1}^K \Gamma(\eta + l(k', s))}{\Gamma(K\eta + \sum_{k'=1}^K l(k', s))} \end{aligned}$$

表 1 ギブズサンプリングに基づく学習アルゴリズム

Table 1 The learning algorithm using Gibbs sampler.

入力: $\mathcal{R}, \alpha, \beta, \gamma, \eta$	出力: Z, W
初期化: Z, W を初期化する．	
ユーザクラス更新ステップ: ランダムに選んだユーザ i の帰属クラス z_i の値を，以下の確率に従ってサンプリングした新たなクラス k^* に更新する．	
$P(z_i = k^* \mathcal{R}, Z_{-i}, W; \alpha, \beta, \gamma, \eta).$ (4)	
アイテムクラス更新ステップ: ランダムに選んだアイテム j の帰属クラス w_j の値を，以下の確率に従ってサンプリングした新たなクラス s^* に更新する．	
$P(w_j = s^* \mathcal{R}, Z, W_{-j}; \alpha, \beta, \gamma, \eta).$ (5)	
クラスの分割・併合: 事後確率値に変化がなくなった場合，ユーザクラスもしくはアイテムクラスの 1 つを分割する，または，ユーザクラスもしくはアイテムクラスのうちの 2 つのクラスを併合する．	
事後確率収束判定: Z, W に関する事後確率が収束基準を満たせば終了．満たさない場合，ユーザクラス更新ステップに戻る．	

$$\times \frac{\alpha^K \prod_{k=1}^K (n_k - 1)!}{\alpha(\alpha+1) \cdots (\alpha+N-1)} \times \frac{\beta^S \prod_{s=1}^S (m_s - 1)!}{\beta(\beta+1) \cdots (\beta+M-1)}.$$

5.3 学習アルゴリズム

無限関係モデルと同様，ギブズサンプリングアルゴリズムに基づき，事後確率 $p(Z, W|\mathcal{R}; \alpha, \beta, \gamma, \eta)$ を最大化する Z, W を求める．アルゴリズムを表 1 に示す．ここで， Z_{-i}, W_{-j} は， z_i, w_j をそれぞれ除いた Z, W を表す．このアルゴリズムは，ユーザとアイテムを適当にクラスタリングした初期状態から，事後確率の意味で良い共クラスタリング状態へと Z, W を更新するアルゴリズムであり，最終的には局所解が求まる．なお，表 1 の式 (4) は，

$$P(z_i = k^* | \mathcal{R}, Z_{-i}, W; \alpha, \beta, \gamma, \eta)$$

$$\propto P(\mathcal{R}|z_i = k^*, Z_{-i}, W; \gamma, \eta) P(z_i = k^* | Z_{-i}; \alpha),$$

より，

k^* が既存クラスである場合：

$$P(z_i = k^* | \mathcal{R}, Z_{-i}, W; \alpha, \beta, \gamma, \eta)$$

$$\propto \frac{n_{-i, k^*}}{N-1+\alpha} \frac{\prod_{s=1}^S \Gamma(\gamma + l_{+i}(k^*, s))}{\Gamma(S\gamma + \sum_{s=1}^S l_{+i}(k^*, s))} \times \frac{\Gamma(S\gamma + \sum_{s=1}^S l_{-i}(k^*, s))}{\prod_{s=1}^S \Gamma(\gamma + l_{-i}(k^*, s))},$$

k^* が新規クラスである場合：

$$P(z_i = k^* | \mathcal{R}, Z_{-i}, W; \alpha, \beta, \gamma, \eta) \propto \frac{\alpha}{N-1+\alpha} \frac{\prod_{s=1}^S \Gamma(\gamma + l_{+i}(k^*, s))}{\Gamma(S\gamma + \sum_{s=1}^S l_{+i}(k^*, s))} \times \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S},$$

と表される．ここで， $n_{-i,k}$ はユーザ i を購買履歴データから取り除いたときのユーザクラス k に所属するユーザ数を表す．また， $l_{-i}(k, s)$ はユーザ i を購買履歴データから取り除いたときの $l(k, s)$ を表し， $l_{+i}(k, s)$ はユーザ i をクラス k に割り当てたときの $l(k, s)$ を表す．表 1 の式 (5) も同様の式で表される．

表 1 のクラスの分割・併合操作は，質の悪い局所解が求まるのを防ぐために導入したヒューリスティクスであり，本論文では，文献 9) で提案されている簡便な方法を利用する．つまり，まず，分割・併合する対象（ユーザ/アイテム）をランダムに選択し，続いて，選択した対象から 2 つのサンプルをランダムに選ぶ．ここで，選んだサンプルが同じクラスに属する場合には選択した 2 つのクラスのみを対象としたギブズサンプリングに基づいてクラスの分割を行い，異なるクラスに属する場合にはクラスの併合を行う．分割・併合の決定はメトロポリス—ヘイスティングスアルゴリズムにより行う²¹⁾．詳細は，文献 9) を参照されたい．本論文の実験では，事後確率の値が 1 万回連続で変わらないたびに分割・併合操作を行い，100 万回連続で事後確率の値が変化しなくなったときに学習アルゴリズムを終了させた．

表 1 の学習アルゴリズムでは，ランダムにクラスタリングした状態を初期状態としており，クラス数を 1 とした状態を初期状態とすることも可能である．しかし，クラス数を 1 とした場合，提案するモデルにおいては必ず互いのユーザクラス/アイテムクラスを選択しあうことになり，現実的な解とはいえない．そこで本論文では，クラス数をある程度大きくした下でランダムにクラスタリングした状態を初期状態とする．また， α, β の値を小さくしすぎると，クラス数が 1 のときの事後確率が最大となる場合があるため， $\alpha, \beta \geq 1$ である値を設定する．

6. 実験

提案法の有効性，具体的には，

- (1) クラス数の推定精度，
- (2) 共クラスタリングの精度，
- (3) 欠損値への適応度，

を検証するため，人工データによる実験と，映画の評価データを用いた実データによる実験を行った．比較対象として無限関係モデルを用いた．

表 2 人工データ作成に用いたパラメータ値 (パターン I)

Table 2 The parameter values for generating the synthetic data (pattern I).

(a) ユーザクラス用パラメータ: $P(z_i = k), \theta_{k,s} \in \Theta$

k	$P(z_i = k)$	$\theta_{k,1}$	$\theta_{k,2}$	$\theta_{k,3}$
1	0.4	0.1	0.8	0.1
2	0.3	0.8	0.1	0.1
3	0.3	0.1	0.1	0.8

(b) アイテムクラス用パラメータ: $P(w_j = s), \phi_{s,k} \in \Phi$

s	$P(w_j = s)$	$\phi_{s,1}$	$\phi_{s,2}$	$\phi_{s,3}$
1	0.4	0.8	0.1	0.1
2	0.3	0.1	0.1	0.8
3	0.3	0.1	0.8	0.1

6.1 人工データ実験

6.1.1 実験設定

真のクラス数が既知であるデータを用いた実験により，クラス数の推定精度，および，共クラスタリングの精度を評価する．実験に用いた人工データは，表 2 (パターン I)，表 3 (パターン II) をモデルパラメータ値とする提案モデルに従って作成した．ここで，パターン I の真のユーザクラス数/アイテムクラス数はそれぞれ 3 であり，パターン II の真のユーザクラス数/アイテムクラス数はそれぞれ 4/5 である．提案モデルが仮定する購入行動に対する 3 ステップに基づき，ユーザ数 N とアイテム数 M の異なる，つまり，データの規模が異なる購買履歴データ R を作成した．

ここで，スパース性 (R の全要素のうち，0 の占める割合) がそれぞれ 80%，85%，90%，95%，99% であるデータ R を各 5 組作成した．スパース性が高いほど欠損値が多いデータであり，スパース性の異なるデータを用いることにより，欠損値への適応度を評価する．共クラスタリング結果の評価尺度には，クラス数の推定値と，異なるクラスタリング結果の類似性を測る尺度である Adjusted Rand Index (ARI)⁷⁾ を用いる．ここで， P 個のクラスにクラスタリングされた結果 1 と， Q 個のクラスにクラスタリングされた結果 2 を比較する場合，ARI は以下の式により計算される．

$$\text{ARI} = \frac{S - E[S]}{S_{\max} - E[S]}.$$

ただし，

表 3 人工データ作成に用いたパラメータ値 (パターン II)

Table 3 The parameter values for generating the synthetic data (pattern II).

(a) ユーザクラス用パラメータ: $P(z_i = k), \theta_{k,s} \in \Phi$

k	$P(z_i = k)$	$\theta_{k,1}$	$\theta_{k,2}$	$\theta_{k,3}$	$\theta_{k,4}$	$\theta_{k,5}$
1	0.2	0.1	0.1	0.6	0.1	0.1
2	0.2	0.1	0.1	0.1	0.6	0.1
3	0.3	0.1	0.6	0.1	0.1	0.1
4	0.3	0.3	0.1	0.1	0.1	0.3

(b) アイテムクラス用パラメータ: $P(w_j = s), \phi_{s,k} \in \Phi$

s	$P(w_j = s)$	$\phi_{s,1}$	$\phi_{s,2}$	$\phi_{s,3}$	$\phi_{s,4}$
1	0.2	0.1	0.7	0.1	0.1
2	0.1	0.1	0.1	0.1	0.7
3	0.2	0.1	0.1	0.7	0.1
4	0.2	0.4	0.1	0.4	0.1
5	0.3	0.1	0.4	0.1	0.4

$$S = \sum_{p=1}^P \sum_{q=1}^Q \binom{d_{p,q}}{2},$$

$$E[S] = \frac{\left[\sum_{p=1}^P \binom{d_p}{2} \sum_{q=1}^Q \binom{d_q}{2} \right]}{\binom{D}{2}},$$

$$S_{\max} = \frac{1}{2} \left[\sum_{p=1}^P \binom{d_p}{2} + \sum_{q=1}^Q \binom{d_q}{2} \right],$$

であり, $d_{p,q}$ ($p=1, 2, \dots, P, q=1, 2, \dots, Q$) は, 結果 1 においてクラス p , 結果 2 においてクラス q に割り付けられたデータの総数を表す. また, $d_p = \sum_{q=1}^Q d_{p,q}$, $d_q = \sum_{p=1}^P d_{p,q}$ であり, D はデータの総数を表す. ARI は, 2 つのクラスタリング結果が完全に一致する場合に 1, 2 つのクラスタリングがランダムに行われた場合に期待値 0 をとる.

ここで, 各手法におけるハイパーパラメータ値は, 予備実験から, 提案法: $\alpha = \beta = 1$, $\gamma = \eta = 0.1$, 無限関係モデル: $\alpha = \beta = 1$, $\gamma_0 = \gamma_1 = 0.1$ とした. また, Z, W の初期化の際のユーザクラス数/アイテムクラス数はそれぞれ 20 とした.

6.1.2 実験結果

提案法と無限関係モデルを適用した結果を表 4 (パターン I) と表 5 (パターン II) に示す. 各表のそれぞれの値は, 1 つの購買履歴データ R に対して各手法をそれぞれ 5 回適用

した際の平均値を示す.

まず, クラス数の推定精度について, 表 4, 表 5 から, いずれの手法においても, スパース性が低い場合には真のクラス数, つまり, 表 4 においては真のユーザクラス数 3/アイテムクラス数 3, 表 5 においては真のユーザクラス数 4/アイテムクラス数 5, に近い値が推定できていることが分かる. これより, クラス数の生成に対して仮定した DP が有効に機能していることが分かる. ただし, スパース性が高くなるにつれてクラス数の推定精度が悪化していることも確認できる. これは, $r_{i,j} = 1$ の数が少なくなるため, より細かなユーザ・アイテムブロックに共クラスタリングしてしまうためと考えられる.

次に, 共クラスタリング精度について, 表 4, 表 5 から, いずれの手法においても, スパース性が低い場合には真の共クラスタリング結果に近い結果が得られていることが分かる. また, クラス数の推定精度の結果と同様に, スパース性が高くなるにつれて共クラスタリング精度は悪化している. 特に, スパース性が 99% の場合には, いずれの手法において, 真の共クラスタリング結果とは相関がないような結果が得られている. クラス数の推定精度が良いと ARI の値も良くなる傾向があり, スパース性が高くなるほど精度良く共クラスタリングすることは困難となることが分かる.

最後に, 欠損値への適応度について, 現実の購買履歴データのスパース性が約 95% であることをふまえ, 該当箇所を比較してみると, 全体的な傾向として, 提案法の方がより精度の良いクラス数, および, 共クラスタリング結果が得られていることが確認できる. ここで, データの規模が小さい場合において, 提案法の方が精度が悪い結果が得られている. これは, データの規模が小さい場合には $r_{i,j} = 1$ となるデータの絶対数が少なくなるため, モデルパラメータの学習が精度良く行えないためと考えられる. データの規模が大きいほど $r_{i,j} = 1$ の絶対数が増え, 提案法の精度が良くなる傾向が確認できることから, 提案法が有効に機能するためにはある程度の $r_{i,j} = 1$ であるデータが必要であることが考察される.

以上より, スパース性が低いデータにおいては, 無限関係モデルと提案法の性能に差はあまりないが, 本論文が対象とする購買履歴データのようなスパース性が高いデータにおいては, 無限関係モデルと比べて, 提案法は, クラス数をより精度良く推定でき, かつ, より精度が良い共クラスタリングが行えることが確認された.

表 4 提案法と無限関係モデル (IRM) の共クラスタリング性能比較 (パターン I)

Table 4 The clustering performances of the proposed method and IRM (pattern I).

(a) ユーザクラスに関する ARI (値が大きいほど良い). () 内はユーザクラス数の推定結果 (真の値 = 3).

スパース性	N = M = 200		N = M = 400		N = M = 600		N = M = 800	
	提案法	IRM	提案法	IRM	提案法	IRM	提案法	IRM
80%	1.00 (3.0)	0.99 (3.2)	0.97 (3.4)	0.96 (3.6)	0.90 (4.1)	0.93 (3.8)	0.93 (3.9)	0.93 (3.9)
85%	1.00 (3.0)	0.98 (3.4)	0.98 (3.4)	0.95 (3.7)	0.95 (3.7)	0.91 (4.4)	0.92 (4.2)	0.90 (4.4)
90%	0.98 (3.0)	0.98 (3.3)	0.95 (3.6)	0.91 (4.3)	0.87 (4.6)	0.89 (4.8)	0.84 (5.3)	0.82 (5.4)
95%	0.22 (3.1)	0.29 (3.3)	0.93 (3.8)	0.86 (4.6)	0.90 (4.8)	0.81 (6.2)	0.83 (6.1)	0.79 (6.6)
99%	0.01 (5.8)	0.01 (4.0)	0.00 (6.9)	0.00 (7.4)	0.01 (9.7)	0.01 (9.3)	0.02 (11)	0.02 (12)

(b) アイテムクラスに関する ARI (値が大きいほど良い). () 内はアイテムクラス数の推定結果 (真の値 = 3).

スパース性	N = M = 200		N = M = 400		N = M = 600		N = M = 800	
	提案法	IRM	提案法	IRM	提案法	IRM	提案法	IRM
80%	1.00 (3.1)	0.99 (3.3)	0.97 (3.5)	0.93 (3.9)	0.95 (3.6)	0.89 (4.4)	0.93 (4.0)	0.95 (3.8)
85%	0.99 (3.0)	0.96 (3.5)	0.96 (3.5)	0.94 (3.6)	0.96 (3.8)	0.90 (4.6)	0.90 (4.7)	0.86 (4.9)
90%	0.98 (3.2)	0.97 (3.3)	0.94 (3.8)	0.92 (4.4)	0.89 (4.4)	0.86 (5.0)	0.83 (5.4)	0.82 (5.5)
95%	0.20 (3.3)	0.33 (3.4)	0.91 (4.0)	0.83 (4.9)	0.89 (4.7)	0.82 (6.2)	0.81 (6.0)	0.78 (6.9)
99%	0.01 (6.0)	0.00 (4.1)	0.01 (7.2)	0.00 (6.8)	0.01 (9.2)	0.01 (9.0)	0.01 (11)	0.02 (11)

6.2 実データ実験

6.2.1 実験設定

実データである MovieLens データ^{*1}を用いて, 提案法と無限関係モデルの欠損値への適応度と共クラスタリングの精度を, 定性的/定量的に評価した. MovieLens データは, ユーザが 5 点満点で映画を評価した評価履歴データであり, 本実験では, 評価が与えられていない (与えられていない) 映画を, 既購入 (未購入) アイテムと見なし, 購買履歴データを作成した. 943 ユーザ, 1,682 アイテムからなる購買履歴データ R であり, スパース性は約 96% である.

ここで, 各手法におけるハイパーパラメータ値は, 予備実験から, 提案法: $\alpha = \beta = 10$, $\gamma = \eta = 0.01$, 無限関係モデル: $\alpha = \beta = 10$, $\gamma_0 = \gamma_1 = 0.01$ とした. また, Z, W の初期化の際のユーザクラス数/アイテムクラス数はそれぞれ 100 とした.

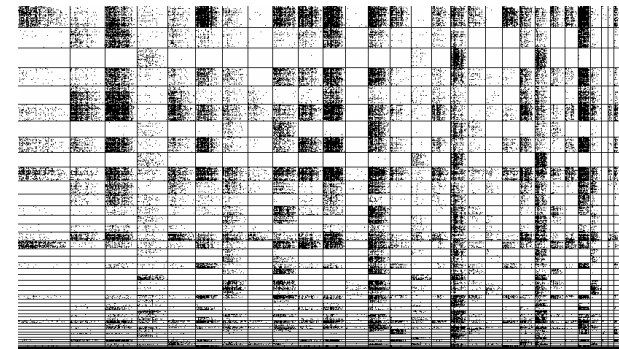


図 1 提案法による共クラスタリング結果

Fig. 1 The co-clustering result by the proposed method.

6.2.2 定性的評価: 欠損値への適応度

購買履歴データ R に対する提案法と無限関係モデルの欠損値への適応度を定性的に比較・評価する. ここで, 図 1, 図 2 は, それぞれ, 提案法と無限関係モデルを適用した結果で

*1 <http://www.grouplens.org/taxonomy/term/14>. ユーザのデモグラフィック情報やアイテムのコンテンツ情報を含む.

表 5 提案法と無限関係モデル (IRM) の共クラスタリング性能比較 (パターン II)

Table 5 The clustering performances of the proposed method and IRM (pattern II).

(a) ユーザクラスに関する ARI (値が大きいほど良い). () 内はユーザクラス数の推定結果 (真の値 = 4).

スパース性	$N = M = 200$		$N = M = 400$		$N = M = 600$		$N = M = 800$	
	提案法	IRM	提案法	IRM	提案法	IRM	提案法	IRM
80%	0.99 (4.0)	1.00 (4.0)	0.99 (4.2)	0.99 (4.2)	0.98 (4.2)	0.97 (4.3)	0.98 (4.5)	0.97 (4.5)
85%	0.99 (4.0)	0.99 (4.1)	0.98 (4.2)	0.98 (4.4)	0.97 (4.5)	0.97 (4.4)	0.95 (5.0)	0.97 (4.5)
90%	0.94 (4.2)	0.98 (4.2)	0.98 (4.5)	0.97 (4.5)	0.97 (4.6)	0.95 (5.0)	0.94 (5.0)	0.96 (4.9)
95%	0.46 (4.8)	0.68 (4.6)	0.92 (4.6)	0.91 (5.2)	0.91 (5.7)	0.92 (5.1)	0.89 (6.1)	0.88 (5.8)
99%	0.02 (5.5)	0.04 (3.8)	0.08 (7.8)	0.16 (7.6)	0.17 (11)	0.22 (11)	0.26 (11)	0.36 (12)

(b) アイテムクラスに関する ARI (値が大きいほど良い). () 内はアイテムクラス数の推定結果 (真の値 = 5).

スパース性	$N = M = 200$		$N = M = 400$		$N = M = 600$		$N = M = 800$	
	提案法	IRM	提案法	IRM	提案法	IRM	提案法	IRM
80%	0.97 (5.0)	0.99 (5.1)	0.99 (5.3)	0.98 (5.3)	0.99 (5.4)	0.99 (5.3)	1.00 (5.5)	0.99 (5.4)
85%	0.94 (5.0)	0.98 (5.1)	0.99 (5.2)	0.99 (5.2)	0.99 (5.4)	0.98 (5.4)	0.95 (6.2)	0.99 (5.3)
90%	0.85 (5.1)	0.94 (5.2)	0.97 (5.4)	0.99 (5.3)	0.96 (5.9)	0.97 (5.8)	0.94 (6.4)	0.97 (5.8)
95%	0.40 (5.1)	0.59 (5.2)	0.86 (5.4)	0.88 (6.1)	0.90 (7.0)	0.93 (6.5)	0.93 (6.8)	0.92 (6.7)
99%	0.02 (5.6)	0.03 (4.2)	0.07 (7.8)	0.12 (7.6)	0.17 (11)	0.19 (11)	0.24 (11)	0.31 (12)

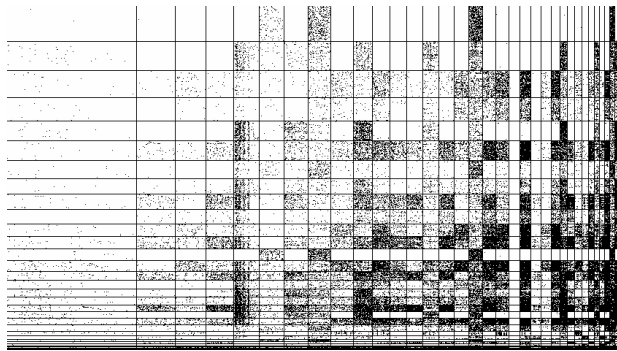


図 2 無限関係モデルによる共クラスタリング結果

Fig. 2 The co-clustering result by the Infinite Relational Model.

ある. 図 1, 図 2 はそれぞれ行列 R を表現しており, 黒い点は $r_{i,j} = 1$ (既購入), 白い点は $r_{i,j} = 0$ (未購入) をそれぞれ表す. ただし, これらの図は, 共クラスタリングによって得られたユーザクラス, アイテムクラスごとにユーザとアイテムをソートしており, クラ

スの区切りを実線で表している. また, 左上から要素数の大きい順に各クラスを並べている. 各手法により得られたユーザクラス数/アイテムクラス数は, 無限関係モデルにおいて 36/36, 提案法において 49/33 であった. 図 1, 図 2 から, どちらの手法によっても, ユーザ・アイテムブロック単位で購買履歴 (黒の部分) が密集している結果が得られることが分かる. ただし, 無限関係モデルでは, 購買履歴の少ないユーザ・アイテム群を一括りにして一番大きなユーザ・アイテムブロック (図 2 の左上部分) を構成してしまっており, 購買履歴データの大部分を占める $r_{i,j} = 0$ の影響を受けていることが分かる. これに対して, 提案法は, $r_{i,j} = 1$ (黒) の部分が全体的に分布しており, 欠損値を含んだデータに適応した結果が得られていることが分かる.

提案法により得られたアイテムクラスごとの映画タイトルの一部を表 6 に示す. クラス番号は, 要素数の大きい順に 1, 2, ... とする. 表 6 より, アイテムクラス 10 は子供・家族向けの映画, アイテムクラス 11 はヒット作, アイテムクラス 25 はホラー映画など特徴のあるクラスが得られていることが分かる. また, ユーザクラス 3 は平均 35 歳で女性の占める割合が高いクラスであり, 履歴が集中しているアイテムクラス 4, 17, 22 などから, ユーザの嗜好の傾向が把握できる.

表 6 提案法により得られたアイテムクラスごとの映画タイトル (一部)

Table 6 The examples of movie titles in the item class obtained by the proposed method.

アイテムクラス 10 (全 69 タイトル)	アイテムクラス 11 (全 62 タイトル)	アイテムクラス 25 (全 33 タイトル)
Pinocchio(1940)	Back to the Future(1985)	Alien(1979)
Pocahontas(1995)	Platoon(1986)	Jaws(1975)
Home Alone(1990)	Stand by Me(1986)	Aliens(1986)
The Jungle Book(1994)	The Lion King(1994)	Psycho(1960)
Swiss Family Robinson(1960)	Beauty and the Beast(1991)	Pulp Fiction(1994)

6.2.3 定量的評価：共クラスタリングの精度

以下の方法により学習用購買履歴データとテスト用購買履歴データ $T = \{T_i\}_{i=1}^N$ を作成し、学習用購買履歴データを用いて共クラスタリングを行った結果に対して、テスト用購買履歴データを用いて精度を定量的に評価する。つまり、購買履歴データ R において、ユーザ i ごとに $r_{i,j}=1$ である履歴のうちからランダムに選んだ 10% を $r_{i,j}=0$ とした学習用購買履歴データを作成し、選ばれた (i,j) 要素をテスト用購買履歴データ T_i とした。テスト用に選択した (i,j) 箇所が欠損箇所にあたり、欠損箇所は未知であるものとする。

ここで、テスト用購買履歴データは購入が確実である履歴であるため、問題設定で述べたとおり、テスト用購買履歴データとして選択された (i,j) 要素は、将来優先して購入される可能性の高いユーザ・アイテムブロック内に割り当てられていることが望ましい。そこで、各ユーザに対して、購入される可能性が高い ($r_{i,j}=1$ の割合が大きい) アイテムクラス順にアイテムを推薦することを想定し、推薦する上位 L 個のアイテムクラス集合 C_L に対するテスト用購買履歴データの再現率を評価尺度とする：

$$\text{Recall}(L) = \frac{1}{N} \sum_{i=1}^N \frac{\sum_{(i,j) \in T_i} I(w_j \in C_L)}{|T_i|}. \quad (6)$$

式 (6) の値が大きいほど、将来優先的に購入される可能性が高いブロックに T_i が割り当てられていることを表し、商品推薦で次の購買を予測する場合に都合の良い共クラスタリング結果であることを意味する。

図 3 に、学習用購買履歴データに対して各手法を 5 回適用したときの式 (6) の平均値を示す。ここで、図 3 の NMF は、非負行列因子分解を適用した際の結果を示し、ユーザクラス/アイテムクラス数は前述の定性的評価での結果から 40 とした。各手法により得られたユーザクラス数/アイテムクラス数の平均値は、無限関係モデルにおいて 30.4/37.2、提案法において 46.4/42.8 であった。図 3 より、提案法が、テスト用データを、購入される可

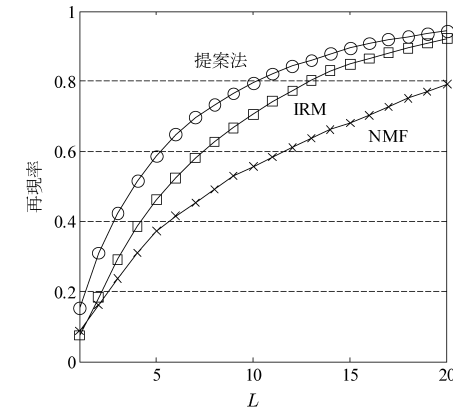


図 3 上位 L アイテムクラスに対する各手法の再現率
Fig. 3 The recall of each method for top- L item classes.

能性のより高いユーザ・アイテムブロックに割り当てていることが分かる。つまり、無限関係モデルや非負行列因子分解に比べて、共クラスタリングの精度が良い。

6.3 実験結果まとめ

人工データを用いた実験、および、実データを用いた実験の結果より、購買履歴データなどの欠損値を多く含んだデータに対し、クラス数の推定精度、共クラスタリングの精度、欠損値への適応度のそれぞれの観点において、無限関係モデルに比べてより良い共クラスタリング結果が得られることが確認できた。

7. ま と め

本論文では、ディリクレ過程混合モデルに基づく離散データに対する共クラスタリング手

法を提案した。提案法は、事前にクラス数を与えることなく、ユーザとアイテムを同時にクラスタリングする。映画の評価データを用いた実験により、無限関係モデルと比べて、より適切な共クラスタリング結果が得られることを確認した。今後は、頻度データに適したモデル化や、大規模データへ適用した際には計算時間が課題となるため、学習の効率化・高速化などについて検討したい。

謝辞 本研究の機会を与えていただいた、株式会社 NTT データ技術開発本部ビジネスインテリジェンス推進センタ関根純センタ長、上島康司部長、松永務課長、ならびに、知的財産室島崎勝美室長に深く感謝いたします。

参 考 文 献

- 1) Blei, D.M. and Jordan, M.I.: Variational Inference for Dirichlet Process Mixtures, *Bayesian Analysis*, Vol.1, No.1, pp.121–144 (2006).
- 2) Chen, G., Wang, F. and Zhang, C.: Collaborative Filtering Using Orthogonal Nonnegative Matrix Tri-factorization, *Proc. 7th IEEE International Conference on Data Mining (ICDM), Workshop on High Performance Data Mining*, Omaha NE, US (2007).
- 3) Deerwester, S., Dumais, S., Furnas, G.W., Landauer, T.K. and Harshman, R.: Indexing by Latent Semantic Analysis, *Journal of the Society for Information Science*, Vol.41, No.6, pp.391–407 (1990).
- 4) Deodhar, M. and Ghosh, J.: A Framework for Simultaneous Co-clustering and Learning from Complex Data, *Proc. 13th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, San Jose, California, US, pp.250–259 (2007).
- 5) Dhillon, I.S.: Co-clustering documents and words using bipartite spectral graph partitioning, *Proc. 7th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, San Francisco CA, US, pp.269–274 (2001).
- 6) Dhillon, I.S., Mallela, S. and Modha, D.S.: Information-Theoretic Co-clustering, *Proc. 9th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, Washington, DC, USA, pp.89–98 (2003).
- 7) Dudoit, S. and Fridlyand, J.: A prediction-based resampling method for estimating the number of clusters in a dataset, *Genome Biology*, Vol.3, No.7, pp.1–21 (2002).
- 8) Hofmann, T.: Latent Semantic Models for Collaborative Filtering, *ACM Trans. Information Systems*, Vol.22, No.1, pp.89–115 (2004).
- 9) Jain, S. and Neal, R.M.: A Split-Merge Markov Chain Monte Carlo Procedure for the Dirichlet Process Mixture Model, *Journal of Computational and Graphical Statistics*, Vol.13, pp.158–182 (2004).
- 10) Kemp, C., Tenenbaum, J., Griffiths, T., Yamada, T. and Ueda, N.: Learning Systems of Concepts with an Infinite Relational Model, *Proc. 21st National Conference on Artificial Intelligence (AAAI)*, Boston, Massachusetts, US, pp.381–388 (2006).
- 11) Kurihara, K., Kameya, Y. and Sato, T.: Discovering Concepts from Word Co-occurrences with a Relational Model, *Trans. Japanese Society for Artificial Intelligence*, Vol.22, No.2, pp.218–226 (2007).
- 12) McAuliffe, J.D., Blei, D.M. and Jordan, M.I.: Nonparametric empirical Bayes for the Dirichlet process mixture model, *Statistics and Computing*, Vol.16, No.1, pp.5–14 (2006).
- 13) Neal, R.M.: Markov chain sampling methods for Dirichlet process mixture models, *Journal of Computational and Graphical Statistics*, Vol.9, pp.249–265 (2000).
- 14) Rashid, A.M., Lam, S.K., Karypis, G. and Riedl, J.: ClustKNN: A Highly Scalable Hybrid Model & Memory-Based CF Algorithm, *Proc. 8th ACM SIGKDD Workshop on Web Mining and Web Usage Analysis (WEBKDD)* (2006).
- 15) Rasmussen, C.E.: The Infinite Gaussian Mixture Model, *Proc. 12th Conference on Neural Information Processing Systems (NIPS)*, Denver, CO, US, pp.554–560 (1999).
- 16) Rege, M., Dong, M. and Fotouhi, F.: Co-clustering Documents and Words Using Bipartite Isoperimetric Graph Partitioning, *Proc. 6th IEEE International Conference on Data Mining (ICDM'06)*, Hong Kong, pp.532–541 (2006).
- 17) Shafiei, M.M. and Milios, E.E.: Latent Dirichlet Co-Clustering, *Proc. 6th IEEE International Conference on Data Mining (ICDM'06)*, Hong Kong, pp.542–551 (2006).
- 18) Symeonidis, P., Nanopoulos, A., Papadopoulos, A. and Manolopoulos, Y.: Nearest-Biclusters Collaborative Filtering, *Proc. 8th ACM SIGKDD Workshop on Web Mining and Web Usage Analysis (WEBKDD)* (2006).
- 19) Teh, Y.W.: A Hierarchical Bayesian Language Model based on Pitman-Yor Processes, *Proc. 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (ACL/COLING)*, Sydney, Australia, pp.985–992 (2006).
- 20) Xu, Z., Tresp, V., Yu, K. and Kriegel, H.-P.: Infinite Hidden Relational Models, *Proc. 22nd International Conference on Uncertainty in Artificial Intelligence (UAI2006)*, Cambridge, MA, US (2006).
- 21) 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).

付 録

A.1 ディリクレ過程

分割 G を,

$$G = \{P(\theta_1), P(\theta_2), \dots, P(\theta_K)\},$$

と表す．ここで, K は分割数を表し, $P(\theta_k)$ ($k = 1, 2, \dots, K$) は各分割の割合を表す．つまり, パラメータ θ_k の出現確率が $P(\theta_k)$ であり, $\sum_{k=1}^K P(\theta_k) = 1$.

ディリクレ過程は, 任意の分割数 K とそれに対応する $\{\theta_k, P(\theta_k)\}_{k=1}^K$, つまり G , を与える分布であり,

$$G \sim \text{DP}(\alpha, G_0),$$

と表現される．ここで, $\alpha (> 0)$ は集中度パラメータ (スカラー値), G_0 は基底分布 (基底測度) をそれぞれ表す．すると, N 個のパラメータの実現値 $\{\theta_1, \theta_2, \dots, \theta_N\}$ が与えられた下での分割 G に対する事後分布の期待値は以下ようになる．

$$\begin{aligned} & E[G|\theta_1, \theta_2, \dots, \theta_N] \\ &= \frac{\alpha}{\alpha + N} G_0 + \frac{1}{\alpha + N} \sum_{i=1}^N \delta(\theta_i), \\ &= \frac{\alpha}{\alpha + N} G_0 + \frac{n_k}{\alpha + N} \sum_{k=1}^K \delta(\theta_k). \end{aligned} \quad (7)$$

ここで, 関数 δ は Dirac のデルタ関数を表し, n_k は $\{\theta_1, \theta_2, \dots, \theta_N\}$ のうちで値が θ_k と一致するパラメータの数を表す．つまり, $n_k = \sum_{i=1}^N \mathbf{1}(\theta_k = \theta_i)$. ここで, $\mathbf{1}$ は引数が真のとき 1, そのほかのとき 0 を返すインジケータ関数を表す．よって, 式 (7) の K は $\{\theta_1, \theta_2, \dots, \theta_N\}$ における θ の異なり数を表す．式 (7) は, $\frac{n_k}{N+\alpha}$ の確率ですでに得られたパラメータ値 θ_k が出現し, $\frac{\alpha}{N+\alpha}$ の確率で基底分布 G_0 から新たにパラメータ値 θ_{K+1} が出現することを意味する．新たにパラメータ値が基底分布 G_0 から出現した場合, 分割数は K から $K+1$ に変わる．

A.2 提案法における Z, W の事後確率の導出

ベイズの定理より,

$$P(Z, W|\mathcal{R}; \alpha, \beta, \gamma, \eta) = \frac{P(\mathcal{R}|Z, W; \gamma, \eta)P(Z; \alpha)P(W; \beta)}{\sum_Z \sum_W P(\mathcal{R}|Z, W; \gamma, \eta)P(Z; \alpha)P(W; \beta)}, \quad (8)$$

となる．ここで, 式 (8) の分母は既知のデータ \mathcal{R} に関する分布であるため固定値となる．そのため, 式 (8) の分子のみについて最大化することを考えればよい．ここで, 式 (8) の分子は,

$$\begin{aligned} & P(\mathcal{R}|Z, W; \gamma, \eta)P(Z; \alpha)P(W; \beta) \\ &= \int_{\Phi} \int_{\Theta} p(\mathcal{R}|Z, W, \Theta, \Phi)p(\Theta; \gamma)p(\Phi; \eta)d\Theta d\Phi \times p(Z; \alpha)p(W; \beta), \end{aligned}$$

と書け, 上式の積分内は,

$$\begin{aligned} & p(\mathcal{R}|Z, W, \Theta, \Phi)p(\Theta; \gamma)p(\Phi; \eta) \\ &= \prod_{r_{i,j} \in \mathcal{R}} p(r_{i,j}|\theta_{z_i}, w_j, \phi_{w_j, z_i}) \times \prod_{k=1}^K \prod_{s'=1}^S p(\theta_{k,s'}; \gamma) \prod_{s=1}^S \prod_{k'=1}^K p(\phi_{s,k'}; \eta), \\ &= \prod_{s=1}^S \prod_{k=1}^K (\theta_{k,s} \phi_{s,k})^{l(k,s)} \times \prod_{k=1}^K \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \prod_{s'=1}^S (\theta_{k,s'})^{\gamma-1} \prod_{s=1}^S \frac{\Gamma(K\eta)}{\Gamma(\eta)^K} \prod_{k'=1}^K (\phi_{s,k'})^{\eta-1}, \\ &= \prod_{k=1}^K \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \prod_{s'=1}^S (\theta_{k,s'})^{l(k,s')+\gamma-1} \times \prod_{s=1}^S \frac{\Gamma(K\eta)}{\Gamma(\eta)^K} \prod_{k'=1}^K (\phi_{s,k'})^{l(k',s)+\eta-1}, \end{aligned}$$

と計算される． $l(k, s)$ は, 式 (3) のとおりである．ここで, ディリクレ分布に関する以下の式,

$$\begin{aligned} & \int_{\theta_k} \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \prod_{s'=1}^S (\theta_{k,s'})^{\gamma-1} d\theta_k = 1, \\ & \Leftrightarrow \int_{\theta_k} \prod_{s'=1}^S (\theta_{k,s'})^{\gamma-1} d\theta_k = \frac{\Gamma(\gamma)^S}{\Gamma(S\gamma)}, \end{aligned}$$

を利用することにより, 以下の積分結果を得る．

$$\begin{aligned} & \int_{\Theta} \prod_{s'=1}^S (\theta_{k,s'})^{l(k,s')+\gamma-1} d\Theta = \frac{\prod_{s'=1}^S \Gamma(\gamma + l(k, s'))}{\Gamma(S\gamma + \sum_{s'=1}^S l(k, s'))}, \\ & \int_{\Phi} \prod_{k'=1}^K (\phi_{s,k'})^{l(k',s)+\eta-1} d\Phi = \frac{\prod_{k'=1}^K \Gamma(\eta + l(k', s))}{\Gamma(K\eta + \sum_{k'=1}^K l(k', s))}, \end{aligned}$$

以上の結果から, 提案法における Z, W の事後確率は,

$$P(Z, W|\mathcal{R}; \alpha, \beta, \gamma, \eta) \propto P(\mathcal{R}|Z, W; \gamma, \eta)P(Z; \alpha)P(W; \beta),$$

$$= \prod_{k=1}^K \frac{\Gamma(S\gamma)}{\Gamma(\gamma)^S} \frac{\prod_{s'=1}^S \Gamma(\gamma + l(k, s'))}{\Gamma(S\gamma + \sum_{s'=1}^S l(k, s'))} \times \prod_{s=1}^S \frac{\Gamma(K\eta)}{\Gamma(\eta)^K} \frac{\prod_{k'=1}^K \Gamma(\eta + l(k', s))}{\Gamma(K\eta + \sum_{k'=1}^K l(k', s))}$$

$$\times \frac{\alpha^K \prod_{k=1}^K (n_k - 1)!}{\alpha(\alpha + 1) \cdots (\alpha + N - 1)} u \times \frac{\beta^S \prod_{s=1}^S (m_s - 1)!}{\beta(\beta + 1) \cdots (\beta + M - 1)}.$$

となる.

(平成 19 年 11 月 22 日受付)

(平成 20 年 1 月 10 日再受付)

(平成 20 年 1 月 30 日採録)



桑田 修平 (正会員)

1977 年生. 2001 年早稲田大学理工学部経営システム工学科卒業. 2003 年同大学院修士課程修了. 同年株式会社 NTT データ入社. 2005 年より 2 年半日本電信電話株式会社 NTT コミュニケーション科学基礎研究所. 現在, 株式会社 NTT データ技術開発本部に所属. 統計的学習, データマイニング等の研究に従事. 電子情報通信学会, 人工知能学会各会員.



山田 武士 (正会員)

1964 年生. 1988 年 3 月東京大学理学部数学科卒業. 同年 NTT 入社. 1996 年より 1 年間英国コベントリー大学客員研究員. 現在, NTT コミュニケーション科学基礎研究所創発環境研究グループリーダ. 主として機械学習, データマイニング, メタヒューリスティクスによる組合せ最適化等の研究に従事. 博士 (情報学). 電子情報通信学会, ACM, IEEE 各会員.



上田 修功 (正会員)

1958 年生. 1982 年大阪大学工学部通信工学科卒業. 1984 年同大学院修士課程修了. 同年 NTT 入社. 1993 年より 1 年間 Purdue 大学客員研究員. 画像処理, パターン認識・学習, ニューラルネットワーク, 統計的学習, Web データマイニング等の研究に従事. 現在, NTT コミュニケーション科学基礎研究所副所長 (企画担当主席研究員). 工学博士. 奈良先端科学技術大学院大学客員教授. 電気通信普及財団賞 (テレコムシステム技術賞) (1997 年, 2006 年), 電子情報通信学会論文賞 (2000 年, 2004 年), 情報処理学会山下記念研究賞 (2006 年) 等受賞. 電子情報通信学会, IEEE 各会員.