

## 混合メンバーシップ・ブロックモデルを用いた 協調フィルタリング

横 峯 樹<sup>†1</sup> 江 口 浩 二<sup>†2</sup>

近年、増え続ける膨大な情報からユーザの嗜好に合った情報を自動的に推薦する仕組みとして協調フィルタリングが注目されている。また、様々なネットワークに対するベイズのアプローチによるモデリングが盛んに研究されている。本稿では、ネットワークに対するノードクラスタリングの手法である混合メンバーシップ・ブロックモデルを用いて、ベイズのアプローチにより協調フィルタリングを行い、有効性を示す。また、アイテム間にも明示的にリンクが存在するデータを対象として MMSB の予備検討を行う。

### Collaborative filtering using mixed membership stochastic block models

TATSUKI YOKOMINE<sup>†1</sup> and KOJI EGUCHI<sup>†2</sup>

Recently, considerable attentions have been paid to collaborative filtering techniques, which can automatically recommend useful information according to a user's preference from vast amounts of information. More recently, various studies have been conducted on Bayesian modeling approach for network analysis. In this paper, we apply the state-of-the-art Bayesian node clustering approach, mixed membership stochastic block models (MMSB), to the collaborative filtering task and demonstrate its effectiveness. Moreover, we explicitly add inter-item links to the network consisting of user-item links and investigate the performance of MMSB for such a situation.

<sup>†1</sup> 神戸大学工学部情報知能工学科

Faculty of Engineering, Kobe University

<sup>†2</sup> 神戸大学大学院工学研究科 情報知能学専攻

Graduate School of Engineering, Kobe University

### 1. はじめに

近年、インターネットなどの情報やそれを扱うユーザがますます増加しており、情報が急激に増加したため代表的な検索技術であるキーワード検索のみでは、条件の絞り込みが甘かったり、適切なキーワードが思いつかないユーザに対して有効な手段を与えていないことなどが課題となっている。また同一のクエリを与えた場合に同一の検索結果を返すのが一般的であり、ユーザの嗜好にあった情報を得ることが容易でないことも課題と言える。そこでユーザごとの履歴情報などから嗜好を推定し、そのユーザの嗜好にあった情報を自動的に推薦する仕組みとして協調フィルタリングが注目されている<sup>1)</sup>。

また、情報の多くはネットワーク構造として表すことができるため、ネットワーク分析の需要がますます高くなってきている。中でも、本稿ではベイズのアプローチによるネットワークのモデリングに着目する。ベイズのアプローチによるネットワークのモデリングには大きく二つの流れがあると考えられる。一つ目は Newman ら<sup>2)</sup> の混合分布を用いたモデルである。これはそれぞれのノードが潜在的なグループの分布を持っているという仮定からネットワークのモデル化を行ったものである。さらに Zhang ら<sup>3)</sup> は Newman ら<sup>2)</sup> の混合分布モデルに事前分布を取り入れた。これは文書解析として有名な Latent Dirichlet Allocation(以下:LDA)<sup>6)</sup> をネットワークに適応したとも言える。二つ目は、ブロックモデルと呼ばれるモデルをベイズ統計学的に拡張した Stochastic Block Models(以下:SB)<sup>4)</sup> である。これはノードをクラスタリングし、同じクラスのノードはすべて相手クラスにリンクを張る確率が同じであるという仮定のもとにネットワークのモデル化を行ったものである。つまり、あるノード  $n_1$  がある別のノード  $n_2$  にリンクを張るとすると、 $n_1$  と同じクラスのノードはすべて  $n_2$  と同じクラスのノードにリンクを張ることとなる。

これら二つの流れをふまえて Airoldi らは Mixed Membership Stochastic Blockmodels<sup>5)</sup>(以下:MMSB) を提案した。これは LDA のようにノードは潜在的なグループの分布を持っており、さらに SB のように、グループ間のリンクの生成確率も考慮したモデルである。MMSB では、各ノードが属する潜在グループを多項分布で表現し、さらにディリクレ事前分布を仮定する。これによって各ノードが複数のグループに属することを許容する柔軟なモデリングが可能となる。また、MMSB においてはグループ対に対するリンクの起こりやすさをベルヌーイ分布で表現するが、これにベータ事前分布を導入し、そのハイパーパラメータを調整することで対象に対する知識を反映させることが可能となっている。つまり、事前知識がある場合でもない場合でも適切にモデルを推定することが可能となる。これによ

り”assortative”, “disassortative”, どちらの特徴を持ったネットワークに対しても適切にモデリングできる.”assortative”とはコミュニティ内のリンクが密であり, コミュニティ間のリンクが疎であるネットワーク構造であり, “disassortative”とはコミュニティ内のリンクが疎であり, コミュニティ間のリンクが密であるネットワーク構造である. これら二つは対極の特徴であると言え, 二つ特徴をうまくモデリングできることは MMSB の表現能力の高さを示している.

本稿では, ネットワークに対するノードクラスタリングの手法として知られる MMSB を協調フィルタリングに適用し従来手法に対する優位性を示す. また, MMSB モデルが二部グラフに限定にしたモデルでないことから, 従来手法では考慮できなかったアイテム間にも明示的にリンクが存在するデータを対象とし, その予備検討を示す.

MMSB モデルを協調フィルタリングに適用した例は, 著者らの調査では未だない.

## 2. 関連研究

### 2.1 協調フィルタリング

本節では協調フィルタリングについて述べる. 協調フィルタリングとは, あるユーザの嗜好に合った情報を推薦するシステムの一種である. 推薦システムのアルゴリズムにおいて, 嗜好の予測の実現方法は大きく二つに分類される. 一つは内容に基づくフィルタリング (content-based filtering) と呼ばれ, もう一つは協調フィルタリング (collaborative filtering) もしくは社会的フィルタリング (social filtering) と呼ばれる. 例として, 映画推薦のシステムを考える. 前者は, あるユーザの好きな監督, 好みのジャンルなどから条件に一致する映画を推薦する方法で, 検索対象の内容を考慮して推薦するアイテムを決定する. 後者は, 映画の趣味が似ている他のユーザの履歴情報を利用して, 映画を推薦する方法であり, 履歴情報を介して他のユーザと「協調」して推薦を実現するものである.

どちらの手法にも長所, 短所が存在する. 例えば, 内容に基づくフィルタリングでは推薦対象が利用者自身が知っている特徴に限られるのでセレンディピティ, つまり目新しさが少ない. 映画の例でみると, 過去に見たジャンルや監督の作品が推薦されることになる. また, それを受け入れることでより一層偏りが生じることが考えられる. それに対して, 協調フィルタリングでは, 自信が知らないジャンルや監督の映画でも, 他の利用者の知識を通じて知ることができる. そうした意外性, セレンディピティのある推薦ができるとされている. 新規のアイテムの場合, 内容に基づくフィルタリングでは利用者のプロフィールがあれば, その特徴ベクトルをもとに適切な推薦をすることが可能であるが, 協調フィルタリング

の場合, そのアイテムを評価した利用者がいなければ適切な推薦をすることは難しい.

現在では, 二つの推薦手法を組み合わせたハイブリッド法が多く提案されている. 他の利用者の情報とアイテムの特徴というのは, 相反する情報ではなく同時に獲得が可能である. つまり, 内容に基づくフィルタリングと協調フィルタリング両方の長所を生かした推薦を行うことが可能であると言える.

### 2.2 Latent Dirichlet Allocation

本節では, LDA に基づくネットワークモデリングについて概要を示す. LDA が最初に紹介されたとき, それは文書解析に用いられた<sup>6)</sup>. LDA は, 文書はある特徴を持った単語の分布 (トピック) の混合分布から生成されるという仮定を前提にし, ある文書, ある単語はトピックから生成されるとしているのでトピックモデルと呼ばれる. Zhang らはこれをネットワークに置き換えた<sup>3)</sup>. つまり, ノードはある特徴を持った他のノードの分布 (グループ, クラスタ) の混合分布から生成されるという仮定をおいたのである. Zhang らはこれを SSN-LDA と呼び, ネットワークのクラスタリングを行った. ネットワークに対する SSN-LDA でのリンクの生成過程を以下に示す. またグラフィカルモデルを図 1 に示す.

- (1) すべてのノード  $v_i$  に対してハイパーパラメータ  $\alpha$  で特定されたディリクレ分布から多項分布  $\theta_{v_i}$  をサンプリング
- (2) すべてのグループ  $g_k$  に対してハイパーパラメータ  $\beta$  で特定されたディリクレ分布から多項分布  $\phi_{g_k}$  をサンプリング
- (3) あるノード  $v_i$  のすべての隣接ノードに対して
  - 多項分布  $\theta_{v_i}$  からグループ割り当て  $z_{ij}$  をサンプリング
  - 多項分布  $\phi_{z_{ij}}$  からノード  $v_j$  をサンプリング

$E = \{e_i\} (i = 1, \dots, N)$  を  $v_i$  に接続した辺の集合とすると, すべての変数とパラメータに関する同時確率 (完全同時確率) は以下の通りである.

$$P(E, Z, \vec{\theta}_{1:N}, \vec{\phi}_{1:K} | \alpha, \beta) = P(\phi | \beta) \sum_{i=1}^N P(\theta_i | \alpha) P(z_i | \theta_i) P(e_i | z_i, \phi) \quad (1)$$

### 2.3 Stochastic Block Models

ブロックモデルとは, 社会科学などの分野で用いられてきた考え方で, 観測された隣接行列を並べ替えることで, ある特徴のあるブロックを見つけ出しクラスタリングを行うモデルである.

Nowicki らは, これにベイズ的アプローチを導入した SB を提案した<sup>4)</sup>. ノードを潜在的

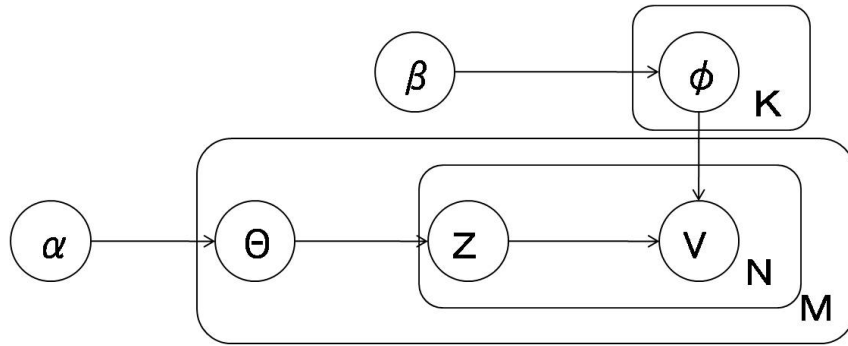


図1 LDAのグラフィカルモデル  
Fig.1 A graphical model of LDA

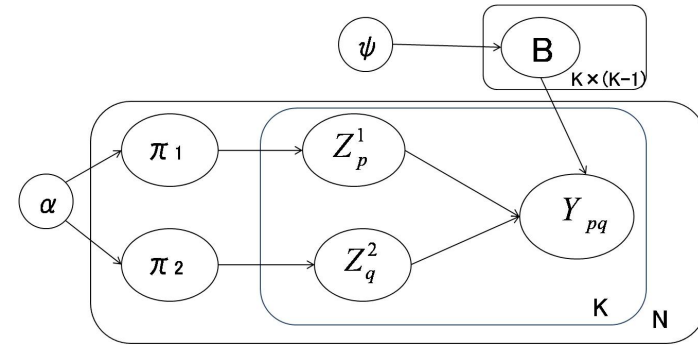


図2 SBのグラフィカルモデル  
Fig.2 A graphical model of SB

なグループによりクラスタリングし、リンクはあるグループとあるグループの間で起こるという仮定を前提にし、ネットワークのモデリングを行った。ネットワークに対するSBでのリンクの生成過程は以下。またグラフィカルモデルを<sup>2)</sup>に示す。

- (1) すべてのノード  $p$  に対して
  - ハイパーパラメータ  $\alpha_1$  で特定されたディリクレ分布から多項分布  $\pi^1$  をサンプリング
  - ハイパーパラメータ  $\alpha_2$  で特定されたディリクレ分布から多項分布  $\pi^2$  をサンプリング
- (2) すべてのグループの対  $(g, h)$  に対して
  - ハイパーパラメータ  $\vec{\psi} = (\psi_1, \psi_2) \in \Psi$  で特定されたベータ分布からベルヌーイ分布  $B(g, h)$  をサンプリング
- (3) すべてのノード対  $(p, q)$  に対して
  - 多項分布  $\pi^1$  から  $z_p^1$  をサンプリング
  - 多項分布  $\pi^2$  から  $z_q^2$  をサンプリング
  - $z_p^1, B, z_q^2$  から  $Y(p, q)$  を生成

### 3. MMSB モデル

本節ではMMSBモデルについて述べる。このモデルはノードに潜在的なグループを割り当て、あるノード対に対してリンクが生成される尤度を推定するモデルである。このモデル

は従来の協調フィルタリングのように二部グラフに限定したのではなく、アイテム間のリンクを考慮することも容易である。

#### 3.1 定義

まずは本稿の以下で用いる定義についてまとめる。グラフを  $G = (N, Y)$  と表し、観測されるデータは隣接行列  $Y(p, q) \in \{0, 1\}$  とする。つまり、モデルは有向グラフで構成されるが、協調フィルタリングにおいてリンクの向きは大きな意味を持たないと考えられるため、本稿では隣接行列が対称な無向グラフでモデルを構築している。あるノード間のリンクはそれぞれのノードの潜在的なグループの多項分布と、グループ対に関するベルヌーイ分布から生成される。それぞれのノードはグループ上の多項分布  $\pi_p$  で特徴づけられるとし、グループ  $g$  に関する多項分布パラメータを  $\pi_{p,g}$  とすると、 $\pi_{p,g}$  はノード  $p$  がグループ  $g$  に属する確率である。つまり、それぞれのノードは複数のグループに属することができる。グループ間の関係はベルヌーイ分布  $B_{K \times K}$  の行列によって定義される。ここで  $B(g, h)$  はグループ  $g$  のノードから、グループ  $h$  のノードへの辺が存在する確率であり、 $K$  はグループ数を示す。指示ベクトル  $\vec{z}_{p \rightarrow q}$  はノード  $p$  からノード  $q$  へリンクが存在するときノード  $p$  に割り当てられる潜在グループを表し（該当するグループの成分が1であり、他が0であるベクトル）、 $\vec{z}_{p \leftarrow q}$  はノード  $q$  に割り当てられる潜在グループで、これら二つのベクトルの集合はそれぞれ  $\{\vec{z}_{p \rightarrow q} : p, q \in N\} = Z_{\rightarrow}$  と  $\{\vec{z}_{p \leftarrow q} : p, q \in N\} = Z_{\leftarrow}$  である。さらに、グループの対、つまり  $Y(p, q)$  に対して  $(\vec{z}_{p \rightarrow q}, \vec{z}_{p \leftarrow q})$  は同じである必要はない。これは、非対称なネットワークにも適応可能であることを示している。

### 3.2 MMSB モデル

本稿では各ノードの分布  $\pi_p$  をディリクレ事前分布による多項分布で表し、グループ間の分布  $B(g, h)$  をベータ分布を事前分布にとるベルヌーイ分布で表す。

以上の定義より、MMSB モデルによってノードは以下の手順にしたがって生成されると仮定する。

- (1) すべてのノード  $p$  に対して
  - ハイパーパラメータ  $\alpha$  で特定されたディリクレ分布から多項分布  $\pi_p$  をサンプリング
- (2) すべてのグループの対  $(g, h)$  に対して
  - ハイパーパラメータ  $\vec{\psi} = (\psi_1, \psi_2) \in \Psi$  で特定されたベータ分布からベルヌーイ分布  $B(g, h)$  をサンプリング
- (3) すべてのノード対  $(p, q)$  に対して
  - 多項分布  $\pi_p$  から指示ベクトル  $\vec{z}_{p \rightarrow q}$  をサンプリング
  - 多項分布  $\pi_q$  から指示ベクトル  $\vec{z}_{p \leftarrow q}$  をサンプリング
  - $\vec{z}_{p \rightarrow q}^T B \vec{z}_{p \leftarrow q}$  から  $Y(p, q)$  を生成

またグラフィカルモデルを図 3 に示す。

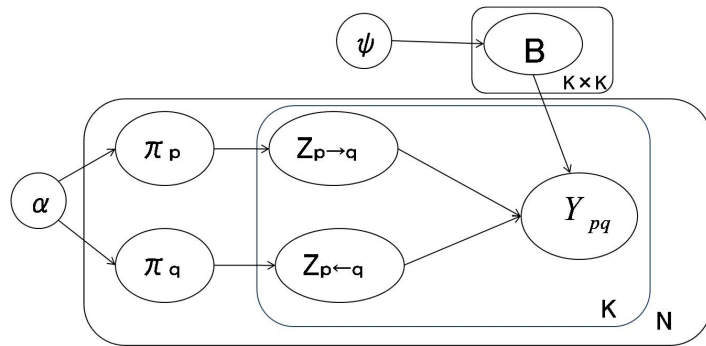


図 3 MMSB のグラフィカルモデル  
Fig. 3 A graphical model of MMSB

このとき、データ  $Y$  と潜在変数  $\pi_{1:N}, Z_{\rightarrow}, Z_{\leftarrow}$  および  $B$  の同時分布（完全同時分布）は

以下のようなになる\*1。

$$P(Y, \pi_{1:N}, Z_{\rightarrow}, Z_{\leftarrow}, B | \alpha, \Psi) \\ = P(B | \Psi) \prod_{p, q, p \neq q} P(Y(p, q) | \vec{z}_{p \rightarrow q}, \vec{z}_{p \leftarrow q}, B) P(\vec{z}_{p \rightarrow q} | \pi_p) P(\vec{z}_{p \leftarrow q} | \pi_q) \prod_p P(\pi_p | \alpha)$$

本稿では MMSB モデルの推定方法として、マルコフ連鎖モンテカルロ法の一種であるギブスサンプリングを用いた。

## 4. MMSB モデルを用いた協調フィルタリング

### 4.1 従来の MMSB モデル

MMSB が最初に提案されたとき、その目的はネットワーク解析、主にノードのクラスタリングであった<sup>5)</sup>。モデルの推定は変分 EM 法で行われており、実験は米国のニューイングランド地方の修道士、ある学校の交友関係ネットワーク、タンパク質の相互作用を対象にクラスタリングを行ったものであった。

### 4.2 MMSB モデルと協調フィルタリング

MMSB は、ベイズ的ネットワーク解析の主流な二つのモデルである LDA と SB、二つの概念を組み合わせたより表現能力の高いモデルである。MMSB は、ネットワークの特徴に依存することなく、ノードクラスタリングタスクにおいて高い性能を発揮することが示されている<sup>5)</sup>。従来は主にノードクラスタリングに用いられてきた MMSB モデルであるが、本稿では協調フィルタリングへの適用を試みる。

従来の協調フィルタリングでは、アイテムとユーザからなる二部グラフを対象に解析を行うのに対して、ノードのクラスタリングに使われてきた MMSB モデルを用いることで二部グラフ以外のデータに対しての協調フィルタリングを試みる。つまり、ユーザはアイテムだけでなく、他のユーザにもリンクを持つことが可能であり、アイテムはユーザだけでなくアイテム同士でリンクを形成することが可能である。ユーザ同士のリンクは友人や知人と捉えると、社会ネットワークであるとみなせる。アイテム間どうしのリンクは、映画を例にすると、同じ俳優あるいは、同じ監督の映画かもしれないし、あるいは同じジャンル、または類似した内容の映画かもしれない。このようなネットワーク表現に基づく情報フィルタリングは、ユーザ・アイテム間リンクのみを用いる一般的な協調フィルタリングの枠組みを超える

\*1 文献<sup>5)</sup> では、ベルヌーイ分布のパラメータを直接推定しているが、本稿では推定にギブスサンプリングを適用するため  $B$  の事前分布にベータ分布を仮定した

ため、ハイブリッド法と捉えることもできる。そして、MMSB はこのような複雑なネットワークを柔軟に扱うことができる。

次節では、従来手法に対する MMSB モデルの優位性を示す。また、アイテム間のリンクをユーザ・アイテム間のリンクに追加した状況で MMSB モデルを適用し、その予備検討を示す。

## 5. 実験

### 5.1 データセット

データセットとして、映画のレビューサイトである MovieLens<sup>\*1</sup>のデータを利用した。データはユーザ数 943、アイテム数 1682 の、10 万件の格付けデータであり、各ユーザは少なくとも 20 のアイテムを格付けしている。本稿ではユーザの格付けの値は考慮せず、格付けしたか、していないかの二値でモデルの推定を行った。本稿ではこのデータから二つのデータを作成し、それぞれで MMSB モデルを推定し実験を行った。一つはユーザ、アイテム間にしかリンクが存在しないデータで MovieLens のユーザとアイテムのデータを適用した。二つ目はアイテム間にもリンクが存在するデータで、アイテム間のリンクを考慮する際には MovieLens の分野情報を採用した。具体的に分野は unknown, Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western という 19 のフィールドからなっており、それぞれ二値で複数回答を許して評価されている。今回はこの分野が完全に一致し、なおかつ二つ以上の分野に属しているものにリンクを仮定した。この場合、アイテム間のリンク数は 9379 である。また、元データの 80 % をトレーニングデータとし、残りの 20 % をテストデータとし、トレーニングデータでモデルを推定し、テストデータで評価を行った。なお、アイテム間のリンクはトレーニングデータのみを追加した。

### 5.2 パラメータの設定

本稿では、モデル推定においてディリクレ分布のハイパーパラメータ  $\alpha = 1.0$  とし、ベータ分布のハイパーパラメータを  $(\psi_1, \psi_2) = (1.0, 1.0)$  と、 $(\psi_1, \psi_2) = (1.0, 6.0)$  と、 $(\psi_1, \psi_2) = (6.0, 1.0)$  の三通りで実験を行った。ベータ分布のハイパーパラメータはどのグループ対にリンクが生成されやすいかの重み付けの役割を持っている。つまり、 $(\psi_1, \psi_2) = (1.0, 1.0)$  のとき、あるグループ間のリンクの生成されやすさは均一であり、 $(\psi_1, \psi_2) = (1.0, 6.0)$  の

きは同じグループ間にリンクが生成されやすく、 $(\psi_1, \psi_2) = (6.0, 1.0)$  のときは違うグループ間にリンクが生成されやすくなる。またグループ数  $K$  は 5, 10, 15, 20, 30, 40, 50 とし、ギブスサンプリングの繰り返し回数はテストセット対数尤度の改善率が確実に 0.1 % 以下になる 500 回とした。

### 5.3 既存手法との比較実験

本節では、アイテム予測における提案手法と既存手法の比較実験の結果を示す。

本実験では、予測精度の比較対象として協調フィルタリングの代表的な手法である Pearson 相関係数に基づく手法（以下、「Pearson」）を用いた。

ここで比較実験の概要を示す。アイテム推薦のシミュレーションを行うため、トレーニングデータによって推定したモデルによって、ユーザごとに（トレーニングデータに含まれたユーザ評価アイテムを除いて）全アイテムをランキングし、テストデータを正解データと見なして、Top-N precision を用いて予測精度を測定した。これらはランキングの精度を測る手法としてよく知られた評価尺度である。Top-N Precision では  $N$  を 1, 5, 10 と変化させた。結果を表 1 に示す。さらに  $(\psi_1, \psi_2) = (1.0, 1.0)$  のときを図 4、 $(\psi_1, \psi_2) = (1.0, 6.0)$  のときを図 5、 $(\psi_1, \psi_2) = (6.0, 1.0)$  のときを図 6 にそれぞれ示す。また、それぞれの  $\psi$  における top-10 precision を図 7 に示す。

表 1 から、 $(\psi_1, \psi_2) = (6.0, 1.0)$  のときを除きすべてのグループ数で MMSB が Pearson 相関を上回った。また、アイテム間にリンクを入れても十分な推薦ができることがわかる。Pearson 相関と比較して、アイテム間リンクを与えない“disassortative”なネットワークに対して、top-1 precision において最大で 152 % 改善し、top-5 precision において最大で 301 %、top-10 precision において最大で 251 % 改善した。また、アイテム間にリンクを与えた“disassortative”でないデータに対しても top-1 precision において最大で 160 % 改善し、top-5 precision において最大で 284 %、top-10 precision において最大で 221 % 改善した。

この結果が示すように MMSB は様々な特徴を持つネットワークのモデリングにおいて、常に高いパフォーマンスを示すと言える。

## 6. 考察

表 1 から MMSB がどんなネットワークに対しても十分なモデリングができ、適切なアイテムの推薦ができるとわかった。また、図 7 からパラメータ  $\vec{\psi}$  によって性能が大きく変わることがわかる。具体的に  $(\psi_1, \psi_2) = (6.0, 1.0)$  のとき、つまり同じグループによりリンク

\*1 <http://grouplens.org/>

表 1 Pearson と MMSB における各グループ数,  $\psi$  の値を変化させたときの Top-N Precision の値 (%)

Table 1 Top-N precision results when the number of groups and  $\psi$  in MMSB are varied.

MMSB グループ数, $\psi_1, \psi_2$	Top-1		Top-5		Top-10	
	nolink	withlink	nolink	withlink	nolink	withlink
5, 1, 1	<b>28.41</b>	<b>29.26</b>	21.20	20.19	19.13	18.14
10, 1, 1	20.46	16.43	19.49	14.50	18.27	14.89
15, 1, 1	23.96	19.61	19.74	21.50	17.96	19.27
20, 1, 1	23.32	17.70	24.15	19.08	<b>21.74</b>	18.19
30, 1, 1	20.78	21.84	21.61	19.93	19.90	18.26
40, 1, 1	26.93	20.78	<b>23.11</b>	20.04	20.26	18.21
50, 1, 1	27.35	18.13	21.95	19.51	20.00	17.93
5, 1, 6	13.36	10.18	9.05	6.76	7.93	6.07
10, 1, 6	4.66	0.10	5.34	0.23	5.05	0.43
15, 1, 6	3.39	0.21	3.60	0.27	3.52	0.41
20, 1, 6	4.24	0.00	3.18	0.27	2.79	0.34
30, 1, 6	3.18	1.16	3.45	1.14	3.16	0.85
40, 1, 6	2.54	0.53	2.37	0.57	2.32	0.69
50, 1, 6	2.43	0.42	2.16	0.72	2.14	0.81
5, 6, 1	0.95	4.24	0.84	3.24	1.00	2.67
10, 6, 1	19.83	23.32	18.40	20.36	17.21	18.04
15, 6, 1	21.20	20.78	18.81	19.72	18.49	17.78
20, 6, 1	18.98	25.34	19.19	<b>22.16</b>	18.34	<b>19.88</b>
30, 6, 1	20.36	19.19	21.08	21.06	18.54	17.97
40, 6, 1	16.75	18.23	19.57	18.81	18.47	16.26
50, 6, 1	13.25	18.23	15.99	18.09	15.58	16.86
Pearson	11.24		5.76		6.18	

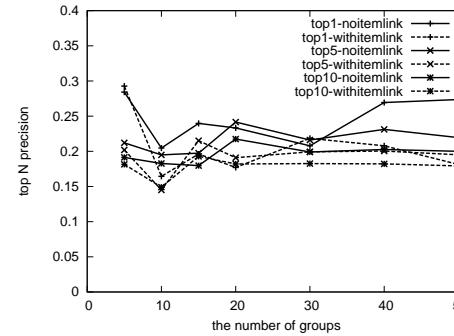


図 4  $(\psi_1, \psi_2) = (1.0, 1.0)$  における各グループ数での Top-N precision  
Fig. 4 Top-N precision and the number of groups when  $(\psi_1, \psi_2) = (1.0, 1.0)$

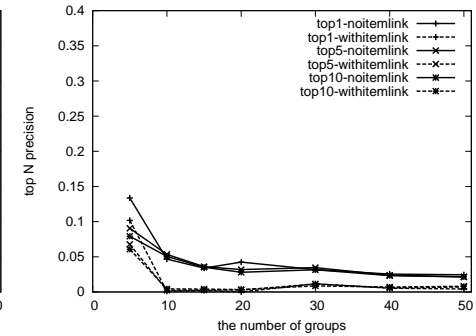


図 5  $(\psi_1, \psi_2) = (1.0, 6.0)$  における各グループ数での Top-N precision  
Fig. 5 Top-N precision and the number of groups when  $(\psi_1, \psi_2) = (1.0, 1.0)$

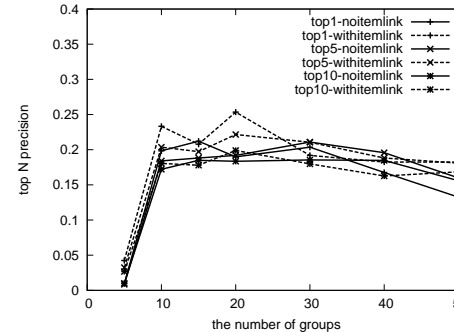


図 6  $(\psi_1, \psi_2) = (6.0, 1.0)$  における各グループ数での Top-N precision  
Fig. 6 Top-N precision and the number of groups when  $(\psi_1, \psi_2) = (6.0, 1.0)$

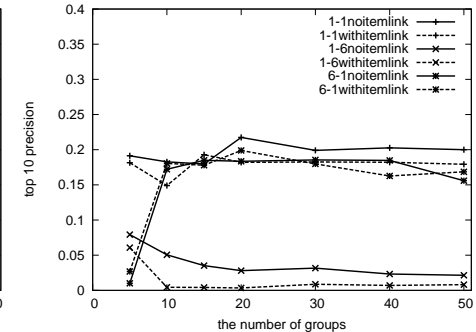


図 7 それぞれの  $\psi$  における各グループ数での Top-10 precision  
Fig. 7 Top-10 precision and the number of groups when each  $\psi$

が張られやすいように事前分布を与えると、グループ数がかなり少ないときに限り 10 % 程度の予測精度を示しているが、グループ数が増えていくとかなり予測精度が小さくなっていく。これは図 5 から、すべての top-N precision で同じ傾向であるとわかる。これはユーザとアイテムが違うグループに割り当てられる確率が高いからであると仮定することができる。つまり、グループ数が少ない内はユーザとアイテムが同じグループに割り当てられることが多いので、同じグループへのリンクを起こりやすくしても予測精度はあまり落ちないが、グループ数が多くなってくるとユーザとアイテムは違うグループとなり、同じグループにリンクが起こりやすいモデルでは予測精度が大きく落ちてしまう。

次に、 $(\psi_1, \psi_2) = (1.0, 6.0)$  のときを見てみると、図 6 から図 5 と逆の振る舞いを見せているように見える。つまり、少ないグループ数のときは 1 % 程度のかかなり小さな値であったがグループ数を増やすと 20 % 程度まで top-N precision の値が上がっている。このことから、先ほどのユーザとアイテムが違うグループに割り当てられやすいという仮定は正しいと言える。つまり、グループ数が少ないと、ユーザとアイテムは同じグループに属していることが多いので違うグループにリンクが起こりやすいモデルでは予測精度がかかなり低く、グループ数が増えてくるとユーザとアイテムは違うグループに割り当てられることが多くなるので、高い予測精度を実現できる。

最後に図 4 を見てみると、グループ数を変化させてもあまり大きく値が変化していない。 $(\psi_1, \psi_2) = (1.0, 1.0)$  のときグループ間、グループ内のリンクの起こりやすさは均一である。つまり一様な事前分布を仮定した場合である。この値から MMSB では知識を導入しない場合でも十分な予測精度を発揮することができると言える。

このように MMSB が様々なネットワークに対して事前知識の有無に関わらず協調フィルタリングタスクにおいて有効であることが観察される。また、その性能がハイパーパラメータの設定によって大きく変わることも今回の実験で確認できた。今後、取り組むべき課題としては、より詳細なハイパーパラメータの設定や、より協調フィルタリングに適したモデルへの改良、具体的にはリンクに属性を持たせることなどが挙げられる。

謝辞 本研究の一部は、科学研究費補助金基盤研究 (B) (20300038) の援助による。

#### 参 考 文 献

- 1) 神嵐 敏弘: 推薦システムのアルゴリズム, 人工知能学会誌 22 巻 6 号, Vol.12, pp. 826-837 (2007).
- 2) M. E. J. Newman: Mixture models and exploratory analysis in networks, proceedings

of the National Academy of Science of the United States of America, Vol.104, No. 23, pp.9564-9569 (2007)

- 3) H. Zhang, B. Qiu, C. L. Giles, H. C. Foley and J. Yen: An LDA-based community structure discovery approach for large-scale social networks, ISI, pp. 200-207 (2007)
- 4) K. Nowicki, T. A. B. Snijders: Estimation and prediction for stochastic block structures, Journal of the American Statistical Association, Vol.96, No.455 (2001)
- 5) Edoardo M. Airoldi, David M. Blei, Stephen E. Fienberg and Eric P. Xing: Mixed membership stochastic block models, The Journal of Machine Learning Research Vol.9, pp.1981-2014 (2008)
- 6) David M. Blei and Andrew Y. Ng and Michael I. Jordan: Latent Dirichlet allocation, Journal of Machine Learning Research, Vol.3, pp.993-1022 (2003)