DBSJ Journal, Vol.12, No.1
June 2013

異種混合メンバーシップ・ブロックモデルと情報推薦への応用

Heterogeneous Mixed Membership Blockmodels and their Application to Item Recommendation

石黒 七海 ♥　　江口 浩二 ♦
横峯 樹 ♠

Nanami ISHIGURO　　Koji EGUCHI
Tatsuki YOKOMINE

ネットワークで表現されたデータの潜在構造をモデル化する有効なアプローチとして，混合メンバーシップ・ブロックモデルが知られている．本研究ではそれを拡張し，複数種類のノードまたはリンクで構成されたネットワークの潜在構造をモデル化する．また，それを用いて，ユーザのアイテム利用履歴に基づいて構成された二部グラフと社会ネットワークなどを合成することによって得られた異種ネットワークに対して適用し，上位 N 推薦問題に応用し，比較評価を行う．

Mixed membership stochastic blockmodels (MMSB) are known as an effective approach to model latent structure underlying data represented as a network. As an extension of MMSB, we attempt to model latent structure in a network consisting of multiple nodes or edges. We apply the proposed models to the heterogeneous networks such as by combining a user-item bipartite graph based on user behavior log and a social network. We demonstrate the effectiveness of the proposed models through experiments on top-N recommendations.

1. はじめに

近年，インターネットの普及に伴い，情報の量とそれを扱うユーザの数が爆発的に増加している．これらの多大な量の情報の中から，自分の必要としている情報を選択することは容易ではない．それらにはネットワーク構造で表される情報も少なくない．その代表的な例が，Facebook や Twitter などのソーシャルメディアである．

♥ 非会員　神戸大学工学部情報知能工学科
ishiguro@cs25.scitec.kobe-u.ac.jp
♦ 正会員　神戸大学大学院システム情報学研究科
eguchi@port.cs.kobe-u.ac.jp
♠ 非会員　神戸大学大学院システム情報学研究科

本論文では，ベイズ的アプローチによるネットワークモデリングに着目する．ベイズ的アプローチによるモデリングには大きく二つの流れがあると考えられる．一つ目は確率論的ブロックモデル（Stochastic Block Models: SBM）[5] とその変形である．これはブロックモデル [2] と同様の，ネットワークデータにおける接続パターンに基づいてノードのクラスタリングを，確率的生成モデルによって実現したものである．ここでは，各ノードを単一のクラスタに割り当て，同じクラスタに属するノードはすべて他のクラスタに属するノードに対してリンクを張る確率が同じであるという仮定が置かれる．また，SBM ではクラスタ数が所与であることを仮定しているが，この仮定を置かないモデルとして無限関係モデル（Infinite Relational Model: IRM）[3] がある．二つ目に，Airoldi らが提案した混合メンバシップ・ブロックモデル（Mixed Membership Blockmodels: MMSB）[1] がある．このモデルでは，ノードに対する潜在的なグループ（またはクラスタ）を多項分布として表現し，グループ対ごとにベルヌーイ分布に従ってリンクが生成されると仮定する．これにより，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になる．社会的ネットワークを例として考えると，勤務先での人間関係と地域コミュニティでの人間関係は一般に異なることが多く，個々人が複数のグループに帰属し，その帰属度合いを表現するモデルは，我々の直感とも合致する．ユーザ利用履歴などに基づいたユーザノードとアイテムノードからなるネットワークを例にとっても，各ユーザが複数の領域にまたがる興味を持つことは現実に即した仮定であると言える．

本論文では，複数種類のノードまたは複数種類のリンクが混在する異種ネットワークに着目する．そのようなネットワークは枚挙に暇がないが，本論文では一例として情報推薦の問題を考える．情報推薦，特に協調フィルタリングでは，ユーザ利用履歴に基づいたユーザ・アイテム間の関係を 2 部グラフとして表現することが多いが，社会的ネットワークなどのようにユーザ間の関係も同時に考慮するには，2 部グラフとして扱うことができない．このとき，ユーザノードとアイテムノードを区別し，これら 2 つの種類のノードからなるネットワークとして表現できる．あるいは，ユーザ・アイテム間の関係とユーザ・ユーザ間の関係などを区別し，複数種類のリンクからなるネットワークとして表現することもできる．ところが，先に述べた従来のネットワークモデリングでは，ネットワーク上におけるノードの種類や，リンクに付与されたラベルを直接的に考慮することができない．そこで本論文では，複数種類のノードまたは複数種類のリンクが混在する異種ネットワークを扱えるよう，MMSB を拡張した異種混合メンバシップ・ブロックモデル（Heterogeneous Mixed Membership Stochastic BlockModel: HMMB）を提案する．リンクの種類に着目するモデルと，ノードの種類に着目するモデルを検討し，補助情報を用いた協調フィルタリングによる上位 N 推薦による実験に基づいて，有効性を評価する．

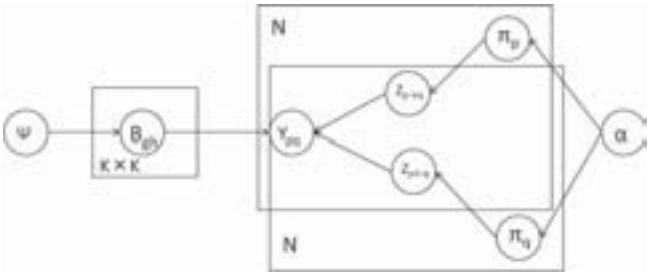

図 1　MMSB のグラフィカルモデル
Fig. 1　Graphical model representation for MMSB.

2.　混合メンバシップ・ブロックモデル

まず，いくつかの定義について説明する．グラフを $G = (N, Y)$ で表し，N をノード集合，Y を隣接行列とする．観測されるデータについて，ノード p，q に対応する隣接行列の要素を $Y(p,q) \in \{0,1\}$ と表す．MMSB において，ある 2 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，それぞれのグループ対における辺の起こりやすさを示すベルヌーイ分布から生成される．すなわち，各ノードはグループ上の多項分布 $Mult(\boldsymbol{\pi}_p)$ 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 に属する確率を示す．このように，それぞれのノードはリンクごとに異なるグループに対応づけることができる．また，グループ間の関係は $K \times K$ 行列で表されたパラメータで決まるベルヌーイ分布 $Bern(B)$ によって定義される．ここで $B(g,h)$ はグループ g に属するノードから，グループ h に属するノードへの辺が存在する確率を示し，K はグループ数を示す．指示ベクトル $\boldsymbol{z}_{p\to q}$ 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 に割り当てられるグループに関する潜在変数を表す（該当するグループの成分が 1 であり，他が 0 である），それに対し $\boldsymbol{z}_{p\leftarrow q}$ はノード q に割り当てられるグループに関する潜在変数を表す．これら二つのベクトルの集合は，それぞれ $Z_{\to} = \{\boldsymbol{z}_{p\to q} \mid p,q \in N\}$，$Z_{\leftarrow} = \{\boldsymbol{z}_{p\leftarrow q} \mid p,q \in N\}$ とする．

このとき，MMSB モデルによるリンクの生成過程は以下のように表すことができる．

1. すべてのノード p に対して
   - ハイパーパラメータ $\boldsymbol{\alpha}$ で特定されたディリクレ分布から多項分布パラメータ $\boldsymbol{\pi}_p$ を選択
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 $\boldsymbol{\psi} = (\psi_0, \psi_1)$ で特定されたベータ分布からベルヌーイ分布パラメータ $B(g,h)$ を選択
3. すべてのノード対 (p,q) に対して
   - 多項分布 $Mult(\boldsymbol{\pi}_p)$ から指示ベクトル $\mathbf{z}_{p\to q}$ を選択
   - 多項分布 $Mult(\boldsymbol{\pi}_q)$ から指示ベクトル $\mathbf{z}_{p\leftarrow q}$ を選択
   - ベルヌーイ分布 $Bern(\mathbf{z}_{p\to q}^{\top} B \mathbf{z}_{p\leftarrow q})$ から $Y(p,q)$ を生成

MMSB のグラフィカルモデルを図 1 で示す．

3.　異種混合メンバーシップ・ブロックモデル

本節では，提案手法である異種混合メンバーシップ・ブロックモデル HMMB-L および HMMB-N について述べる．HMMB-L ではリンクの種類に着目し，HMMB-N はノードの種類に着目したモデルである．

3.1　HMMB-L

MMSB を拡張したモデルである HMMB-L について述べる．グラフを $G = (N, Y, Q)$ と表し，隣接行列 Y におけるノード p，q に関する要素を $Y(p,q) = \delta$，そのリンクの種類を $Q(p,q) = t$ と表す．ただし，$\delta \in \{0,1\}$ であり，リンクの有無を表す．また，$t \in \{1,2,\ldots,T\}$ であり，リンクの種類を表す．各ノード対の間のリンクは，MMSB と同様にして二種類の分布から生成される．すなわち，ノード p がグループ g に属する確率 $\pi_{p,g}$ と，グループ g に属するノードからグループ h に属するノードへの辺が存在する確率 $B(g,h)$ である．さらに，ある種類のリンクのグループ間の関係を，$K \times K \times T$ の 3 次のテンソル M で定義されたグループ対ごとの多項分布によって表す．ここで K はグループ数であり，T はネットワークにおけるリンクの種類の数を表す．よって，$M(g,h,t)$ はリンクの種類 t に関して，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を表す．また，指示ベクトル $\boldsymbol{z}_{p\to q}$，$\boldsymbol{z}_{p\leftarrow q}$ は，それぞれノード p，q に割り当てられるグループに関する潜在変数を示す．これら二つのベクトルの集合は，それぞれ $Z_{\to} = \{\boldsymbol{z}_{p\to q} \mid p,q \in N\}$ と $Z_{\leftarrow} = \{\boldsymbol{z}_{p\leftarrow q} \mid p,q \in N\}$ とする．

これらの定義から，HMMB-L モデルによるリンクの生成過程を以下に示す．

1. すべてのノード p に対して
   - ハイパーパラメータ $\boldsymbol{\alpha}$ で特定されたディリクレ分布から多項分布パラメータ $\boldsymbol{\pi}_p$ を選択
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 $\boldsymbol{\psi} = (\psi_0, \psi_1)$ で特定されたベータ分布からベルヌーイ分布パラメータ $B(g,h)$ を選択
   - リンクの種類 $t \in \{1,\ldots,T\}$ に対して
     - ハイパーパラメータ $\boldsymbol{\phi}$ で特定されたディリクレ分布から多項分布パラメータ $M(g,h,t)$ を選択
3. すべてのノード対 (p,q) に対して
   - 多項分布 $Mult(\boldsymbol{\pi}_p)$ から指示ベクトル $\mathbf{z}_{p\to q}$ を選択
   - 多項分布 $Mult(\boldsymbol{\pi}_q)$ から指示ベクトル $\mathbf{z}_{p\leftarrow q}$ を選択
   - ベルヌーイ分布 $Bern(\mathbf{z}_{p\to q}^{\top} B \mathbf{z}_{p\leftarrow q})$ から $Y(p,q)$ を生成
     - $Y(p,q) = 1$ のとき，多項分布 $Mult(\mathbf{z}_{p\to q}^{\top} M \mathbf{z}_{p\leftarrow q})$ から $Q(p,q)$ を生成

このとき，データ Y, Q と潜在変数 $\boldsymbol{\pi}_{1:N}, Z_{\to}, Z_{\leftarrow}$ および M, B の完全同時分布は次式で表すことができる．

$$
\begin{aligned}
&P\left(Y, Q, Z_{\to}, Z_{\leftarrow}, \boldsymbol{\pi}_p, \boldsymbol{\pi}_q, B, M | \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\phi}\right) \\
&= \prod_p P(\mathbf{z}_{p\to q}|\boldsymbol{\pi}_p) P(\boldsymbol{\pi}_p|\boldsymbol{\alpha}) \prod_q P(\mathbf{z}_{p\leftarrow q}|\boldsymbol{\pi}_q) P(\boldsymbol{\pi}_q|\boldsymbol{\alpha}) \\
&\quad \prod_g \prod_h P(Y(p,q)|\mathbf{z}_{p\to q}, \mathbf{z}_{p\leftarrow q}, B) P(B|\boldsymbol{\psi})
\end{aligned}
$$

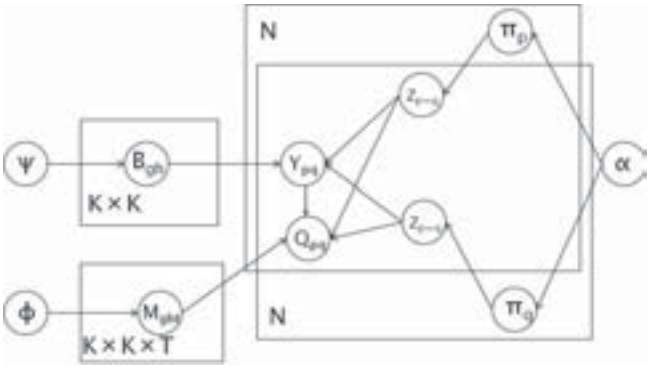

図 2　HMMB-L のグラフィカルモデル
Fig. 2　Graphical model representation for HMMB-L.

$$P(Q(p,q)|\mathbf{z}_{p\to q},\mathbf{z}_{p\leftarrow q},M)P(M|\boldsymbol{\phi}) \tag{1}$$

また，完全条件付確率分布は以下のようになる．

$$\begin{aligned} &P\left(z_{p\to q}=g,z_{p\leftarrow q}=h \mid Y,Q,Z_{\to},Z_{\leftarrow};\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\phi}\right)\\ &\propto \frac{C(p,g)+\alpha_g}{\Sigma_g C(p,g)+\Sigma_g\alpha_g}\frac{C(q,h)+\alpha_h}{\Sigma_h C(q,h)+\Sigma_h\alpha_h}\\ &\quad\frac{C(g,h,\delta)+\psi_\delta}{C(g,h,0)+C(g,h,1)+\Sigma_\delta\psi_\delta}\frac{C(g,h,t)+\phi_t}{\Sigma_t C(g,h,t)+\Sigma_t\phi_t} \end{aligned} \tag{2}$$

ここで，$C(p,g)$ はノード p がグループ g に割り当てられた頻度である．$C(g,h,\delta)$ は，観測されたネットワーク全体における任意のノード対に関して，一方のグループが g であり，他方のグループが h であり，かつ，リンクの有無それぞれに関する頻度を表し，$\delta\in\{0,1\}$ はリンクの有無を示す．また，$C(g,h,t)$ は，任意のノードがグループ g に割り当たり，その隣接ノードがグループ h に割り当たり，かつ，リンクの種類が t である頻度を表す．HMMB-L のグラフィカルモデルを図 2 に示す．

3.2　HMMB-N

次に，ノードの種類を考慮した HMMB-N について述べる．グラフを $G=(N,Y,L)$ と表し，隣接行列 Y におけるノード p，q に関する要素を $Y(p,q)=\delta$，そのノードの種類を $L(p)=x$ と表す．ただし，$\delta\in\{0,1\}$ であり，リンクの有無を示す．また，簡単のため，$x\in\{U,I\}$ であるとし，それぞれユーザ・ノード，アイテム・ノードを示すとする．各ノード対の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，それぞれのグループ対における辺の起こりやすさを表すベルヌーイ分布から生成される．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，これは任意の種類のノード p がグループ g に属する確率を表す．また，グループ間の関係は $K\times K$ 行列で表されたパラメータで決まるベルヌーイ分布 $Bern(B^{x_1,x_2})$ に従う．ここで，$B^{x_1,x_2}(g,h)$ はグループ g に属する種類 x_1 のノードから，グループ h に属する種類 x_2 のノードへの辺が存在する確率を表す．また，指示ベクトル $\boldsymbol{z}_{p\to q},\boldsymbol{z}_{p\leftarrow q}$ は，ノード p，q に割り当てられるグループに関する潜在変数を表す．これら二つのベクトルの集合は，それぞれ $Z_{\to}=\{\boldsymbol{z}_{p\to q}\mid p,q\in N\}$ と $Z_{\leftarrow}=\{\boldsymbol{z}_{p\leftarrow q}\mid p,q\in N\}$ とする．

これらの定義から，HMMB-N モデルによるリンクの生成過程は以下のようになる．

1. 隣接するノードの種類が U であるすべてのノード p に対して
   - ハイパーパラメータ $\boldsymbol{\alpha}$ で特定されたディリクレ分布から多項分布パラメータ $\boldsymbol{\pi}_p$ を選択
2. 隣接するノードの種類が I であるすべてのノード q に対して
   - ハイパーパラメータ $\boldsymbol{\alpha}$ で特定されたディリクレ分布から多項分布パラメータ $\boldsymbol{\pi}_q$ を選択
3. ノードの種類 (U,I) の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 $\boldsymbol{\psi}^{UI}=(\psi_0^{UI},\psi_1^{UI})$ で特定されたベータ分布からベルヌーイ分布パラメータ $B^{UI}(g,h)$ を選択
4. ノードの種類 (I,U) の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 $\boldsymbol{\psi}^{IU}=(\psi_0^{IU},\psi_1^{IU})$ で特定されたベータ分布からベルヌーイ分布パラメータ $B^{IU}(g,h)$ を選択
5. ノードの種類 (U,U) の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 $\boldsymbol{\psi}^{UU}=(\psi_0^{UU},\psi_1^{UU})$ で特定されたベータ分布からベルヌーイ分布パラメータ $B^{UU}(g,h)$ を選択
6. ノードの種類 (I,I) の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 $\boldsymbol{\psi}^{II}=(\psi_0^{II},\psi_1^{II})$ で特定されたベータ分布からベルヌーイ分布パラメータ $B^{II}(g,h)$ を選択
7. ノードの種類 (U,I) のすべてのノード対 (u,i) に対して
   - 多項分布 $Mult(\boldsymbol{\pi}_u)$ から指示ベクトル $\mathbf{z}_{u\to i}$ を選択
   - 多項分布 $Mult(\boldsymbol{\pi}_i)$ から指示ベクトル $\mathbf{z}_{u\leftarrow i}$ を選択
   - ベルヌーイ分布 $Bern(\mathbf{z}_{u\to i}^{\top}B_{gh}^{UI}\mathbf{z}_{u\leftarrow i})$ から $Y(u,i)\in\{0,1\}$ を生成
8. ノードの種類 (I,U) のすべてのノード対 (i,u) に対して
   - 多項分布 $Mult(\boldsymbol{\pi}_i)$ から指示ベクトル $\mathbf{z}_{i\to u}$ を選択
   - 多項分布 $Mult(\boldsymbol{\pi}_u)$ から指示ベクトル $\mathbf{z}_{i\leftarrow u}$ を選択
   - ベルヌーイ分布 $Bern(\mathbf{z}_{i\to u}^{\top}B_{gh}^{IU}\mathbf{z}_{i\leftarrow u})$ から $Y(i,u)\in\{0,1\}$ を生成
9. ノードの種類 (U,U) のすべてのノード対 (u,u') に対して
   - 多項分布 $Mult(\boldsymbol{\pi}_u)$ から指示ベクトル $\mathbf{z}_{u\to u'}$ を選択
   - 多項分布 $Mult(\boldsymbol{\pi}_{u'})$ から指示ベクトル $\mathbf{z}_{u\leftarrow u'}$ を選択
   - ベルヌーイ分布 $Bern(\mathbf{z}_{u\to u'}^{\top}B_{gh}^{UU}\mathbf{z}_{u\leftarrow u'})$ から $Y(u,u')\in\{0,1\}$ を生成
10. ノードの種類 (I,I) のすべてのノード対 (i,i') に対して
    - 多項分布 $Mult(\boldsymbol{\pi}_i)$ から指示ベクトル $\mathbf{z}_{i\to i'}$ を選択
    - 多項分布 $Mult(\boldsymbol{\pi}_{i'})$ から指示ベクトル $\mathbf{z}_{i\leftarrow i'}$ を選択
    - ベルヌーイ分布 $Bern(\mathbf{z}_{i\to i'}^{\top}B_{gh}^{II}\mathbf{z}_{i\leftarrow i'})$ から $Y(i,i')\in\{0,1\}$ を生成

このとき，データ Y と潜在変数 $\boldsymbol{\pi}_{1:N},Z_{\to},Z_{\leftarrow}$ および B の完全同時分布は次式のようになる．

$$\begin{aligned} &P\left(Y,L,Z_{\to},Z_{\leftarrow},\boldsymbol{\pi}_p,\boldsymbol{\pi}_q,B|\boldsymbol{\alpha},\boldsymbol{\psi}\right)\\ &=\prod_{p,x_1}P(\mathbf{z}_{p\to q}|\boldsymbol{\pi}_p)P(\boldsymbol{\pi}_p|\boldsymbol{\alpha})\prod_{q,x_2}P(\mathbf{z}_{p\leftarrow q}|\boldsymbol{\pi}_q)P(\boldsymbol{\pi}_q|\boldsymbol{\alpha})\\ &\quad\prod_{g,h,x_1,x_2}P(Y(p,q)|\mathbf{z}_{p\to q},\mathbf{z}_{p\leftarrow q},B^{x_1,x_2})P(B^{x_1,x_2}|\boldsymbol{\psi}^{x_1,x_2}) \end{aligned}$$

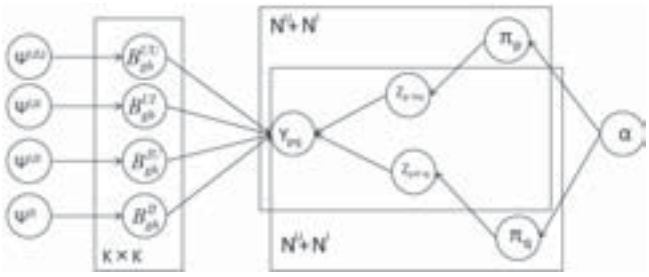

図 3　HMMB-N のグラフィカルモデル
Fig. 3　Graphical model representation for HMMB-N.

また，種類 U のノードと種類 I のノード間のリンクに関する完全条件付確率分布は以下のようになる（他のリンク種類の対に関しても同様である）．

$$P\left(z_{p\to q}=g, z_{p\leftarrow q}=h \mid Y, L, Z_{\rightarrow}, Z_{\leftarrow}; \boldsymbol{\alpha}, \boldsymbol{\psi}\right) \propto \frac{C(p,g)+\alpha_g}{\Sigma_g C(p,g)+\Sigma_g \alpha_g} \frac{C(q,h)+\alpha_h}{\Sigma_h C(q,h)+\Sigma_h \alpha_h} \frac{C^{UI}(g,h,\delta)+\psi_\delta^{UI}}{C^{UI}(g,h,0)+C^{UI}(g,h,1)+\Sigma_\delta \psi_\delta^{UI}} \tag{3}$$

ここで，$C(p,g)$ はノード p がグループ g に割り当てられた頻度である．$C^{UI}(g,h,\delta)$ は，観測されたネットワークにおいて，一方の種類が U のノードのグループが g，他方の種類が I のノードのグループが h であり，かつ，リンクの有無それぞれに関する頻度を表す．HMMB-N のグラフィカルモデルを図 3 に示す．なお，図中の N^U，N^I はそれぞれ種類 U と I のノード数を示す．

4.　実験

4.1　推定

本論では，HMMB-L，HMMB-N，および比較対象の MMSB における未知パラメータと潜在変数を，周辺化ギブスサンプリングで推定する．また，ハイパーパラメータ $\boldsymbol{\alpha}$ は対称ディリクレを仮定し，$\alpha=0.1$ とした．$\boldsymbol{\psi}$ については非対称ベータを仮定し，不動点反復法により更新する [4]．HMMB-L においては次式による．

$$\psi_1^{new}=\psi_1 \frac{\sum_{g,h}\Psi(C(g,h,1)+\psi_1)-\Psi(\psi_1)}{\sum_{g,h}\Psi(C(g,h,1)+(1-\rho)C(g,h,0)+\sum_\delta \psi_\delta)-\Psi(\sum_\delta \psi_\delta)}$$

$$\psi_0^{new}=\psi_0 \frac{\sum_{g,h}\Psi((1-\rho)C(g,h,0)+\psi_0)-\Psi(\psi_0)}{\sum_{g,h}\Psi(C(g,h,1)+(1-\rho)C(g,h,0)+\sum_\delta \psi_\delta)-\Psi(\sum_\delta \psi_\delta)}$$

また，HMMB-N では次式を用いる．

$$\psi_1^{UI,new}=\psi_1^{UI} \frac{\sum_{g,h}\Psi(C^{UI}(g,h,1)+\psi_1^{UI})-\Psi(\psi_1^{UI})}{\sum_{g,h}\Psi(C^{UI}(g,h,1)+(1-\rho^{UI})C^{UI}(g,h,0)+\sum_\delta \psi_\delta^{UI})-\Psi(\sum_\delta \psi_\delta^{UI})}$$

$$\psi_0^{UI,new}=\psi_0^{UI} \frac{\sum_{g,h}\Psi((1-\rho^{UI})C^{UI}(g,h,0)+\psi_0^{UI})-\Psi(\psi_0^{UI})}{\sum_{g,h}\Psi(C^{UI}(g,h,1)+(1-\rho^{UI})C^{UI}(g,h,0)+\sum_\delta \psi_\delta^{UI})-\Psi(\sum_\delta \psi_\delta^{UI})}$$

他の種類の ψ についても同様である．また，すべての実験においてグループ数は $K=10$ とする．

なお，HMMB-N において，各ノードが多項分布 $Mult(\boldsymbol{\pi}_p)$ で特徴づけられると仮定した場合と，各ノードが隣接ノードの種類 $x\in\{U,I\}$ によって区別された多項分布 $Mult(\boldsymbol{\pi}_p^x)$ で特徴づけられると仮定した場合を予備実験によって比較し，より安定的な結果をもたらした後者を用いた．

4.2　データセット

本論文では二つのデータセットを用いて実験を行う．

一つ目は，映画のレビューサイトである MovieLenz のデータ[1]である．ここでは，ユーザと映画タイトル（アイテム）をノードの種類 $\{U,I\}$ とする．ユーザ数は 943, アイテム数は 1682 で，リンク数は 121694 である．ユーザ・アイテム間には，各ユーザとそのユーザが与えた評価値が平均以上であるアイテムの間にリンクが存在すると仮定する．また，アイテム間にも，同じ監督であるか同じ俳優（作品ごとの重要度に関して上位 10 名の俳優のいずれか）が出演している場合にのみリンクが存在すると仮定し，ユーザ間にはリンクが存在しないと仮定する．このデータではユーザ間にはリンクが存在しないと仮定する．

二つ目は，音楽に関するソーシャルメディアである Last.fm のデータ[2]である．ここでは，ユーザと音楽アーティスト（アイテム）をノードの種類 $\{U,I\}$ とする．ユーザ数は 1101，アイテム数は 524，リンク数は 76088 である．ユーザ・アイテム間にはユーザとそのユーザが視聴したアイテムにリンクが存在し，ユーザ間には Last.fm 上で繋がりがあるものにユーザ間リンク存在すると仮定する．ユーザ間には音楽的嗜好が似た者同士にリンクが生まれやすいと考えられる．このデータではアイテム間にはリンクが存在しないと仮定する．

4.3　評価方法

ユーザごとにユーザ・アイテム間のテスト用リンクを尤度の高いものから順にランキングするような上位 N 推薦問題を想定する．4.2 節のデータを 5 分割し，5-fold cross validation によって評価を行う．

HMMB-L の場合，テスト用リンクとその種類に関する尤度は，以下の式で求められる．

$$P(Y \mid Z_{\rightarrow}, Z_{\leftarrow}, \boldsymbol{\pi}_p, \boldsymbol{\pi}_q, B, M, Q; \boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\phi}) = \prod_{p,q}\sum_{g,h}(1-\rho)\Big(\frac{C(p,g)+\alpha_g}{\Sigma_g C(p,g)+\Sigma_g\alpha_g}\frac{C(q,h)+\alpha_h}{\Sigma_h C(q,h)+\Sigma_h\alpha_h} \frac{C(g,h,\delta)+\psi_\delta}{C(g,h,0)+C(g,h,1)+\Sigma_\delta\psi_\delta}\frac{C(g,h,t)+\phi_t}{\Sigma_t C(g,h,t)+\Sigma_t\phi_t}\Big) \tag{4}$$

ここで，ρ はスパース性のモデリングのためのパラメータ [1] であり，ここでは以下のように定義する．

$$\rho = 1-\sum_{p,q} Y(p,q)/(N\times N) \tag{5}$$

HMMB-N の場合，種類 U のノード p と種類 I のノード q 間のテスト用リンクに関する尤度は，以下の式で求められる（他の

[1] http://www.grouplens.org/node/73
[2] http://www.grouplens.org/node/462

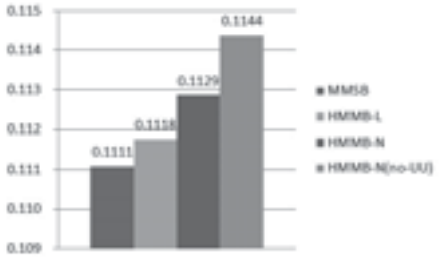

図 4　平均 10 位精度（MovieLenz データセット）
Fig. 4　Mean top-10 precision with MovieLenz dataset.

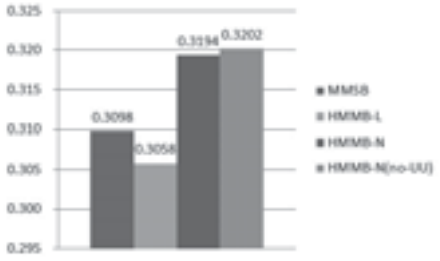

図 5　平均逆順位（MovieLenz データセット）
Fig. 5　Mean reciprocal rank with MovieLenz dataset.

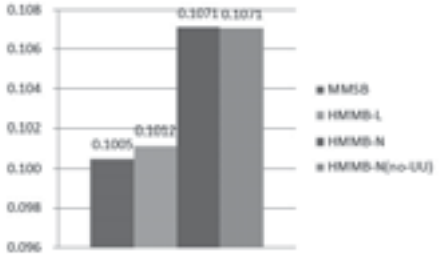

図 6　平均精度（MovieLenz データセット）
Fig. 6　Mean average precision with MovieLenz dataset.

種類のリンクに関しても同様である）．

$$P(Y \mid Z_{\rightarrow}, Z_{\leftarrow}, \boldsymbol{\pi}_p, \boldsymbol{\pi}_q, B, L; \boldsymbol{\alpha}, \boldsymbol{\psi}) = \prod_{p,q} \sum_{g,h} (1-\rho^{UI}) \Big(\frac{C(p,g)+\alpha_g}{\Sigma_g C(p,g)+\Sigma_g \alpha_g} \frac{C(q,h)+\alpha_h}{\Sigma_h C(q,h)+\Sigma_h \alpha_h} \frac{C^{UI}(g,h,\delta)+\psi_\delta^{UI}}{C^{UI}(g,h,0)+C^{UI}(g,h,1)+\Sigma_\delta \psi_\delta^{UI}} \Big) \quad (6)$$

ここで，ρ^{UI} は以下のように定義する．

$$\rho^{UI} = 1 - \sum_{p \in U, q \in I} Y(p,q)/(N^U \times N^I) \quad (7)$$

N^U，N^I はそれぞれ種類 U と I のノード数を示す．

4.4　評価結果

本実験では，上位 N 推薦を想定し，訓練データによって推定したモデルによって，テストデータに含まれる全アイテムに対してユーザごとにランキングを行い，以下の 4 つの方法を用いて予測性能を測定した．

- 上位 10 件に含まれる正解アイテムの割合の平均をとる平均 10 位精度（mean top-10 precision）
- 最上位に出現する正解アイテムの順位の逆数の平均をとる平均逆順位（mean reciprocal rank）
- 正解アイテムが見つかった順位における精度を計算し平均をとる平均精度（mean average precision）

表 1　Wilcoxon 符号付順位検定の結果（MovieLenz データセット）
Table 1　Wilcoxon signed rank test with MovieLenz dataset.

Evaluation metrics	Models	p-value
mean top-10 precision	HMMB-N	0.1708
	HMMB-N(no-UU)	0.1014
mean reciprocal rank	HMMB-N	2.376×10^{-5}
	HMMB-N(no-UU)	6.511×10^{-14}
mean average precision	HMMB-N	2.200×10^{-16}
	HMMB-N(no-UU)	2.200×10^{-16}

表 2　Wilcoxon 符号付順位検定の結果（Last.fm データセット）
Table 2　Wilcoxon signed rank test with Last.fm dataset.

Evaluation metrics	Models	p-value
mean top-10 precision	HMMB-N	0.08375
mean reciprocal rank	HMMB-N	0.1789
mean average precision	HMMB-N	2.200×10^{-16}

MovieLenz のデータを用いた結果を図 4，図 5，図 6 に示す．この実験において，MovieLenz のデータにユーザ間リンクがないことから，ユーザ間にグループ割り当てをしないという設定でも実験を行った．これを「HMMB-N(no-UU)」と表す．これに対し，HMMB-N はユーザ間リンクが無くてもユーザ間にグループ割り当てを行う．図 4，図 5，図 6 においては，提案モデルである HMMB-N は，MMSB や HMMB-L よりも有効であると言える．また，これらについて，Wilcoxon の符号付順位検定 [6] を用いて，MMSB を基準とした有意差の検定を行った結果を表 1 に示す．本検定においては，提案モデルである HMMB-N の評価値と MMSB のそれとを比較している．表 1 から，HMMB-N，HMMB-N(no-UU) は平均逆順位および平均精度に関して $p < 0.01$ であるため，このとき MMSB に対して 0.01 の有意水準で有意に差があると言える．

Last.fm のデータを用いた結果を図 7，図 8，図 9 に示す．この実験において，Last.fm のデータにアイテム間リンクがないことから，アイテム間にグループ割り当てをしないという設定でも実験を行った．これを「HMMB-N(no-II)」と表す．これに対し，HMMB-N はアイテム間リンクが無くてもアイテム間にグループ割り当てを行う．図 8，図 9 においては，今回の提案モデルすべ

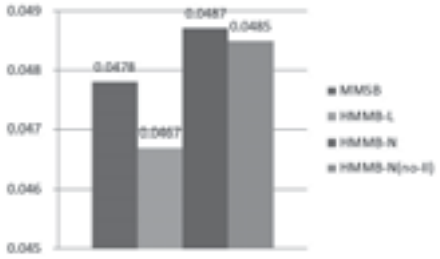

図 7　平均 10 位精度（Last.fm データセット）
Fig. 7　Mean top-10 precision with Last.fm dataset.

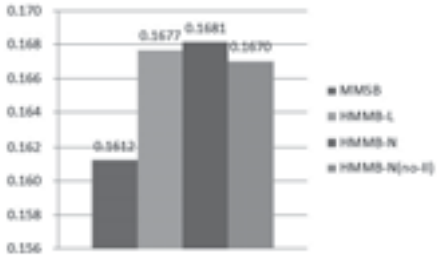

図 8　平均逆順位（Last.fm データセット）
Fig. 8　Mean reciprocal rank with Last.fm dataset.

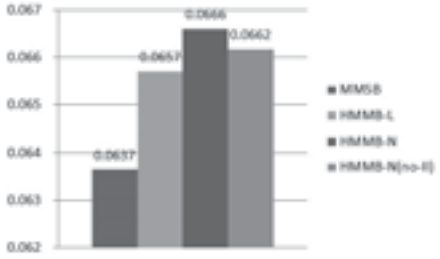

図 9　平均精度（Last.fm データセット）
Fig. 9　Mean average precision with Last.fm dataset.

てが MMSB を上回っている．また，これらについて，Wilcoxon の符号付順位検定により，MMSB を基準とした有意差の検定を行った結果の一部を表 2 に示す．表 2 から，HMMB-N は平均精度のみに関して $p < 0.05$ であるため，このとき MMSB に対して 0.01 の有意水準で有意に差があると言える．

4.5　考察

MovieLenz，Last.fm いずれのデータセットを用いても，上位 N 推薦問題に関して提案モデルである HMMB-N が従来の MMSB よりも有効であることがわかる．これは，同一ネットワーク上に複数の異なる種類のノードが存在するとき，ノードの種類を考慮することが有効であるということを示唆する．ここで，平均 10 位精度に関して有意性が見られないのは，この評価指標の性質上，ゼロ値が多いことが原因であると考えられる．一方で，もうひとつの提案手法である HMMB-L の性能は MMSB に及ばない場合が少なくなかった．その原因として，HMMB-L で用いる 3 次のテンソルがスパースであることが考えられる．

5.　おわりに

本論文では，異種ネットワークのための混合メンバシップブロックモデルとして，リンクの種類に着目したモデル HMMB-L と，ノードの種類に着目したモデル HMMB-N を提案した．上位 N 推薦に関する実験により，HMMB-N は HMMB-L や従来の MMSB よりも概ね有効であることを確認した．本論文ではこれらのモデルの基本的な挙動を理解するための実験を行ったが，より詳細な評価は今後の課題とする．また，発展的な課題として，リンクに付与された複数のラベルを考慮するモデルについて検討する．例えば，本論文で実験に用いた Last.fm では，ユーザは音楽アーティストに対してその音楽ジャンルなどを表すタグを付与できる．このような補助情報をモデリングに反映させることにより，モデル推定精度や推薦性能が改善すると期待される．

【謝辞】

本研究の一部は，科学研究費補助金基盤研究 (B)（23300039）の援助による．

【文献】

[1] E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

[2] K. Faust and S. Wasserman. Blockmodels: Interpretation and evaluation. *Social Networks*, 14:5–61, 1992.

[3] C. Kemp, J. Tenenbaum, T. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. *Proc. of the 21st National Conference On Artificial Intelligence*, volume 1, pages 381–388, 2006.

[4] T. Minka. Estimating a Dirichlet distribution. Technical report, 2000.

[5] K. Nowicki and T. Snijders. Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96(455):1077–1087, 2001.

[6] F. Wilcoxon, S. K. Katti, and R. Wilcox. *Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test.* American Cyanamid, 1963.

石黒 七海　Nanami ISHIGURO
株式会社ヒミカ勤務．平成 25 年神戸大学工学部情報知能工学科卒業．

江口 浩二　Koji EGUCHI
神戸大学大学院システム情報学研究科准教授．博士（工学）．情報検索，統計的機械学習，データマイニングの研究に従事．

横峯 樹　Tatsuki YOKOMINE
株式会社ガイアックス勤務．平成 22 年神戸大学工学部情報知能工学科卒業．平成 24 年同大学大学院システム情報学研究科情報科学専攻博士前期課程修了．