評価と購買の両履歴データを用いるアスペクトモデルの パラメータ推定法に関する研究

情報数理応用研究　　5212C009-2　大井貴裕
指導教員　後藤正幸

A Study of Parameter Estimation Method Based on the Aspect Model Estimated by Combining Both Evaluation and Purchase Histories

OI Takahiro

1　はじめに

近年，インターネット上の電子商取引サイト (以下 EC サイト) は増加の一途をたどり，扱われるアイテム数も膨大となっている．EC サイトでは，ユーザの購買履歴や購買アイテムに対する評価履歴がデータベース上に蓄積されるため，これらの膨大なデータを活用したプロモーションが可能である．このうち，特に各ユーザの嗜好や特性に合わせて自動で推薦を行う推薦システム [1] の web マーケティングツールとしての重要性が増している．その代表的な手法として，ユーザ間の購買履歴や評価履歴の類似性から被推薦ユーザの好むアイテムを予測し提示する協調フィルタリング (以下 CF)[2] がある．CF では，類似した嗜好を持つユーザは類似したアイテムを好むという仮定に基づいており，その推定に購買履歴または評価履歴のいずれか一方を活用する．ここで，評価履歴とはユーザが購入したアイテムに対して付与した評価値の履歴であり，評価値とは希望するユーザが購入した商品に対して満足度の程度を付与した点数を指す．

他方，現実世界の EC サイト上での購買活動を考えると，商品を購入する利用者の大半は購入した商品に対する評価値をわざわざ投稿しない．また，頻繁に評価値情報を投稿する利用者も少数存在するが，このような利用者も全ての購入商品に対して評価値を付与するわけではない．このとき，商品アイテムが購入されたという履歴は残るものの，そのユーザがどの程度満足したかを表す評価値は得られないことになる．この点を考慮すると，EC サイトのデータベースに蓄積されている評価履歴データに比べて，購買履歴データは遥かに膨大な量が蓄積されていると想定される．しかしながら，前述の通り既存の CF においては，ユーザの嗜好を推定する際にいずれか一方の履歴のみを想定したモデルが用いられている．購買履歴のみを活用する場合，評価値はユーザ自身による満足度の評価であり，購買履歴と比べてユーザの嗜好をより直接的に表現しているにも関わらず活用されていないことになる．また，評価履歴のみを活用する場合，得られる評価履歴データは比較的少数であることから，嗜好を推定するために十分な評価履歴データを確保できるとは限らない．その一方で，購買履歴データは評価履歴データに比べて非常に多く蓄積されており，嗜好が類似したユーザは類似したアイテムを購入する傾向があるという意味で情報を有しているため，嗜好の推定に活用されるべきである．以上のことから，少数の評価履歴データのみではなく，膨大な購買履歴データも同時に用いることができれば，嗜好の推定に必要な評価履歴データの不足分を補うことができ，推薦精度の向上が期待できる．

そこで本研究では，確率的潜在クラスを用いた CF による評価値予測問題を対象に，評価履歴データに加え，購買履歴データも同時に用いたパラメータ推定法を提案する．具体的には，潜在クラスモデルの一つである Aspect Model(以下 AM)[3] に着目し，AM の確率モデルに従って生起した評価と購買の両履歴データを用いたパラメータの推定式を導出する．ベンチマークデータを用いた実験により，両履歴データを用いてパラメータを推定する本提案が，評価履歴のみを用いて学習を行う従来法に比べ推薦精度の面で優れていることを示す．

2　準備

2.1　推薦システム

推薦システム [1] は，ユーザの購買履歴，または評価履歴を基にユーザの未購買アイテムに対する嗜好を予測するシステムである．いま，ユーザ集合を $\mathcal{Y}=\{y_1,y_2,\cdots,y_I\}$，推薦対象である被推薦ユーザを $y_i\in\mathcal{Y}$ とし，アイテム集合を $\mathcal{X}=\{x_1,x_2,\cdots,x_J\}$ とする．離散値をとる評価値を r とし，その定義域を $\mathcal{V}=\{1,2,\cdots,R\}$ とする．いま，ユーザ $a\in\mathcal{Y}$ がアイテム $b\in\mathcal{X}$ を購入したという事象を，この組合せを用いて (a,b) と表し，システムにおける既知の N 件の購買データの集合を $\mathcal{G}=\{(a_l,b_l)\}_{l=1}^{N}$ とする．また，全ての購買データは M 件の評価値を持つ購買データと $(N-M)$ 件の評価値を持たない購買データに分けられるものとする．いま，$c_l\in\mathcal{V}$ とするときに M 件の評価値を持つ購買データ集合を $\mathcal{D}_E=\{(a_l,b_l,c_l)\}_{l=1}^{M}$ とし，これらを「評価付き購買データ」と呼ぶ．また，評価値を持たない購買データは評価値に関する情報を持たないので，本研究ではこれを評価値の欠損とみなし，r_{mis} と定義する．すなわち，$(N-M)$ 件の評価値を持たない購買データ集合は $\mathcal{D}_P=\{(a_l,b_l,r_{mis})\}_{l=M+1}^{N}$ となり，これらを「評価なし購買データ」と呼ぶ．例えば，ユーザ y_i がアイテム x_j を購買したが，評価を付けなかった場合，ユーザ y_i のアイテム x_j に対する評価値は r_{mis} となり，このデータを (y_i,x_j,r_{mis}) として表現する．

2.2　評価付き購買履歴を用いた Aspect Model

AM はユーザの嗜好とアイテムの特徴を推定するために用いられる確率的潜在クラスモデルの一つである．AM は購買履歴データを対象としたモデルとして提案され，後に評価付き購買履歴データを対象としたモデルに拡張された [4]．評価付き購買履歴 $\mathcal{D}_E$ を用いた AM では，ユーザとアイテムと評価値の 3 次元情報のみを扱う．このモデルでは，潜在クラスの集合を $\mathcal{Z}=\{z_1,z_2,\cdots,z_K\}$ とした下で，ユーザとアイテムと評価値の背後に潜在クラスが仮定されており，類似した嗜好を持つユーザが同じ潜在的なセグメントに所属するという仮定を置いている．AM のグラフィカルモデルは図 1 で示される．ユーザ y_i がアイテム x_j に対して評価値 r を付与する事象を (y_i,x_j,r)

と表すとき，AM による確率モデルは式 (1) で示される．

$$P(y_i,x_j,r)=\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k) \quad (1)$$

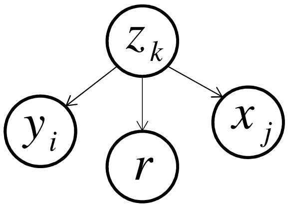

図 1:AM のグラフィカルモデル

いま，式 (1) の各確率分布 $P(y_i|z_k)$，$P(x_j|z_k)$，$P(r|z_k)$，$P(z_k)$ は多項分布に従うものとする．ただし，$\sum_{i=1}^{I}P(y_i|z_k)=1$，$\sum_{j=1}^{J}P(x_j|z_k)=1$，$\sum_{r=1}^{R}P(r|z_k)=1$, $\sum_{k=1}^{K}P(z_k)=1$ である．これらは観測できない変数 z_k を含むため，EM アルゴリズム [5] によってそれぞれを推定する．AM における対数尤度関数を L とし，以下で定義する．

$$L=\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r=1}^{R}\delta_{ijr}\log\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k) \quad (2)$$

ここで，δ_{ijr} はユーザ y_i がアイテム x_j を購買し，評価値 r を付与したとき，すなわち $\mathcal{D}_E$ に対して 1 を，それ以外のときに 0 の値をとるインジケータ関数である．このようにして推定された $\hat{P}(y_i|z_k)$，$\hat{P}(x_j|z_k)$，$\hat{P}(r|z_k)$，$\hat{P}(z_k)$ を用いて，アクティブユーザ y_i の未評価アイテム x_j に対する予測評価値を以下の式 (3) で算出する．

$$\hat{r}(y_i,x_j)=\sum_{r=1}^{R}r\frac{\sum_{k=1}^{K}\hat{P}(y_i|z_k)\hat{P}(x_j|z_k)\hat{P}(r|z_k)\hat{P}(z_k)}{\sum_{r=1}^{R}\sum_{k=1}^{K}\hat{P}(y_i|z_k)\hat{P}(x_j|z_k)\hat{P}(r|z_k)\hat{P}(z_k)} \quad (3)$$

3　提案手法

3.1　概要

本研究では推薦精度向上のため，評価付き購買履歴 $\mathcal{D}_E$ に加えて評価なし購買履歴 $\mathcal{D}_P$ を用いて，AM のパラメータを推定する方法を提案する．評価なし購買データであっても，ユーザとアイテムの共起情報を持つことから，$P(y_i|z_k)$，$P(x_i|z_k)$ の推定には寄与可能であると考えられる．これにより，潜在クラスとユーザ，アイテムの関係性を推定するための情報が増加し，推定精度の向上が期待できる．しかし，従来の評価付き購買履歴を用いた AM では，評価値の欠損を含まずに観測された完全データ (y_i,x_j,r) を対象としているため，評価付きと評価なしの両履歴データを用いて学習を行うことができない．

そこで提案手法においては，観測された評価付き購買データ (y_i,x_j,r) と評価なし購買データ (y_i,x_j,r_{mis}) の双方を用いたパラメータの推定式を導出する．導出にあたり，双方の購買データ全体に対して対数尤度関数を新たに定義し，極値問題を解く．各パラメータは EM アルゴリズムによって推定され，推定されたパラメータによって予測評価値が算出される．

3.2　推定式の導出

評価なし購買データは式 (1) の確率モデルから評価値 r に関する情報が欠損した状態で観測されたデータとみなす．そこで，式 (1) の確率モデルをユーザ y_i，アイテム x_j について周辺化を行うと以下の式 (4) を得る．

$$P(y_i,x_j)=\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(z_k) \quad (4)$$

提案手法において，評価なし購買データ (y_i,x_j,r_{mis}) は，式 (4) の確率分布に従い生起したと考える．よって提案手法においては，観測された評価付き購買データに対しては式 (1) を，評価なし購買データに対しては式 (4) を用いて対数尤度関数 L_h を以下の式 (5) によって定義する．

$$L_h=\sum_{i=1}^{I}\sum_{j=1}^{J}\Bigg\{\sum_{r=1}^{R}\delta_{ijr}\log\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k)+\delta_{ijr_{mis}}\log\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(z_k)\Bigg\} \quad (5)$$

ただし，式 (5) における $\delta_{ijr_{mis}}$ はユーザ y_i がアイテム x_j を購買し，評価をしなかったとき，すなわち $\mathcal{D}_P$ に対して 1 を，それ以外のときに 0 の値をとるインジケータ関数である．δ_{ijr} は評価付き購買データ集合 $\mathcal{D}_E$ に対して，$\delta_{ijr_{mis}}$ は評価なし購買データ集合 $\mathcal{D}_P$ に対してのみ値をとるため互いに背反であり，観測された購買データに対していずれか一方に 1 が代入される．

本研究では，L_h を最大化するパラメータ推定法を EM アルゴリズムにより構成する方法を導出する．E-step においては，各パラメータの値を既知とした元で潜在変数 z_k の分布を推定する．M-step では E-step で算出した潜在変数の分布を固定した元で各パラメータの値を更新する．E-step と M-step を繰り返すことで尤度を最大にするパラメータを算出できる．

まず E-step において，評価付き購買データをモデルに取り込むためにユーザ，アイテム，評価値に関するパラメータを固定したときの潜在変数の分布 $P(z_k|y_i,x_j,r)$，および評価なし購買データをモデルに取り込むためにユーザ，アイテムに関するパラメータを固定したときの潜在変数の分布 $P(z_k|y_i,x_j)$ を以下の式により求める．

$$P(z_k|y_i,x_j,r)=\frac{P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k)}{\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k)} \quad (6)$$

$$P(z_k|y_i,x_j)=\frac{P(y_i|z_k)P(x_j|z_k)P(z_k)}{\sum_{k=1}^{K}P(y_i|z_k)P(x_j|z_k)P(z_k)} \quad (7)$$

次の M-step においては，$P(z_k|y_i,x_j,r)$, $P(z_k|y_i,x_j)$ を固定したときのパラメータの値を求める．いま，
$\alpha_{ijrk}=P(y_i|z_k)P(x_j|z_k)P(r|z_k)P(z_k)$,
$\beta_{ijk}=P(y_i|z_k)P(x_j|z_k)P(z_k)$ とおくと，対数尤度を最大化する各パラメータの更新式を導出するために以下のように式 (5) を展開する．

$$L_h=\sum_{i=1}^{I}\sum_{j=1}^{J}\Bigg\{\sum_{r=1}^{R}\delta_{ijr}\log\Bigg(\sum_{k=1}^{K}P(z_k|y_i,x_j,r)\frac{\alpha_{ijrk}}{P(z_k|y_i,x_j,r)}\Bigg)+\delta_{ijr_{mis}}\log\Bigg(\sum_{k=1}^{K}P(z_k|y_i,x_j)\frac{\beta_{ijk}}{P(z_k|y_i,x_j)}\Bigg)\Bigg\}$$

$$\leq \sum_{i=1}^{I}\sum_{j=1}^{J}\Bigg\{\sum_{r=1}^{R}\delta_{ijr}\sum_{k=1}^{K}P(z_k|y_i,x_j,r)\log\bigg(\frac{\alpha_{ijrk}}{P(z_k|y_i,x_j,r)}\bigg)+ \delta_{ijr_{mis}}\sum_{k=1}^{K}P(z_k|y_i,x_j)\log\bigg(\frac{\beta_{ijk}}{P(z_k|y_i,x_j)}\bigg)\Bigg\} \quad (8)$$

右辺第 1 項から第 2 項への展開は Jensen の不等式による．式 (8) の右辺をさらに変形し，定数項を省略したものを式 (9) の L'_h とし，これを最大化する．

$$L'_h = \sum_{i=1}^{I}\sum_{j=1}^{J}\Bigg\{\sum_{r=1}^{R}\delta_{ijr}\sum_{k=1}^{K}P(z_k|y_i,x_j,r)\log\alpha_{ijrk}+ \delta_{ijr_{mis}}\sum_{k=1}^{K}P(z_k|y_i,x_j)\log\beta_{ijk}\Bigg\} \quad (9)$$

いま，式 (9) を最大化するため，ラグランジュの未定乗数法を用いる．未定乗数をそれぞれ λ_k，η_k，ξ_k，π とし，ラグランジュ関数を

$$f = L'_h + \sum_{k=1}^{K}\lambda_k\Bigg(1-\sum_{i=1}^{I}P(y_i|z_k)\Bigg)+\sum_{k=1}^{K}\eta_k\Bigg(1-\sum_{j=1}^{J}P(x_j|z_k)\Bigg) +\sum_{k=1}^{K}\xi_k\Bigg(1-\sum_{r=1}^{R}P(r|z_k)\Bigg)+\pi\Bigg(1-\sum_{k=1}^{K}P(z_k)\Bigg) \quad (10)$$

と定義する．これを偏微分して 0 とおくことにより，各パラメータ $P(y_i|z_k)$, $P(x_j|z_k)$, $P(r|z_k)$, $P(z_k)$ の更新式は式 (11)–(14) のように表される．

$$P(y_i|z_k) = \frac{\sum_{j=1}^{J}\phi_{ijrk}}{\sum_{i=1}^{I}\sum_{j=1}^{J}\phi_{ijrk}} \quad (11)$$

$$P(x_j|z_k) = \frac{\sum_{i=1}^{I}\phi_{ijrk}}{\sum_{j=1}^{J}\sum_{i=1}^{I}\phi_{ijrk}} \quad (12)$$

$$P(r|z_k) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\delta_{ijr}P(z_k|y_i,x_j,r)}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r=1}^{R}\delta_{ijr}P(z_k|y_i,x_j,r)} \quad (13)$$

$$P(z_k) = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\phi_{ijrk}}{\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\phi_{ijrk}} \quad (14)$$

ただし，

$$\phi_{ijrk} = \Bigg\{\sum_{r=1}^{R}\delta_{ijr}P(z_k|y_i,x_j,r)+\delta_{ijr_{mis}}P(z_k|y_i,x_j)\Bigg\}$$

とする．以上の式 (11)–(14) が提案手法におけるパラメータの更新式であり，これらの推定値を代入することによって再度 E-step の式 (6),(7) が更新される．

3.3 推定，予測アルゴリズム

3.2 節の結果に基づき，提案するパラメータ推定アルゴリズムは具体的には，以下の 5 ステップで与えられる．

Step1) 全てのパラメータにランダムな初期値を与える．

Step2) [E-step] 式 (6)，(7) により，$P(z_k|y_i,x_j,r)$，$P(z_k|y_i,x_j)$ をそれぞれ更新する．

Step3) [M-step] 式 (11)-(14) により，パラメータ $P(y_i|z_k)$，$P(x_j|z_k)$，$P(r|z_k)$，$P(z_k)$ を更新する．

Step4) 式 (5) の対数尤度関数 L_h の値が収束するまで Step2 と Step3 を繰り返す．収束後のパラメータを推定量 $\hat{P}(y_i|z_k)$，$\hat{P}(x_j|z_k)$，$\hat{P}(r|z_k)$，$\hat{P}(z_k)$ とする．

Step5) $\hat{P}(y_i|z_k)$，$\hat{P}(x_j|z_k)$，$\hat{P}(r|z_k)$，$\hat{P}(z_k)$ を用いて，式 (3) により，ユーザ y_i の未評価アイテム x_j に対する予測評価値を算出する． □

4 評価実験

提案手法の有効性を示すために，評価付き購買データと評価なし購買データの数を変化させ，パラメータを推定したときの推薦精度の比較を行う．また，パラメータの推定精度の推移の確認を行う．

4.1 実験条件

本実験では，推薦システムのベンチマークであるデータセット MovieLens[1] を用いた．このデータは 1997 年 9 月から 1998 年 4 月までに集められた映画の評価付き購買データである．ユーザ数 $I = 943$，アイテム数 $J = 1682$，評価値は 5 段階評価 $(R = 5)$，総データ数 10 万件である．当該データにおいて，各ユーザは全てのアイテムのうち最低 20 件以上に評価を与えている．このデータを無作為に学習用データ 8 万件とテスト用データ 2 万件に分ける操作を 5 回繰り返し 5 つのデータセットを得た．全てのデータには，評価値 r が含まれているため，以下では欠損評価値 r_{mis} を作成するため，学習用データのうち無作為に S 件の評価付き購買データ (y_i,x_j,r) を抽出して $\mathcal{D}_E$ とし，残りの $(80000-S)$ 件の評価値を欠損させ，評価なし購買データ (y_i,x_j,r_{mis}) を作成し $\mathcal{D}_P$ とした．ただし，各ユーザは必ず 1 つ以上のアイテムを評価しているものとし，各アイテムは必ず 1 つ以上評価値が付与されているものとする．評価なし購買データ $(80000-S)$ 件から無作為に 1000 件，10000 件，40000 件を抽出し，評価付き購買データ S 件とともに学習に用いた．比較手法としては評価付き購買履歴 $\mathcal{D}_E$ を用いた AM を用い，S 件の評価付き購買データのみを用いた．比較手法の対数尤度関数 L，提案手法の対数尤度関数 L_h 共に変化率が 10^{-5} 以下となった場合に，収束とみなした．なお実験の繰り返し回数は各データセット毎に 5 とし，その平均を結果とした．

4.2 評価手法

評価付き購買履歴を用いた推薦システムの評価指標として一般的な MAE(mean absolute error) を用いて，各手法の評価を行う．MAE は，以下の式により求められる．

$$\mathrm{MAE} = \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{ij}\mid r_{ij}-\hat{r}(y_i,x_j)\mid}{\sum_{i=1}^{I}\sum_{j=1}^{J}\varphi_{ij}} \quad (15)$$

ただし，r_{ij} はテストデータの評価値，φ_{ij} はテストデータにおいてユーザ y_i がアイテム x_j を評価しているときに 1 を，その他の場合に 0 を出力するインジケータ関数である．$\hat{r}(y_i,x_j)$ は式 (3) で算出される予測評価値である．

また，評価付き購買データ 80000 件をそのまま全て用いて比較手法を行ったときに算出される $P(y_i,x_j,r)$ の推定値を $P^*(y_i,x_j,r)$ とする．本実験では，この 80000 件からランダムに評価値を欠損させ，評価なし購買データ

を疑似的に生成していることから，この $P^*(y_i, x_j, r)$ に分布がどれだけ近いかは，学習の精度を測る一つの尺度となり得る．本稿では，$P^*(y_i, x_j, r)$ と $\hat{P}(y_i, x_j, r)$ との差を測るために，KL 情報量を用いる．$P^*(y_i, x_j, r)$ からみた分布 $\hat{P}(y_i, x_j, r)$ の平均 KL 情報量は式 (16) で示される．

$$\mathrm{D_{KL}}(P^*||\hat{P}) = \frac{1}{W}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{r=1}^{R}P^*(y_i, x_j, r)\log\frac{P^*(y_i, x_j, r)}{\hat{P}(y_i, x_j, r)} \quad (16)$$

ただし，テストデータにのみ存在するアイテムがあるため，提案手法によって算出される $\hat{P}(y_i, x_j, r)$ と $P^*(y_i, x_j, r)$ において確率が共に 0 でない (y_i, x_j, r) の組についてのみを算出に用い，その数を W とした．平均 KL 情報量が小さくなるとき，分布の差異が小さくなっていると解釈できる．

4.3 実験結果

表 1 に学習に用いる評価付き購買データ数を S 件に固定した上で，評価なし購買データの数を増加させたときの MAE を示す．

表 1:MAE の比較 ($K = 5$)

		評価なし購買データ数			
		0(比較)	1000	10000	40000
購買データ数	評価付き 4000	0.957	0.954	0.913	**0.884**
	8000	0.918	0.905	0.888	**0.877**
	12000	0.880	0.884	0.889	**0.879**
	16000	**0.854**	0.858	0.872	0.874
	20000	**0.838**	0.848	0.861	0.866
	24000	**0.830**	0.835	0.857	0.865

表 2:MAE の比較 ($K = 10$)

		評価なし購買データ数			
		0(比較)	1000	10000	40000
購買データ数	評価付き 4000	0.985	0.959	0.927	**0.888**
	8000	0.955	0.946	0.904	**0.878**
	12000	0.919	0.914	0.891	**0.874**
	16000	0.876	0.881	0.878	**0.871**
	20000	**0.849**	0.859	0.865	0.870
	24000	**0.830**	0.841	0.861	0.869

各表において，網掛けは比較手法に比べて提案手法の MAE が優れていること，太字は最良値を示す．表 1，2 から評価なし購買データを同時に用いることが推薦精度を向上させるか低下させるかの境界は，$K = 5$ のとき 12000 件前後，$K = 10$ のとき 16000 件前後であることが分かる．

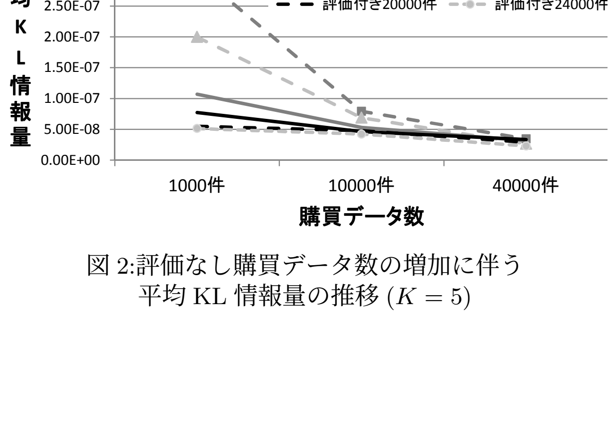

図 2:評価なし購買データ数の増加に伴う
平均 KL 情報量の推移 ($K = 5$)

また，$P^*(y_i, x_j, r)$ からみた分布 $\hat{P}(y_i, x_j, r)$ との平均 KL 情報量を算出した結果を図 2 に示す．平均 KL 情報量は，評価なし購買データ数の増加とともに減少していくことが分かった．これにより，評価なし購買データ数が増加すると分布 $\hat{P}(y_i, x_j, r)$ と分布 $P^*(y_i, x_j, r)$ の差異が小さくなっていくことが示された．

4.4 考察

表 1,2 より評価付き購買データ数が比較的少ない場合は，評価なし購買データ数を増加させることで，評価値予測精度が向上することが明らかとなった．これは，図 2 のように $P^*(y_i, x_j, r)$ と $\hat{P}(y_i, x_j, r)$ の平均 KL 情報量が評価なし購買データ数の増加とともに減少していくことから，評価なし購買データの活用によりパラメータ全体の推定精度が向上したことで評価値予測精度が向上したと考えられる．また，表 1,2 より評価なし購買データの活用が推薦精度を向上させるか低下させるかを分ける評価付き購買データ数の境界は潜在クラス数が $K = 5$ のときよりも $K = 10$ のときの方が大きい値となった．本手法は評価付き購買データ数が少なく，評価なし購買データ数が多いときに高い精度を示している．現実世界の EC サイトにおいては購買されるアイテム数が非常に膨大であり，かつユーザは一部のアイテムにしか評価値を付与しないことを考慮すると，本提案は実データに対して非常に有効な手法となると考えられる．

5 まとめと今後の課題

本研究では，AM の学習に対して，評価付き購買データ，評価なし購買データの両履歴データを用いる学習方法を提案し，ベンチマークデータを用いたシミュレーション実験によってその有効性を示した．今後の課題として，評価付き購買データ数が多いときに評価なし購買データの活用が推薦精度の低下を招く原因について理論的に検証をすることが考えられる．より膨大な評価付き購買データ数のデータセットを用いたシミュレーション実験も今後の課題である．

参考文献

[1] Resnick. P, Iacovou. N, Suchak. M, Berstrom. P, Riedl. J, “An Open Architecture for Collaborative Filtering of Netnews,” *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work,* pp.175-186, 1994.

[2] Herlocker. J. L, Konstan. J.A, Terveen. L, Riedl. J, “Evaluating Collaborative Filtering Recommender-Systems,” *Journal of ACM Transactions on Information Systems,* Vol.22, No.1, pp.5-53, 2004.

[3] Hofmann. T, and Puzicha. J, “Latent Class Models for Collaborative Filtering,” *Proceedings of 16th International Joint Conference on Artificial Intelligence,* pp.688-693, 1999.

[4] Hofmann. T, “Latent Semantic Models for Collaborative Filtering,” *ACM Trans. Information Systems*, 22(1), pp.89-115, 2004.

[5] Dempster. A, Laird. N, and Rubin. D, “Maximum Likelihood from Incomplete Data via the EM Algorithm,” *J. Royal Statist*, Soc. pp.1-38, 1977.