転移学習を利用した集団協調フィルタリング

Collective Collaborative Filtering Introducing Transfer Learning Techniues

神嶌 敏弘　　赤穂 昭太郎
Toshihiro Kamishima　　Shotaro Akaho

*1産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

Collaborative filtering suffers from a so-called start-up problem, which is the degradation of prediction accuracy due to the small sample size of preference data. To alleviate this problem, preference data in other sites are exploited by introducing a transfer learning technique. In this article, we consider issues regarding privacy, adaptability to each site, and communication efficiency, which arise when exploiting these data,

1. はじめに

本論文では，文献 [神嶌 07] にて提案した，複数の参加サイトからデータを集めて推薦をするサイト適応型集団協調フィルタリングについて論じる．この文献で示していた大域潜在変数を用いたモデルについての予備実験の結果を示す．

推薦システムは利用者が好むであろうアイテムを予測し，それを利用者に提示することで，情報過多の問題に対処するためのシステムである [神嶌 08]．協調フィルタリングは，この推薦システムを実現するための枠組みの一つである．これは,「口コミ」の過程を自動化したもので，過去の嗜好パターンが類似している利用者群を見つけ出し，彼らが好むものを推薦する．

しかし，運用上の問題はいくつも存在する．その中で，推薦システムの利用者数の少なさに起因する start-up 問題（cold-start 問題などともいう）[Schein 02] に注目する．大規模なネットショッピングモールでは，十分な数の利用者がシステムを利用している．だが，中小規模のサイトでは，利用する顧客数も少ないので，必然的にシステムの利用者数も少なくなる．このような状況では次の二つの問題を生じる．一つは，だれにも評価されてない，すなわち，好きかどうかの情報を与えられていないアイテムが頻繁に存在してしまう．これは，各利用者はアイテム集合の一部しか評価せず，また，評価されるアイテム群は一部のものに集中する傾向があるためである．他の利用者の意見を参考にして推薦アイテムを決める協調フィルタリングでは，こうした誰にも評価されていないアイテムは推薦の対象には決してなることはなく，推薦の被覆率が低下する．もう一つの問題は，現在利用している利用者（活動利用者）と類似した嗜好をもつデータベース中の利用者（標本利用者）が見つからない問題である．利用者数が少ないシステムでは，類似した嗜好をもつ利用者がシステム内にいないため，ノイズのように扱われてしまい，その嗜好が適切に推薦に反映されることはない．

この問題に対する解決法として，小規模なシステムの利用者のデータを集積して，十分な数のデータを集めることが考えられる．しかし，この方法には以下に述べる三つ問題がある．第一に，利用者の嗜好データは，秘匿すべき個人情報であるため，全てを集積して利用することは個人情報保護の観点から問題を生じる．この問題への対処であるプライバシー協調フィルタリング [Canny 02] の枠組みでは，利用者集団の嗜好のパターンからは個人それぞれの嗜好パターンは復元できないため個人情報ではないとの前提に立つ．そして，個人情報である嗜好データを秘匿したまま，全体の嗜好パターンを表すモデルを獲得することで，個人情報を保護しつつ推薦を行う．Canny [Canny 02] では，暗号の応用技術である秘密計算を用いた手法を提案しているが，その計算量は大きく，大規模・高速化には困難が伴う．そこで，個人単位で嗜好データを保護するのではなく，個人情報管理が可能な信頼できる小規模サイトでデータを蓄積することを想定する．そして，各サイトごとに，局所的な嗜好パターンのモデルを計算し，それらを集積して推薦を行う．局所的なモデルからは，個人ごとの情報は復元できないので，プライバシーの問題は生じない．また，この枠組みでは秘密計算に伴う計算量の問題も生じない．こうして集めたローカルモデルを要約して大域的なモデルを生成すれば，多数の利用者の意見を反映した推薦ができる．この局所的サイトのモデルを集めて，集団的に推薦の精度を高める枠組みを，集団協調フィルタリングと呼んだ [神嶌 07].

本論文では，Hofmann の，probabilistic latent semantic analysis (pLSA) [Hofmann 99a] モデル（aspect モデルとも呼ばれる）を利用した協調フィルタリング [Hofmann 99b] を対象に，この集団協調フィルタリングを考える．このモデルのパラメータは EM アルゴリズムで計算できるが，これは密結合な分散環境では容易に並列計算できる [Forman 00]．pLSA を用いた協調フィルタリングを，EM アルゴリズムを並列計算で解く試みとして Das らの研究 [Das 07] がある．ここで二つ目の問題が生じる．嗜好データはそれぞれ異なるサイトで保持されており，サイト間のネットワークは疎結合である．よって，サイト間の通信量は抑制する必要があり，これらの手法は適さない．

第三の問題は，サイトごとの利用者の個性である．中小サイトはそれぞれ特徴あるアイテムを扱い，また，利用者集団にも固有の特徴があるだろう．そうした集団に，集積した大域的なモデルを適用しても，十分にその特殊性を反映したモデル構築はできない．よって，各サイトごとの特徴を表現したモデルの獲得を行う**サイト適応型集団協調フィルタリング (site adaptive collective CF; SACCF)** が必要となる．ここでは，異なるドメインの知識を利用して予測精度の改善を行う転移学習 [神嶌 09, Pan 08] を利用する．

2. 節では，pLSA を用いた基本的な CF とその分散環境での実行について，3. 節では，サイト適応型協調フィルタリング手法を提案し，4. 節では予備的な実験結果を示す．最後に 5. 節でまとめを述べる．

連絡先: 神嶌 敏弘，http://www.kamishima.net/

2. pLSA による協調フィルタリングとその分散環境での実行

まず，pLSA による協調フィルタリング [Hofmann 99b] の，形式的定義と問題設定を述べる．利用者とアイテムをそれぞれ確率変数 x と y で表す．x は $\mathcal{X} = \{1, \ldots, i, \ldots, n\}$ 中の値を，y は $\mathcal{Y} = \{1, \ldots, j, \ldots, m\}$ 中の値をとる多値の確率変数である．ここで，このモデルで重要な役割をはたす潜在変数 z を導入する．これも多値変数で $\mathcal{Z} = \{1, \ldots, l\}$ 中の値をとり，潜在的な嗜好のパターンを表す．

ここでは，評価値を使わない，変数 x と y の共起関係だけを使うモデルを示す．嗜好データを，利用者 i がアイテム j を閲覧または，購入した場合を考える．これは，$x = i$ という事象と，$y = j$ という事象が同時に観測されることに相当する．このモデルでは購入しないという行為が無関係なので，未評価と不支持が区別できない暗黙的評価での問題が生じない利点がある．この x と y の同時確率分布を，潜在変数を導入して次のように表すのが pLSA モデルである．

$$\Pr[x, y] = \sum_{z \in Z} \Pr[x|z] \Pr[y|z] \Pr[z] \tag{1}$$

このモデルでは z が与えられたときに x と y が条件付独立であることを仮定して，モデルのパラメータの総数を減らしている．モデルのパラメータは $\theta = (\{\Pr[z|x]\}, \{\Pr[y|z]\}, \{\Pr[z]\})$ であるが，これらを最尤推定で求める．すなわち，N 個の共起データ $\mathcal{D} = \{(x_k, y_k)\}_{k=1}^{N}$ に対する対数尤度 $\mathcal{L}(\mathcal{D}; \theta) = \sum_{(i,j) \in \mathcal{D}} \log \Pr[x = i, y = j; \theta]$ を最大化するパラメータを求める．この推定は，以下の E と M のステップを交互に繰り返す EM アルゴリズム [Bishop 08] によって行う．E ステップは次式：

$$\Pr[z|x, y] = \frac{\Pr[z] \Pr[x|z] \Pr[y|z]}{\sum_{z'} \Pr[z'] \Pr[x|z'] \Pr[y|z']} \tag{2}$$

M ステップでは次式：

$$\Pr[x|z] = \frac{\sum_y n(x, y) \Pr[z|x, y]}{\sum_{x', y} n(x', y) \Pr[z|x', y]} \tag{3}$$

$$\Pr[y|z] = \frac{\sum_x n(x, y) \Pr[z|x, y]}{\sum_{x, y'} n(x, y') \Pr[z|x, y']} \tag{4}$$

$$\Pr[z] = \frac{\sum_{x,y} n(x, y) \Pr[z|x, y]}{l} \tag{5}$$

ただし，$n(x, y)$ は，$x = i$ かつ $y = j$ となる $\mathcal{D}$ 中の対の数である．

以上の手続きを反復し，収束した $\Pr[x|z]$，$\Pr[y|z]$，および $\Pr[z]$ が計算できれば，利用者 i がアイテム j をどれくらい好きかを，利用者が i が与えられたときアイテム j を好む条件付き確率 $\Pr[y = j|x = i]$ の大きさによって測れる．この条件付き確率分布は次式で計算できる．

$$\Pr[y|x] = \frac{\sum_z \Pr[z] \Pr[x|z] \Pr[y|z]}{\sum_{y', z} \Pr[x|z] \Pr[y'|z] \Pr[z]} \tag{6}$$

この量は利用者 i の各アイテムへの嗜好の度合いを示すので，次のアイテム y^* をその利用者に推薦すればよい．

$$y^* = \arg\max_{y \in \mathcal{Y}} \Pr[y|x = i] \tag{7}$$

表 1: テスト集合の利用者上の $\Pr[y|x]$ の総和

テスト集合	平均確率		人数
全体	**0.0695**		189
20 歳未満	0.0664	×	77
20 歳代	0.0747	○	332
30 歳代	0.0706	○	240
40 歳代	0.0593	×	168
50 歳以上	0.0610	×	125

この pLSA モデルの計算は，容易に分散環境で実行できる [Das 07]．簡単に述べると，利用者やアイテムのデータを複数のサイトで分散保持させる．E ステップには，x や y に関する和はないため，他のサイトのデータを参照せずに，局所的に計算が可能である．一方の M ステップには x や y に関する和が存在する．しかし，これらの和は，各サイトがが保持するデータに対して，サイト内で局所的部分和を計算し，これらの部分和を中心サイトに集めて総和をとれば容易に分散環境で計算できる．よって各反復の M ステップの前に同期をとって，このようにして和を求めれば，サイト数にほぼ比例する割合で計算を高速化できる．

3. サイト適応型集団協調フィルタリング

3.1 分散 pLSA の問題点

1. 節では，(1) 個人情報への配慮，(2) サイト間の通信量の抑制，そして，(3) サイトごとの利用者の特性への適応という三つの課題を示した．これらの課題は，単純に pLSA を分散環境で並列化するだけでは対処できない点について論じる．

サイト適応型の推薦の必要性に関する文献 [神嶌 07] の予備解析の結果をもう一度簡単に示しておく．協調フィルタリングの代表的なベンチマークである MovieLens データ [MovieLens data] を対象に予備実験を行った．このデータでは，943 人の利用者が，1682 種の映画について採点法で 5 段階の 10 万個の評価値を与えている．利用者 i が，映画 j について 5 段階で 4 か 5 の好意的評価なら $(x = i, y = j)$ のデータが観測されたとみなした．利用者を訓練用とテスト用に分け，訓練用利用者の全嗜好データと，テスト用の半分嗜好データから 2. 節の方法でモデルを獲得する．そして，テスト用の各利用者 i の残り半分のデータに含まれる映画 j について $\Pr[y = j|x = i]$ を求め，これら値の各利用者ごとの総和を計算した．これらの和の全テスト利用者について平均である次式で予測精度を評価した．

$$\frac{1}{|\mathcal{X}_t|} \sum_{i \in \mathcal{X}_t} \sum_{(i,j) \in \mathcal{D}_t} \Pr[y = j|x = i] \tag{8}$$

ただし，$\mathcal{X}_t$ はテスト用利用者で，$\mathcal{D}_t$ はテスト用データである．実験結果を表 1 に示す．「テスト集合」の列にはテスト用利用者を選択した基準を示した．全体とは，利用者の 20%をランダムにサンプリングしてテスト用利用者とした場合の結果である．これは，データ全体の平均的な利用者を反映したベースラインである．それ以外では，年齢によってテスト用利用者を選択した．テスト用ではない利用者全てが訓練用利用者である．それぞれの条件に該当する利用者の人数は「人数」の列に示した．「平均確率」は，各実験での，式 (8) のスコアであり，ベースラインである「全体」の結果を上回るものに○を，そうでないものに×を付けてある．この結果では，人数が多くこのデータの中核となっている 20〜30 代の利用者では予測精度は良いが，少数派の集団では悪くなる．すなわち，たとえ複数のサイトから大量にデータを集積できたとしても，大域的に均一なモ

デルを用いたのでは，少数派のサイトの利用者には良い推薦ができないことを示唆している．以上のことから，複数サイトによる集団協調フィルタリングでは，各サイトに適応したモデルを，集積したデータから獲得する必要がある．

また，各サイトは広域分散しており，サイト間の通信は抑制する必要がある．しかし，2. の分散 EM アルゴリズムでは通信量を抑制できない．これは，EM アルゴリズムの各反復ごとに総和を求めるため，中心サイトへデータを送る必要が生じるためである．

プライバシーに関しても問題がある．ある利用者の，異なるアイテムに関する情報が分散保持されている場合を想定する．この場合，2. の分散 EM アルゴリズムでは，利用者個人の嗜好データは単独では外部サイトには送信されない．しかし，複数のサイトの利用者が同一であるかどうかという情報は計算のために必要であり，個人情報保護の観点からはやや問題である．

3.2 大域潜在変数を導入したサイト適応集団協調フィルタリングモデル

前節で述べたように，分散 pLSA では 1. 節で述べた三つの課題を解決できない．そこで，これらの問題を解決できるような分散モデルを示す．まず，各サイト利用者の個性を考慮できない原因は，購買パターンを示す潜在変数 z が全サイトで共通であることである．そこで，各サイトそれぞれに潜在変数 z_k があり，さらに全サイトに共通なパターンを表す w を導入した階層モデルを考える．この w は z_k の超パラメータに相当する．こうした階層的なモデル化は，転移学習 [神嶌 09, Pan 08] の一つであるマルチタスク学習 [Caruana 97] の代表的な方法である [Raina 06]．転移学習とは，ある問題を解くために，その問題自体のデータに加え，それと類似した問題のデータも利用して，より精度のよいモデルを獲得するアプローチである．ここでは，他のサイトでの嗜好予測という類似した問題のデータを利用するため転移学習が使われる．転移学習の中でも，複数の問題に共通する因子を見つけ出し，これら複数問題を同時に改善しようとする枠組みをマルチタスク学習という．複数のサイトでの予測精度を同時に改善したい SACCF 問題は，このマルチタスク学習の問題として論じることができる．

しかし，単純に階層モデルにしただけでは，プライバシーと，サイト間の通信量の問題には対処できない．そこで，図 1 のようなモデルを採用した．この例では，サイト数は二つで，各サイトの潜在変数 z_1 と z_2 が大域的な潜在変数 w に依存している．そこで，各サイトで局所的に pLSA を実行し，求めたパラメータを一つの中心サイトに集めて，集団協調フィルタリングを実行する．pLSA のパラメータには，$\Pr[x|z]$，$\Pr[y|z]$，および $\Pr[z]$ がある．このうち最初の $\Pr[x|z]$ は利用者個人に依存した個人情報なので，中心サイトには送信できない．よって，$\Pr[y|z]$ と $\Pr[z]$ のみを中心サイトに送信し，改良したパラメータ $\hat{\Pr}[y|z]$ と $\hat{\Pr}[z]$ を返させることで，個人情報の問題に対処している．さらに，EM アルゴリズムの反復は各サイト内で実行されるため，各反復ごとにサイト間で通信をする必要がなく，通信量を抑制できる．

以下，中心サイトでの学習手法の詳細を述べる．一般の階層モデルの超パラメータを最尤推定する問題は，経験ベイズやエビデンス近似 [Bishop 08] と呼ばれ，EM アルゴリズムで計算できる．しかし，このモデルでは，モデルの層の間で参照できる値に制限があるため，以下のような手続きで計算する．各サイト k からはデータ $\mathcal{D}_k$ に含まれるアイテムの個数 $n_1(y_1)$ と，分布パラメータ $\Pr_1[y_1|z_k]$ を中心サイトに集める．これらを使って中心サイトのモデルのパラメータを求めることになる．中心サイトでサイト 1 からのデータが生じる確率は次式で表

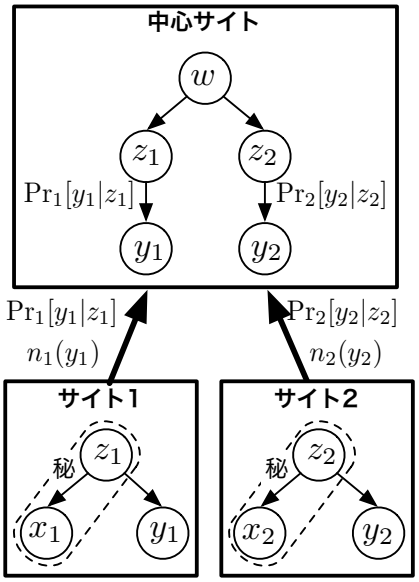

図 1: 大域潜在変数を導入した SACCF モデル

せる．

$$\Pr[y_1] = \sum_{z_1,w} \Pr_1[y_1|z_1] \Pr[z_1|w] \tag{9}$$

ここで，添え字のある Pr で表した，$\Pr_1[y_1|z_1]$ はパラメータではなく，サイト 1 から送られてきたものであり，定数である．するとサイト 1 でのデータの中心サイトのモデルでの対数尤度は次式になる．

$$\log \mathcal{L}_1 = \sum_{y_1} n_1(y_1) \log \left[\sum_{z_1,w} \Pr_1[y_1|z_1] \Pr[z_1|w] \Pr[w] \right] \tag{10}$$

サイト 2 以降に関しても対数尤度は同様であり，全体の対数尤度は $\log \mathcal{L} = \sum_k \log \mathcal{L}_k$ となる．このパラメータも EM アルゴリズムで解くことができる．すなわち，サイト k での E ステップは次式で，

$$\Pr[z_k, w|y_k] = \frac{\Pr_1[y_k|z_k] \Pr[z_k|w] \Pr[w]}{\sum_{z'_k,w'} \Pr[y_k|z'_k] \Pr[z'_k|w'] \Pr[w']}$$

M ステップは次式となる．

$$\Pr[z_k|w] = \frac{\sum_{y_k} n_k(y_k) \Pr[z_k, w|y_k]}{\sum_{y_k,z'_k} n_k(y_k) \Pr[z'_k, w|y_k]}$$

$$\Pr[w] = \frac{\sum_k \sum_{y_k,z_k} n_k(y_k) \Pr[z_k, w|y_k]}{\sum_k \sum_{y_k,z_k} n_k(y_k)}$$

こうして，中心サイトでのモデルが求まれば，各サイトの潜在変数の事前分布は次式で計算できる．

$$\Pr^{new}[z_k] = \sum_w \Pr[z_k|w] \Pr[w] \tag{11}$$

こうして求めた $\Pr^{new}[z_k]$ を送り返せば，サイト k では，自身がもつ $\Pr_k[y_k|z_k]$ と $\Pr_k[x|z_k]$ によって推薦ができる．

表 2: 全データ集合の利用者上の $\Pr[y|x]$ の総和

低年齢グループ

A サイト		B サイト	
$\|w\|$	平均確率	$\|w\|$	平均確率
orig	0.286818	orig	0.317305
2	0.286815	2	0.317307
3	0.286815	3	0.317307
5	0.286815	5	0.317307

高年齢グループ

A サイト		B サイト	
$\|w\|$	平均確率	$\|w\|$	平均確率
orig	0.299104	orig	0.274708
2	0.299105	2	0.274708
3	0.299105	3	0.274708
5	0.299105	5	0.274708

4. 実験

前節で提案した手法の有効性を検証するために予備的な実験を行った．データには 3.1 の MovieLens データを利用した．サイトはまず，10 と 20 代で構成される低年齢と，30 代以上の高年齢で分けた．人数はそれぞれ 411 と 536 人であった．さらに各年代の利用者をほぼ半分ずつに分け，それぞれを A と B サイトと呼ぶ．すなわち，全部で四つのサイトがある．こうすることで，低年齢 A サイトの予測性能は，別の類似した利用者構成の低年齢 B サイトのデータが転移されることで向上すると考えた．サイト傾向パターンの数，すなわち，大域潜在変数が取り得る値の種類数 $|w|$ を 2，3，5 と変化させてみた．各サイトで，各利用者が評価したアイテム集合を二つに分け，一方を訓練集合とし，もう一方を予測する，すなわち，2 分割の交差確認を実行した．評価尺度は，式 (8) と同じものを用いた．ただし，表 1 の結果ではテスト用のアイテムについてのみの和であった．だが，ここでは交差確認により全てのアイテムについて $\Pr[y = i|x = j]$ が計算されているので，テスト用の $\mathcal{D}_t$ ではなく，全データ $\mathcal{D}$ について和をとった．

結果を表 2 に示す．平均確率とは，上記の全データ集合の利用者上の $\Pr[y|x]$ の総和を，全利用者について平均したものである．$|w|$ は，大域潜在変数 w のとりうる値の数である．ただし，“orig” は，集団協調フィルタリングを利用せずにサイト内で局所的に予測した結果である．残念ながら，集団的協調フィルタリングをしない orig の結果と，他の結果を比較すると，効果はほとんど見られなかった．これは，各サイトに返された z_k の分布自体が，元のそれとあまり差がないという状況であった．

現状では，データに依存した問題か，モデル自体の問題であるかについてもまだ検証はできていない．まずデータに関しては，サイトの条件や，データ分布の条件を変えても提案手法に有効性がないのかを検証する必要がある．例えば，サイトの規模に差がある場合には，小さい方のサイトでは，予測精度が向上することも考えられる．また，もっと多数のサイトがある場合なども，多数のサイトから少量ずつの知識が転移されることで予測精度が向上することも考えられる．一方，手法に関しても，改良を試みる必要があるだろう．今回のモデルでは，z_k の分布のみを更新した．そこで，$\Pr[z_1]$ を中心サイトのモデル $\Pr^{new}[z_1]$ に固定して，ローカルパラメータ $\Pr_1[y_1|z_1]$ と $\Pr_1[x_1|z_1]$ を次回のデータ集約時に更新し，中心サイトに送るという改良も考えられる．また，サイト 1 と 2 で評価されるアイテムをそれぞれ y_1 と y_2 として，中心サイトの補助のもと $\Pr[y_1 = j, y_2 = j|x = i]$ を最大にするアイテム j を利用者 i に推薦する方法も考えられる．

5. まとめ

本稿では，小規模サイトのデータを集積することで，start-up 問題に対処する集団的協調フィルタリングの枠組みについて論じた．この実行にあたっては，(1) 個人情報は各サイト内でのみ扱いプライバシーに配慮し，(2) サイト間の通信量を抑制し，(3) 各サイトごとの個性に対応できる必要があることを論じた．本稿では，大域的な潜在変数を導入したモデルについて述べ，予備的な実験を行った．予想に反し，集団的協調フィルタリングを採用しても，他のサイトから知識は転移されず，予測精度にはほとんど影響がなかった．今後は実験を継続し，問題の性質を明らかにし，新たなモデルを考案する予定である．

参考文献

[Bishop 08] Bishop, C. M.: パターン認識と機械学習 上下 — ベイズ理論による統計的予測, シュプリンガー・ジャパン (2007–2008), (監訳：元田 浩 他；翻訳：神嶌 敏弘他)

[Canny 02] Canny, J.: Collaborative Filtering with Privacy, in *Proc. of the 2002 IEEE Symposium on Security and Privacy*, pp. 45–57 (2002)

[Caruana 97] Caruana, R.: Multitask Learning, *Machine Learning*, Vol. 28, pp. 41–75 (1997)

[Das 07] Das, A., Datar, M., Garg, A., and Rajaram, S.: Google News Personalization: Scalable Online Collaborative Filtering, in *Proc. of The 16th Int'l Conf. on World Wide Web*, pp. 271–280 (2007)

[Forman 00] Forman, G. and Zhang, B.: Distributed Data Clustering Can Be Efficient and Exact, *SIGKDD Explorations*, Vol. 2, No. 2 (2000)

[Hofmann 99a] Hofmann, T.: Probabilistic Latent Semantic Analysis, in *Uncertainty in Artificial Intelligence 15*, pp. 289–296 (1999)

[Hofmann 99b] Hofmann, T. and Puzicha, J.: Latent Class Models for Collaborative Filtering, in *Proc. of the 16th Int'l Joint Conf. on Artificial Intelligence*, pp. 688–693 (1999)

[神嶌 07] 神嶌 敏弘, 赤穂 昭太郎：参加システムの嗜好パターンが異なる場合の集団協調フィルタリング, 人工知能学会研究会資料, SIG–FPAI–A702–03 (2007)

[神嶌 08] 神嶌 敏弘：推薦システムのアルゴリズム (1)〜(3), 人工知能学会誌, Vol. 22, No. 6 〜 Vol. 23, No. 2 (2007–2008)

[神嶌 09] 神嶌 敏弘：転移学習のサーベイ, 人工知能学会研究会資料, SIG–DMSM–A803–06 (2009)

[MovieLens data] MovieLens data, : http://www.grouplens.org/node/12#attachments

[Pan 08] Pan, S. J. and Yang, Q.: A Survey on Transfer Learning, Technical Report HKUST-CS08-08, Dept. of Computer Science and Engineering, Hong Kong Univ. of Science and Technology (2008)

[Raina 06] Raina, R., Ng, A. Y., and Koller, D.: Constructing Informative Priors using Transfer Learning, in *Proc. of The 23rd Int'l Conf. on Machine Learning*, pp. 713–720 (2006)

[Schein 02] Schein, A. I., Popescul, A., Ungar, L. H., and Pennock, D. M.: Methods and Metrics for Cold-Start Recommendations, in *Proc. of The 25th Annual ACM SIGIR Conf. on Research and Development in Information Retrieval*, pp. 253–260 (2002)