6M-06

クロスドメイン推薦に向けたユーザ嗜好の予測手法の提案

吉井　和輝†　　青野　雅樹‡　　立間　淳司‡
豊橋技術科学大学　情報・知能工学課程†　　豊橋技術科学大学　情報・知能工学系‡

1. はじめに

ユーザが好みそうな情報を提示する情報推薦システムにおいて，異なったドメインの情報を補助的に用いるクロスドメイン推薦の研究が盛んに行われている(図 1)．これにより,様々なサービスによって蓄積された大量の異種データを有効に活用でき，推薦精度の向上や既存手法の問題緩和が期待できるとされている．

Loni ら[1]は，因子分解により要素間の相互作用を考慮して回帰予測を行うことの出来る手法である Factorization Machines[2](FM)に着目し，補助ドメインの情報を取り入れた特徴ベクトルを提案して FM の入力として用いることでユーザ嗜好の予測を行っている．

本研究では，クロスドメインデータを用いたユーザ嗜好の予測精度向上を図る．そのため，ドメインをまたがるすべてのアイテム間に新たに類似度を導入し，これにより Loni らの手法を改良する．

2. 従来手法

2.1 Factorization Machines

Factorization Machines(FM)では，以下の回帰式で予測を行う．

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} < v_i, v_j > x_i x_j$$

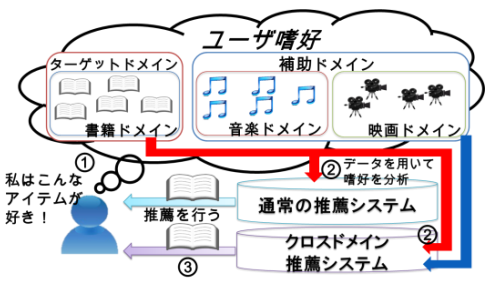

図 1 通常の推薦とクロスドメイン推薦

A Method for Predicting User Preferences toward Cross-Domain Recommendation
† Kazuki YOSHII　‡ Masaki AONO　‡ Atsushi Tatsuma
†‡ Dept. of Computer Science and Engineering, Toyohashi University of Technology

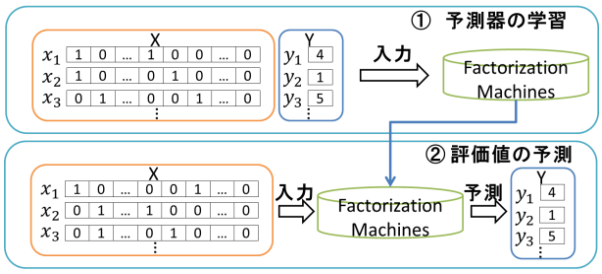

図 2 Loni らの手法の流れ

$\hat{y}(x)$は実数値ベクトルxに対する予測値を，$< v_i, v_j >$はベクトルv_iとv_jの内積を表す．実数値ベクトルxとそれに対応する値yを与え，パラメータ$\theta = \{w_0, w_1, \ldots, w_n, v_{1,1}, \ldots, v_{n,k}\}$を学習する．学習アルゴリズムは確率的勾配法やマルコフ連鎖モンテカルロ法などが適用できる．

v_iが実数値ベクトルのi番目の要素を因子数kで因子分解したベクトルであり，このベクトルの内積を計算することにより，要素間の相互作用を用いて予測を行うことが出来る．

2.2 Loni らのユーザ嗜好予測手法

Loni らは，ユーザの嗜好情報をまとめた特徴ベクトルとそれに対応する評価値を FM で訓練し，それに未知なユーザ特徴ベクトルを与えその評価値を予測する手法を提案した(図 2)．

Loni らが提案するユーザ特徴ベクトルxは以下の括弧のベクトルを並べたもので表現できる．

$$\begin{cases} a(u) = (0, \ldots, 0, 1, 0, \ldots, 0) \\ b(i) = (0, \ldots, 0, 1, 0, \ldots, 0) \\ z_j(u) = (\varphi_j(u,p), \varphi_j(u,q), \ldots) \end{cases}$$

$$\varphi_j(u,p) = \frac{r_j(u,p)}{|s_j(u)|} \quad (1)$$

aとbはそれぞれ「どのユーザ」が「どのアイテム」を評価したのかを表す．評価を行ったユーザ u，評価されたアイテムiに対応する要素が 1 となり，それ以外が 0 となるベクトルである．

$z_j(u)$はドメインD_jでのユーザuの嗜好を表す$(1 \leq j \leq m)$．$r_j(u,p)$はユーザuがアイテムpにつけた評価値を，$s_j(u)$はユーザuがドメインD_j内で評価したアイテムの集合を表す．ユーザのドメイン毎の評価数の偏りを考慮するため，ドメイン毎に行った評価数で正規化をする．

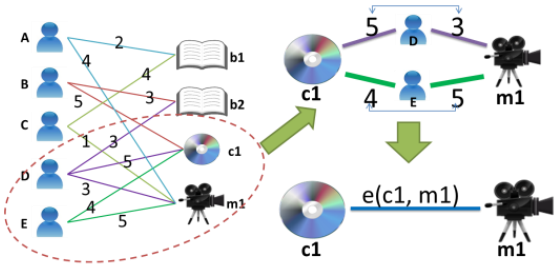

図 3 アイテム間の重みの付け方の例

3. 提案手法

本研究では，新たにドメインをまたがる全てのアイテム間に類似度を導入し，(1)式に掛け合わせることで特徴ベクトルを改良するアイデアを提案する．

3.1 内積類似度法

内積類似度法では，どのユーザがどのアイテムにどのくらいの評価値をつけたかを表すユーザ-アイテム行列とその転置行列の行列積を取ることでアイテム間の類似度を算出する．ユーザ-アイテム行列をRとすると，アイテム間の類似度を表す行列Fは以下のように表せる．

$$F = R^T D^{-1} R$$

D^{-1}は正規化のための対角行列を表す．ユーザの評価付けの傾向を考慮するために，ユーザのつけた評価値の和により正規化を行う．

3.2 アイテムグラフ類似度法

アイテムグラフ類似度法(以降，アイテム類似度法と呼ぶ)では，ユーザ-アイテム行列を隣接行列とする重み付き二部グラフの構造からアイテム間の類似度を算出する．二つのアイテムを共に評価するユーザの付けた評価値をもとにしてアイテム間に重み付きエッジを付与し，ノードがアイテムのみとなるアイテムグラフを作成する (図 3).

アイテムpとqのエッジの重みは以下のように与える．

$$e(p,q) = \frac{|U_{p,q}|}{\sum_{u \in U_{p,q}} ||r(u,p) - r(u,q)|| + 1}$$

$r(u,p)$はユーザuがアイテムpにつけた評価値を，$U_{p,q}$はアイテムpとqを共に評価したユーザの集合を，$|X|$は集合Xの要素数を，$||Y||$は式Yの絶対値を表す．ここで算出したエッジの重みをアイテム間の類似度とした．

3.3 複合類似度法

3.1，3.2 より計算した類似度をそれぞれ掛け合わせたものを，複合類似度法とする．

4. 実験

米 Amazon.com の購入履歴データ[3]より，複数ドメインのアイテムを評価しているユーザを 2505 人，そのユーザが評価をしているアイテムを 17000 件(書籍 6000 件，CD5000 件，DVD3000 件，VHS3000 件)選定し，4 つのドメインからなるデータセットを作成した．実験では書籍を推薦対象のアイテムを表すターゲットドメイン，CD，DVD，VHS を補助的な情報として用いる補助ドメインに設定した．このデータのうちターゲットドメインの情報の 25%，または 80%のデータをランダムに取り除いて正解データとした． 残り 75%，20%のデータと補助ドメインのデータを訓練データ(TR75，TR20)として学習を行い，評価値の予測を行った．評価尺度は平均絶対誤差(MAE)，二乗平均誤差の平方根(RMSE)を用いた．学習には libFM[4]を用い，学習反復回数は 500 回， 因子分解の因子数は 8，学習アルゴリズムはマルコフ連鎖モンテカルロ法とした．

実験結果を表 1 に示す．われわれの提案手法が Loni らの手法の精度を上回っていることがわかる．

表 1 ユーザ嗜好の予測実験結果

	TR75		TR20	
	MAE	RMSE	MAE	RMSE
Loni らの手法	0.4914	0.6678	0.7064	0.9200
内積類似度法	0.3966	0.5584	0.6593	0.8553
アイテムグラフ法	**0.2996**	**0.4458**	**0.5235**	**0.7128**
複合類似度法	0.4150	0.5906	0.6415	0.8592

5. まとめ

本稿では,ドメインをまたがるアイテム間の嗜好関係の類似度を導入することで，ユーザ嗜好の予測精度が向上することを示した．今後の課題は，類似度算出の計算量を削減すること，作成したアイテムグラフや類似度行列を用いて効果的な推薦アイテムを決定することが挙げられる．

参考文献

[1] B. Loni, Y. Shi, M.Larson, A. Hanjalic: Cross-Domain Collaborative Filtering with Factorization Machines, ECIR2014, pp.656-661. (2014).
[2] S. Rendle: Factorization Machines, ICDM2010, pp.995-1000, (2010).
[3] Amazon product co-purchasing network metadata http://snap.stanford.edu/data/amazon-meta.html
[4] S. Rendle: Factorization Machines with libFM, ACM Trans. Intell. Syst. Technol., 3(3), pp.57:1-57:22, May, (2012)