

一括予測型協調フィルタリング

桑田 修平, 上田 修功

NTT コミュニケーション科学基礎研究所

確率モデルに基づく新たな協調フィルタリング手法を提案する。提案法は、評点済みのデータから算出されるユーザごとの評点分布、アイテムごとの評点分布、および全評点に対する評点分布が、予測対象の未評点データに対して同様に算出されるそれぞれの評点分布と類似しているとの仮定に基づき、評点済みのデータに対する各評点分布と未評点データに対する各評点分布との間のKLダイバージェンスを最小化することにより予測値を求める。映画に対する評価データを用いた実験の結果、予測精度は従来の代表的手法とほぼ同程度であるが、計算時間の面で顕著な優位性を確認した。

One-shot Collaborative Filtering

SHUHEI KUWATA and NAONORI UEDA

NTT Communication Science Laboratories

We propose a new collaborative filtering method based on a probabilistic approach. In the proposed method, we assume that the empirical marginal distributions of the ratings over users and/or items are similar to these corresponding distributions of unrated data. Based on this assumption, we try to predict unobserved ratings by minimizing the Kullback-Leibler divergence between both the rated and unrated rating distributions. We evaluate the prediction performance and the computational time of our method by using real movie rating data, and confirmed that the proposed method could provide prediction errors comparable to those provided by the conventional top-level methods, but could significantly reduce the computational time.

1. はじめに

協調フィルタリング (Collaborative filtering: CF) は、推薦対象ユーザの未購入アイテムに対する評価値 (評点) を予測することでそのユーザに対する推薦アイテムを決定する技術である。ここで、従来法は評点予測をユーザ・アイテムごとに独立に行っている。換言すれば、予測対象のアイテム (ターゲットアイテム) の評点を予測する際、既知である評点履歴のみを用いている。これに対し我々は、ターゲットアイテム以外の未評点アイテムに対する評点予測値も、ターゲットアイテムの評点予測に対して何らかの有用な情報を持つと考える。つまり、あるターゲットアイテムの評点を予測する際、既知である評点履歴に加え、ターゲットアイテム以外の未評点アイテムに対する予測値も考慮することで予測精度が向上すると考える。そこで本稿では、予測対象の評点に対する予測値を互いに依存させ、予測値を一括して求める CF 手法を提案する。

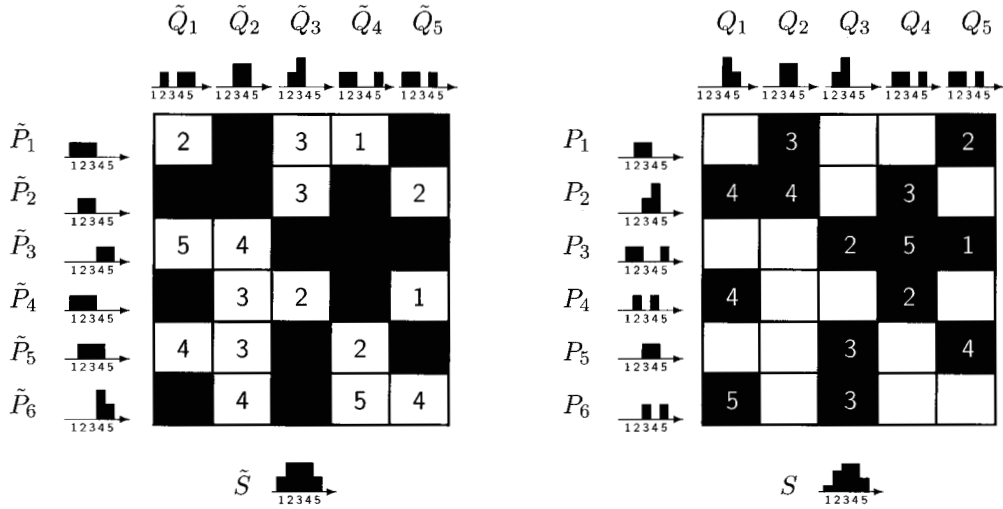
本稿の構成は以下のとおりである。まず、2 で、本稿で扱う協調フィルタリング問題を定義する。次に、3 で提案法の詳細を述べ、4 で評価実験と考察を行う。最後に 5 でまとめる。

2. 問題設定

2.1 定義

N 人のユーザが M 個のアイテムに対して与えた評点からなる N 行 M 列の評点行列を R とする。 R の (i, j) 要素 $r_{i,j}$ は、 i 番目のユーザの j 番目のアイテムに対する評点を表す。評点 $r_{i,j}$ は $\{1, 2, \dots, V\}$ の離散値をとり、値が大きいほど良い評価を意味するものとする。ただし、未評価である $r_{i,j}$ の値は 0 とする。また、 $\hat{r}_{i,j}$ は $r_{i,j}$ の予測値を表す。以降、 i をユーザインデックス、 j をアイテムインデックスとする。

評点行列 R において、既に評価が与えられた (未評価の) 評点の集合を \mathcal{R}_{obs} (\mathcal{R}_{mis}) と表す。さらに、予測対象の評点集合を \mathcal{R}_{tar} ($\mathcal{R}_{tar} \subseteq \mathcal{R}_{mis}$) とおく。ここで、既評価の評点数 $\#\{\mathcal{R}_{obs}\}$ は、未評価の評点数 $\#\{\mathcal{R}_{mis}\}$ に比べてはるかに小さいことに注意 ($\#\{\mathcal{R}_{obs}\} \ll \#\{\mathcal{R}_{mis}\}$)。 $\#\{\}$ は集合の要素数を表す。CF においてベンチマークデータとしてよく用いられる映画の評価データにおいては、評点行列 R の約 95% 以上の評点が未評価である。ただし、予測対象とする評点数は既評価の評点数とほぼ同等であるものとする ($\#\{\mathcal{R}_{tar}\} \approx \#\{\mathcal{R}_{obs}\} \ll \#\{\mathcal{R}_{mis}\}$)。



(a) 評点済みの評点集合 \mathcal{R}_{obs} (□) に対する周辺分布

(b) 未評点の評点集合 \mathcal{R}_{mis} (■) に対する周辺評点分布

図 1 評点行列 R における各周辺評点分布 ($N = 6, M = 5, V = 5$). (a) 評点済みと (b) 未評点の分布間の KL ダイバージェンスを最小化することで、未評点箇所の値を予測する。

2.2 目的

本稿における目的を、“評点行列 R を用いて \mathcal{R}_{tar} の値を予測する”と設定する。ここで、評点を予測する際に評点以外の情報を用いる手法もあるが、本稿においては評点履歴のみを用いる。また、予測精度の評価尺度として、正規化された平均絶対誤差 (Normalized Mean Absolute Error: NMAE) を用いる¹⁾。NMAE は、評点をランダムに予測した場合の MAE の期待値で MAE を除した値であり、値が小さいほど良い。NMAE が 1 より大きい (小さい) 場合、ランダムに予測する手法と比べて悪い (良い) と解釈できる。

3. 提案法

3.1 ランダムな欠損

文献 1) で指摘されているように、全ての CF 手法は評点行列 R に対して欠損はランダム (Missing At Random: MAR) という仮定を陰に置いている。つまり、評点行列 R において、未評点箇所はランダムに存在していると仮定している。提案法も従来法と同様に MAR の仮定を置く。すると、MAR の仮定の下では、評点済みの評点が従う評点分布と、未評点の評点が従う評点分布は類似していると考えられる。従って、評点済みの評点が従う評点分布と、未評点の評点が従う評点分布との類似性を測ることにより、予測対象の評点に対する予測値の妥当性を評価することができると考えられる。

3.2 周辺評点分布に基づく評点予測

上記考えに基づき、提案法は、評点済みの評点が従う評点分布と未評点の評点が従う評点分布間の距離 (非類似性) を最小にすることで予測を行う。具体的には、まず最初に、評点行列 R における評点済み評点の集合 \mathcal{R}_{obs} を用いて、

- ユーザ周辺評点分布 (行方向で周辺化した分布)、
- アイテム周辺評点分布 (列方向で周辺化した分布)、
- ユーザ・アイテム評点分布 (全体の分布)、

の 3 種類の評点分布を求める (図 1(a) 参照)。ここで、評点済みユーザ周辺評点分布を $\tilde{P}_i (i = 1, 2, \dots, N)$ 、評点済みアイテム周辺評点分布を $\tilde{Q}_j (j = 1, 2, \dots, M)$ 、評点済みユーザ・アイテム評点分布を \tilde{S} とする。それぞれの評点分布の算出は、例えば評点分布 \tilde{S} の場合、以下の式により計算する。

$$\tilde{S}(v) = \frac{\sum_{r_{i,j} \in \mathcal{R}_{obs}} \delta(r_{i,j} = v) + \eta}{\sum_l \sum_{r_{i,j} \in \mathcal{R}_{obs}} \delta(r_{i,j} = l) + \eta V}, \quad (1)$$

ここで、 $v, l = 1, 2, \dots, V$ であり、 $\tilde{S}(v)$ は評点 v の生起確率を表す。明らかに $\sum_v \tilde{S}(v) = 1$ 。 η は各評点の生起確率が 0 となることを防ぐためのスムージングパラメータを表し、本稿では $\eta = 1$ とした。また、 $\delta()$ は引数が真のとき 1、それ以外の場合は 0 を出力するインジケータ関数を表す。

次に、未評点データに対する上記 3 種類の (周辺) 評点分布、 $P_i (i = 1, 2, \dots, N)$ 、 $Q_j (j = 1, 2, \dots, M)$ 、

S_i が、最初に求めた3種類の(周辺)評点分布に類似するように未評点データの評点の組み合わせを求める(図1(b)参照)。なお、各(周辺)評点分布、 P_i 、 Q_j 、 S_i は、式(1)と同様の計算式でそれぞれ求める。

それぞれの評点分布間の非類似性は、2つの分布間の偽距離を表す *Kullback-Leibler (KL) ダイバージェンス* によって測る。ここで、例えば評点分布 \tilde{P}_i と P_i の KL ダイバージェンスを $KL(\tilde{P}_i||P_i)$ と表すものとする、評点予測問題は、予測対象となる未評点データ集合 \mathcal{R}_{tar} を変数とする以下の目的関数 J 、

$$J(\mathcal{R}_{tar}) = \frac{1}{N} \sum_{i=1}^N KL(\tilde{P}_i||P_i) + \frac{1}{M} \sum_{j=1}^M KL(\tilde{Q}_j||Q_j) + KL(\tilde{S}||S), \quad (2)$$

を最小化する最適化問題として定式化される。

3.3 予測アルゴリズム

目的関数 J (式(2)) を最小化する方法として、種々の最適化手法を適用することが考えられるが、本稿では、実用性を重視し、目的関数値が比較的小さい解を短時間で求めるヒューリスティックな予測アルゴリズムを提案する。この予測アルゴリズムは、評点を離散値と見なし、目的関数値が比較的小さい解からスタートして、目的関数値が減少する限り評点を個別に更新し続ける。以下に詳細を示す。

[予測アルゴリズム]

入力: \mathcal{R}_{obs}

出力: \mathcal{R}_{tar}

初期化 \mathcal{R}_{tar}

for all $r_{i,j} \in \mathcal{R}_{tar}$ do

以下を計算、

$$\alpha_{i,j}^P \leftarrow \arg \min_{r_{i,j}} KL(\tilde{P}_i||P_i),$$

$$\alpha_{i,j}^Q \leftarrow \arg \min_{r_{i,j}} KL(\tilde{Q}_j||Q_j),$$

$$\beta_{i,j}^{P+Q} \leftarrow \min_{r_{i,j}} KL(\tilde{P}_i||P_i) + \min_{r_{i,j}} KL(\tilde{Q}_j||Q_j).$$

end for

$\{r_{i,j} | r_{i,j} \in \mathcal{R}_{tar}\}$ を、 $\beta_{i,j}^{P+Q}$ の値で昇順にソート。

while 目的関数値 $J(\mathcal{R}_{tar})$ が減少 do

ソートした順に評点を更新、

$$r_{i,j} \leftarrow \alpha_{i,j}^P \quad \text{only if} \quad \alpha_{i,j}^P = \alpha_{i,j}^Q.$$

end while

ここで、 $\arg \min_x f(x)$ は、 $f(x)$ を最小にする引数 x を表す。また、 \mathcal{R}_{tar} の初期化については、例えば $r_{i,j}$ を初期化する場合、以下の式で表される、ユーザ i の

表 1 MovieLens と EachMovie データセット

	ML1	ML2	EM
ユーザ数	943	6,040	35,280
アイテム数	1,682	3,706	1,622
評点数	100,000	1,000,209	2,314,777
Sparsity	93.7%	95.5%	96.0%
評点のスケール	5	5	6
E[MAE]	1.6	1.6	1.944
学習用ユーザ数	800	5,000	30,000
テストユーザ数	143	1,040	5,280

平均 \bar{r}_i とアイテム j の平均 \bar{r}_j の平均とした。

$$(\bar{r}_i + \bar{r}_j)/2. \quad (3)$$

ただし、初期化する際には整数値に丸めた値を用いる。

4. 評価実験

4.1 データセットと実験設定

CF におけるベンチマークデータとして広く用いられている映画の評価データである、*MovieLens* (ML1, ML2) と *EachMovie* (EM) の3種類の実データを用いて、代表的な従来法と性能を比較した。ただし、EM に関しては、評点数が20以上のユーザのみを用いた。表1の上部に各データセットの詳細を示す。

実験設定は文献1)と同様にした。つまり、全ユーザを学習用ユーザとテストユーザに分け、(a) 弱汎化性能(既存ユーザに対する予測精度)と(b) 強汎化性能(新規ユーザに対する予測精度)の2つの性能を評価した。学習用ユーザとテストユーザのそれぞれの分割数を表1の下部に示す。また、既知の評点を評点済みと未評点の2つに分ける方法として、いずれの設定においても AllBut n プロトコルを用いた。このプロトコルは、ユーザごとにランダムに n 個の評点を未評点の評点として選択する。実験では、 $n = 10\%, 20\%$ とし、それぞれ、“AllBut 10%”, “AllBut 20%” と表す。

4.2 従来法

提案法と比較を行う従来法として、最近傍法に基づくアプローチである、 k 近傍法 (k NNs)^{2),3)}、Unified method with Similarity Fusion (SF)⁴⁾、確率モデルに基づくアプローチである、多項分布モデル (MULTI)¹⁾、混合多項分布モデル (MIXMULTI)¹⁾、Aspect Model (AM)¹⁾、User Rating Profile Model (URP)¹⁾ を用いた。

4.3 結果と考察

Dual Xeon 3.60GHz CPU、2GB メモリの計算機で実験した結果を表2と表3に示す。表2は予測精度、表3は予測に要した計算時間である。ただし、類似した結果を示したことから(b) 強汎化性能評価結果は省略する。ここで、表中の BASE LINE は、式(3)

表 2 各手法の予測精度。値が小さいほど予測精度が良いことを意味する。

(a) 弱汎化性能

		提案法	kNNs	SF	MULTI	MIXMULTI	AM	URP	BASE LINE
ML1	AllBut10%	0.456	0.468	0.477	0.488	0.489	0.470	0.487	0.498
	AllBut20%	0.457	0.470	0.479	0.490	0.492	0.486	0.492	0.499
ML2	AllBut10%	0.443	0.440	0.459	0.464	0.445	0.416	0.464	0.481
	AllBut20%	0.446	0.441	0.459	0.465	0.444	0.419	0.468	0.482
EM	AllBut10%	0.445	0.419	0.451	0.463	0.409	0.403	0.465	0.469
	AllBut20%	0.447	0.421	*	0.463	0.411	0.405	0.465	0.469

表 3 各手法の予測に要した計算時間。要した時間が短いほどスケラビリティが良いことを意味する。

(a) 弱汎化性能

		提案法	kNNs	SF	MULTI	MIXMULTI	AM	URP	BASE LINE
ML1	AllBut10%	0.67 s	1.24 m	5.10 m	0.22 s	11.0 s	5.74 m	4.42 m	0.22 s
	AllBut20%	1.00 s	2.19 m	8.67 m	0.21 s	8.33 s	6.40 m	4.39 m	0.22 s
ML2	AllBut10%	11.9 s	2.36 h	22.4 h	4.22 s	10.3 m	1.65 h	1.20 h	4.33 s
	AllBut20%	15.0 s	3.82 h	40.4 h	4.11 s	12.8 m	1.67 h	4.00 h	4.22 s
EM	AllBut10%	41.3 s	12.2 h	220 h	13.9 s	3.98 h	5.35 h	3.54 h	14.1 s
	AllBut20%	50.6 s	21.5 h	*	13.8 s	4.71 h	5.48 h	3.18 h	14.2 s

による予測精度を示している。ただし、BASE LINE では値の丸め操作は行わない。表中の“*”は、計算時間が2週間(336h)以上かかったため、計算を中止させたことを示す。ここで、表3中の h, m, s はそれぞれ時間、分、秒を表す。表中の各値は、学習用ユーザとテストユーザの分割を各データセットごとに3パターン作成し、さらに、各パターンに対して評点済みと未評点の分割を3パターン作成して得られたデータに対する結果の平均値を表す。

4.3.1 予測精度に関する考察

表2から、データの規模が大きくなるほど各手法の予測精度が向上することが確認できる。この傾向はクラス構造を持つ確率モデル(MIXMULTI, AM)で顕著である。これは、規模が大きいほど情報が多く、ユーザは嗜好性に基づいてクラス化されるという直感的な理解と一致する。提案法は、データの規模によらず安定的な値を示しており、特にデータの規模が小さいときに最も良い性能を示している。これはデータの規模が小さく、評点に関する全体的な情報が少ない場合には、予測値を互いに依存させることで、予測精度が向上することを示している。

4.3.2 計算時間に関する考察

表3に示すとおり、提案法は、他手法に比べてかなり高速であることが確認できる。ここで、MULTIやBASE LINEは提案法よりもさらに高速であるが、その差は、それら以外の手法の計算時間との差に比べれば大差ではない。また、MULTIやBASE LINEは、予測性能面で、全データにおいて提案法よりも劣っている。提案法とは対照的に、その他の手法は、データの規模に合わせて急激に計算時間が増大している。

4.3.3 考察のまとめ

実世界で扱われているデータの規模は、今回の実験で用いたデータの規模に比べはるかに大きい(数百万ユーザ、アイテム)。提案法は、従来法と同程度の予測精度を短時間で得られるという点で実用上の価値があると言える。

5. まとめ

予測対象の評点を互いに依存させ、全評点を一括して予測する協調フィルタリング手法を提案した。性能評価実験の結果、提案法は従来法と比べて大幅に短い計算時間で、従来法とほぼ同等の予測精度を得ることを確認した。今後は超大規模なデータ(数百万ユーザ)を対象に更に効果的なCF手法を検討する予定である。

参考文献

- 1) Marlin, B.: Collaborative Filtering: A Machine Learning Perspective, Master's thesis, University of Toronto (2004).
- 2) Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P. and Riedl, J.: GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of ACM CSCW1994*, pp. 175-186 (1994).
- 3) Sarwar, B., Karypis, G., Konstan, J. and Riedl, J.: Item-Based Collaborative Filtering Recommendation Algorithms, *Proceedings of ACM WWW2001*, pp. 285-295 (2001).
- 4) Wang, J., de Vries, A. P. and Reinders, M. J.: Unifying User-based and Item-based Collaborative Filtering Approaches by Similarity Fusion, *Proceedings of ACM SIGIR2006* (2006).