

不完全なユーザ情報を用いた情報推薦

中野 知彦[†] 増田 宏[‡] 大和 裕幸^{*} 安藤 英幸^{*}

東京大学大学院工学系研究科[†]

東京大学人工物工学研究センター[‡]

東京大学大学院新領域創成科学研究科^{*}

1. 背景

大量のデータのなかから必要な情報を効率よく見つけ出すための技術として、情報検索や情報フィルタリングがある。これらの技術を応用して各個人に適した情報を推薦するシステム（推薦システム）は商業的にも利用され[1]、近年注目を集めている。推薦システムでは広く協調フィルタリングという手法が用いられる。協調フィルタリングを用いた情報推薦を行うためには、各ユーザの各情報に対する評価情報（ユーザ情報）が必要になる。これは、アンケートなどの明示的な方法やユーザ行動の観察などによる暗黙的な方法[2]によって取得されるが、必要十分な量のユーザ情報を集めることは容易ではない。そこで、不完全なユーザ情報から高精度な推薦を行うための仕組みが必要とされている[3]。

2. 協調フィルタリングによる情報推薦の問題点

協調フィルタリングにおいて必要となる「ユーザ情報」を、本稿では以下のように定義する。

$$U_j = \{r_{1j} \ r_{2j} \ \dots \ r_{ij} \ \dots \ r_{nj}\}$$

ただし U_j : ユーザ j のユーザ情報

r_{ij} : 情報 i に対するユーザ j の評価

n : 情報の総数

また、「ユーザ情報の完全さ」を本稿では以下のように定義する。

$$p_j = \frac{\text{評価が既知である情報の数}}{n}$$

ただし p_j : ユーザ j のユーザ情報の完全さ

実際に集められたユーザ情報が表1であったときのことを考えてみると、ユーザ A はすべての情報に関する評価が既知であるが、ユーザ B は全情報のうちの一部についてしか評価が取得できていない。ユーザ B のユーザ情報の完全さは 0.5、ユーザ C のユーザ情報の完全さは 0.25 である。協調フィルタリングでは、ユーザ間の相関を利用して他ユーザの評価を重み付きで足し合わせて該当ユー

ザの評価値を推測する。この例のように不完全なユーザ情報をもとにして協調フィルタリングを行うと、ユーザ A とユーザ B はたった二つの情報に関する評価をもとに相関が決定されてしまい、推薦結果が不正確になってしまう可能性がある。また、ユーザ C は他ユーザとの相関を求めることができず、推薦することができない。ユーザ間の相関を利用して推薦する情報を決定する協調フィルタリング手法にとって、ユーザ情報が不足することは致命的な弱点となりうる。

	情報 1	情報 2	情報 3	情報 4
ユーザ A	2	3	1	4
ユーザ B		3		2
ユーザ C				4

表 1 ユーザ情報の例

3. 提案手法

そこで、不完全なユーザ情報をそのまま利用して協調フィルタリングを行うのではなく、情報の特徴量を利用して k 近傍法によりユーザ情報の評価が未知である部分（欠損部）を補完し、補完されたユーザ情報を利用して、協調フィルタリングを行う手法を提案する。手順は以下になる。

(1) あるユーザ j について、評価値が未知である情報 i とその他の情報すべてとの非類似度（小さいほど類似性が高いことを表す非負の実数）を求める。非類似度計算には、情報の種類に応じた特徴量間のユークリッド距離・内積・インターセクションなどを用いる。

(2) 非類似度の小さい k 個の情報（これを i の neighbor と呼ぶ）に対するユーザ j の評価値を重み付きで足し合わせて評価値を推測する。

$$r_{ij} = \frac{1}{\sum_{l=1}^k W_{il} \delta} \sum_{l=1}^k W_{il} r_{lj} \delta$$

$$\text{ただし } W_{il} = 1/d_{il}^2$$

d_{il} : 情報 i と情報 l の非類似度

$$\delta = \begin{cases} 1 & (\text{if } l \text{ is } i\text{'s neighbor}) \\ 0 & (\text{if } l \text{ is not } i\text{'s neighbor}) \end{cases}$$

(3) (1)(2)を評価値が未知であるユーザ、情報について繰り返す。

(4) 協調フィルタリングによる評価値の推測を行う。

4. 評価実験

東京国立博物館のウェブサイト

Information recommendation using incomplete user profiles
Tomohiko Nakano[†], Hiroshi Masuda[‡], Hiroyuki Yamato^{*},
Hideyuki Ando^{*}

[†]School of Engineering, The University of Tokyo

[‡]Research into Artifacts, Center for Engineering, The University of Tokyo

^{*}Graduate School of Frontier Sciences, The University of Tokyo

(<http://www.tnm.go.jp/>)では、保管する文化財の写真約2万6千枚をフィルムからスキャナー等を用いて入力した画像データ(画像サイズの大小の2種類)が公開されている。各々の写真には、対応する17種類の属性データが付与されている。ユーザはこれらの属性データを用いた検索により、利用したい情報にアクセスできるようになっている。また、ユーザの閲覧行動はウェブサーバのログに記録されている。これらのデータを利用して本稿で提案した手法の評価実験を以下の手順により行った。

(1) ログを利用して暗黙的にユーザ情報を取得

ログには、IPアドレス、アクセスしたファイル、アクセス時刻が記録されている。同一IPアドレスからの(30分以上の無アクセス状態を挟まない)連続したアクセスは同一ユーザによるアクセスであると仮定し、アクセスした画像サイズの違い・アクセス回数により重みをつけて、各画像に対する評価とした。尚、ユーザによる点数のばらつきの影響をなくすために、ユーザごとに平均0標準偏差1に正規化した。

(2) ユーザ情報の補完

ログから作成したユーザ情報には、欠損部分が多く含まれる。そこで、前節の手法を用いて補完を行った。画像データはR,G,B各16段階のヒストグラムインターセクションの逆数を非類似度として利用した。属性データは主要な属性7種類(品名,作者,時代,地域,国,材料,登録)に絞った上で、属性データごとに形態素解析して得られた品詞一つをtermとした、 $tf(term\ frequency)$ の内積の逆数を非類似度として利用した。

(3) 協調フィルタリングによる推薦と評価

前項で得られた補完されたユーザ情報を用いて協調フィルタリングを行い、各情報に対する各ユーザの評価値を推測し、(1)で得られている既知の評価値と比較して誤差を計算して精度評価を行った。精度評価の指標にはMAEを用いた。

実験の結果を図1、図2、図3に示す。図1より、ユーザ情報の完全さが低くなるに従って推薦誤差が大きくなり、推薦可能な情報の割合が小さくなっていることが確かめられる。図2は、補完なし、平均点を用いた補完、属性データを利用したk近傍法による補完、画像データを利用したk近傍法による補完、画像データ・属性データを利用したk近傍法による補完の5通りについて、協調フィルタリング計算および評価を行った結果を示す。この結果より、

- ・ 補完を行うことにより推薦誤差が減少する
- ・ 平均値による補完よりも、k近傍法による補完を利用したほうが、推薦誤差が減少する
- ・ 画像データ・属性データを併用する補完のとき

に、最も推薦誤差が減少することが分かった。図3は、補完の有無による推薦可能な情報の割合の違いを示す。提案手法により大半の情報が推薦可能になることが分かった。

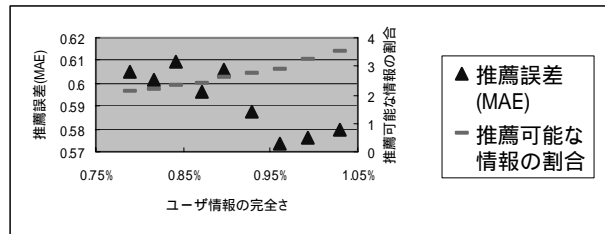


図1 ユーザ情報の完全さによる、推薦誤差と推薦可能な情報の割合の変化(補完なし)

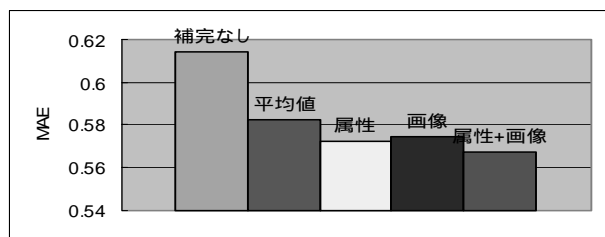


図2 補完手法による推薦誤差の違い

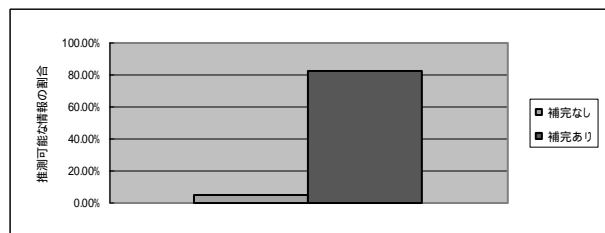


図3 補完の有無による推薦可能な情報の割合の違い

5. 結論

本稿では、ログデータ等から得られた不十分なユーザ情報を元に情報推薦を行う場合の問題点を解決するために、情報の特徴量を利用してユーザ情報の欠損部を補完し、補完されたユーザ情報によって協調フィルタリングを行う手法を提案した。また、東京国立博物館のデータを利用して評価実験を行い、提案した手法によって、

- ・ より多くの情報の推薦が可能になること
- ・ 推薦精度を高められること

が分かった。

参考文献

[1] J. Ben Schafer, Joseph Konstan, John Riedl, "Recommender Systems in E-Commerce", ACM Conference on Electronic Commerce, 1999
 [2] Mark Claypool, Phong Le, Makoto Waseda, David Brown, "Implicit Interest Indicators", In Proceedings of the International Conference on Intelligent User Interfaces, 2001
 [3] Prem Melville, Raymond J. Mooney, Ramadass Nagarajan, "Content-Boosted Collaborative Filtering", In Proceedings of the SIGIR-2001 Workshop on Recommender Systems, 2001