D_029

ユーザ間の類似性に着目したユーザ格付け

User Ranking Method Using Similarity Between Users

辻 隆生† 柳本 豪一† 福永 邦雄†
Takao Tsuji Hidekazu Yanagimoto Kunio Fukunaga

1. まえがき

ネットワーク上から入手できる情報は膨大であり，その情報は個人が発信することも多く，玉石混淆である．このため，ユーザが必要な情報を探す際に，数多くの不要な情報にも目を通さなければならないという問題が起こる．ここで，何らかの基準を用いて，ユーザにとって必要な情報を選別することができれば，上記の問題を解決することができる．これを実現するために，情報を発信したユーザを評価することで，情報の取捨選択を行うことを考える．ユーザの評価方法としては，多くのユーザから寄せられる信頼が利用できる．例えば，多くのユーザから情報を求められるユーザを評価することがあげられる．なぜなら，このようなユーザは有益な多くの情報を持っている人として他の人から信頼されているとみなせるからである．

本稿では，興味の類似性の高いユーザに情報を求めやすいと仮定し，興味の類似性を用いてユーザ全体をネットワークで表現し，このネットワーク構造を用いて上記のユーザを見つける方法を提案する．興味の類似性は，ユーザが過去に評価した情報への評価を用いて類似度として定義し，ユーザのネットワークは類似度を重みとした重みつき有向グラフを作成することで表現される．そして，このネットワークを行列で表現して，この行列の固有値を求めることでユーザの評価を行う．最後に評価実験を行い，提案手法より得られたユーザの評価を検討し，仮定を満たしたユーザが高い評価となることを確認する．

2. ユーザの類似性を用いた格付け

2.1 ユーザ全体のネットワーク表現

従来，ユーザが過去に評価した情報を用いて，ユーザの興味という観点から2者間の関係性が議論されてきた．例えば，小塩らは，ユーザ間の類似性を数値化するため，個々のユーザがブックマークしたウェブページを用いる方法を提案している [1]．この手法は，ブックマークはユーザの興味に応じて評価された情報であり，この評価された情報を用いてユーザの類似度を計算していると考えられる．本稿では，過去にユーザが評価した情報からユーザの興味の類似度を求め，ユーザ全体のネットワークを表現する．

今，ユーザ i からユーザ j への類似度を ω_{ij} とし，ω_{ij} をユーザ i とユーザ j が過去に評価した情報から決定することを考える．評価が似ている人は興味が似ており，そのような人からは有益な情報が得られると期待できるので，ω_{ij} はユーザ i とユーザ j の類似度を表すだけでなく，ユーザ i がユーザ j から新しい情報を得られる期待を表しているとみなすこともできる．具体的に類似度 ω_{ij} を検討する．まず，ユーザが行った評価の一致数に着目する．評価の一致数が多い人は自分と興味が似ていると考えられるので，ユーザ間の類似度の計算に利用できると思われるからである．この方法によって，式 (1) により ω_{ij}^1 を定義する．次に，評価した情報の一致数に着目する．類似度 ω_{ij}^1 では，ユーザの評価が一致しているかに着目していたが，実際にユーザは閲覧する情報を選択していると考えられるので，評価した情報自体もユーザの興味が反映されていると考えられる．したがって，情報に対する評価を考慮せず，評価した情報の一致数をもとに，ユーザの興味の判定を行うこととする．この方法によって，式 (2) により ω_{ij}^2 を定義する．最後に，ω_{ij}^1 と ω_{ij}^2 双方を組み合わせた類似度 ω_{ij}^3 を定義する．つまり，ω_{ij}^3 では，評価した情報の一致数が多く，かつその評価が一致しているユーザ同士は大きな類似度となる．式 (3) により ω_{ij}^3 を定義する．以下では，3種類の類似度を用いて実験を行う．

$$\omega_{ij}^1 = \frac{\text{ユーザ } i \text{ とユーザ } j \text{ の評価の一致数}}{\text{ユーザ } i \text{ の全評価数}} \quad (1)$$

$$\omega_{ij}^2 = \frac{\text{ユーザ } i \text{ とユーザ } j \text{ の評価した情報の一致数}}{\text{ユーザ } i \text{ の全評価数}} \quad (2)$$

$$\omega_{ij}^3 = \frac{\omega_{ij}^1 + \omega_{ij}^2}{2} \quad (3)$$

ただし，，$i = j$，すなわち自分自身との類似度は ω_{ij}^1，ω_{ij}^2，ω_{ij}^3 ともに0とする．以上のような類似度を用いて，ノードをユーザ，アークをユーザ間の連結，アークの重みを類似度とみなすことにより，ユーザ全体を重みつき有向グラフで表現する．

2.2 ユーザのネットワークを用いた評価値の導出法

本手法は，興味の類似性の高いユーザから有益な情報を得られやすいと仮定しているので，多くのユーザから情報を求められるユーザを高く評価する必要がある．今，前節で作成したユーザのネットワークを用いて考えると，あるユーザは有向グラフの重みにしたがって，他のユーザに対して情報を要求すると考えられる．そして，次のユーザもまた有向グラフの重みにしたがって，他のユーザに対して情報を要求すると考えられる．上記の操作を繰り返すことによって，頻繁に情報を求められるユーザと情報をあまり求められないユーザに分類されると考えられる．そして，頻繁に情報を求められるユーザは高い評価値とし，あまり情報を求められないユーザは低い評価値とする．このような評価値を求める方法は，重みつき有向グラフを推移確率行列として表現し，その推移確率行列に基づいて移動を繰り返したときの定常状態を求める方法と一致する．この定常状態は推移確率行列の最大固有値の固有ベクトルに対応するので，ユーザの評

†大阪府立大学大学院 工学研究科, Graduate School of Engineering, Osaka Prefecture University

表 1: 評価値の順位の比較

ユーザ	ランキング順位 R^1	ランキング順位 R^2	ランキング順位 R^3	評価の一致数順位	映画の一致数順位
450	1	3	1	1	3
13	4	1	2	4	1
276	2	4	3	2	4
416	3	5	4	3	5
655	8	2	5	8	2
303	5	7	6	5	7

価値の計算は，推移確率行列の固有値問題と一致する．本稿では評価値を確率として表現するので，評価値の高いユーザは，多くの他ユーザに情報を与える期待の高いユーザとなる．

以上から推移確率行列を利用してユーザの評価値を求める方法について具体的に説明する．まず，ユーザのネットワークを推移確率行列に変換する必要がある．ユーザのネットワークを行列として表現するため，まず隣接行列 $\boldsymbol{K}$を作成する．これは，行列 $\boldsymbol{K}$の成分を k_{ij} として，$k_{ij} = \omega_{ij}$ とすることで作成できる．次に隣接行列 $\boldsymbol{K}$を転置し，各列成分の総和が 1 になるように正規化を行ない，推移確率行列 $\boldsymbol{M}$を作成する．転置する理由としては，評価値を決定する際に，どれだけ新しい情報をもらえるかよりもどれだけ新しい情報を与えるかを重視するためである．そして，固有ベクトル $\boldsymbol{V}$を求め，$\boldsymbol{V}$を総和が 1 となるように正規化した各成分を，各ユーザの評価値とする．

本稿では類似度を 3 種類定義したので，ω^1_{ij} を使用したランキング値 R^1，ω^2_{ij} を使用したランキング値 R^2，ω^3_{ij} を使用したランキング値 R^3 を導出する．

3. 実験

本手法で用いる実験用のデータセットとして，GroupLens 提供の 100K MovieLens Dataset を使用した．100K MovieLens Dataset は 943 人のユーザ，1,682 本の映画に対して 100,000 個の評価がつけられたデータセットである．ユーザは最低 20 本の映画に対して評価を行なっており，その評価は 1 から 5 までの 5 段階となっている．本手法では 2 段階評価とし，評価値が 3 以上を 1，3 未満を 0 として 2 段階の評価とした．

本手法は，興味の類似性の高いユーザに対して情報を求めやすく，多くのユーザから情報を求められるユーザは信頼されていると仮定する．そこで，この仮定を満たす評価値が求められているか確認した．まず，3 つの類似度を用いて各ユーザの評価値を求めた．そして，他のユーザ全員との評価の一致数の総和を他のユーザ全員との評価した映画の一致数の総和の順位との比較を行った．

まず，上位ユーザ陣に注目する．その結果を表 1 に示す．表 1 は R^1，R^2，R^3 の上位 5 名のユーザをあげ，それぞれの順位を示している．表 1 から，R^1 では，ランキング順位と他ユーザ全員との評価の一致数の総和順位が，R^2 では，ランキング順位と他のユーザ全員との映画の評価数の総和順位が完全に一致した．また，R^3 では双方の順位が共に高いものが上位となっている．これより，仮定を満たす評価が行えたことが確認された．

次に，全体に目を向けると，ランキング値の順位と他ユーザ全員の評価の一致数の総和の順位は完全に一致しないものがあった．その一例として表 2 をあげる．表 2 では，ランキング値 15 位のユーザを A，16 位のユーザを B で表し，各要素は順位を表している．なお，表中の上位 30 名とは，本手法より求められたランキング上位 30 名に限定して，評価の一致数の総和順位を求めたものである．表 2 から分かるように，上位 30 名に限定した順位ではランキング順位と一致している．この理由として，本手法は推移確率行列を使用しているため，高い評価値をもつユーザ同士との類似度の大きさが影響したと思われる．すなわち，ユーザ A はユーザ B よりも上位ユーザとの興味の類似性が高かったため，ランキング順位でユーザ A が上位となったと考えられる．また，R^2，R^3 に対しても同様の傾向が確認された．これより，本手法は単純に他ユーザ全員の評価の一致数の総和から順位づけをしているのではなく，加えて，上位ユーザと興味の類似性が高いユーザが上位となるといえる．一般に，信頼されているユーザが情報を求められるユーザも信頼されていると考えられるので，上位ユーザと興味の類似性が高いユーザが上位になることは，評価として有益であると考えられる．

表 2: R^1 における散布図結果と一致しないユーザの例

ユーザ	ランキング値	他のユーザ全員	上位 30 名
A	**15**	16	**15**
B	16	**15**	16

4. まとめ

本稿では，興味の類似性の高いユーザに対して情報を求めやすいと仮定し，ユーザ間のつながりを興味の類似性で定義することにより，ユーザの格付けを行なう手法を提案した．実験で，提案手法より得られたユーザの評価に関して検討を行い，仮定を満たしたユーザが高い評価となることが確認でき，仮定が正しかったことがいえた．また，本手法はそれに加え，上位ユーザと興味の類似性が高いユーザが上位にくることもわかった．今後の課題として，本手法を利用した情報推薦システムの実装を行なっていく．

参考文献

[1] 小塩 力也, 横山 大作, 田浦 健次朗, 近山 隆, “利用者間の協調による検索エンジンのページランキングの修正” 第 65 回情報処理学会全国大会