

口コミ型情報推薦システムの提案と評価

○保科 一明 武田 利浩 平中 幸雄
山形大学大学院理工学研究科 情報科学専攻 平中・武田研究室

1.はじめに

インターネットの普及が進み、誰でも様々な情報が容易に得られるようになった。その一方で、膨大な情報の中から目的のものを見つからない「情報過多」と呼ばれる問題が起きている[4]。その解決のために、検索エンジンが開発され、普及してきた。しかし、検索キーワードをどこから入手するのか、という問題が残っている。

新しいキーワードを知ってもらうには、テレビ等のマスメディアが有効であり、テレビCMなどで検索キーワードを広告することが広く行われている。この方法では、各ユーザが自分の興味のあるキーワードに巡り合うまでに時間がかかってしまうことや、偶然性に依存するという問題が残る。

本研究では、各ユーザが興味・関心のある情報や、検索エンジンで利用するキーワードの手軽な入手を目的として、「口コミ型情報推薦システム」を提案する。相手に何か情報を伝える際、相手が欲しがる情報を伝えていることはよくあるが、これは情報を推薦していると考えることが出来る。提案システムはこのような口コミの伝搬メカニズムに基づいている。

本稿では、提案システムについて簡単に解説し、シミュレーションによる評価実験の結果を報告する。

2. 口コミ型情報推薦システム

図1に口コミ型情報推薦システムの概要を示す。提案システムでは、各ユーザが自分のプロファイルと知人のプロファイルを持つ。自分のプロファイルは、ユーザの興味・関心のあるタグを登録することで作成する。ここで登録されたタグを「有効タグ」と呼ぶ。知人のプロファイルは情報交換により追加・更新される。また、推薦情報にその内容を表すタグを付加する。推薦の際には、各プロファイルと情報に付いているタグをマッチングすることでフィルタリングを行う。フィルタリングは人間が口コミを伝える「理由に」対応した形で行う。以下にその理由を示す。

- (1) 友人が好みそうな情報を伝える
- (2) 自分が好む情報を伝える
- (3) 流行事その他

情報のタグと、知人のプロファイル内のタグが一致した場合、理由(1)に対応したと見なして情報を推薦する。また、情報に付いているタグと、自分のプロファイル内のタグが一致した場合、理由(2)に対応したと見なす。最後に、自分と知人のプロファイルどちらにもマッチしなかった場合は、理由(3)が適用される。ただし、この理由で推薦してしまうと、推薦の精度が低下してしまう。そこで、理由3の判断には制限を設ける。

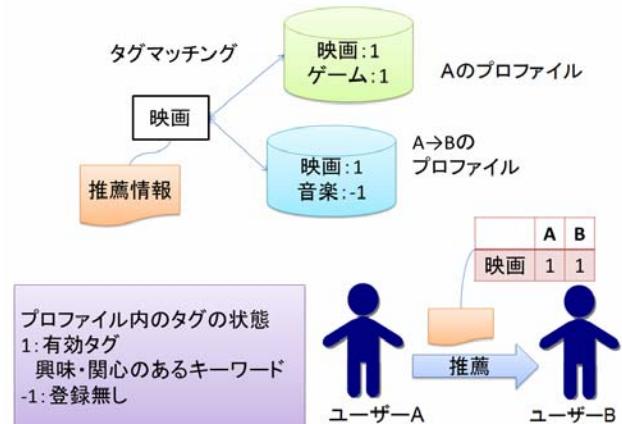


図1 口コミ型情報推薦システム

以上のフィルタリングの結果を、推薦情報と一緒に送信する。この事により、推薦と同時にプロファイルの交換を実現する。ここで送信されたAが持つBのプロファイルに誤りがあった場合、ユーザBは正しい値をユーザAに返信する。

3. シミュレーションによる評価実験

提案システムを評価するために、マルチエージェントシミュレーションによる評価実験を行う。本研究では、Repast Symphony[3]を利用した。

評価指標は、Noveltyな情報の適合率 $P(\text{Novelty})$ と再現率 $R(\text{Novelty})$ 、Serendipityなタグの適合率 $P(\text{Serendipity})$ と再現率 $R(\text{Serendipity})$ を用いる。

ここで Novelty とは、ユーザの嗜好に一致し、かつ新規性のある内容のものを指す。また Serendipity とは、ユーザにとって意外性のある内容のことを指し、こういったタグの入手により、検索エンジンの利用を補助出来ると考えることが出来る。

• Noveltyな情報の適合率と再現率

エージェント i が推薦された情報の集合を $L_m(i)$ 、エージェント i の Novelty な情報の集合を $F_N(i)$ 、エージェント総数を n とすると、 $P(\text{Novelty})$ と $R(\text{Novelty})$ は以下の式で表される。

$$P(\text{Novelty}) = \frac{1}{n} \sum_i^n \frac{|F_N(i) \cap L_m(i)|}{|L_m(i)|} \dots (1)$$

$$R(\text{Novelty}) = \frac{1}{n} \sum_i^n \frac{|F_N(i) \cap L_m(i)|}{|F_N(i)|} \dots (2)$$

• Serendipity なタグの適合率と再現率

エージェント i が推薦されたタグの集合を $L_{tag}(i)$ 、エージェント i の Serendipity なタグの集合を $F_{tag}(i)$ 、エージェント総数を n とすると、 $P(Serendipity)$ と $R(Serendipity)$ は以下の式で表される。

$$P(Serendipity) = \frac{1}{n} \sum_i^n \frac{|F_{tag}(i) \cap L_{tag}(i)|}{|L_{tag}(i)|} \dots (3)$$

$$R(Serendipity) = \frac{1}{n} \sum_i^n \frac{|F_{tag}(i) \cap L_{tag}(i)|}{|F_{tag}(i)|} \dots (4)$$

4. アルゴリズムを用いたデータセットの作成

評価実験には、(1) 人間関係を表すソーシャルグラフ、(2) 伝搬する情報、(3) 各ユーザの興味・関心の表すプロファイルの 3 つのデータセットが必要になる。

我々はデータセットのアルゴリズムによる生成を試みた。以下にその作成手順を示す。

(1) ソーシャルグラフの作成方法

ソーシャルグラフは、CNN モデルで作成する。CNN モデルは現実の友人関係を表すモデルである。確率 P で人と人が友達になる。

(2) 推薦情報の作成方法

推薦情報は、タグの組み合わせで表現する。本研究では、タグをランダム英数字で生成して利用した。また、出来るだけ現実世界の単語に近づくように、タグ同士に上位下位の階層構造を作成した。情報に付くタグは、ツリー上で基準のタグから一定範囲のものとする。こうすることで、ある特定分野の情報であることを示している。また、ある情報に異なる分野のタグが混ざり込むことも考えられる。そこで、本研究では、一定確率で基準のタグから離れたタグも情報に混ぜる。この確率を「ブレンド率」とする。

(3) プロファイルの作成方法

プロファイルも推薦情報と同様に、タグの組み合わせで表現する。推薦情報で利用したタグを各ユーザに割り当てていくことで、プロファイルを作成していく。ここで、ソーシャルグラフと各ユーザの嗜好について考えてみると、嗜好の近いユーザ同士が結び付いてソーシャルグラフが形成されると考えられる。そこで、ユーザのプロファイルにあるタグのいくつかをランダムに選択し、隣接ユーザにコピーすることで嗜好の共有を再現した。ここで隣接ユーザにコピーする割合をタグ共有率 α とする。

表 1 α の変化に伴う推薦の振舞い

(a) 提案システム

α	P(Nov)	R(Nov)	P(Seren)	R(Seren)
0	11.19%	17.42%	0.43%	24.22%
10	21.05%	28.28%	1.11%	37.34%
20	29.07%	36.68%	1.50%	45.43%

(b) ランダム推薦

α	P(Nov)	R(Nov)	P(Seren)	R(Seren)
0	3.31%	10.12%	0.41%	35.31%
10	4.16%	10.25%	0.62%	36.81%
20	5.01%	10.21%	0.83%	38.33%

5. 評価結果

提案システムの評価実験結果を論じる。前章に記述したデータセットの方法に従い、エージェント数 647、情報数 1000 のデータセットを作成した。ソーシャルグラフ作成の際の P は 60%、ブレンド率は 30% で設定している。

5.1 α の変化に伴うシステムの振舞い

まず、隣接ユーザ間のタグ共有率 α を 0 ~ 20 % まで変化させデータセットを作成し、それぞれのネットワークにおいて、提案システムの振舞いがどう変化するか調査する。この結果を表 1(a)に示す。また、同じ条件のデータセットを利用したランダム推薦の結果を表 1(b)に示す。

両結果を比較すると、提案システムは全てのデータセットにおいてランダム推薦以上の適合率と再現率を示した。この事から、提案システムは情報推薦として一定の効果があると言える。また、 α が向上するにつれ、適合率や再現率が向上している。これは、隣接ユーザ間の興味が近いネットワークにおいて、提案システムの効果がより発揮されるということを示している。

しかし、 $R(Novelty)$ の結果を見ると分かるように、 α が 20% のネットワークでも約 36% と、65% の興味のある情報がユーザに届いていないことになる。この原因として、コミュニティの分断が考えられる。コミュニティとは、ある分野に関心のある人達がリンクして形成される集まりである。ソーシャルグラフ全体でこのコミュニティが分断していると、単純な情報の伝播だけでは再現率が低下してしまう。

5.2 コミュニティ接続の検討

分断したコミュニティ間で情報を伝播させるための振舞いを検討する。新しい振舞いとして、次の 2 つを評価する。

- 1) ホップ数の利用
- 2) 中継タグの利用

1) は、情報のホップ、すなわち推薦の回数に注目する。次に推薦するユーザが興味のない情報だったとしても、それまで推薦された回数分は中継伝搬させる、という方法

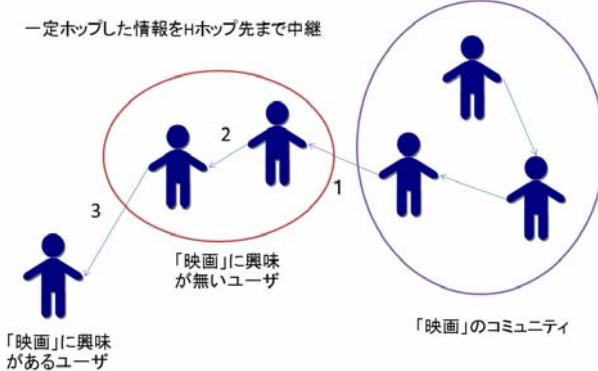


図2 ホップ数を利用した中継

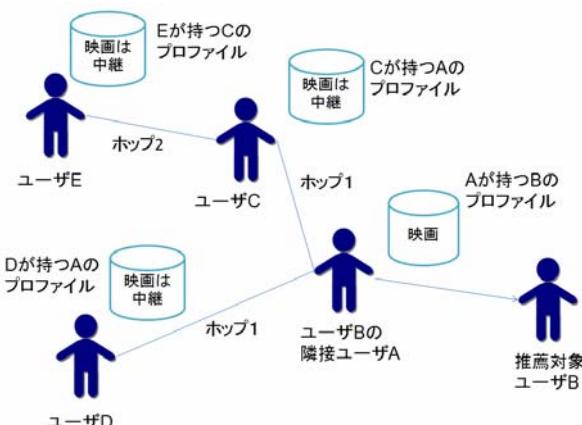


図3 中継タグを利用した中継

である。この際、中継するユーザーには情報を見せずに、自分の友人に伝搬するだけにする。これにより、各ユーザーが興味のない情報を大量に推薦することを防ぐ。エラー! 参照元が見つかりません。は、映画に関するコミュニティ内を2ホップした情報の扱いを示している。通常ではこのコミュニティから出た時点で伝搬が止まってしまうが、既にコミュニティ内を2ホップしているため、2つ先のユーザーまでは伝搬するといった処理を行う。

2)の方法は、自分が持つ友人のプロファイルにあるタグを、他のユーザーに教える、という方法である。このイメージを図3に示す。

推薦対象Bの友達Aは、Aが持つBのプロファイルにあるタグを、B以外の友人CやDに教える。通常の伝搬では、Aが持つBのプロファイルに映画タグが入っていたとしても、Aが映画に興味がなければ、CやDはAに情報を推薦しない。それを防ぐために、Aの先にいるユーザーに伝えるべき情報を表すタグを友人に教える。このタグを「中継タグ」と呼ぶ。また、中継タグをホップさせる範囲を変化させることでより遠くのユーザーから中継してもらうことが可能になる。

表2 コミュニティ接続手法の効果

(a) ホップ数を利用した中継

P(Nov)	R(Nov)	P(Seren)	R(Seren)
26.12%	45.07%	0.91%	37.67%

(b) 中継タグを利用した中継

ホップ	P(Nov)	R(Nov)	P(Seren)	R(Seren)
1	30.23%	42.65%	0.86%	37.34%
2	33.32%	48.15%	0.78%	33.51%
3	33.75%	49.99%	0.78%	35.73%

以上2種類の方式の評価を行う。用いるデータセットは $\alpha = 10\%$ のものである。ホップ数利用の中継の結果を表2(a)に、中継タグの結果を表2(b)に示す。

ホップ数利用の方式は、情報がより広く広がるようになるため、R(Novelty)が15%ほど向上している。それに伴い、P(Novelty)も5%の向上が見られた。この方法は、R(Novelty)を向上させるために効果があると考えられる。

中継タグの結果は、中継タグをホップする範囲を1から3まで変化させて実験している。この結果を見ると、ホップ数1の時点でP(Novelty)もR(Novelty)も10%以上向上している。この事から、Noveltyな情報を集めることに大きな効果があると考えられる。中継タグのホップを増やしていくと、より広い範囲から情報を集めるようになるため、R(Novelty)の向上も示すことができた。

5.3 既存手法との比較

評価実験の最後に、既存の情報推薦手法との比較を行う。既存手法の代表例として、「コンテンツベースフィルタリング」と、「協調フィルタリング」がある。

・コンテンツベースフィルタリング

コンテンツベースフィルタリングは、情報の内容と、各ユーザーのプロファイルの内容を比較し、一致したものを推薦する方式である。各ユーザーの興味のある情報が集めやすいが、Serendipityな推薦は難しいと考えることが出来る。

・協調フィルタリング

協調フィルタリングは、まず推薦対象ユーザーと嗜好が似ている類似ユーザー群を特定する。そのユーザー群が高評価を付けていて、かつ推薦対象ユーザーが未評価のものを推薦する手法である。協調フィルタリングはSerendipity性のある推薦が期待されるが、新規ユーザーが十分な推薦を受けられないコールドスタート問題が指摘されている。

以上の2種類との比較実験を行う。この実験も $\alpha = 10\%$ のデータセットで行う。まず表3提案システムのコールドスタートの結果とコンテンツベースフィルタリングの結果を示す。

表3 提案手法とコンテンツベースフィルタリング(CBF)の比較

	P(Nov)	R(Nov)	P(Seren)	R(Seren)
--	--------	--------	----------	----------

口コミ	17.06%	22.85%	1.14%	34.52%
CBF	100.00%	96.08%	0.60%	8.11%

表 4 提案手法と協調フィルタリング(CF)との比較

	P(Nov)	R(Nov)	P(Seren)	R(Seren)
口コミ	20.59%	31.98%	1.09%	35.61%
CF	10.19%	7.69%	0.90%	15.51%

前節の実験において、提案システムは初期値として、各ユーザの有効タグの一部を、隣接ユーザを提示していた。

表 3 の比較では、その初期値無しで実行している。

5.1 節の同条件のものと比較すると多少適合率や再現率が低下するが、十分コールドスタートで実行可能だと考えられる。一方コンテンツベースフィルタリングは、P(Novelty)や R(Novelty)が非常に高い値になっている。コンテンツベースフィルタリングは、ユーザが欲しいと示したタグが付いている情報をのみを集めるので、高い値になるのは当然である。しかし、Serendipity な推薦に関しては、提案システムの評価値が上回っている。適合率に大きな差はないが、再現率は 4 倍ほどの差がある。この点で、提案システムはコンテンツベースフィルタリングに比べて、Serendipity 性のある推薦に優れると言える。

次に、表 4 に提案手法と協調フィルタリングとの比較結果を示す。用いるデータは同じだが、提案手法を協調フィルタリングの動作に合わせるように条件を変更したため、提案手法の結果が他のものと異なっている。

表 4 より、全ての項目で提案手法が優れるという結果になった。Serendipity な推薦に関しても、コンテンツベースフィルタリングとの比較同様に、P(Serendipity) は大きな差はないが、R(Serendipity) は提案手法が高い値になり、Serendipity なタグをより提示できることを示している。

6. まとめ

本稿では、各ユーザが興味・関心のある情報や、検索エンジンで利用するキーワードの入手を目的として、「口コミ型情報推薦システム」を提案した。

提案システムの評価のために、適合率や再現率を指標としたマルチエージェントシミュレーションを行った。実験に用いるデータセットはアルゴリズムを用いて生成した。

ランダム推薦との比較結果より、提案システムは情報推薦として一定の効果がある。また、興味が近いユーザ同士がリンクしているネットワークにおいて、提案システムの推薦はより効果的になる。

しかし、ソーシャルグラフ上でコミュニティが分断している場合、単純な情報伝播だけでは上手くいかない場合がある。そこで、情報のホップ数や、中継タグを利用したコミュニティの接続を試みた。結果、Novelty な情報の適合率や再現率が向上し、情報の幅広い伝搬が可能になった。

最後に、既存の情報推薦手法であるコンテンツベースフィルタリングや協調フィルタリングと比較実験を行った。まずコンテンツベースフィルタリングとの比較において、Novelty な情報を推薦する点ではコンテンツベースフィル

タリングが優れるが、Serendipity なタグを推薦する点では提案システムが優れていた。協調フィルタリングとの比較では、全ての結果で提案手法が優れるという結果になった。

今後の課題として、別のデータセットを利用したシミュレーションが挙げられる。本稿で利用したデータセットでは、提案システムが優れるという結果になったが、異なるデータを用いた場合、この結果が変わってくる可能性がある。データセット作成の際のパラメータ変更や、実データを利用した比較実験を行う必要がある。

参考文献

- [1] 保科一明, 武田利浩, 平中幸雄, "口コミ型情報推薦システムのシミュレーションによる解析", FIT2010(第 10 回情報科学技術フォーラム), 第 2 分冊 529-530, F-058 (2010).
- [2] 保科一明, 吉津翔太, 武田利浩, 平中幸雄, "口コミ型情報推薦システムの提案", FIT2009(第 8 回情報科学技術フォーラム), 第 2 分冊 405-406, F007 (2009).
- [3] Repast Agent Simulation Toolkit
<http://repast.sourceforge.net/>
- [4] 神鳴敏弘, "協調フィルタリングの課題: プライバシー, サクラ攻撃, 評価値のゆらぎ", 情報処理, Vol.48, No.9 (2007).
- [5] 土方嘉徳, "嗜好抽出と情報推薦技術", 情報処理, Vol.48, No.9 (2007).
- [6] 小川祐樹, 諏訪博彦, 山本仁志, 岡田勇, 太田敏澄, "動的なトピック分類に基づく Novelty を考慮した推薦アルゴリズムの提案", 情報処理学会論文誌, Vol.50, No.6, 1636-1648, (2009).
- [7] 清水拓也, 土方嘉徳, 西田正吾, "発見性を考慮した協調フィルタリングアルゴリズムに関する複数方式の検討", 人工知能学会論文誌, Vol.23, No.5, 330-343, (2008).
- [8] Parag Singla, Matthew Richardson, "Yes, There is a Correlation - From Social Network to Personal Behavior on the Web", www2008, (2008)
- [9] Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", WWW10 (2001).
- [10] 吉田匡史, 伊藤雄介, 沼尾正行, "口コミによる分散型情報推薦システム WAVE を起こそう -Word-of-mouth-Assisting Virtual Environment", MACC2001 論文集(2001).