

ロコミ型情報推薦システムの提案 Proposal of The Word-of-Mouth Recommendation System

保科 一明[†] 吉津 翔太^{*} 武田 利浩[†] 平中幸雄[†]
Kazuaki Hoshina Syouta Kitsu Toshihiro Taketa Yukio Hiranaka

1. はじめに

近年インターネットの普及が進み、社会の中の情報量は膨大になってきている[1]。そのため人々は検索エンジンを利用して自分が好む情報を入手するようになった。しかし検索するにはキーワードが必要であり、そのキーワードが不明確な場合や、そもそも知らない場合には十分な情報収集が行えない。そこでユーザに新しいキーワードを提供する情報推薦への期待が高まっている。

ここで我々は「ロコミ」に注目した。ロコミはその特徴から情報推薦の一種であるといえ、ロコミを用いて情報推薦を行うことで、直接の友人から有効な情報を入手することが可能になる。

本稿では興味のある情報や検索に利用する新しいキーワードの入手手段としてロコミ型情報推薦を提案し、その概要を述べた後、シミュレーションによる評価方法とその結果を示す。

2. ロコミ型情報推薦システム

現実世界におけるロコミを情報推薦の視点で分析すると、(1)情報の入手、(2)伝える相手と理由の判断、(3)情報の伝達、(4)プロフィール(自分や相手の嗜好を表す情報)の更新の4つのステップから成ると考えられる。

ロコミ型情報推薦システムの実現のために、(2)「判断」をどのように行うかが問題になる。我々はこの判断をタグマッチングにより実現したモデル化を試みた。タグとは文章中のキーワードであり、その情報の特徴を表すものである。本稿で提案するロコミ型情報推薦システムのモデルを図1に示す。図1(a)は推薦が推薦されるまでの処理で、図1(b)は情報が推薦された後の処理である。

各ユーザは自分のプロフィールと知人のプロフィールを持っている。自分のプロフィールはユーザの興味・関心のあるタグを登録することで構成される。ここで登録されたタグを「有効タグ」と呼ぶ。知人のプロフィールは情報の交換で追加・更新されていく。本稿で行うシミュレーションでは、友人同士はお互いの興味のある程度把握しているという仮定で、ある程度相手の有効タグが登録してある状態から始める。

- (1)人から聞く、自分で生成するなどして情報を入手する。
- (2)情報についているタグと「自分自身のプロフィール」や「Aが持つBのプロファイル」のタグとマッチングする。
- (3)(2)を受けて情報が推薦される。ここで情報と一緒にタグマッチングの結果を送る。
- (4)で受信者Bが情報を受信する。
- (5)では情報についている「Aが持つBの評価」と「B自身のプロフィール」を比較する。

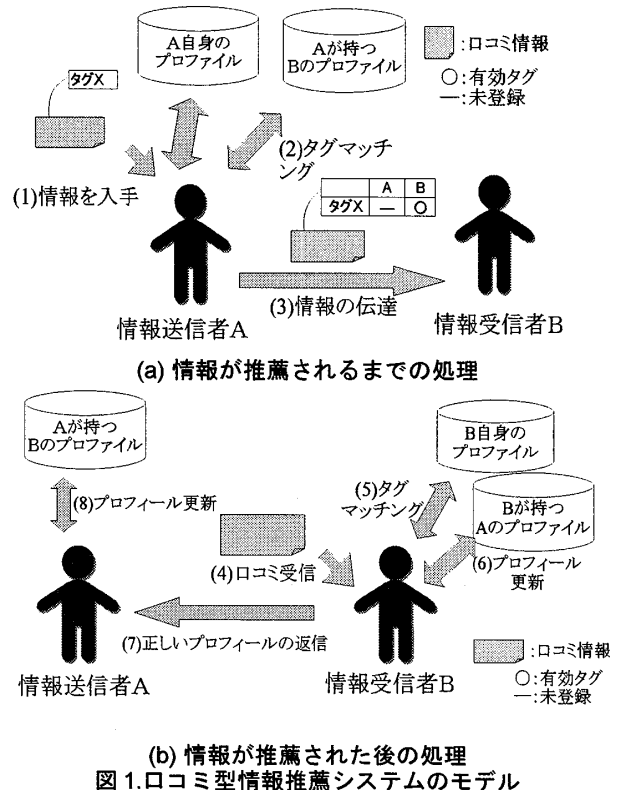


図1.ロコミ型情報推薦システムのモデル

ここで相違があれば、「Aが持つBのプロファイル」に誤りがあるということになるので、(7)で正しい評価を送信し、(8)で相手に更新してもらう。

(6)では情報についている「A自身の評価」と「Bが持つAのプロファイル」を比較する。ここで相違があれば、Aの正しい評価を登録する。

以上のように情報の送受信と同時にプロフィールの更新を行う、というのが本稿で提案するシステムの大きな特徴になる。

3. シミュレーションによる評価

3.1 シミュレーションの目的と方法

提案システムの有効性を検証するために、本稿ではマルチエージェントシミュレーションによる評価を行う。マルチエージェントシミュレーションとは判定アルゴリズムの異なるエージェントを多数設定し、エージェント同士の相互作用をシミュレーションできるといったもので、本研究では Repast Symphony[2]を利用した実験を行う。Repast Symphony はシカゴ大学で開発されたシミュレータで、社会科学系の研究に広く利用されている。

評価項目として、ユーザに Novelty な情報や Serendipity なキーワードを推薦できるかを調べる。ここで Novelty と

[†]山形大学 Yamagata University
^{*}東京コンピュータサービス株式会社 TOKYO
 COMPUTER SERVICE Co.,LTD

はそのユーザの嗜好に一致し、かつ新規性のある内容のことを指す。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、こういったキーワードの入手により検索エンジンに利用を補助できると考えられる。これらの適合率と再現率を求め、評価を行う。

3.2 評価指標

提案モデルの有効性の評価のために、ユーザにとって Novelty な情報や Serendipity なキーワードを用意する。Novelty な情報は、ユーザの有効タグが含まれる情報で、そのユーザがまだ受け取っていない情報とする。Serendipity なキーワードに関しては、まずユーザの有効タグの一部をそのユーザが未知なものとして扱う。シミュレーションによりそのタグを推薦できたとき、Serendipity なキーワードを推薦したと扱う。

・ Novelty な情報についての適合率と再現率

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 $F_N(i)$ 、エージェントの総数を n とすると、本研究で評価に用いる適合率 $P(\text{novelty})$ と再現率 $R(\text{novelty})$ は以下の式で表される。

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

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

・ Serendipity なキーワードについての適合率と再現率

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 なキーワードの集合を $F_{tag}(i)$ 、エージェントの総数を n とすると、本研究で評価に用いる適合率 $P(\text{serendipity})$ と再現率 $R(\text{serendipity})$ は以下の式で表される。

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

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

3.3 データセット

シミュレーションに用いるデータセットに必要な項目は、(1)ソーシャルグラフ、(2)各ユーザのプロファイル、(3)情報である。以下で順に説明する。

(1)ソーシャルグラフはスケールフリーネットワークモデルで表すことができる。本研究ではスケールフリーネットワークの生成モデルのうち、最も代表的な BA モデルを利用してソーシャルグラフを作成した。

(2)プロファイルはランダムな英数字文字列を生成し、それをユーザの有効タグとすることで作成する。ここでソーシャルグラフの各ユーザの嗜好について考えると、嗜好の近いユーザ同士が結びついてソーシャルグラフが形成されていることが考えられる。そこであるユーザの有効タグを隣接ユーザにコピーしていくことで、嗜好の共有を再現している。この共有はユーザの既知の有効タグと未知の有効タグそれぞれ別に行う。

表 1.シミュレーションの条件

α	β	n	t
0.4	0.2	300	120

表 2.シミュレーション結果

P(novelty)	R(novelty)	P(serendipity)	R(serendipity)
88.5%	96.6%	0.6%	61.6%

ここで既知の有効タグの共有率を α 、未知の有効タグの共有率を β とする。

(3)推薦される情報は各ユーザの有効タグをランダムに 2~10 個選択し、組み合わせることで作成する。シミュレーションの評価に必要なのは情報についているタグのみなので、情報の本文などは省略する。

4. シミュレーション結果・考察

シミュレーションの条件を表 1、結果を表 2 に示す。ここで表 1 の n はシミュレーションのエージェント数、 t はシミュレーション時間を表す。

結果を見ると Novelty な情報に関する評価値は共に高い値なので、興味・関心にあった情報を推薦可能と考えられる。一方 Serendipity なキーワードに関する評価値は、再現率に比べて適合率が極端に少ない値となっている。この原因として、今回利用したデータセットの作成方法に問題がある、もしくは意外性のあるキーワードの発生率はこの程度であるといったことが考えられるが、現状では適切な比較対象がないので厳密な判断ができない。

これらの結果の妥当性を検証するための今後の課題として、現実世界のソーシャルグラフと嗜好の関係を分析することや、同じデータを利用して協調フィルタリング [3]を行った場合の式(1)~(4)の結果を今回の結果と比較すること、別のパラメータで作成したデータセットを利用したシミュレーション結果と今回の結果を比較する、といったことが挙げられる。

5. まとめ

本稿では、ロコミに倣った情報推薦システムを提案した。また適合率や再現率を指標にした評価をマルチエージェントシミュレーションにより行った。

結果は Novelty な情報に着目した適合率と再現率はそれぞれ 88.5%と 96.6%で、Serendipity なキーワードに着目した適合率と再現率はそれぞれ 0.6%と 61.6%となった。

今後の課題として、現実世界のソーシャルグラフと嗜好の分析や、今回のデータセットを用いて他の推薦システムを利用したときの結果と本稿の結果の比較を行わなければならない。

参考文献

- [1]神宮敏弘, "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).
- [2]Repast Agent Simulation Toolkit <http://repast.sourceforge.net/>
- [3]Badrul Sarwar, George Karypis, Joseph Konstan, John Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", WWW10 (2001)
- [4]土方嘉徳, "嗜好抽出と情報推薦技術", 情報処理, Vol.48, No.9 (2007).
- [5]吉田匡史, 伊藤雄介, 沼尾正行 "ロ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).