

## FriendRank : SNS における友人推薦システム

FriendRank : Friend Recommender System for SNS

谷川 恭平      大坪 正典      土方 嘉徳      西田 正吾  
 Kyohei Tanikawa    Masanori Otsubo    Yoshinori Hijikata    Shogo Nishida

大阪大学 基礎工学研究科  
 Graduate School of Engineering Science, Osaka University

Recently there is a big growth of social networking service (SNS). This service allows users to introduce their friends to others. We propose a system which provides friend recommendation for SNS. In this system, firstly we focus on the target user's friends on SNS. Then we select their friends on SNS, we mean friends of the target user's friends, as recommendation candidates. This system gives a score to the strength of the relationship between the user and each candidate, considering the link structure and the content of the friend introduction description. We call this score *FriendRank*. Finally it finds person who may be the user's acquaintance or person who may share common interests with the user from the candidates.

## 1. はじめに

近年ソーシャルネットワーキングサービス (以下, SNS) というサービスが爆発的な広がりを見せている. SNS とは, 友人同士のつながりを利用したコミュニティ型の web サービスであり, 登録者は互いに友人を紹介し合うことができる.

本研究ではこの SNS 上で, 新しい友人を推薦するシステムを提案する. 推薦候補はユーザの友人の友人, つまりユーザからの距離が 2 であるユーザとする. 本研究では, 推薦を受けるユーザと推薦候補との関係の強さに対してスコアを付け, このスコアを *FriendRank* と呼ぶ.

*FriendRank* の算出では, まずユーザと友人, 推薦候補の 3 者から構成されるリンク構造を調べる. 次に各ユーザ同士の人間関係を推定し, 同時にこの推定がどの程度正しいかを表す信頼度を求める. そして, リンク構造と人間関係, 推定の信頼度を用いて *FriendRank* を計算する. 人間関係とは 2 者間の関係を表すものであり「学校つながり」や「趣味つながり」などのラベルとして表す.

## 2. トポロジー

まずユーザ周辺のリンク構造について説明する. まず, 推薦を受けるユーザを  $P$ , 推薦候補を  $T$  とする. そして  $P$  と  $T$  に共通の友人の集合を  $M = \{M_1, M_2, \dots, M_n\}$  とする. 我々は  $P$  と  $T$ ,  $M$  から構成される 4 つのトポロジーを定義した (図 1 参照).

まず  $P$  と  $T$  に共通の友人が 1 人だけいる場合をトポロジー 1 とする. また  $P$  と  $T$  に共通の友人が 2 人いて, この 2 人の友人同士の間にリンクがない場合をトポロジー 2 とし, 友人同士の間にリンクがある場合をトポロジー 3 とする. 最後に,  $P$  と  $T$  に共通の友人が 3 人以上いる場合をトポロジー 4 とする.

## 3. ラベルの種類と分類

*FriendRank* の算出では, ユーザ間のリンクに対してラベルを推定する必要がある. このラベルについて, SNS 上の実際の友人紹介文を対象とする調査を行うことで, その種類と分

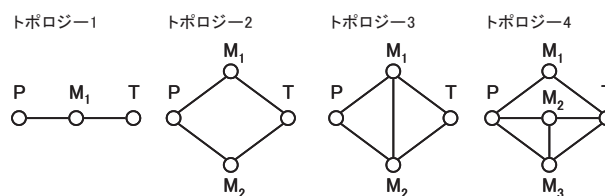


図 1: トポロジーの例

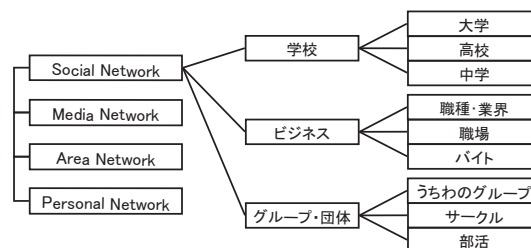


図 2: ラベルの分類の一部

類を決定した. 参加者は 5 人で, 400 件の実際の友人紹介文を調査した.

調査の結果, 第 1 階層として以下の 4 つの大きなカテゴリができた.

- Social Network: ある組織の中での関係を表すラベル
- Media Network: ある媒体 (趣味, インターネット) を介した関係を表すラベル
- Area Network: ある場所やイベントを介した関係を表すラベル
- Personal Network: 個人同士の関係を表すラベル

図 2 にラベルの分類の一部を示す. 図 2 より, まず第 1 階層として上記の 4 つのカテゴリがあり, 第 2 階層, 第 3 階層になるにつれてより細かいラベルに分かれていくのが分かる.

連絡先: 〒 560-8531 大阪府豊中市待兼山町 1-3

大阪大学 基礎工学研究科 西田研究室

TEL:06-6850-6382 FAX:06-6850-6341

#### 4. リンクへのラベル付与

リンクに対するラベルの付与は、機械学習を用いて自動で行う。本研究ではマーケットバスケット分析 [1] を用いている。実際の SNS 上からランダムに抽出した友人紹介文 3,000 件を用いて学習を行うことにより、友人紹介文を各ラベルに分類するためのルールを作成した。この学習により生成されるルールは、「単語  $w$  が出現すれば、信頼度  $x$  でラベル  $L$  である」という比較的単純なものである。このルールにより、リンクに対してラベルとその信頼度を付与することができる。

#### 5. FriendRank の算出方法

トポロジー 1 から説明する。まずラベルの集合を  $\mathbf{L} = \{L_1, L_2, \dots, L_n\}$  とする。そして各ラベル  $L_i$  のパス  $P_{M_1-T}$  に対する信頼度  $P_{PM_1T}(L_i)$  を以下のように定義する。

$$P_{PM_1T}(L_i) = P_{PM_1}(L_i) \times P_{M_1T}(L_i) \quad (1)$$

ただし、 $P_{PM_1}(L_i)$  は  $PM_1$  間のリンクに対するラベル  $L_i$  の信頼度を表す。次に  $\mathbf{L}$  の中で、ラベルの信頼度  $P_{PM_1T}(L_i)$  を最大とするラベルを  $L_R$  とし、このときの信頼度をパススコア  $PS_{PM_1T}$  と定義すると、以下ようになる。

$$PS_{PM_1T} = P_{PM_1T}(L_R) \quad (2)$$

トポロジー 1 の場合は、このパススコアを *FriendRank* とする。

$$FriendRank = PS_{PM_1T} \quad (3)$$

次にトポロジー 2 とトポロジー 3 を考える。まず両トポロジーについてパススコアの和  $SPS_{1,2}$  を定義する。添え字の 1 と 2 は友人  $M_1$  と  $M_2$  に対応している。トポロジー 2 の場合、 $SPS_{1,2}$  は以下の式で求める。

$$SPS_{1,2} = PS_{PM_1T} + PS_{PM_2T} \quad (4)$$

トポロジー 3 の場合、 $SPS_{1,2}$  は以下の式で求める。

$$SPS_{1,2} = PS_{PM_1T} + PS_{PM_2T} + \alpha(PS_{PM_1M_2T} + PS_{PM_2M_1T}) \quad (5)$$

そしてトポロジー 2 とトポロジー 3 の場合の *FriendRank* は上記の  $SPS_{1,2}$  を用いて以下のように計算する。

$$FriendRank = \log_x(SPS_{1,2} + 1) \quad (6)$$

トポロジー 4 では、友人  $\mathbf{M} = \{M_1, M_2, \dots, M_n\}$  の中から 2 人 ( $M_i$  と  $M_j$ ) を選ぶと、 $P$  と  $T$  を含んだ 4 者によってリンク構造が作られ、このリンク構造のパススコアの和は  $SPS_{i,j}$  (ただし  $i \neq j$ ) で表される。トポロジー 4 の場合の *FriendRank* は、この  $SPS_{i,j}$  を用いて以下の式で計算する。

$$FriendRank = \log_x \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n SPS_{i,j} + 1 \right) \quad (7)$$

#### 6. 関連研究

本研究では、推薦を受けるユーザ周辺のリンク構造とコンテンツ (友人紹介文) を用いて、新たな友人を推薦する方法を提案している。

松尾ら [2] は、web 上から人間関係を抽出する研究を行っている。この研究は、研究者間の協働関係 (共著関係, 同研究室関係など) を web ページ上から抽出してくるものである。機械学習のアルゴリズムとして C4.5 を用いて、「単語  $A$  が出れば共著の関係である」というようなルールを生成する。そして Web 上で研究者 2 名の名前を AND で検索し、その検索結果の上位数ページにおいて上記のルールを適用することで、研究者間の協働関係を抽出している。この研究では Web 上の共起関係を利用して研究者同士をリンクで結んでいるが、我々の研究では、すでに存在するリンク構造と、コンテンツを利用して知り合いである可能性のあるユーザの推薦を行う。

また松村ら [3] は、オンラインコミュニティ内でやり取りされたコミュニケーションデータから、参加者のプロフィールを自動的に抽出する手法を提案している。この研究では、ある発言者のコメントに含まれる語の影響量を計算する。コメントに含まれる語の影響量は、このコメントに対する他者からの返信コメントの中で、その語が使われた回数が多いほど大きくなる。そして、このように他者への「語の伝播」に基づいて求めた影響量が大きい語をその発言者のプロフィールとしている。この研究の目的は、オンラインコミュニティを活性化させることにあり、本研究の、知り合いである可能性のあるユーザを推薦するという目的とは異なる。

#### 7. おわりに

本稿では、ユーザと推薦候補との関係の強さを *FriendRank* という指標で表し、これに基づいて推薦を行う手法を提案した。この *FriendRank* の算出では、ユーザ周辺の局所的なリンク構造と、この構造内のリンクに付与されるラベルの信頼度の 2 つを用いる。

本稿ではまずユーザ周辺のリンク構造を 4 つのトポロジーに場合分けした。続いて、構造内のリンクに付与するラベルを分類するために行った調査について述べた。次に、リンクに対してラベルを付与する方法について述べた。そして最後に、各トポロジーに対する *FriendRank* の算出方法について述べた。

今後の課題は、まず *FriendRank* 算出方法の評価実験を行い、その後実際の友人推薦システムを実装を行うことである。

#### 参考文献

- [1] Michael J.A.Berry, Gordon Linoff: Data Mining Techniques, WILEY COMPUTER PUBLISHING (1997)
- [2] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満: Web 上の情報からの人間関係ネットワークの抽出, 人工知能学会論文誌, 20 巻 1 号 E, pp.46-56(2005)
- [3] 松村真宏, 大澤幸夫, 石塚満: 影響の普及モデルに基づくオンラインコミュニティ参加者のプロフィールング, 人工知能学会論文誌, 18 巻 4 号 A, pp.165-172(2003)