

FriendRank : SNS における友人推薦システム

学籍番号 : 90134104 西田研究室 谷川恭平

1 はじめに

ソーシャルネットワーキングサービス (以下, SNS) というサービスが現在広く普及しつつある. SNS は, インターネット上で登録者が互いに友人を紹介し合うことができるサービスである. 本研究では 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

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. 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 人の友人同士の間リンクがない場合をトポロジー 2 とし, 友人同士の間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. 図 1 にトポロジーの例を示す.

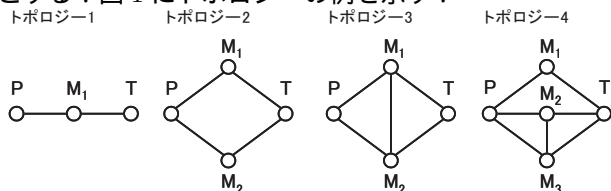


図 1: トポロジーの例

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

これらの分類の一部を図 2 に示す.

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を用いている. 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.

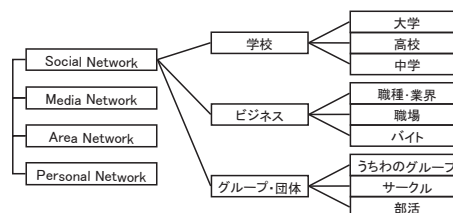


図 2: ラベルの分類の一部

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 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 $M = \{M_1, M_2, \dots, M_n\}$ 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 おわりに

本研究では, SNS 上でリンク構造と友人紹介文のコンテンツを用いて友人推薦を行うシステムを提案した. そしてリンク構造やラベルの調査を行い, *FriendRank* の算出方法を考えた. 今後は, *FriendRank* 算出方法の評価実験とシステムの実装を行う予定である.