

円滑な人脈形成のための SNS 投稿に着目した仲介者探索手法

小松 恭子[†] 中澤 昌美[†] 池田 和史[†] 服部 元[†] 滝嶋 康弘[†]

[†] KDDI 研究所 〒356-8502 埼玉県ふじみ野市大原 2-1-15

E-mail: [†] {ky-komatsu, ms-nakazawa, kz-ikeda, gen, takisima}@kddilabs.jp

あらまし 本稿では、面識のない2者を円滑に引き合わせるための仲介者を探索する手法を提案する。従来研究では、SNS 上のソーシャルグラフを用いることで2者を繋げる直接的な経路を探索する手法が提案されている。しかし、面識のない2者の場合、必ずしも直接的な経路が存在するとは限らない。また、間接的な経路を発見したとしても、実世界における2者間の繋がりが親密でない場合、仲介は非現実的となる。本稿では、これらの課題を解決するため、立場や環境が異なる面識のない2者を円滑に繋げる経路探索手法を提案する。提案手法では、ソーシャルグラフだけでなく、投稿内容から実世界での繋がりを抽出すると共に、投稿行為から2者の親密度合いを推定することが特徴である。提案手法における有効性を検証するため、社内の SNS データを使用して評価実験を行った。

キーワード Web マイニング, ソーシャルネットワーク, 経路探索, テキストマイニング

1 はじめに

本稿では、ビジネスシーンにおいて面識のない2者を円滑に引き合わせるための仲介者を探索する手法を提案する。既存のサービスでは、SPYSEE[1]における人間関係抽出や、Facebook における友人推薦など、Web 上で公開されている情報から繋がりを提示する機能が提供されている。しかし、SPYSEE は関連性の強いユーザを可視化することを主目的としているため、面識のないユーザ同士を引き合わせる用途に最適とは言えない。また、Facebook や他の SNS サイトで見られる友人推薦などにおいては、複数の共通情報を持つユーザを推薦し提示するものであり、趣味や仕事が異なるなど、共通点の少ない者同士の仲介には適さない。

そこで、本稿ではこれらの課題を解決するため、立場や環境が異なる面識のない2者を円滑に繋げる経路探索手法を提案する。図1は社内で想定される一例である。実務上、これまで交流のなかった人と話をする場面は多々ある。研究者であれば、自分の専門外のことは同僚の専門家に聞くのが正確かつ効率よく情報収集できる。しかし、今ひとつ誰の得意分野なのか分からないことや、普段関わりのない部署へ用事があってもアクセス方法が分からないことなどが問題点として挙げられる。そこで、2者 from・to に人名や組織名、技術名などを入力すると、2者を繋げる経路を提示するシステムを想定し、仲介者探索手法を提案する。

2者を引き合わせる従来研究では、SNS 上のソーシャルグラフを用いることで2者を繋げる直接的な経路を探索する手法が提案されている。しかし、想定するサービスでは、2者を円滑に繋げるために、より実務

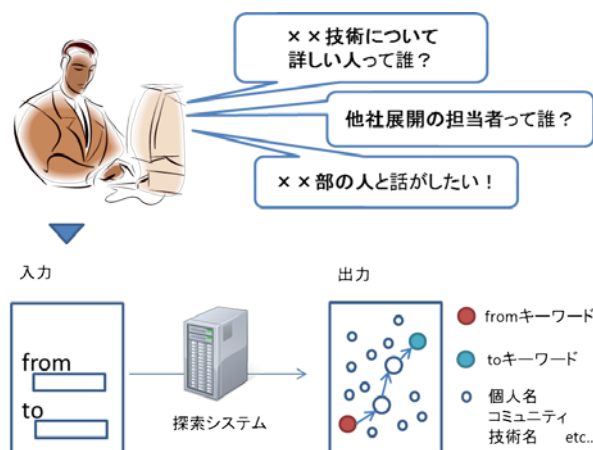


図1 利用シーンの一例

に即した経路や、頻繁にコミュニケーションを取るユーザを仲介する経路を提示する必要がある。

このため、提案手法では、ソーシャルグラフだけでなく、投稿内容から実世界での繋がりを抽出すると共に、投稿行為から2者の親密度合いを推定することが特徴である。提案手法における有効性を検証するため、社内 SNS に蓄積された実活動を表記する週報データを使用して評価実験を行う。

2 関連研究

コミュニケーションを活性化させるために、同空間に居合わせた2者に共通する話題を提供する試みがなされている[2][3]。共通の話題は、文献[2]では予めシステムに登録したユーザ個人のプロフィールや HP な

どの Web サイトを用い、文献[3]では 2 者それぞれのブログや SNS などのソーシャルグラフを用いて個人情報や趣味を表す用語を取得することにより発見する。具体的にはサッカーや料理などが提示される。また、ソーシャルグラフにより共通の知人がいる 2 者を対象に、潜在的な共通話題を推定するものがある[4]。ソーシャルグラフを利用する際、人同士の関係に限らず、社会的なイベントやコミュニティをノードとして含めることにより、直接関わりはないが共通の知人が存在する 2 者に共通する話題を推定する。

しかしながら、文献[2]ではユーザが自ら情報を登録する必要がありシステム利用ユーザに負担がかかる構造となっている。また、文献[3]は、全く異なる環境または立場にある 2 者が居合わせた場合、共通情報が現れず実世界における 2 者の円滑なコミュニケーションが期待できないことを課題としている。また、文献[4]では、共通の知人がいない 2 者の関係は対象としておらず、本稿で目的とする立場や環境が異なる 2 者間の間接的な経路提示は期待できない。

そこで、本稿ではこれらの課題を解決するために、SNS 上のソーシャルグラフ情報に加えて、SNS への投稿内容から実世界での繋がりを抽出すると共に、投稿行為から 2 者の親密度合いを利用ユーザの負担なく推定することで、立場や環境が異なる面識のない 2 者を円滑に繋げる経路探索手法を提案する。

3 提案手法

提案手法の概要を図 2 に示す。本稿では、SNS 上の友人関係、投稿内容および投稿に対するコメントの応答を解析し、SNS を利用する全ての登録者における関係データ D を生成する。関係データとは、各登録者について SNS 上で関わりの強い個人・コミュニティ・技術名などを表す文字列(キーワード)と、当該登録者との関連の強さを表す関連度 c を付与したものを指す。生成された登録者数分の関係データに基づいて 2 者の仲介経路を探索する。経路探索にはダイクストラのアルゴリズム[4]を用いることにより、2 者 $from \cdot to$ に対して、関連度を重みとする重み付き最短経路の提示を可能とする。

以下、3.1 節で従来手法である SNS 上の友人関係解析手法 M_α について、3.2 節で投稿内容解析手法 M_β について、3.3 節で投稿に対するコメント応答解析手法 M_γ について述べ、3.4 節で関連度算出方法について説明する。

3.1 友人関係解析手法

従来手法 M_α として、ソーシャルグラフの繋がりにより仲介経路を探索する手法がある。ソーシャルグラフ

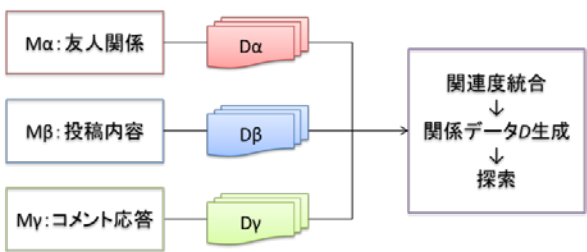


図 2 提案手法の概要

表 1 User1 の関係データ例

	具体例
D_α	$U_2:1.0, U_3:1.0, U_5:1.0, \dots, U_N:1.0$
D_β	A部:0.6, $U_K:0.5$, 対話システム:0.7, S社:0.3, B部:0.5
D_γ	$U_3:1.0$ 20 0.5, $U_5:1.0$ 2 10, ..., $U_K:1.0$ 10 2
D	$U_1(U_2:1.0, U_3:1.0, U_5:2.0, U_K:1.5, \dots, U_N:1.0, A部:0.6, 対話システム:0.7, S社:0.3, B部:0.5)$

表 2 投稿内容解析フロー及び具体例

フロー	具体例
User1の投稿内容	A部K氏の紹介で、対話システム「商品P」に関するプレスリリースに関心を持ったS社のB部より問い合わせを受けたため、個別のデモ・説明を行った。
重要語抽出	A部、K氏、対話システム、商品P、プレスリリース、S社、B部、デモ
辞書	社名、部署名、人名、技術名
キーワード選択	A部、K氏、対話システム、S社、B部

上の繋がりの有無により、繋がりがあれば関係データ D_α に列挙する。表 1 に User1 の関係データ D の例を示す。User1 が User2、User3、User5、UserN と SNS 上で友人関係にある場合、表 1 の関係データ D_α のように列挙される。関連度 c_α は一定の値とする。

3.2 投稿内容解析手法

提案手法 M_β として、投稿内容により投稿者と関わりのあるキーワードを抽出し、関係データ D_β を生成する。

投稿内容解析フロー及び具体例を表 2 に示す。まず SNS 上の投稿を収集し、TFIDF を用いて投稿内容より重要語句を複数語抽出する。次にユーザ関連情報として扱える語句を登録した辞書に基づき、抽出した重要語句から投稿者と関係のある語句を分別して選択する。辞書にはユーザ関連情報として扱える語句として、社名、部署名、個人名、技術名など実務に即した投稿者と関係のある語句を登録する。この際、表記のゆれについては、互いに表記の揺れの関係にある語を同一の語として扱える仕組みを用いる。選択されたキーワードは表 2 の関係データ D_β のように列挙する。関連度 c_β には TFIDF 値を用いる。このとき投稿者ごとに投稿

内容を解析することで SNS 登録者ごとの関係データ D_β を生成することが可能である。

3. 3 コメント応答解析手法

提案手法 M_γ として、SNS 上の投稿に対するユーザ間のコメント応答の解析を行う。まずコメントのやり取りがあるユーザを表 1 の関係データ D_γ のように列挙する。関連度 c_γ は一定の値とする。また関係データ D_γ については、コメント頻度 f (回) とその応答時間 t (時間) の平均値を算出し、記録する。User1 について User3 とコメントのやり取りをしており、コメント回数が 20 回、その平均応答時間が 0.5 時間であった場合の例を表 1 の関係データ D_γ に示す。

3. 4 関連度算出方法

探索に使用する登録者ごとの関係データ D は関係データ D_α 、 D_β 、 D_γ により生成される。関係データ D の関連度 c は、関係データ D_α 、 D_β 、 D_γ の関連度 c_α 、 c_β 、 c_γ 、およびコメント頻度 f 、応答時間 t の 5 つを用いて、(1) 式により算出する。

$$c = (\alpha c_\alpha + \beta c_\beta + \gamma c_\gamma) * \begin{cases} \gamma_1 \frac{1}{f} + \gamma_2 \frac{t}{t_{Max}} & (c_\gamma = 0 \text{ のとき}) \\ \gamma_3 & (c_\gamma = 1 \text{ のとき}) \end{cases} \quad (1)$$

(1) 式において、 f はコメントのやり取りがある場合のみの値であるため 1 以上の値を取る。このため $1/f$ は $(0,1]$ となる。また、平均応答時間の基準値 t_{Max} は使用データにより決定する値である。ここでは、基準値 t_{Max} を、全コメント平均応答時間の 90% を含む値とし、基準値より応答時間の長いものは全て $t = t_{Max}$ として扱い、 $t/t_{Max} = [0,1]$ となるようにする。また、変数 α 、 β 、 γ は各手法に重み付けを行う際に用いる。例えば実関係に基づく繋がりを重要視した経路提示を期待する場合、 β を α 、 γ より大きい値にすることが考えられる。また、変数 γ_1 、 γ_2 は仲介時に求める項目により重み付けを行う際に用いる。例えばスピードを重要視した経路提示を期待する場合、 γ_2 を γ_1 より大きい値に設定する。変数 γ_3 は γ_1 、 γ_2 の値を考慮して設定する。例えば $\gamma_1 = \gamma_2 = 1$ としたとき、 $1/f = (0,1)$ かつ $t/t_{Max} = [0,1]$ であるため $\gamma_1/f + \gamma_2 * t/t_{Max}$ の値は $(0,2]$ とする。これを踏まえて γ_3 の値を 2 以上にするなどして、コメントのやり取りがある場合とそうでない場合で関連度に差をつける。

4 性能評価実験

表 3 手法ごとの経路数と平均ホップ数
(従来手法： M_α 提案手法： $M_{\alpha+\beta}$ 、 $M_{\alpha+\gamma}$ 、 $M_{\alpha+\beta+\gamma}$)

ホップ数	M_α	$M_{\alpha+\beta}$	$M_{\alpha+\gamma}$	$M_{\alpha+\beta+\gamma}$
1	2190	2383	1351	1574
2	26850	26500	12633	13668
3	33374	31911	27157	27250
4	4274	5859	20248	19351
5	136	647	4966	4815
6	—	33	457	619
7	—	3	12	58
8	—	—	—	7
経路無し	1036	518	1036	518
平均ホップ数	2.60	2.64	3.24	3.17

表 4 手法ごとの主観評価実験結果平均値

		M_α	$M_{\alpha+\beta}$	$M_{\alpha+\gamma}$	$M_{\alpha+\beta+\gamma}$
Q1: 応答の迅速さ	[-1,1]	0.48	0.46	0.66	0.53
Q2: 仲介者との親密さ	[-1,1]	0.40	0.42	0.40	<u>0.24</u>
Q3: 経路の適切さ	[0,4]	2.53	2.48	2.62	2.57

本章では、関連研究において問題点として取り上げた、実務に即した経路が選択されない点と普段からコミュニケーションをとる経路が提示されない点について、従来手法と提案手法の比較を行う。

4. 1 実験方法

本稿では、企業の社内 SNS データを対象に解析を行った。登録者数は 261 人、収集期間は 2011 年 3 月から 2013 年 4 月を利用した。投稿内容解析手法に用いる辞書登録語は人物名のみを対象とした。収集期間の対象データをもとに 応答時間の基準値 $t_{Max} = 72$ (時間) とし、本実験では各手法を均等に扱い $\alpha = \beta = \gamma = 1$ 、 $\gamma_1 = \gamma_2 = 1$ 、 $\gamma_3 = 3$ として関連度を計算した。

ソーシャルグラフのみを用いる従来手法を M_α 、 M_α に加えて投稿内容に基づく関係を用いる提案手法を $M_{\alpha+\beta}$ 、同様に投稿に対するユーザ間のコメントの有無及びその頻度と応答時間を考慮した提案手法を $M_{\alpha+\gamma}$ 、全てを合わせた提案手法を $M_{\alpha+\beta+\gamma}$ とする。

まず、登録者 261 人における全経路を探索する定量評価を行う。従来手法では発見できなかった実務上の経路について、提案手法で発見することが可能か検証する。

次に、全経路探索により提示された経路が普段のコミュニケーションを考慮した円滑なものであるかについて、被験者による主観評価実験を行う。被験者 (from) から被験者と直接的な友人関係にない人物 (to) につい

て、各手法を用いた経路を提示する。仲介者が増すほど、提示された経路を被験者が正確に評価することが難しくなるため、本稿では 2 ホップ、すなわち仲介者が 1 人となる経路を対象とする。また、比較する各手法で最短経路のホップ数が同一である経路のみを選択する。対象 SNS 登録者 3 名に対し、条件を満たす合計 45 経路について、評価実験を行う。また、to は自身と直接関わりのない人の話に耳を傾ける人ではないため仲介者が必要であることを前提とした。

被験者は、提示された仲介者に関する 2 つの質問、Q1：迅速な対応をしてくれそうであるか、Q2：実際に良く関わる人であるかどうか、に対し 3 段階評価（思わない・1・普通 0・思う 1）を行う。また、Q3：仲介人の適切さを 4 段階でポイント付けする、ことにより各手法の有効性を比較する。

4. 2 実験結果

4. 2. 1 定量評価

経路探索を行った結果を表 3 に示す。表 3 は、各手法における、ホップ数別経路数と平均ホップ数である。

ホップ数 1 に注目すると従来手法 M_α の 2190 に対し、提案手法 $M_{\alpha+\beta}$ は 2383 と増加した。これは、実務に即した経路を新たに発見することが可能になったことにより、1 ホップで到達可能な経路が増加したことを表し、提案手法における投稿解析手法が有益に作用したことを示している。一方、提案手法 $M_{\alpha+\gamma}$ は 1351 に減少した。これは、SNS 上では友人関係だけでは実世界での繋がりがカバーできていないことに加え、SNS 上では友人関係にある人であっても SNS 上でのやりとりは低頻度なため、第 3 者を経由する 2 ホップ以上のより円滑と思われる経路が提示されていることを表す。

また従来手法 M_α では経路が発見できなかった 1036 パターンに対し、提案手法 $M_{\alpha+\beta}$ ではその半分の 518 パターンにおいて、実務に基づく繋がりが考慮されたことにより経路を発見することができた。一方、提案手法 $M_{\alpha+\gamma}$ では新たな経路を発見することができなかった。これはコメントを行うには記事を閲覧するために友人関係にあることが前提となる SNS の性質を表していると考えられる。また、人物のみを取り扱った実験では、ホップ数を増加させても経路を発見できないケースが存在するため、辞書登録には個人名だけでは不十分だということがわかる。具体的には、辞書に部署名や技術名などの情報を追加することで、投稿には、特筆していない個人に対して、実務上の関わりを推測することにより提案手法 M_β が作用し、より実際に関係のある人を結ぶ経路を発見可能となる。

一方、平均ホップ数に着目すると従来手法 M_α 2.60 に対し、提案手法 $M_{\alpha+\beta}$ では 2.64、 $M_{\alpha+\gamma}$ では 3.24、 $M_{\alpha+\beta+\gamma}$ で

は 3.17 と増加した。これは、提案手法では、従来手法 M_α では密な関係ではない仲介者を経由していた経路が実際の関係に基づいて仲介者を多く通した経路に変更されたためと考えられる。

4. 2. 2 主観評価

結果を表 4 に示す。Q1、Q2 については 3 段階評価（思わない・1・普通 0・思う 1）を実施したため、数値が大きい方が良い評価と言える。Q1「迅速な対応をしてくれそうであるか」については従来手法 M_α 0.48 に対し、提案手法 $M_{\alpha+\gamma}$ 0.66 が最も高い評価を得た。また、Q2「実際に良く関わる人であるかどうか」については従来手法 M_α 0.40 に対し、提案手法 $M_{\alpha+\beta}$ 0.42 と良い評価を得たが、提案手法 $M_{\alpha+\beta+\gamma}$ は 0.24 と減少している。これより、提案手法 M_γ は、従来手法 M_α では考慮されていない対応スピードを考慮した仲介者の提示が来ていると考えられる。同時に提案手法 M_β は、従来手法 M_α では得られない実関係を補足し得ると考えられる。しかしながら、それらを組み合わせた提案手法 $M_{\alpha+\beta+\gamma}$ において高評価を得られなかったのは、本実験において各手法を均等に扱い $\alpha = \beta = \gamma = \gamma_1 = \gamma_2 = 1$ としたためであり、実際の仲介提示を行う際に重要とする項目、すなわち対応スピードや実業務での関わりの強さなどを配慮した重み付けが必要であると考ええる。

また Q3 により本実験において提示経路が仲介に最も適している手法は $M_{\alpha+\gamma}$ となった。

5 まとめ

本稿では、立場や環境が異なる面識のない 2 者を円滑に繋げる経路探索手法を提案した。提案手法は、ソーシャルグラフだけでなく、投稿内容から実世界での繋がりを抽出すると共に、投稿行為から 2 者の親密度合いを推定することが特徴である。性能評価実験によって、提案手法 M_β は、従来手法では発見できなかった実世界のつながりに即した経路を新規に発見可能であることが確認された。また、被験者による主観実験の結果、提案手法 M_γ によって提示された仲介経路が従来手法 M_α よりも適した仲介者を提示できたと言える。

しかしながら、提案手法 M_β と M_γ を組み合わせた手法 $M_{\alpha+\beta+\gamma}$ では各々の利点を活かせず良い結果が得られなかった。被験者による主観評価のコメントより、仲介者は世代が近いことが望ましいことや、確実にコネクションすることが重要なのかスピードを重要とするのかによって求める仲介者が異なることが判明した。これより、今後の課題として仲介時の目的に応じて各手法に重みを付与した関連度の算出が不可欠であり、各手法の重みの最適値を設定する必要があると言える。

また、投稿内容における人物名は姓のみの表記が多い。今回同姓名については一番環境に近い人物を割り当てたが、同姓同名の区別を行う仕組みを用いる必要があると考える。また、文献[6]において要求情報を見せることにより普段あまり行動をともしないような疎な関係にある知人との間での会話が促進される可能性があると言及されている。このため疎な知人に対しては共通情報を提示するなどの仕組みを用いることで別の経路を発見可能であると考ええる。

参 考 文 献

- [1] SPYSEE: <http://spysee.jp/>
- [2] 岡本昌之, 中西英之, 西村俊和, 石田亨, “Silhouettell: 実空間での出会いにおけるアウエアネス支援”, マルチメディア、分散、協調とモバイルシンポジウム論文集, pp.701-708, 1998.
- [3] 藤本義治, 星亮輔, 高宮浩平, 井口真朝, 岡本誠, 松原仁, “MAKOTO: ソーシャルグラフを用いたコミュニケーション支援システムの提案”, 情報処理学会シンポジウム論文集, no. 3, pp.703-706, 2011.
- [4] 與語一史, 新熊亮一, 笠井裕之, 山口和泰, 高橋達郎, “人同士の潜在的な共有関係を推定するためのソーシャルグラフ表現方法”, 電子情報通信学会技術研究報告, ネットワークシステム, vol. 110, no. 372, pp. 81-84, 2011.
- [5] E. W. Dijkstra, “A Note on Two Problems in Connexion with Graphs”, Numerische Mathematik, vol. 1, no. 1, pp. 269-271, 1959.
- [6] 松田完, 西本一志, “HuNeAS: 大規模組織内での偶発的な出会いを利用した情報共有の促進とヒューマンネットワーク活性化支援の試み”, 情報処理学会論文誌, vol. 43, no. 12, pp.3571-3581, 2002.