

ソーシャルネットワークとコラボレーション支援

Social Network and Collaboration Support

森 純一郎*¹
Junichi-Lo Mori

松尾 豊*²
Yutaka Matsuo

橋田 浩一*²
Kôiti Hasida

石塚 満*¹
Mitsuru Ishizuka

*¹東京大学

The University of Tokyo

*²産業技術総合研究所

National Institute of Advanced Industrial Science and Technology

A social network receives much attention in computer science, especially in Web technologies. Blogs and SNSs become popular world-wide. At the same time, collaboration networks such as coauthorship and user communication are analyzed in several research. In this paper, we address the our current technologies related to social network and the future development in a ubiquitous environment.

1. はじめに

近年、ソーシャルネットワーク（社会ネットワーク）が注目を集めている。SNS（ソーシャルネットワーキングサービス）は昨年から、Blog は一昨年から急激に日本でもユーザ数を伸ばした。ユーザは自分の身近にいる人が何に興味を持っているか、何を考えているのか気になるものであり、情報収集や情報の整理といった点からも、ユーザの関係性に着目することは重要な方向性であろう。

一方で、近年では、ユビキタス情報環境に関する研究が盛んに行われている。サイバースタディーズ研究センター [中島 01] では、「今、ここで私が」使える情報やサービスの提示をめざし、さまざまな研究が行われてきた。RFID や携帯機器をはじめとしたユビキタスコンピューティング技術が進展する中で、我々は特に人の関係性に基いたアプローチを模索している。

ユーザにとって価値が高い情報を提示するには、ユーザの性質をどう捉えるかが鍵となる。ユーザモデリングと呼ばれる分野では、ユーザをどうモデル化するかについてさまざまな手法があるが、我々は、情報提供において重要なユーザの性質の中には、ユーザ個人の性質に還元できず、ユーザ相互の関係性の性質として捉えなければならないものがあるのではないかと考えている。社会学者の安田氏はその著書のなかで次のように述べている。

『自分探し』という言葉は、私はどうしても好きになれない。この言葉は、職業選択、結婚など、人生で大切な何かとのかかわりを決定するときに、自分の内面を見つめる行為に対して使われることが多いようである。だが、自分とは自分の中に存在するものなのだろうか。自分というものは、個人が孤独に己自身について考えることによって見出せるものではないように私には思える。人が誰であり、何であるのかは、他者の力を借りずに知ることはできない。（中略）人であれ組織であれ、そのアイデンティティは他者の認知に依存する部分が必ずあるのだ（人脈作りの科学 [安田 04] より）

本稿では、我々のソーシャルネットワークに関する研究を紹介するとともに、ユビキタス情報環境におけるコラボレーション

支援のために、ソーシャルネットワークに着目したアプローチの可能性について議論する。

ソーシャルネットワークといっても、どのようなコミュニティのソーシャルネットワークをどのような情報源から何のために構築するかによって、さまざまな形があり得る。ユビキタスコンピューティング技術をどのように利用するかという観点からは、一例として次のような可能性が考えられるだろう。

- ソーシャルネットワークを獲得するためにユビキタス情報技術を利用する。例えば、[西村 04] などである。これまで、Web 上の情報や E-mail を利用する [Adamic 03]、Blog の情報を利用する [古川 05] などの研究があった。センサ情報を組み合わせることで、そのコンテキストにおける関係性の情報を、特に匿名で得ることも重要な技術である。
- 得られたソーシャルネットワークを、実世界においてユーザの情報提示に利用する。例えば、ソーシャルネットワーク上で近い人で、かつ近くにいる人の講演に関する情報を提示するなどである。その試みのいくつかは、イベント空間情報支援プロジェクト [西村 03] で行われている。

以下では、それぞれの一例として、位置情報からソーシャルネットワークを抽出する試みについて、および学会でのコラボレーション支援のために、人に適切なメタデータを付与する試みについて述べる。

2. 位置情報を利用したソーシャルネットワークの抽出

イベント空間情報支援プロジェクトでは、2003 年から人工知能学会の全国大会で学会の参加者の学会参加やコミュニケーションを活発にするための学会支援システムを構築・運用している。我々は特に、研究者のソーシャルネットワークに着目した「人間関係ネットワーク支援システム」を構築してきた。イベント空間情報支援プロジェクトでは、学会会場内にセンサを配置し、ユーザの位置を表示する、会場内で知り合いを検索するサービスなどを行った。JSAI2003 では、34 個のセンサを配置し、400 個のカード型端末 (CoBIT) を配布した。

この位置情報を利用すると、誰と誰が出会ったかという「出会いネットワーク」を作ることができる。これによって、学会での交流が活発に行われているか、また、知り合いで今回話をしていない人がどこにいるかなどのサービスにつなげることができるだろう。

連絡先: 森 純一郎、東京大学大学院情報理工学研究所, 東京都文京区本郷 7-3-1, 03-5841-6755, jmori@miv.t.u-tokyo.ac.jp

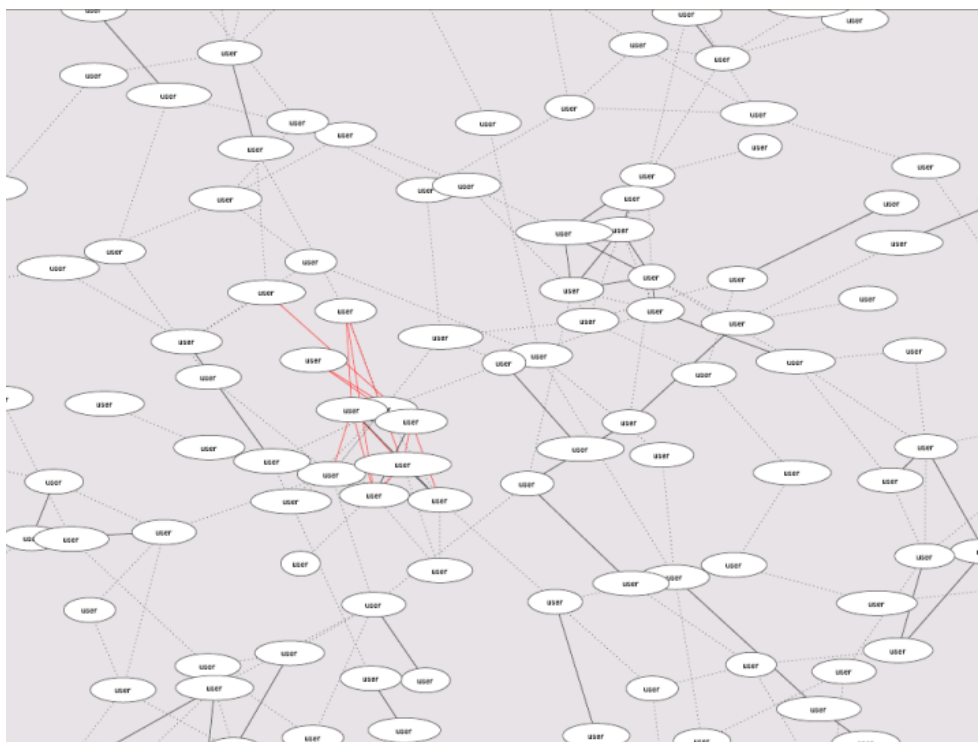


図 1: 位置情報からのソーシャルネットワーク

位置情報からの出会いネットワークの構成は次のように行う。まず、「出会う」ことを「同一のセンサで 30 秒以内に検出されること」と定義する。そして、次のように出会い Jaccard 係数を定義する。

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|}$$

$|X \cap Y|$ は、X さんが検知された 30 秒後までに Y さんが検知された、もしくは Y さんが検知された 30 秒後までに X さんが検知された回数であり、 $|X \cup Y|$ は、X さんまたは Y さんがセンサーに検知された回数である。30 秒という値はパラメータである。 $J(X, Y)$ が閾値を超えればエッジを張ることで、ネットワークを形成することができる。

これを示したものが図 1 である。次数の極端に多いノードがなく、比較的均等にエッジが張られていることが分かる。

JSAI2003 の後に行ったアンケート調査では、表 1 のような結果が得られた。アンケート項目のうち「JSAI2003 で会った」にチャックをしたペアを正解として、精度を再現率を示したものである。これによると、再現率が非常に低くなっているが、これはある程度やむを得ないと考えられるだろう。もちろん、学会の会場外で話をすることもあるし、センサーのないところで話をすることも普通である。しかし、着目したいのは精度の高さで、同じセンサーで 30 秒以内に検出されたという簡単な指標だけでも、それが統計的に値が高ければその 2 人が会っている確率が高いということである。これを利用して、話した人の論文や関連の情報が後から送られてくるなどのサービスに生かすことができると考えられる。

表 1: 出会いネットワークの精度・再現率

閾値	Precision	Recall
0.1	92.3% (12/13)	2.9% (12/416)
0.035	84.8% (28/33)	6.7% (28/416)
0.001	76.4% (68/89)	16.3% (68/416)

3. 人物や関係性のメタデータ付与

研究者のコラボレーションを支援するための学会支援システムにとっても、また一般の SNS や Blog などのシステムにとっても、Web 上の情報をユーザが利用しやすい形に加工しておくことは重要な課題のひとつである。例えば、ある情報に関して詳しい人を探したいといったときに、人に対して適切なメタデータを付与することが必要になる。このため、我々は Web 上の情報を用いて、研究者の情報をメタデータとして自動的に抽出する手法を研究している。ここでのメタデータとは、研究者の活動を表すために重要なキーワードであり、例えば、所属組織、研究テーマ、共著者、プロジェクト名である。これらは FOAF [Brickley 04] のようなオントロジーを用いて記述可能である。

メタデータ抽出手法の基本的なアイデアは、Web ページにおける語の共起に着目し Web 上で氏名と特に強く共起する語をその人物のメタデータとするものである。ある人物の名前とよく共起する語は、その人物が何で知られているかというメタデータになるであろうし、その人物の書いた Blog に良く含まれる語は、その人物が何に興味を持っているかのメタデータになる。

我々は、氏名と語の共起の強さを検索エンジンを用いて求める。提案手法では語の共起の強さを表す指標として、Jaccard 係

数を用いる。例えば「石塚満」と「高速推論」の共起の強さは、「石塚満」を検索クエリーとしたときの検索ヒット数を $|A|$ 、「高速推論」を検索クエリーとしたときの検索ヒット数を $|B|$ 、「石塚満 AND 高速推論」を検索クエリーとしたときの検索ヒット数を $|A \cap B|$ とすると、Jaccard 係数 $|A \cap B| / (|A| + |B| - |A \cap B|)$ で測ることができる。Jaccard 係数が大きいほど、2つの集合の重なりが大きく、関連が強いことになる。

さらに、目的に沿ったキーワード抽出を行うために、コンテキストワードと呼ぶ語を用いる。コンテキストワードは、抽出するメタデータの文脈を特定するものであり、文脈に合致しない語を除外する働きがある。例えば、ある人物の人工知能に関連する活動を知りたい場合、コンテキストワードを「人工知能」とし、その人物の氏名と強く共起する語の中でも、特に「人工知能」と共起する語だけに絞り込むことにより、「人工知能」に関連した語を抽出する。

メタデータの抽出法は以下の通りである。

- まず、検索エンジンを利用して人物の検索を行い、検索結果の上位 Web ページを取得する。
- 取得した Web ページに対して html タグの除去および形態素解析を行い、Termex^{*1} を用いて用語を抽出する。
- Web ページから抽出した語と名前との共起の強さを Jaccard 係数を用いて求める。氏名 n を含む Web ページ集合を N 、語 w を含む Web ページ集合を W として、 n と w の単独のヒット数および n と w の AND 検索のヒット数をそれぞれ $|N|$ 、 $|W|$ 、 $|N \cap W|$ として表す。この時、Jaccard 係数 $J(n, w)$ は次のように計算できる。

$$J(n, w) = \frac{|N \cap W|}{|N| + |W| - |N \cap W|}$$

次にコンテキストワード c と語 w の共起の強さを、同様に求める。 c を含む Web ページ集合を C とすると、 $J(c, w)$ は次のように計算できる。

$$J(c, w) = \frac{|C \cap W|}{|C| + |W| - |C \cap W|}$$

語 w のコンテキストワード c のもとでのスコア $Score(n, c, w)$ を、次のように与える。

$$Score(n, c, w) = J(n, w) + \alpha J(c, w)$$

$Score(n, c, w)$ が高いほど語 w は人物 n と関連の深いメタデータとなる。

なお、 α はコンテキストワードの Jaccard 係数の重みであり、コンテキストの影響度合いを調整するものである。 $J(n, w)$ が小さいにもかかわらず $J(c, w)$ が大きくなるような語を除くために $J(n, w)$ がある閾値 k_0 以下となる語はキーワードから除外する。また、 $|W|$ や $|N \cup W|$ などが閾値 k_1 以下の場合はキーワード除外する。重みや閾値はヒューリスティックに与えている。

コンテキストワードを「人工知能」とした場合の石塚満氏のキーワードを表 2 および 3 に示す。ここで、 $\alpha = 1$ とし閾値は、それぞれ $k_0 = 0.001$, $k_1 = 6$ とした。抽出されたキーワー

表 2: コンテンツワード「人工知能」に対する「石塚満」の上位語とスコア

人工知能学会	0.170
岡崎直観	0.081
松尾豊	0.070
土肥浩	0.065
キャラクターエージェント	0.065
自然言語	0.055
擬人化インタフェース	0.053
人工知能学会誌	0.048
擬人化エージェント	0.047
自然言語処理	0.044
東京大学工学部電子情報工学科	0.044
高速仮説推論	0.043
黒橋禎夫	0.040
仮説推論	0.036
マルチモーダルメディア	0.034
谷内田正彦	0.033
高速仮説推論システム	0.033
人工知能学会全国大会	0.032
空間共有コミュニケーション	0.031
仮説推論システム	0.031
マルチエージェント	0.030

表 3: 「松尾豊」と「石塚満」の関連語

松尾豊-石塚満
石塚満
松尾豊
産業技術総合研究所-
サイバーアシスト研究センター
人工知能
石塚研究室
岡崎直観
土肥浩
大澤幸生
橋田浩一
松村真宏
人間関係ネットワーク

ドを見ると人物名、組織名、研究内容などに関する多様な語が抽出されていることがわかる。

提案手法は、特定のコンテキストにおける人物のメタデータを抽出するものであるが、コンテキストとして人物の氏名を用いることもできる。この場合、メタデータは 2 者間の関係を表す語となる。例えば、表 2 は松尾豊氏と石塚満氏の関連語である。共通の研究室名や両者をつなぐ人物名が得られている。

図 2 は抽出されたメタデータをもとに作成した研究者の FOAF ファイルの例である。例えば、人物名、組織名、研究テーマ、研究プロジェクトなどは foaf:knows, foaf:current-organization, foaf:interest, foaf:currentproject などの属性と対応付けられる。

この手法は、研究者のつながりに基づいて研究者を検索する研究者ネットワーク検索エンジン POLYPHONET (図 3) に利用されている。

```
<foaf:Person>
<foaf:mbox rdf:resource="" />
<foaf:name>石塚満 </foaf:name>
<foaf:interest
  rdfs:label="擬人化エージェント" rdf:resource="" />
<foaf:currentProject
  rdfs:label="マルチモーダル擬人化インタフェース" rdf:resource="" />
<foaf:workplaceHomepage
  rdfs:label="東京大学" rdf:resource="" />
<foaf:knows>
<foaf:Person>
<foaf:mbox rdf:resource="" />
<foaf:name>松尾豊</foaf:name>
.....
```

図 2: 抽出されたキーワードを利用した FOAF ファイルの例

*1 東京大学中川研究室、横浜国立大学森研究室で開発された用語抽出システム。http://gensen.dl.itc.u-tokyo.ac.jp/win.html



図 3: 研究者検索システム

4. おわりに

学会において、一体誰と誰が出会えば素晴らしい研究につながるのか、自分が一体誰と出会えばよいのかは難しい問題である。社会ネットワーク分析の分野では、例えば、ネットワークのノードがどのような影響力を持っているのか、どのようなクラスタ構造があるのかなどを分析することができるが、全体としてどのようなネットワークを作ればそのコミュニティの目的に寄与するのは、対象となるドメインや目的に深く依存している。

我々は生活の多くの部分を他者との関係性の中で過ごしている。ユビキタス情報環境においても、ユーザの関係性はユーザにとって重要な要素であろう。

参考文献

- [Adamic 03] Adamic, L. A. and Adar, E.: Friends and Neighbors on the Web, *Social Networks*, Vol. 25, No. 3, pp. 211–230 (2003)
- [Brickley 04] Brickley, D. and Miller, L.: in *FOAF: the 'frined of a friend' vocabulary*, <http://xmlns.com/foaf/0.1/> (2004)
- [中島 01] 中島, 橋田, 森, 伊東, 本村, 車谷, 山本, 和泉, 野田: 情報インフラに基づくグラウンディングとその応用 – サイバーアシストプロジェクトの概要 –, コンピュータソフトウェア, Vol. 18, No. 4, pp. 48–56 (2001)
- [西村 03] 西村 拓一, 橋田 浩一, 中島 秀之: イベント空間支援プロジェクト, 人工知能学会全国大会, No. 3E1-01 (2003)
- [安田 04] 安田 雪: 人脈づくりの科学, 日本経済新聞社 (2004)
- [古川 05] 古川 忠延, 松澤 智史, 松尾 豊, 内山 幸樹, 武田 正之: Weblog におけるユーザの繋がりと閲覧行動の分析, 電子情報通信学会論文誌 (2005)
- [西村 04] 西村 拓一, 中村 嘉志, 松尾 豊, 坂本 和彌, 宮崎伸夫: 高密度で存在する多数オブジェクトのトポロジー推定, 第 5 回計測自動制御学会 システムインテグレーション部門講演会 (2004)