

位置情報と人間関係ネットワーク

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

E-mail: y.matsuo@aist.go.jp

1. はじめに

近年、多くのコンピュータやセンサが環境や機器に埋め込まれ、多様な情報通信インフラがシームレスに接続されるユビキタスネットワークを実現するための研究が行われている。特に、ユーザの好みや行動履歴に基づきサービスをカスタマイズするパーソナライズドシステム技術は重要な研究課題である。産業技術総合研究所サイバーアシスト研究センターでは、センサやアクチュエータが環境中に配置された、ユビキタス環境のもとでの情報支援という研究を行っており、特に私は、人と人との関係に着目した情報支援の研究を行っている。本報告では、Webからの人間関係ネットワークの抽出と、位置情報からの人間関係ネットワークの抽出について述べる。

2. Webからの人間関係ネットワークの抽出と評価

2.1. 手法の概要

近年、Web上にある多様で大量の情報から、隠された構造や重要な情報を見つけ出すWebマイニングの研究が盛んに行われている。Webのリンク関係から重要なページを発見したり (Brin & Page, 1998)、あるトピックに関するWeb上のコミュニティ (Kleinberg, 1998; Kumar, Raghavan, Rajagopalan, & Tokins, 1999) を発見する、特定の2人の人間をつなぐ知り合い関係のパスを抽出したり (Kautz, Selman, & Shah, 1997)、参照の1共起性からコミュニティを発見する (村田, 2001)、またあるページの評判情報を抽出する研究 (Rafiei, Mendelzon, 2000) など、さまざまな研究が行われている。このような研究のひとつとして、我々は、名前の共起関係から特定のコミュニティにおける社会ネットワークを取り出す研究を行っている。詳細は第1回の友部による報告を参照いただきたいが、基本的には、「松尾豊 安田雪」などのように名前のANDをとって検索エンジンにクエリーを出し、そのヒット件数(すなわち、両者の名前を含むWebページの数)をもって関係の強さの推測値とするものであ

る。

また、社会的関係の種類として、研究分野に特有の次のようなクラスを定める。これらが、ネットワークのエッジに付与されるラベルの種類となる。

共著関係 共著の論文がある関係。

同研究室関係 同じ研究室や研究所のメンバーなど所属が同じである(あった)関係。

同プロジェクト関係 同じプロジェクトや委員会など、組織をまたがる同グループに所属している(いた)関係。

同発表関係 同じ研究会で発表する(した)関係。

これを、ヒットしたページのさまざまな属性から判別ルールにより自動で判別する。

2.2. JSAI2003 におけるサービス

我々は、第17回人工知能学会全国大会(JSAI2003)において、得られたネットワークを会場内に設置されたKIOSK端末およびWeb上で表示するサービスを行った。図1はその様子である。2003年および過去4年の人工知能学会で発表した著者、共著者は合計1560人であり、そのなかで単独でのヒット件数が30件に満たない、もしくは他のノードとのSimpson係数が閾値(0.5)以下で孤立ノードとなってしまうノードは除外し、結果的に266ノードとなった。

ネットワークは、SVGで出力され、SVG viewerにより閲覧することができる。SVGは、W3Cによって作成された規格であり、ベクトル表現によるXML形式のグラフィック記述言語である。Javascriptが埋め込まれているので、ノードをドラッグしてつながり具合を確かめ

図1



ることができる。各ノードには丸印と星印のアイコンがあり、それぞれスケジューリング支援システム、CoBITによる位置情報表示システムと連携している。（各システムの概要は西村, 濱崎, 松尾, 大向, 友部, 武田, 2004 参照。）エッジラベルとして、“共”（共著）、“研”（研究室）、“プ”（プロジェクト）、“発”（発表）がそれぞれ 243 本、243 本、92 本、192 本のエッジに付与されており、それらをクリックすると、その判断の根拠となったページへジャンプする。初期配置では、エッジの長さが $R(X, Y)$ （の逆数）をできるだけ反映するような配置となっている。¹

JSAI2003 の会場で運用を行うことによって、旧姓の併用の問題、外国人名の問題など、いくつか問題点も明らかになったが、人間関係ネットワークを表示するページへのアクセスも多く、分かりやすく面白いシステムであった、研究者の全体的な関係を理解するのに役立つなどの声も聞かれた。

2.3. 評価

表 1 は、275 の訓練例を 5 群に分け、クロスバリデーションを行った平均エラー率である。研究室クラスに対するエラー率が高いが、他のクラスでは 10%程度もしくはそれ以下のエラー率である。また、実際に得られたラベルの適合率、再現率を知るために、ルールを生成する際に用いた 275 の訓練例とは別にランダムに 200 個のエッジを選び、そのラベルを手で判定し、適合率、再現率を求めた。なお、この評価はネットワークを生成した後で行ったものであり、評価のための 200 例はルール生成には利用していない。

表 1 ラベルのエラー率、適合率と再現率

クラス	エラー率*	適合率	再現率	エラー率**
共著	4.1%	91.8% (90/98)	97.8% (90/92)	5.0%
研究室	25.7%	70.9% (73/103)	86.9% (73/84)	20.5%
プロジェクト	5.8%	74.4% (67/90)	91.8% (67/73)	14.5%
発表	11.2%	89.7% (87/97)	67.4% (87/129)	26.0%

* : 225 訓練例のクロスバリデーションによるエラー率

** : 別の 200 例に与えたときのエラー率

¹ Graphviz <http://www.research.att.com/sw/tools/graphviz/> を使い、ばねモデルによる初期配置を求めている。

表2 JSAI2003 ネットワークにおけるエッジラベルのアンケートによる評価

クラス	適合率	再現率
共著	89.0% (81/91)	32.1% (81/252)
研究室	78.3% (72/92)	18.7% (72/385)
プロジェクト	50.0% (9/18)	3.0% (9/300)
発表	79.5% (35/44)	6.5% (35/538)

表3 抽出された全エッジラベルのアンケートによる評価

クラス	適合率	再現率
共著	78.5% (135/172)	53.6% (135/252)
研究室	55.6% (109/198)	28.3% (109/385)
プロジェクト	20.3% (60/296)	20.0% (60/300)
発表	39.9% (222/556)	41.3% (222/538)

共著クラスは、適合率、再現率ともに高い一方、研究室クラスの適合率は低い。これは個人の業績のページで、人間が見れば共著関係か研究室関係か(その両方か)は区別できても、ここで用いている属性だけでは判別しづらいためであると考えられる。プロジェクトクラス、発表クラスでは、クロスバリデーションのエラー率と比較し、200例に対する適合率、再現率から算出したエラー率は高い。これは、プロジェクトや発表を表すページが比較的多様であるためであろう。

JSAI2003 の後、我々は人間関係ネットワークに関するアンケート調査を行った。調査対象者は、JSAI2003 に参加登録した人の中から選んだ 141 人とした。² CGIによるアンケートシステムを作成し、アンケートの協力をお願いするメールを対象者に送付した。82 名から回答を得、回収率は 58%であった。

² 2JSAI2003 にスケジューリングシステムにメールアドレスを登録した 231 人のうち、スケジューリングシステムにおける know リンクの数と本システムにおけるエッジの数の和が 10 人に達しない 90 人を除いた、141 人全員を対象者とした。アンケート送付は 2003 年 12 月 4 日であり、その後約 2 週間で回答を締め切った。

アンケートでは、各被験者に対して、スケジューリング支援システムで know リンクを張った / 張られた人から 10 人、さらに我々のシステムにおいて共起の閾値つき Simpson 係数 $R(X, Y)$ に応じたルーレット選択で 10 人を抽出し、一人あたり各 20 人の相手との関係を質問した。質問は、一人の相手あたり各 15 問であり、「共著の論文がある（既に公になっているものに限る）」、「同じ研究室や部署など 30 人規模の組織に同時期に所属している、またはしていた」、「同じプロジェクトや委員会に所属している、またはしていた」、「JSAI2003 以外の研究会や国際会議で会ったことがある」などの項目を含む。それぞれ、共著、研究室、プロジェクト、発表の関係の有無を問う意図で設定した質問項目である。

表 2 に、JSAI2003 で表示したネットワークのエッジラベルに対して、アンケートから得た回答を正解とした場合の適合率および再現率を示す。また、表 3 は、抽出した全関係に対する適合率および再現率であり、ネットワーク中にエッジラベルとして表示していないものも含む。言い換えれば、関係の強さ $R(X, Y)$ が閾値以下のノードペアに対するラベルも含んでいる。

表 1 と比較して、表 2、表 3 は、適合率、再現率ともに低い値になっている。この理由として考えられるのは、回答者が共著やプロジェクトの関係を忘れている、記述もれしている、プロジェクトの定義としてより広いものを想定しているなど、アンケートの回答に関する問題である。また、我々の手法は、検索でヒットした上位 5 ページのテキストを分析したものであるため、関係を取り逃す場合もあること、すべての情報が Web 上にあるわけではないことなどの原因にもよる。この適合率、再現率は今後の改良で精度が上がっていくと思われる。

3. 位置情報と人間関係ネットワーク

ユビキタス環境では、特定の領域においてユーザがどこにいるかという位置を取得することができる。この情報を用いて、誰と誰が近くにいるか、誰と誰が話している可能性が高いかなどの、接触によるネットワークを作ることができる。

JSAI2003 において得られたユーザの位置情報から、ある手法によって、誰と誰が出会った可能性が高いかというネットワークを構築することができる。その評価を行った結果、ユーザが話した人を 8 割から 9 割程度の適合率であてることができた。ただし、当然、センサの前で話していない場合、会場外で話している場合もあるため、適合率は非常に低くなっている。

こういったコミュニケーションのネットワークと、共著や研究室といったネットワークがどういう関わりをもっているのか、どういうところでコミュニケーションが行われているの

かという視点は、今後の分析の視点として面白いと考えている。

4. おわりに

本報告では、Web から人間関係ネットワークの抽出、特にその評価に関してと、位置情報から人間関係ネットワークの抽出について述べた。後者は、まだ研究段階のものが多く含まれるため、ここでは概要の記述にとどめた。今後は、研究者のコミュニティにどのような社会ネットワークがあって、どのように交流しているのか、それがどのように変化し、研究業績に結び付いていくのかといった視点から分析を進めていきたいと考えている。

参考文献

- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Proceedings of the Seventh International World Wide Web Conference*.
- Kautz, H., Selman, B., & Shah, M. (1997). The hidden web. *AI Magazine*, 18(2), 27-35.
- Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. *Proceedings of ACM-SIAM Symposium on Discrete Algorithms*, 668-677.
- Kumar, S. R., Raghavan, P., Rajagopalan, S., & Tokins, A. (1999). Trawling the web for emerging cyber communities. *Proceedings of the Eighth International World Wide Web Conference*.
- 村田剛志 (2001)「参照の共起性に基づく Web コミュニティの発見」『人工知能学会誌』16(3), 316-323.
- Rafiei, D., & Mendelzon, A. (2000). What is this page known for? Computing web page reputations. *Proceedings of the Ninth International World Wide Web Conference*.
- 西村拓一, 濱崎雅弘, 松尾豊, 大向一輝, 友部博教, 武田英明 (2004)「2003 年度人工知能学会全国大会支援統合システム」『人工知能学会誌』19(1), 43-51.

赤門マネジメント・レビュー編集委員会

編集長 新宅 純二郎

編集委員 阿部 誠 粕谷 誠 片平 秀貴 高橋 伸夫 藤本 隆宏

編集担当 西田 麻希

赤門マネジメント・レビュー 5巻1号 2006年1月25日発行

編集 東京大学大学院経済学研究科 ABAS/AMR 編集委員会

発行 特定非営利活動法人グローバルビジネスリサーチセンター

理事長 高橋 伸夫

東京都文京区本郷

<http://www.gbrc.jp>