論 文

Technical Papers

Web上の情報を用いた企業間関係の抽出

Extracting Inter-business Relationship from World Wide Web

金 英子
YingZi Jin
東京大学大学院 情報理工学研究科 電子情報学専攻
Graduate School of Information Science and Technology, The University of Tokyo
eiko-kin@mi.ci.i.u-tokyo.ac.jp

松尾 豊
Yutaka Matsuo
独立行政法人 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
y.matsuo@aist.go.jp, http://ymatsuo.com/

石塚 満
Mitsuru Ishizuka
東京大学大学院 情報理工学研究科 創造情報学専攻 / 電子情報学専攻
Graduate School of Information Science and Technology, The University of Tokyo
ishizuka@i.u-tokyo.ac.jp, http://www.miv.t.u-tokyo.ac.jp/ishizuka/

keywords: WWW, social network, information extraction, search query, relation extraction

Summary

Social relation plays an important role in a real community. Interaction patterns reveal relations among actors (such as persons, groups, companies), which can be merged into valuable information as a network structure. In this paper, we propose a new approach to extract inter-business relationship from the Web. Extraction of relation between a pair of companies is realized by using a search engine and text processing. Since names of companies co-appear coincidentaly on the Web, we propose an advanced algorithm which is characterized by addition of keywords (or we call *relation words*) to a query. The relation words are obtained from either an annotated corpus or the Web. We show some examples and comprehensive evaluations on our approach.

1. ま え が き

企業間にはさまざまな関係があるが，企業間の関係が総体として織り成すネットワーク構造を分析することで，例えば，ある企業が他の企業と競争関係であるか，どうすればネットワーク上で優位な立場に位置することができるかなど，企業の競争力の分析や取るべき戦略の決定に用いることができる．また，全体的なネットワーク構造の特徴を分析することで，その産業分野全体におけるネットワークの安定性，成長性も分析することができる．経済学や社会学の分野では，このようにネットワークを分析し，関係構造の中に隠された知見を読み解く**社会ネットワーク分析**と呼ばれる研究が行われている [相馬 05, 安田 97, 金光 03]*1.

一方，近年では多種多様な情報が Web 上に公開されている．個人の Blog から政府の公開情報まで，ますます多くの新しい情報が Web 上に存在するようになってきており，Web から有用な知識を抽出しようとする研究が盛んに行われている [佐藤 01, 藤井 04, 立石 04]．さらに最近では，Blog や SNS の分析 [Adar 04, 湯田 05]，また研究者ネットワークを抽出する研究 [松尾 05, Matsuo 06b] など，Web 上の情報からネットワーク分析に必要なデータを抽出し分析する研究が行われている．多様なデータに対して社会ネットワーク分析の手法が適用可能になっている．

Web 上では企業同士の関係に関わる情報も公開されている．企業間の共同開発，合併・買収，あるいは訴訟などの情報は，プレスリリースや報道などの形で素早く Web 上に公開されることが多い．本研究は，Web 上に公開されている情報から，企業間の関係を抽出する手法を提案する．日々変化する複雑な社会の関係性を俯瞰することは，社会の動向を見渡し，その構造を深く理解して新たな活動につなげる上で重要であり，社会学や俯瞰工学といった分野で研究されている [増田 06, 坂田 05]．これまで，Web 上の情報から企業ネットワークを抽出する研究は行われておらず，複数の企業の俯瞰的な情報を得たり，社会ネットワーク分析の手法を利用した構造的な分析を行う上で重要な技術である．企業間には，取引，提携，役員派遣，訴訟など様々な関係があるが，本研究では，特に提携関係と訴訟関係に焦点を当てて，抽出手法

*1 国際的には，INSNA(International Network for Social Network Analysis: 1978 年に Barry Wellman によって創設）という団体が，毎年 Sunbelt conference という国際会議を開いている．また，Social Networks というジャーナルが刊行されている．

図 1 業務提携に関する報道

図 2 訴訟に関するニュース

を構築する．

本論文は，2 章では，Web から企業間の関係を抽出する概要について説明し，3 章では，関係の抽出において重要な，関係語を特定する方法について述べる．4 章と 5 章では，システム全体の流れと実験考察について説明し，6 章で関連研究との比較を行い，7 章で結論を述べる．

2. 企業間関係の抽出

2·1 検索エンジンを利用した関係抽出

Web 上には，企業の関係を表す多くの情報があるが，それらを大別すると次の 2 つに分けて考えることができる．ひとつは，企業間の関係を整理してまとめてあるサイトで，例えば，Yahoo!ニュース[*2]や日経のプレスリリースサイト[*3]などである．こういった整形された情報源は Web に限らず，会社四季報などの書籍からも企業間の関係について知ることが可能である．

もうひとつは，企業が自ら公開するニュースリリースや Web 上のニュースサイトなど，分散した情報源である．例えば，図 1 は企業のホームページ上に掲載された業務提携に関するリリース[*4]であり，図 2 はビジネスニュースサイトの訴訟に関するニュース[*5]である．これらは特定のサイトに整形されてまとめて置かれているのではなく，さまざまなサイトに分散して存在する．

情報源として考えると，前者の情報は，何らかの編集作業を経ているため，信頼性が高く，統一的な基準であるという意味で質が高い．一方，後者の情報は，基準は一定でなく，信頼できない情報が含まれているかもしれないが，中小企業や海外の企業などの情報も得られるという点で網羅性が高く，また新しい情報をそのまま取り扱うことも可能である．今後の Web の発展を考えると，後者のように整形されていない情報を収集し情報源とすることは，重要な技術になると考えられる．

そこで，本研究では Web 全体の情報を用いて企業間関係を抽出することを目的とする．特定の企業名を与えたとき，その間の関係を Web 中の情報から同定する．そのために検索エンジンを用いて Web ページを集め，文書処理を行う．

2·2 関係を特定する関係語

検索エンジンを利用して，2 つの企業の関係を調べるときに考えられるのは，その 2 つの企業名を並べてクエリとすることであろう．例えば，松下電器とジャストシステムの関係を調べたいとすると，

「松下 AND ジャストシステム」

と検索エンジンにクエリを投げれば，425,000 件[*6]もヒットする．上位のページには，昨年話題になった松下とジャストシステムの訴訟に関するページ[*7]が多く含まれる．したがって，このテキストを分析すれば，松下とジャストシステムの訴訟関係は同定できそうである．

ところが，松下とジャストシステムは，実は 2001 年にナレッジマネジメントにおける提携を行っている．この Web ページ[*8]は，ヒットした文書中 124 件目であった．このように企業間の関係はさまざまなものがあり，大企業同士の提携や訴訟関係など，注目される関係は多くのページに書かれているが，逆に話題にならなかった関係は，ごく少数のページにしか情報がないこともある．

ここで，仮に，検索クエリを「松下 AND ジャストシステム AND 訴訟」，あるいは，「松下 AND ジャストシステム AND 提携」としてみよう．すると，それぞれ松下とジャストシステムの訴訟関係や提携関係に関するページが上位に来ることになる．これは，キーワードスパイス [Oyama 04] に近いアイディアであり，目的とするページを上位にヒットさせるために，クエリに適切なキーワードを加えるものである．本研究では，このような関係を特定するためにクエリに加える語を**関係語**（Relation word）と呼ぶことにする．

関係語は，対象となる企業間の目的とする関係をうまく特定できる方が望ましく，また，関係がある場合にはできるだけ漏れが少なくページを見つけられる方がよい．したがって，関係語を加えることにより，関係を特定する精度と再現率の両方について考慮する必要がある．これについては，3.1 節で詳しく説明する．

*2 http://headlines.yahoo.co.jp/hl
*3 http://release.nikkei.co.jp
*4 http://www.hitachi.co.jp/New/cnews/040309.html
*5 http://www.worldtimes.co.jp/ の 2005 年 2 月 16 日の記事．
*6 2005 年 5 月 22 時点での Google による検索結果，以下の例でも同様．
*7 http://pc.watch.impress.co.jp/docs/2005/0201/just2.htm
*8 http://ascii24.com/news/ の 2001 年 2 月 19 日の記事．

2·3 本研究で取り扱う関係

企業間の関係としては，株式の持ち合いや子会社・グループ会社といった資本的な関係，業務上での提携関係や取引関係，役員等の人的な関係，訴訟・係争関係，競合関係などさまざまなものがある．社会学におけるネットワーク分析 [安田 97, 金光 03, Scott 00] では，企業間の紐帯（ネットワーク分析では関係を紐帯と呼ぶ）の種類や強さ，またその成長や淘汰といった時間的変化も重要な分析材料になる．

例えば，企業間の訴訟を考えると，訴訟関係が永続するわけではなく，いずれ和解や判決により決着がつく．したがって，係争中の訴訟関係なのか，和解した訴訟関係なのかという区別をつけることは重要である．また，企業間の提携関係では，製品の共同開発やサービスの共同提供といった業務提携と，資本参加を含めた合併・買収や営業譲渡などの資本提携の関係がある．前者よりも後者の方が強い関係である．

本研究では，企業間の関係として，提携関係と訴訟関係を扱う．それぞれ，企業間の友好的関係，敵対的関係の代表的なものである．さらに，提携関係は，業務提携と資本提携，訴訟関係は係争関係と和解関係という 4 種類を扱うことにする．提携関係に対しての業務提携や資本提携の関係，また訴訟関係に対しての係争関係や和解関係を，**詳細関係**と呼ぶことにする．

3. 関係語の抽出

本章では，求めたい企業間関係が記述されたページを見つけるために，検索クエリに加える関係語を得る方法について述べる．学習データを用いる方法と，Web の共起を用いる方法を提案する．

3·1 学習データからの関係語の獲得

関係語を得るためには，企業間の関係が含まれた多くの Web ページを準備して，そのページに共通する語を求めればよい．つまり，学習データから特定性の高い関係語を学習する．

まず，業務提携や資本提携などの企業間関係について書かれた Web ページと，そうでないページを集め学習データを作る．各 Web ページから語が出現するかしないかという属性を生成し，分類を学習する．

本来は，語の出現を含むページのさまざまな特徴を属性とする分類問題になるが，現実的には，検索エンジンに複雑なクエリを入力するのは難しいため[*9]，単語 1 語，もしくはそのうち 2 語の連言による組み合わせだけを調べる．学習の評価に F 尺度を用いることにすると，この分類問題は，各単語（もしくはその組み合わせ）が出現するかどうかによって F 尺度がどう変化するかを調べればよいことになる．

関係が含まれたページを正解，ある語 w が含まれているページを出力として，F 尺度は

$$\mathbf{F}_{Rel}(w) = \frac{2\,\mathbf{P}_{Rel}(w)\mathbf{R}_{Rel}(w)}{\mathbf{P}_{Rel}(w)+\mathbf{R}_{Rel}(w)} \qquad (1)$$

と定義される．$\mathbf{P}_{Rel}(w)$ は，単語 w を含むページのうち関係が正しく記述されたページの割合であり，$\mathbf{R}_{Rel}(w)$ は，関係が記述されたページのうち単語 w が含まれるページの割合である．一般的に，学習データに対して最も分類精度のよい仮説を選ぶと過学習が起こる可能性があるが，ここでは関係語として単語 1 語もしくは 2 語に限定しているので，その影響は少ない．

F 値が高い語を関係語として用いると，企業関係について書かれたページが得られる可能性が高くなるが，確実に得られるわけではない．複数の関係語を用いて，検索クエリを複数生成し検索することで，より網羅的に関係を抽出することができる．したがって，F 値が上位の複数の関係語を用いる．また，検索されたテキストの内容から企業間関係が実際に存在するかを判断するルールの中でも，この関係語を利用する．

3·2 Web を用いた関係語の抽出

一方，関係語を得るために学習データを用意するのは手間がかかる．そこで，本研究では，Web を用いて関係語を抽出することを考える．この方法は [森 05a] と同様の方法であり，関係を特定する単語が与えられたときに，それに関連する他の語を得ることができる．また，少数の学習データが与えられたときに，そこから多くの関係語を得ることも可能である．

基本的なアイディアは，例えば提携関係を調べたいのであれば,「提携」という語を直接クエリに加えればよいというものである．さらに,「提携」とよく共起する語も，提携関係を把握する手がかりになりそうである．そこで,「提携」という語とよく共起する語を Web 上から獲得しようというものである．

ここでは,「提携」などの語を w_{Rel} とする．そして，Web 上でのヒット件数を用いて Jaccard 係数

$$J_{w_{Rel}}(w) = \frac{|w_{Rel} \cap w|}{|w_{Rel} \cup w|} \qquad (2)$$

を計算し，これが高い語 w を関係語として用いる．ただし，$|w_{Rel} \cap w|$ は「w_{Rel} AND w」をクエリにした場合のヒット件数，$|w_{Rel} \cup w|$ は「w_{Rel} OR w」をクエリにした場合のヒット件数である．

なお，全ての語に対してこの値を計算するのは現実的でないので，ここでは企業間関係について書かれたページを用意し，そこから候補をとなる単語を切り出す．例えば，提携関係であれば，日経のプレスリリースカテゴリから取得する．しかし，学習データを使う方法と違っ

*9 検索エンジンによっては，クエリ内の単語の数が制限されていたり，NOT や OR のオペレータが必ずしも正確な結果を返さない場合がある．

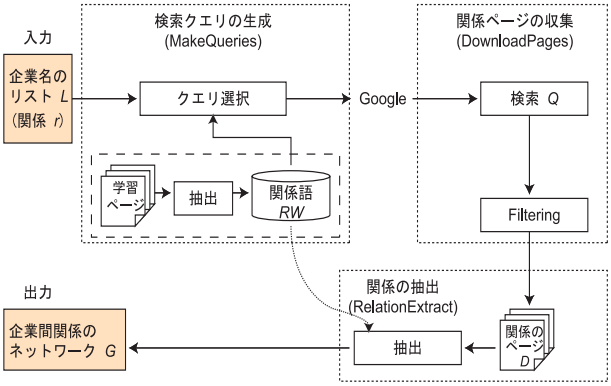


図 3 システム全体の流れ

て，各ページごとに実際に提携関係について記述されているかというラベルを付与する必要はない．

企業間の詳細関係を得るために，詳細関係に応じた関係語を取得する必要がある．この場合も，上に述べた学習データによる方法，Web から取得する方法のいずれを用いることもできる．

4. システム全体の流れ

システムの全体を図 3 に示す．具体的な処理の流れは図 4 の擬似コードとして記述する．前節に述べた方法で，あらかじめ提携関係，訴訟関係などの関係ごとに関係語 RW のリストを取得しておく．そして，システムに企業名のリスト L が入力されると，それぞれの企業 (x とする）を取り出し，自分以外の企業（y とする) との関係 r の有無を調べてエッジを生成することで，企業間の関係のネットワーク G を出力する．全体の処理は，検索クエリの生成，関係ページの収集，関係の抽出という大きく３つに分けられる．

検索クエリの生成フェーズ（$MakeQueries$）では，関係語の上位 $n_{queries}$ 個を氏名 x，y に加えることで検索クエリ集合 Q を生成する．関係ページの収集フェーズ（$DownloadPages$）では，生成された検索クエリ集合 Q を検索エンジンに入力し，上位にヒットしたページ n_{pages} 件をダウンロードして関係のページ集合 D を得る．最後の関係の抽出フェーズ（$RelationExtract$）では，ダウンロードしたページの内容を調べ，２つの企業に関する関係の記述があるかどうかを判断する．

(1) 収集した関係ページ集合 D に含まれる全ての文のリスト S を収集する．

(2) ２つの企業の名前と関係語 $rw(rw \in RW)$ が同時に出現する文 $s(s \in S)$ に対して，s に出現する関係語のスコアを足し合わせてその文のスコア $score_s$ とし，すべての文の中で最もスコアの高いものをその企業間関係のスコア $score_{xy}$ とする．

(3) $score_{xy}$ が閾値 $score_{thre}$ を超えれば，２つの企業は関係があると判断する．この部分は，構文解析や意味解析などより深い処理を行うことも可能であるが，本研究ではできるだけ簡単な方法にするため，このようなシンプルなルールを用いている．本研究では，$n_{queries} = 2$，$n_{pages} = 5$ とした．

```
Input: a list of firm names L and relation r
Output: a network of firms G
  given thresholds score_thre, n_pages, n_queries

for each x ∈ L
    do InsertNode(G, x)

RW ← GetRelationWords(r)
for each rw ∈ RW
    w_rw ← GetWeight(r, rw)
    RW_query ← top n_query weighted words in RW
for each x ∈ L and y ∈ L where x ≠ y do
    Q ← MakeQueries(x, y, RW)
    D ← DownloadPages(Q)
    score_xy ← RelationExtract(D, x, y, RW)
    if score_xy > score_thre then
        InsertEdge(G, x, y)
done
return G

/* RW を用いて x と y の関係を調べるクエリ集合を得る*/
function MakeQueries(x, y, RW)
    Q ← {}
    for each rw ∈ RW_query
        Q ← {"x AND y AND rw"}∪Q
    return Q

/* クエリ集合 Q から検索される（上位の）ページ集合を返す*/
function DownloadPages(Q)
    D ← {}
    for each q ∈ Q
        D ← GoogleTop(q, n_pages)∪D
    return D

/* D と RW を用いて x と y の関係のスコアを計算する*/
function RelationExtract(D, x, y, RW)
    score_xy ← 0
    S ← GetSentences(D)
    for each s ∈ S do
        if s contains "x" and s contains "y" then
            score_s ← Σ_{rw contained in s} w_rw
            if score_s > score_xy then
                score_xy ← score_s
    done
    return score_xy
```

- InsertNode(G,x): ノード x をネットワーク G に追加する．
- InsertEdge(G, x, y): x と y をつなぐエッジをネットワーク G に追加する．
- GetWeight(r,rw): 関係 r に対する関係語 rw の重みを返す．
- GetRelationWords(r) : 3.1 節と 3.2 節の手法から得られる関係 r の関係語の集合を返す．
- GoogleTop(q, n_{pages}): クエリ q で検索してヒットする上位 n_{pages} 件のページ集合を返す．
- GetSentences(D): 関係のページ集合 D に含まれる全ての文をリストで返す．

図 4 システム全体の擬似コード

なお，閾値 $score_{thre}$ は，予備実験において企業の関係の有無を判断しておいた学習データから F 尺度が最大になるような値にしている[*10]．閾値を高くすると，特定性の高い関係語を多く含むような記事でない限り２つの企業は関係がないと判断されてしまうので再現率が下がる．逆に，閾値を低く設定すると，特定性の低い語を含んで実際に関係がない企業に対しても関係があると判断されることが多くなるので適合率が下がる．

*10 学習データから獲得した関係語を用いる場合，提携関係とその詳細関係である業務提携と資本提携の閾値は，それぞれ 1.1984, 3,3429, 0.6598 である．Web を用いて抽出した関係語を用いる場合，提携関係とその詳細関係の閾値はそれぞれ，0.5044, 0.8786, 0.2575 で，訴訟関係とその詳細関係である係争段階と和解段階の閾値はそれぞれ，0.5626, 1.0217, 1.996 である．

表 1 用いた企業名

松下電器産業，ジャストシステム，ニッポン放送，ライブドア，日本電産株式会社，日本ビクター株式会社，日本 IBM，キヤノンスター，株式会社ニデック，株式会社カイノス，東和薬品，三井住友カード，東京電力，エレコム，フジテレビ，富士通株式会社，富士通インフォソフトテクノロジ，コネクトテクノロジーズ，バイオマティクス，サムスン SDI，イーバンク銀行，株式会社ニコン，ＡＳＭＬ社，日本マックストア，ミネベア株式会社，株式会社東芝，韓国ハイニックス，米 SCO，米 IBM，株式会社トランスウエア，Opera Software，米 Agere，米 Intersil，米インテル，米ブロードコム，LG 電子株式会社，セイコーエプソン，上海中材，シスコシステムズ，サン・マイクロシステムズ・インク，日本電気株式会社，KDDI 株式会社，日立製作所，ソースネクスト株式会社，東京エレクトロン株式会社，ルネサステクノロジ，シックス・アパート，ニイウス株式会社，楽天株式会社，サイバーブレインズ，全日本空輸，ニフティ株式会社，松下電子工業，京セラ株式会社，株式会社サイバード，株式会社 JIMOS，日本セラテック，セランクス，日本信販株式会社，株式会社 UFJ カード

5. 評価実験と考察

この章では，本論文で提案した手法の評価を行う．まず，抽出された企業間関係の具体例を示し，企業間の関係をどの程度的確に抽出できるかというシステム全体の評価を行う．その後，関係語取得に関する部分の具体例と評価を行う．

情報，通信，放送，電機などの産業分野を中心に 60 社を選び，その企業間関係を Web から抽出した．具体的に用いた企業名を表 1 に示す．大企業から情報系の新しい企業まで幅広く含んでいる．

5·1 企業間関係抽出の具体例と評価

4 章に述べた方法で，関係の有無を判断する．つまり 60 社の組み合わせ，つまり ${}_{60}C_2$ の 1770 通りに対して，それぞれ関係の有無を判断する．抽出された関係の適合率および再現率の評価を表 2 に示す．表 2(a) は，3.1 節で提案している学習データから獲得した関係語を利用して抽出した提携関係（およびその詳細関係）の結果である．1770 組の企業間に実際には提携関係が 113 組存在するが，本手法では 68 組抽出することができた．また提携関係の詳細関係である資本提携と業務提携は実際 21 組と 100 組存在することに対し，本手法ではそれぞれ 11 組と 58 組を抽出することができた．なお，正解データは Web 上からそれぞれの企業の関係を人手で調べることで作成している．Web に書かれていない外部知識は利用しないため，原理的には 100 % の適合率，再現率を取りえる．(b) と (c) は，3.2 節で提案している Web を用いて抽出した関係語を利用して抽出した提携関係と訴訟関係の結果である．それぞれの実験で検索クエリに利用した関係語は，表 5，表 6 と表 7 の各関係においての上位 2 語である．学習データから獲得した関係語を利用した結果と Web を用いて抽出した関係語を利用した結果を比較すると，再現率はほぼ差がないが，前者のほうの適合率が低いことが分かる．これは，学習データから獲得した関係語のスコアは学習データに偏った表現が多いので関係（特に詳細関係）を正確に特定できないことである．なかでも資本提携の精度が低くなっているのは，学習データにおいて「提携」「合意」「提供」といった業務提携を特定するスコアが高い関係語が，資本提携においても高いスコアを持っているので，関係の抽出段階で実際には業務提携であるのに資本提携と誤って判断されることが多いことが原因であった．また，訴訟関係とその詳細関係は，提携関係よりも正確に抽出することができた．これは，訴訟関係は「提訴」「判決」「訴訟」「和解」といったある程度決まった用語を使うことが多いことに対し，提携関係は，サービス提供や共同研究，販売提携，合併・買収などに関する多様な表現が用いられるので，再現率が低くなっている．

表 2 本手法により抽出された関係の評価

(a) 学習データから得られた関係語を利用（提携関係）

関係・詳細関係	適合率	再現率
提携関係	55.7% (68/122)	60.2% (68/113)
資本提携	23.9% (11/46)	52.4% (11/21)
業務提携	55.2% (58/105)	58.0% (58/100)

(b) Web の共起関係から得られた関係語を利用（提携関係）

関係・詳細関係	適合率	再現率
提携関係	60.9% (70/115)	62.0% (70/113)
資本提携	75.0% (9/12)	42.9% (9/21)
業務提携	67.4% (60/89)	60.0% (60/100)

(c) Web の共起関係から得られた関係語を利用（訴訟関係）

関係・詳細関係	適合率	再現率
訴訟関係	61.5% (16/26)	100% (16/16)
係争段階	63.6% (14/22)	87.5% (14/16)
和解段階	72.7% (8/11)	88.9% (8/9)

表 3 Web サイトに含まれている関係の評価

関係・詳細関係	適合率	再現率
提携関係	100.0% (27/27)	23.8% (27/113)
資本提携	100.0% (6/6)	28.6% (6/21)
業務提携	100.0% (21/21)	21.0% (21/100)
訴訟関係	100.0% (11/11)	68.8% (11/16)
係争段階	100.0% (11/11)	68.8% (11/16)
和解段階	100.0% (6/6)	66.7% (6/9)

なお，現実的に Web 上の情報から企業間関係を取得することを考えた場合，特定の Web サイトにまとめている情報を利用することができる．そこで，手法自体の比較対象ではないが，提携関係は日経のプレスリリースサイトから，訴訟関係は知的財産局の訴訟ニュースから，この 60 社の関係を調べたものが表 3 である．これらのサイトは，もちろん適合率は 100% であるが，例えば，日経のプレスリリースでは半年間のニュースしか公開しないし，知財局のサイトでも最大 2,008 件のニュースだけを検索対象にするなど，情報の期間が限定されたり，情報の量を制限したりしているので，すべての企業の情報

表 4 関係の抽出の例

企業名ペア	正解		web site		本手法	
	訴	和	訴	和	訴	和
松下電器 vs LG 電子株式会社	○	○	○	○	○	○
米インテル vs 米ブロードコム	○	○	○	○	○	○
日本電産 vs 日本マックストア	○	○	○	○	○	○
日本電産 vs ミネベア	○	○	○	○	○	○
富士通株式会社 vs サムスン SDI	○	○	○	○	−	○
ニコン vs ASML社	○	○	○	○	○	−
ライブドア vs フジテレビ	○	○	−	−	○	○
ライブドア vs イーバンク銀行	○	○	−	−	−	○
エレコム vs エプソン	○	○	−	−	○	○
松下電器 vs ジャストシステム	○	−	○	−	○	○
松下電器 vs サムスン SDI	○	−	○	−	○	−
東芝 vs 韓国ハイニックス	○	−	○	−	○	−
米 SCO vs 米 IBM	○	−	○	−	○	−
米 Agere vs 米 Intersil	○	−	−	−	○	−
ライブドア vs トランスウエア	○	−	−	−	○	−
日本電産 vs 日本ビクター	○	−	○	−	○	−
米 IBM vs 日立製作所	−	−	−	−	○	−
ニッポン放送 vs フジテレビ	−	−	−	−	○	−
ジャストシステム vs 日本 IBM	−	−	−	−	○	−
松下電器 vs 日立製作所	−	−	−	−	○	−
日本マックストア vs ミネベア	−	−	−	−	○	○
ライブドア vs Opera Software	−	−	−	−	○	−
日立製作所 vs ルネサステクノロジ	−	−	−	−	○	○
ニッポン放送 vs ライブドア	−	−	−	−	○	−

を網羅できないことから，再現性は低い．

表 4 は，訴訟関係の詳細関係である提訴段階と和解段階について，本システムで実際に抽出した具体例と正解，および日経／知財局の Web サイトに含まれている記事の具体例を示したものである．「○」は関係があることを表し，「−」は関係がないことを表す．例えば，「エレコム」と「エプソン」の間には，訴訟関係があって，それが提訴段階から和解段階に至っているのが正確な関係だが，これは 2000 年と 2002 年の古いニュースであることから，知財局のサイトには載っていないが，提案手法ではこの関係を抽出することができる．しかし，実際にない関係もあると出力される場合がある．例えば，表 4 に示されている日立製作所と米 IBM の間の実際に存在しない訴訟関係が抽出された．これは “日立製作所と米 IBM の HDD 合弁会社、特許侵害で中国企業を提訴” という文が原因で，実際には中国企業との係争であることを示す文であるが，文内に複数の企業名（「日立制作所」「米 IBM」「中国企業」）と複数の関係語（「特許」「侵害」「提訴」「会社」「企業」）を含んでいるために誤ったものである．これは適切な係り受けの解析を行うことで対処できると考えられ，4.3 節で述べた関係の抽出フェーズを改良することでさらに精度が改善される可能性がある．また，実際に関係があるが抽出できなかった場合として，特定性の高い関係語が記事の中に出現しないケースがある．例えば，「ライブドア vs イーバンク銀行」では，提携関係を行おうとしたあとトラブルが起き，イーバンクがライブドア社長を刑事公訴し，すぐに訴訟を取り上げて和解となった．厳密には訴訟関係があったのだが，記事中では「騒動」や「トラブル」と表現され，うまく取り出すことができなかった．抽出したい関係を詳細化し，「公訴」といった語が使われる場合もあることをシステムは認識する必要がある．こういった係り受け解析と関係の詳細化が今後の課題のひとつである．

図 5 は，企業名をノードとし，抽出された関係[*11]をリンクで繋ぐことで訴訟関係と提携関係のネットワークを生成したものである．黒線が提携関係を示し，点線が訴訟関係を示す．点線のなかでも太線は資本提携を，そうでないものは業務提携である．また，点線で，破線は係争段階を表し，細い点線は和解段階を表す．ネットワーク図からは，電機・電力の大手企業を中心に連携が活発に行われていることが分かる．特に，活発に周りと連携している松下や日立，積極的に合併・買収などに取り組んでいるライブドアといった姿を理解することができる．また，社会ネットワークの分析手法を適応することで，他企業と似たような紐帯（関係）を持っている企業同士[*12]や，企業間の連携において媒介的な役割をする企業，さらにネットワーク全体の密度や傾向などを分析することが可能である．

5·2 関係語抽出の具体例と考察

本研究では，最終的な企業間の関係を抽出するために，関係語をいかに得るかが重要な部分であるが，本節ではその具体例を示す．

3.1 節で述べた，学習データを作り F 尺度の高い関係語を示したものが表 5 である．提携関係について，F 尺度が高いものは「事業 AND リリース」「ニュース AND 事業」などであった．また，提携関係の中でも，資本提携および業務提携の 2 つの詳細関係については，それぞれ「通信 AND 合意」「合意」，「提携 AND 今回」，「提携 AND 提供」などが上位であった．

この結果を考察すると，「リリース」「ニュース」などの語は情報源を特定する働きが強く，「事業」「開始」などは具体的な提携関係を示す語である．この組み合わせである「事業 AND リリース」や「ニュース AND 事業」は，情報源を特定しながら提携関係を把握するよい関係語である．しかし 5 位の「発表」や 9 位の「リリース」などの関係語は，提携関係を示す具体的な語が入っておらず，この学習データに偏った結果であると考えられる．詳細関係に対しては，さらにこの傾向が強い．

つぎに，3.2 節で述べた，Web を用いた関係語の取得について具体例を示す．提携関係に関しては w_{Rel} を「提携 AND 株式会社」（与えた企業名はすべて株式会社であるため），訴訟関係に関しては「侵害 AND 訴訟」としたものが表 6 と表 7 である．それぞれ，提携関係と訴訟関係，およびそれらの詳細関係に対して，Jaccard 係数が高い上位 10 個の関係語とそのスコアを示している．提携関係の業務提携関係においては「提携 AND 企業」，「提携 AND 事業」などの語がスコアが高いことに対し，資本提携においては，「資本 AND 経営」，「資本 AND 企業」などの語のスコアが高い．また，訴訟関係の係争段階においては「特許 AND 提訴」，「提訴 AND 技術」の

*11 ここでは，Web を用いて抽出した関係語を利用して抽出した提携関係と訴訟関係 (表 2 の (b) と (c)) を用いている．

*12 ネットワーク分析では構造同値と呼び，このような企業同士は競争関係になりやすい

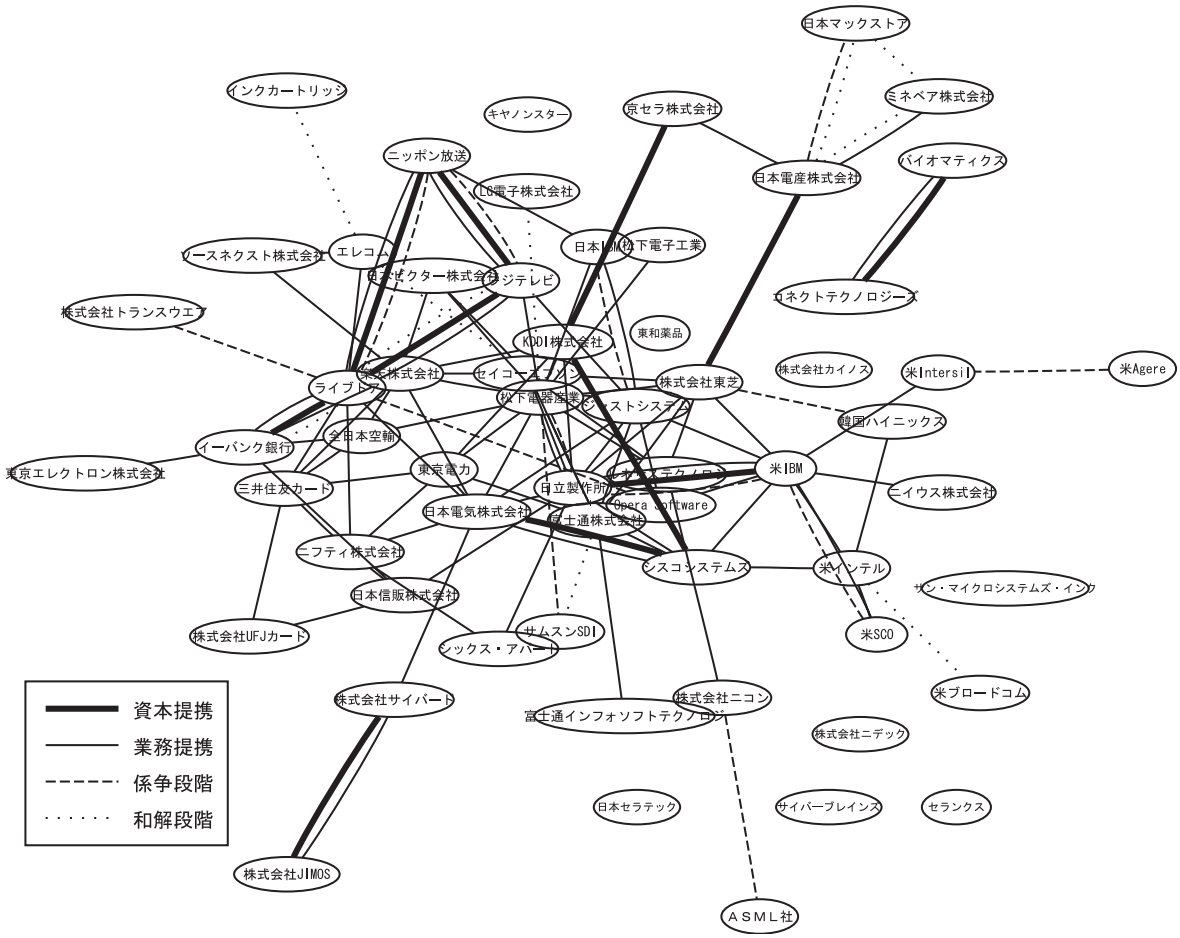


図 5 企業間関係のネットワークの抽出例

ような語がスコアが高いことに対し，和解段階においては「和解 AND 会社」,「和解 AND 発表」のような語のスコアが高い．Web 全体においての語の共起を用いているので，学習データを用いる場合よりもロバストで納得性の高い結果が得られていることが見て取れる．

5·3 関係語付与による関係抽出の評価

最後に，関係語が，企業の関係を示すページを探すことに対してどのくらい有効であるかを評価する．ここでは，表 4 で正解として示した 16 組の訴訟関係のペア（そのうち，8 組が既に和解になっている）を取り上げる．単純に企業の名前のペアをクエリにして検索した場合と，それに関係語を加えて検索した場合で，上位にヒットするページでどのくらい企業間関係の情報を含んでいるかを比較した．

ここでは，次の 5 つを比較している．

noRW 関係語を全く用いない．

RW1 関係語の上位 1 位だけ用いる．

RW2 関係語の上位 2 位だけ用いる．

RW1+ RW2 関係語の上位 1 位と 2 位を用い，2 回検索する．

RW1+RW2+noRW 上位 1 位，2 位， 関係語を用いない場合の 3 回検索する．

複数回検索する場合も，取得するページ数は合計で 10 ページになるようにしている．図 6 にその結果を示す．適合率は，検索結果の上位 k 位中，どれだけのページがその企業間の訴訟関係について書かれているか（16 組について平均）を示したものであり，横軸は k である．k を大きくしていくと，さまざまなほかの情報も含まれるようになるので，徐々に適合率は下がる．カバレージは，検索結果の上位 k 位中，その訴訟関係が含まれれば 1，そうでなければ 0 として，16 組について平均をとったものである．k が大きくなれば 1 に近くなる．

このグラフで分かるのは，複数個の関係語で複数回の検索を行った方が，1 個の関係語あるいは関係語を入れない場合よりページの適合率が高いことである．また，抽出する関係によっては（特に注目される訴訟関係などは）関係語を加えない場合にもカバレージは高いことが分かった．

6. 議論と関連研究

6·1 F 尺 度

3.2 節で定義している Jaccard 係数の式は，ある近似のもとで F 尺度を最大化する語を見つけていることになる．まず，語 w の適合率は，ページに関係が記述されて

表 5 学習データから得られた提携関係の関係語

提携関係	$F_{\text{提携}}$	業務提携	$F_{\text{業務提携}}$	資本提携	$F_{\text{資本提携}}$
事業 AND リリース	0.4688	通信 AND 合意	0.4490	提携 AND 今回	0.4444
ニュース AND 事業	0.4522	合意	0.4483	提携 AND 提供	0.4364
事業 AND 発表	0.4279	グループ AND 合意	0.4243	提供 AND 株式	0.4333
事業 AND 開始	0.4274	事業 AND 合意	0.4151	事業 AND 株式	0.4211
発表	0.4271	株式会社 AND 合意	0.3956	合併 AND 予定	0.4071
事業 AND 記事	0.4242	合意 AND 目指す	0.3956	合併 AND 合意	0.4043
提携 AND 提供	0.4242	記事 AND 合意	0.3953	株式 AND リリース	0.4040
事業 AND 提供	0.4228	発表 AND 合意	0.3918	携帯 AND 譲渡	0.4000
リリース	0.4224	提携 AND 合意	0.3871	株式 AND 今回	0.4000
記事	0.4224	合併 AND 合意	0.3863	提携	0.3979

表 6 Web の共起関係から得られた提携関係の関係語

提携関係	$J_{W_{\text{提携}}}$	業務提携	$J_{W_{\text{業務提携}}}$	資本提携	$J_{W_{\text{資本提携}}}$
提携 AND 株式会社	1.0000	提携 AND 業務	1.0000	事業 AND 資本	1.0000
提携 AND 株式	0.8776	提携 AND 企業	0.4747	資本 AND 経営	0.5528
提携 AND 会社	0.7036	提携 AND 事業	0.4588	資本 AND 企業	0.5483
提携 AND システム	0.5654	提携 AND 開発	0.4367	資本	0.5431
提携 AND ビジネス	0.5339	提携 AND 会社	0.4315	資本 AND 管理	0.5331
提携 AND サービス	0.5329	提携 AND 提供	0.4292	開発 AND 資本	0.5193
提携 AND 事業	0.5255	提携 AND 経営	0.4238	資本 AND 利用	0.5100
提携 AND 管理	0.5200	提携 AND 株式	0.4190	業務 AND 資本	0.5058
提携 AND 開始	0.5183	提携 AND サービス	0.4092	販売 AND 資本	0.5049
提携 AND 対応	0.5071	提携 AND 販売	0.4089	資本 AND 会社	0.4940

表 7 Web の共起関係から得られた訴訟関係の関係語

訴訟関係	$J_{W_{\text{訴訟}}}$	係争段階	$J_{W_{\text{係争段階}}}$	和解段階	$J_{W_{\text{和解段階}}}$
侵害 AND 訴訟	1.0000	侵害 AND 提訴	1.0000	訴訟 AND 和解	1.0000
侵害 AND 請求	0.5142	特許 AND 提訴	0.5332	和解 AND 会社	0.6479
侵害 AND 判決	0.4900	提訴 AND 技術	0.4860	和解 AND 発表	0.6456
侵害 AND 裁判所	0.4582	提訴 AND 開発	0.4825	和解 AND 開発	0.6410
侵害 AND 賠償	0.4441	提訴 AND 関連	0.4685	和解 AND 製品	0.6404
侵害 AND 会社	0.4335	提訴 AND 会社	0.4639	和解 AND 関連	0.6290
侵害 AND 発表	0.4267	提訴 AND 販売	0.4627	和解 AND 技術	0.6165
侵害 AND 損害	0.4251	提訴 AND 企業	0.4562	和解 AND 情報	0.5989
侵害 AND 企業	0.4183	提訴 AND 発表	0.4528	和解 AND 問題	0.5843
侵害 AND 裁判	0.4181	提訴 AND 情報	0.4448	和解 AND 企業	0.5799

いることを Rel という記号で表すと，簡略的に次のように表される．

$$\mathbf{P}_{Rel}(w) = P(Rel|w) \tag{3}$$

つまり，語 w が出現するページのうちで，関係が記述されている割合である．また，再現率は

$$\begin{aligned}\mathbf{R}_{Rel}(w) &= P(w|Rel) \\ &= \frac{P(Rel,w)}{P(Rel)} = \frac{P(Rel|w)P(w)}{P(Rel)}\end{aligned} \tag{4}$$

となる．ここで，w_{Rel} を関係 Rel を最も適切に表す語とすると，それぞれの確率は，

$$\mathbf{P}_{Rel}(w) \sim \frac{|w_{Rel} \cap w|}{|w|} \tag{5}$$

$$\mathbf{R}_{Rel}(w) \sim \frac{|w_{Rel} \cap w|}{|w_{Rel}|} \tag{6}$$

となり，最終的に

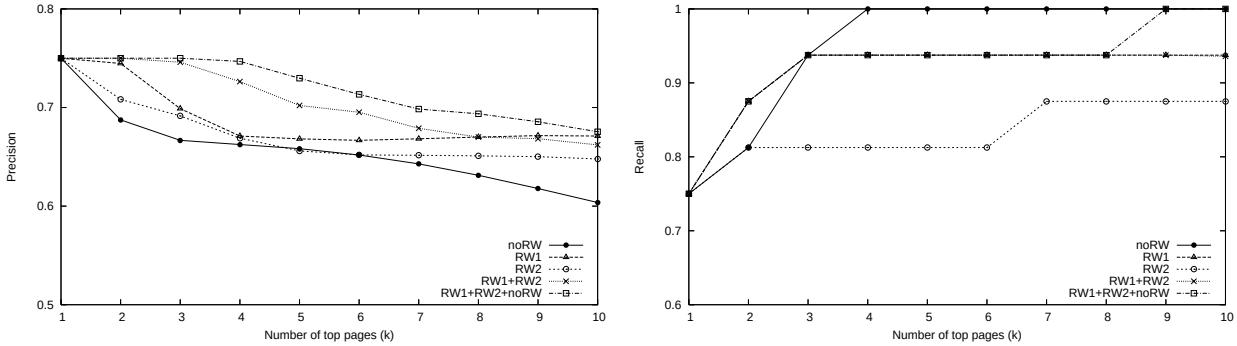
$$\mathbf{F}_{Rel}(w) \sim \frac{2\,|w_{Rel} \cap w|}{|w| + |w_{Rel}|} \tag{7}$$

となる．ここで，$|w| + |w_{Rel}| \sim |w \cup w_{Rel}|$ のとき（通常，$|w \cap w_{Rel}| \ll \max(|w|, |w_{Rel}|)$ であるのでこれが成り立つ），Jaccard 係数が最大の語を求めることは，w_{Rel} が出現する文書を正解としたときの F 値が最大の語を求めていることになる．

[森 05a] の研究では，このような文脈を特定する単語を「コンテキストワード」と呼んでいる．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，氏名と共にクエリに加えることで，検索されたページからその人の人工知能に関連した語を抽出している．表 6 や表 7 の結果からも分かるように，単なる「訴訟」，あるいは「提携」のような 1 語より，2 つ以上の単語のペアが関係語として特定性が高く，Web 上でクエリとして相応しいことがわかる．

6·2 Web 上から企業間関係を抽出する可能性と限界

企業の関係で重要なものには，本論文で取り上げた提携関係や訴訟関係のほかに，株式の持ち合いや子会社・グループ会社といった資本的な関係，取引関係，役員派遣などの人的な関係，競合関係などがある．本論文で取り上げたのは，新製品開発やサービス提供開発における業務上の連携と，事業統合，営業譲渡，合併・買収などの資本的な関係，および，訴訟・係争関係であり，これら



(a) 関係ページの適合率　　(b) 関係のカバレージ

図 6　訴訟関係の関係語の評価

はニュースとして報道されることが多いため抽出が可能である．他にも，競合関係は製品の比較サイトなどで分かるかもしれないが，本論文とは異なるアルゴリズムになるだろう．取引関係や資本関係，人的な関係は，Web に書かれていることもあるが，そうでないことも多いと予想されるので，本研究では取り上げなかった．

本論文で取り上げなかった関係についても，どの程度 Web から抽出可能であるか，今後アルゴリズムの拡張を行っていきたいと考えている．具体的には，より多様な企業間の関係のページをヒットさせるための検索クエリを OR や NOT 等の条件も加えながら探索的に見つけていく手法の構築，さらに収集されたページから企業間の関係の有無をより正確に判断するために文の係り受け解析や意味解析と組み合わせていくこと，表形式のページに対応することなどが考えられる．

6·3 関 連 研 究

社会ネットワークは，セマンティック Web における情報の信頼性の計算 [Golbeck 05, Massa 05]，クチコミマーケティングの分析 [Leskovec 05]，情報の共有・推薦 [Mori 05b, Ghita 05]，コミュニティ抽出 [Newman 04]，オントロジー抽出 [Mika 05b] など，近年多くの研究で着目されている．

Referral Web[Kautz 97a, Kautz 97b] では，2 人の人間関係の強さを，Web 上における 2 人の氏名の共起頻度の強さによって計算している．Mika らが開発した Flink というシステム [Mika 05a] では，Email，FOAF，書誌情報および Web 全体から関係情報を調べ，社会ネットワークを視覚化している．松尾らは，POLYPHONET というシステムを開発し，研究者のネットワークを用いて学会等でのコミュニケーション支援に用いている [Matsuo 06b, Matsuo 06a]．これらの研究では研究者が対象になっており，検索ヒット件数をそのまま用いてもある程度有効であるが，企業の場合は Web 上のメディア効果というべき現象が顕著で，注目される関係とそうでない関係の差が激しい．したがって，より詳細に企業間にどういった関係があるかを同定し，個々の関係を総体的にみて紐帯の強さを測ることが必要であろう．本研究は，このような方向に向けたひとつのアプローチを示している．

クエリにどういった語を加えればよいかについては，[Oyama 04] らの研究がある．特定領域の情報だけを検索するために，あらかじめその領域特定の検索語（キーワードスパイスと呼ばれる）を学習しておいて，入力されたクエリにそれぞれの検索語を加えることで，ドメイン限定の検索を可能にする．キーワードスパイスは学習用のページを集め，決定木を用いて学習する．Mika らの Flink システムでは名前のあいまいさを解消するために，「*Semantic Web* OR *ontology*」というキーワードをクエリに加えている．Bollegara らの研究 [Bollegara 06] はこのキーワードを自動で獲得するため，氏名で検索してヒットされる上位のページをクラスタリングすることで，同姓同名の問題を解決している．

企業のネットワークを抽出して分析する研究にはさまざまな研究がある．例えば，稲岡らは金融機関の振替による決済記録から資金取引ネットワークを抽出して，金融システムの安定性や特徴を分析している [稲岡 03]．相馬らは大株主データを用いて，上場企業もしくは店頭登録企業に関して，日本の株所有ネットワークの遷移と特徴について分析している [相馬 05]．本研究では，手軽にアクセスできる Web 上の公開情報を用いて企業関係のネットワークを抽出しており，企業の動きを早期に的確に捉える分析方法としての可能性を秘めていると考えている．時系列的な変化については，今後の課題のひとつである．

7. む　す　び

本稿では，Web 上の情報から企業間関係を抽出する手法について述べた．入力された企業リストに対して，検索エンジンを利用して Web 中にある関係のページを収集し，関係のネットワークを構成する．企業間の特定の関係に絞るために，関係語と呼ぶ語を検索クエリに加え，目的の関係だけを抽出する．本稿では，特に訴訟と提携という関係に焦点を当てて手法を述べたが，基本的には企業間の多様な関係を抽出することが可能である．今後

は，企業間の関係を定期的に抽出することにより，業界や地域の企業間関係の変化や動向を分析する手法につなげていきたいと考えている．

◇ 参 考 文 献 ◇

[Adar 04] Adar, E., Zhang, L., Adamic, L. A., and Lukose, R. M.: Implicit Structure and the Dynamics of Blogspace, in *WWW 2004 Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics* (2004)

[Bollegara 06] Bollegara, D., Matsuo, Y., and Ishizuka, M.: Extracting key phrases to disambiguate personal names on the Web, in *Proc. CICLing 2006* (2006)

[Ghita 05] Ghita, S., Nejdl, W., and Paiu, R.: Semantically Rich Recommendations in Social Networks for Sharing, Exchanging and Ranking Semantic Context, in *Proc. ISWC05* (2005)

[Golbeck 05] Golbeck, J. and Hendler, J.: Inferring Trust Relationships in Web-Based Social Networks, *ACM Transactions on Internet Technology*, Vol. 7, No. 1 (2005)

[稲岡 03] 稲岡 創, 二宮 拓人, 清水 季子, 高安 秀樹：金融機関の資金取引ネットワーク, Technical Report ワーキングペーパー 2003-J-2, 日本銀行金融市場局 (2003)

[Kautz 97a] Kautz, H., Selman, B., and Shah, M.: The Hidden Web, *AI magazine*, Vol. 18, No. 2, pp. 27–35 (1997)

[Kautz 97b] Kautz, H., Selman, B., and Shah, M.: Referral Web: Combining Social Networks and Collaborative Filtering, *Communications of the ACM*, Vol. 40, No. 3, pp. 63–65 (1997)

[Leskovec 05] Leskovec, J., Adamic, L. A., and Huberman, B. A.: The Dynamics of Viral Marketing (2005), http://www.hpl.hp.com/research/idl/papers/viral/viral.pdf

[Massa 05] Massa, P. and Avesani, P.: Controversial Users demand Local Trust Metrics: an Experimental Study on Epinions.com Community, in *Proc. AAAI-05* (2005)

[松尾 05] 松尾 豊, 友部 博教, 橋田 浩一, 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, Vol. 20, No. 1E, pp. 46–56 (2005)

[Matsuo 06a] Matsuo, Y., Hamasaki, M., Takeda, H., Mori, J., Bollegala, D., Nakamura, Y., Nishimura, T., Hasida, K., and Ishizuka, M.: Spinning Multiple Social Networks for Semantic Web, in *Proc. AAAI-06* (2006)

[Matsuo 06b] Matsuo, Y., Mori, J., Hamasaki, M., Takeda, H., Nishimura, T., Hasida, K., and Ishizuka, M.: POLYPHONET: An advanced social network extraction system, in *Proc. WWW 2006* (2006)

[Mika 05a] Mika, P.: Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks, *Journal of Web Semantics*, Vol. 3, No. 2 (2005)

[Mika 05b] Mika, P.: Ontologies are us: A unified model of social networks and semantics, in *Proc. ISWC2005* (2005)

[森 05a] 森 純一郎, 松尾 豊, 石塚 満：Web からの人物に関するキーワード抽出, 人工知能学会論文誌, Vol. 20, No. 5, pp. 337–345 (2005)

[Mori 05b] Mori, J., Ishizuka, M., Sugiyama, T., and Matsuo, Y.: Real-world Oriented Information Sharing Using Social Networks, in *Proc. ACM GROUP'05* (2005)

[Newman 04] Newman, M. E. J. and Girvan, M.: Finding and evaluating community structure in networks, *Physical Review E*, Vol. 69, p. 026113 (2004)

[Oyama 04] Oyama, S., Kokubo, T., and Ishida, T.: Domain-Specific Web Search with Keyword Spices, *IEEE TKDE*, Vol. 16, No. 1, pp. 17–27 (2004)

[佐藤 01] 佐藤 理史：ワールドワイドウェブを利用した住所探索, 情報処理学会論文誌, Vol. 42, No. 1, pp. 59–67 (2001)

[Scott 00] Scott, J.: *Social Network Analysis: A Handbook (2nd ed.)*, SAGE publications (2000)

[相馬 05] 相馬 亘：経済における複雑系ネットワーク –日本の経済ネットワークは特殊か？–, 人工知能学会誌特集, Vol. 20, No. 3, pp. 289–295 (2005)

[立石 04] 立石 健二, 石黒 義, 福島 俊一：インターネットからの評判情報検索, 人工知能学会学会誌, Vol. 19, No. 3 (2004)

[安田 97] 安田 雪：社会ネットワーク分析 –何が行為を決定するか–, 新曜社 (1997)

[金光 03] 金光 淳：社会ネットワーク分析の基礎 –社会的関係資本論にむけて–, 勁草書店 (2003)

[坂田 05] 坂田 一郎, 柴田 尚樹, 小島 拓也, 梶川 裕矢, 松島 克守：地域経済圏の成長にとって最適な地域ネットワークとは— Small-World Networks の視点による４地域クラスターの比較分析—, 一橋ビジネスレヴュー, Vol. 53, No. 3, pp. 182–195 (2005)

[増田 06] 増田 直紀, 今野 紀雄：「複雑ネットワーク」とは何か, University of Toronto (2006)

[湯田 05] 湯田 聴夫, 藤原 義久：SNS における人のネットワーク構造～その地平線の超え方～, Web が生み出す関係構造と社会ネットワーク分析ワークショップ (2005)

[藤井 04] 藤井 敦：百科事典としてのWWW, 人工知能学会誌, Vol. 19, No. 3, pp. 296–301 (2004)

〔担当委員：伊藤 公人〕

2006 年 6 月 3 日　受理

著 者 紹 介

金　英子(正会員)

2001 年（中国上海）華東師範大学物理学部卒業．同年騰龍計算機軟件（上海）有限公司入社．2006 年東京大学大学院情報理工学系研究科修士課程終了．現在，同大学院博士課程在学中．Web マイニング，言語処理等に興味がある．言語処理学会会員

松尾　豊(正会員)

1997 年 東京大学工学部電子情報工学科卒業．2002 年 同大学院博士課程修了．博士（工学）．同年より，産業技術総合研究所 情報技術研究部門 勤務，2005 年 10 月よりスタンフォード大学客員研究員．人工知能、特に高次 Web マイニングに興味がある．人工知能学会、情報処理学会，AAAI の各会員．

石塚　満(正会員)

1971 年東京大学工学部電子卒，1976 年同大学院博士修了．工博．同年 NTT 入社，横須賀研究所勤務．1978 年東大大学生産技術研究所・助教授,（1980-81 年 Purdue 大学客員準教授），1992 年東京大学工工学部電子情報・教授，2001 年情報理工学系研究科・電子情報学専攻，2005 年同創造情報学専攻（電子情報学専攻兼任）．研究分野は人工知能，Web インテリジェンス，次世代 Web 情報基盤，生命的エージェントによるマルチモーダルメディア．IEEE, AAAI, 情報処理学会，電子情報通信学会，映像情報メディア学会，画像電子学会，等の会員．