

大域知能：Webからのコミュニティ研究分野マイニング

石塚 満 （協力者：浅田 洋平，松尾 豊）
情報理工学系研究科 電子情報学専攻

概要

情報流通，共有の基盤になってきた WWW(Web) に知的能力を付与し，社会の広域知能基盤に成長させることに向けての以下のような研究を進めた． Web テキスト文からのキーワード抽出，要約， Web 情報の変化の検知と変化分の要約（ETTS システム， ChangeSummarizer システム）， Web からの用語説明のマイニング， 自然言語に近い表現で推論機能をもつ知識表現言語 KRNL による Web テキスト情報の知識化と利用， Web から人間関係ネットワークの構築．

本報告では Web から人間関係ネットワーク構築に関連して行った，研究者がどのような研究分野で活動しているかを見出しクラスタ化し，Web 情報から研究者のクラスタとその研究分野をマイニングする手法と，作成したシステムについて紹介する．

1 まえがき

仕事や研究といった知的生産活動を効率的に進める上で「この仕事を誰に任せるか」「このテーマで共同研究するなら誰か」ということは，重要な問題である．効率の良い共同研究とは，それぞれの研究者がそれぞれに違う分野の詳しい知識を持ち，一人の研究者では足りない知識を他の研究者によって補う，という相補的なものであろう．すなわち，研究を効率よく進めるためには，自分の研究分野以外の領域において，誰がどのような知識を持っているか，ということについても把握しておく必要がある．

ところが，研究者にとって，自分の研究分野において誰がどのようなことを研究しているか，ということ把握できても，自分の研究分野以外の研究分野で誰がどのようなことを研究しているか，ということまで把握することは困難である．

一方，近年，学会のプログラムや論文誌の著者情報が Web 上に公開されたり，研究者が自分の研究をホームページで紹介したりするようになってきているので，Web 上には研究者とその研究分野に関する情報が最新のものを含めて豊富に存在している．これらの情報から特定の研究分野において誰がどのような研究をしているのか，という情報を抽出できれば，その研究分野の最新の研究動向を把握するのに役立つと考えられる．

本稿では，Web 上の情報を用いて研究分野を発見する手法を提案する．提案手法では，Web ページにおける研究者名と研究キーワードの共起に着目し，統計的な処理およびクラスタリングを行うことで，研究者のクラスタとして研究分野を発見する．

2 提案手法

本節では，Web 上の情報を用いて研究分野を発見する手法（以下「提案手法」）について説明する．

2.1 基本的な考え方

ある研究者がどのような研究分野の知識を持っているか，ということを探るとき，Web を用いて次のようなことを行ってみることはよくあるだろう．まず「石塚満 AND 仮説推論」というクエリで Web 検索を行うと 219 件のページが検索される．次に「石塚満 AND ロボティクス」というクエリで Web 検索を行うと 47 件のページが検索される．同一 Web ページに出現することを Web 上での共起と呼ぶことにすると「石塚満」「仮説推論」の方が「石塚満」「ロボティクス」よりも Web 上でよく共起していることが分かる．このことから，石塚満氏はロボティクスよりも仮説推論に関係の強い人，すなわちロボティクスよりも仮説推論の知識を持った研究者ではないかと推測できる．実際，

石塚満氏は仮説推論の研究を長年行っているのので、この推測は当たっている。

このように、提案手法の基本的な考え方は、研究者名と Web 上でよく共起する語は、研究者の研究分野と関係が深い、というものである。

以上の議論における「仮説推論」や「ロボティクス」のような研究内容を表す単語を「研究キーワード」と呼ぶことにする。

2.2 提案手法の概要

以上のような考え方に基づく提案手法の概要は、次のようなものである。

1. Web 検索エンジンを用いた研究者名と研究キーワードの共起行列の抽出
2. χ^2 値による共起の重みの計算
3. 研究者のクラスタリング

それぞれの部分について、以下に詳しく説明する。

2.3 Web からの研究者名と研究キーワードの共起行列の抽出

情報検索では、全文書がしばしば文書行列として表される。これはベクトル空間モデル [4] としてよく知られているアプローチである。ベクトル空間モデルでは、全文書から特徴として内容語を抽出し、各文書をこの特徴との関係の深さを表す重みベクトルとして表現する。一般的に、文書 i における単語 j の重み w_{ij} は、式 (1) のように tfidf 法を用いて計算される。

$$w_{ij} = tf_{ij} \times \left(\log \frac{N}{df_j} + 1 \right) \quad (1)$$

ただし、 tf_{ij} は文書 i における単語 j の出現頻度 (term frequency)、 df_j は、単語 j の出現する文書数 (document frequency)、 N は全文書数である。定性的には、 tf の効果により、文書中によく出現する単語ほど重みは大きくなり、 idf の効果により、特定の文書中に局所的に出現する単語ほど、その文書における重みは大きくなる。

同様に、研究者を研究キーワードの重みベクトルとして表すことを考える。文書の場合には、文書中

に出現する語の出現頻度によってベクトルとして表現する手法が一般的に多く用いられているが、研究者の場合にはこのような直接的な方法は使えない。そこで、提案手法では、あらかじめ研究キーワードのリストを用意しておき、Web における研究キーワードとの共起に基づいて研究者を表現することを考える。

2.1 節で述べたように、提案手法では、研究者名と Web 上でよく共起する研究キーワードは、研究者の研究分野と深い関係がある、と考えるので、研究者名と研究キーワードが共起する Web ページ数を重みとして、研究者を研究キーワードの重みベクトルで表現する。この重みは「研究者名 AND 研究キーワード」をクエリとして Web 検索を行ったときのヒット件数として得ることができる。

2.4 χ^2 値による共起の重みの計算

Web 検索におけるヒット件数をそのまま重みにすると、一般的な語に大きな重みを与えすぎるという問題がある。例えば、岡崎直観氏は著者の研究室の先輩で、自動要約の研究を主になさっている方だが、Web 検索エンジンで「岡崎直観 AND 要約」で検索すると 52 件、「岡崎直観 AND 画像」で検索すると 43 件のページがヒットし、ヒット件数の上ではあまり差がない。これは「画像」というキーワードが Web 上によく出現する一般的な語であるということが原因である。

そこで、提案手法では、研究キーワードとの共起ヒット件数そのものではなく、そのヒット件数が他の研究者との共起ヒット件数に比べてどの程度多いのか、という共起の偏りの強さによって研究キーワードを重みづけする。

偏りの強さを表す統計量として、最も一般的なものの一つに、 χ^2 値がある。 χ^2 値は、一般的に式 (2) のように定式化される。

$$\chi^2 = \frac{(\text{観測値} - \text{期待値})^2}{\text{期待値}} \quad (2)$$

本問題の場合に即して、観測値、期待値について説明する。研究者名と研究キーワードの共起行列が図 1 のように得られたとする。 i 番目の研究者名 $name_i$ と j 番目の研究キーワード $keyword_j$ が共起する Web ページ数 O_{ij} が、それぞれの観測値である。

図 1 における s_{ni} 、 s_{kj} はそれぞれ行ごとの和、列ごとの和であり、 S は全要素の和であり、式 (3)、(4)、

	keyword ₁	keyword ₂	...	keyword _j	...	total
name ₁	O ₁₁	O ₁₂	...	O _{1j}	...	s _{n1}
name ₂	O ₂₁	O ₂₂	...	O _{2j}	...	s _{n2}
⋮	⋮	⋮		⋮		⋮
name _i	O _{i1}	O _{i2}	...	O _{ij}	...	s _{ni}
⋮	⋮	⋮		⋮		⋮
total	s _{k1}	s _{k2}	...	s _{kj}	...	S

図 1: 共起行列

(5) のように書ける .

$$s_{ni} = \sum_j O_{ij} \quad (3)$$

$$s_{kj} = \sum_i O_{ij} \quad (4)$$

$$S = \sum_{ij} O_{ij} = \sum_i s_{ni} = \sum_j s_{kj} \quad (5)$$

共起行列の列ごとの和 s_{kj} を見れば、与えられた研究者に対して一般的によく共起する語ほど、大きな値になっているので、そのような語との共起ヒット件数が増えることが期待される . そので、観測値 O_{ij} に対する期待値 E_{ij} を、 s_{ni}, s_{kj}, S を用いて式 (6) のように計算する .

$$E_{ij} = s_{ni} \times \frac{s_{kj}}{S} \quad (6)$$

以上により、共起行列の重み χ_{ij}^2 は、観測値 O_{ij} と期待値 E_{ij} を用いて式 (7) のように計算できる .

$$\chi_{ij}^2 = \frac{(O_{ij} - E_{ij})^2}{O_{ij}} \quad (7)$$

ただし、式 (7) において、 $O_{ij} - E_{ij}$ が負になる場合にも、 χ^2 値は正の値を取るが、本研究の目的は、研究者を、その人の研究分野に関係したキーワードで重み付けることであるから、期待値よりも観測値が小さい場合の χ^2 値は 0 とした . すなわち、式 (8) のようになる .

$$\chi_{ij}^2 = \begin{cases} \frac{(O_{ij} - E_{ij})^2}{O_{ij}} & O_{ij} - E_{ij} > 0 \\ 0 & O_{ij} - E_{ij} \leq 0 \end{cases} \quad (8)$$

2.5 研究者のクラスタリング

ベクトル空間モデルでは、一般的に、文書の類似度を文書ベクトルのコサインで定義する . 同様に、本問題の場合では研究者が研究キーワードの重みベクトルとして表現されているので、研究者の研究分野の点での類似度をベクトルのコサインで定義できる . すなわち、研究者 a, b のベクトルを v_a, v_b と書くと、研究分野の点での類似度 $Sim(a, b)$ は、式 (9) のように書ける .

$$Sim(a, b) = \frac{v_a \cdot v_b}{|v_a||v_b|} \quad (9)$$

研究者の類似度が定義されると、最大距離法や重心法などの一般的な文書クラスタリング手法や、行と列を同時にクラスタリングする Co-clustering [3] など手法を用いて、研究者をクラスタリングすることができる .

3 評価実験

この節では提案手法の有効性を評価する実験について説明する .

3.1 研究者と研究キーワード

対象とした研究者は、JSAI2004 (2004 年度の人工知能学会全国大会) の参加者である . このリストは、Web 上に公開されており ¹、簡単に入手できる .

また、研究キーワードは、JSAI2003 の論文タイトルから形態素解析システム ChaSen ² によって名詞および未知語と判定されたものについて、出現回数が 3 回以上の 148 個を選んだ (図 2) .

3.2 実験 1: 研究キーワードの重み付け方法の比較

以上の研究者 540 人と研究キーワード 148 語に対して、共起の強さを計算するのに χ^2 を用いることの有効性を評価するために、研究キーワードの重みを

- 共起ヒット件数とした場合
- 文書行列の idf の考え方を導入して計算した場合

¹http://www-kasm.nii.ac.jp/jsai2004_schedule/person_list.html

²<http://chasen.aist-nara.ac.jp/hiki/ChaSen/>

支援, システム, 情報, 学習, データ, エージェント, 知識, Web, モデル, 環境, 対話, 構築, 型, 手法, ロボット, 抽出, 検索, 動作, 移動, 利用, 設計, 法, 評価, 分析, 行動, 物語, 協調, 生成, 概念, 発見, 空間, イベント, 制御, アルゴリズム, ネットワーク, ユーザ, 研究, テキスト, コミュニティ, 間, 知, 検討, 機能, 画像, 系列, 方法, 提案, 表現, 実現, 分類, 類似, 人間, サービス, タラ, イン, 関係, ク, 処理, 歩行, 共有, 機構, 考察, メタ, マイニング, コミュニケーション, 拡張, 時, 考慮, 自動, ション, 認識, 構造, 応用, 創造, 構成, マルチ, 適応, 作成, 階層, 障害, 自律, 日常, 議論, パターン, クラスタリング, 開発, 決定, 実, 意味, 視覚, 演奏, 認知, オントロジー, 要約, 上, 選, 扱, 論, 管理, 連携, 知能, 言語, 遺伝, センサ, デザイン, インタラクティブ, 方式, コ, コンテンツ, 地図, ゲーム, 物体, 解, 意図, 獲得, 問題, 携帯, 強化, 映像, 配置, 推論, 知, 的, データベース, 戦略, 導入, 系, 運動, 確率, 判断, 活性, 検出, 状況, フレーム, 適, 用, 擬人, 言説, ページ, 文脈, 記録, 活動, 人工, 意思, アニメーション, 複数, 音声, 端末, 履歴, 統合, 文書, 度

図 2: JSAI2003 の論文タイトルから抽出した研究キーワード (148 語)

• χ^2 値で計算した場合

の 3 通りで共起行列を作成した。

ここで, 文書行列の idf と類似した方法とは, 式 (1) のアナロジーで式 (10) のように研究者 $name_i$ に対する研究キーワード $keyword_j$ の重み w_{ij} を定義したものである。

$$w_{ij} = O_{ij} \times \left(\log \frac{N}{pf_j} + 1 \right) \quad (10)$$

ただし, pf_j は, 研究キーワード $keyword_j$ が共起する研究者数 (*person frequency* とでも呼ぶべき値), N は全研究者数である。

ここで, それぞれの共起行列を用いた場合の研究者の研究分野の点での類似度が, どの程度, 現実合っているかを, JSAI2004 のセッションを用いて調べることを考える。学会のセッションは, 類似した発表をまとめたものであるので, 同じセッションに割り当てられた研究者は研究分野の点で類似度が高いはずである。逆に言えば, 研究分野の点で類似度の高い研究者ほど, 同じセッションに割り当てられる確率が高いはずである。

そこで, それぞれの共起行列を用いて, 研究者の類似度を行ベクトルのコサインによって計算し, 類似度に閾値 Sim_{th} を設定したときに, 類似度がその閾値以上の研究者のペアが, JSAI2004 で同じセッションで発表している確率 $P(Sim_{th})$ を計算した (JSAI2004 のセッションは, Web 上に公開されているものを用いた³⁾。

すなわち, $P(Sim_{th})$ は式 (11) のように計算できる。

$$P(Sim_{th}) = \frac{Sim_{th} \text{以上かつ, JSAI2004 で同じセッションのペアの数}}{Sim_{th} \text{以上のペアの数}} \quad (11)$$

³http://www-kasm.nii.ac.jp/jsai2004_schedule/paper_list.htm

類似度の閾値 Sim_{th} を 0 から 1 まで 0.01 刻みで変化させ, それぞれの共起行列を用いて $P(Sim_{th})$ を計算した。

3.3 実験 2: クラスタリング

χ^2 値によって重みを計算した共起行列を用いて, 類似度をコサインで計算し, 最大距離法によって研究者のクラスタリングを行い, クラスタ数が 30 個になった時点でクラスタリングを終了した。クラスタ数が 30 個であることは特に意味はないが, JSAI2004 ではセッション数が約 30 個であったので, 研究分野を発見する際にセッションよりも細かく分ける必要はないと考え, これとした。

4 結果と考察

本節では, 前節の評価実験の結果を示し, 考察を行う。

4.1 研究キーワードの重みを χ^2 値で計算することの有効性

実験 1 の結果を図 3 に示す。図 3 では, 研究キーワードの重みを共起ヒット件数とした場合 (hit) と, idf の考え方を導入して計算した場合 (idf) と, χ^2 で計算した場合 (χ^2) の 3 つの系列を重ねて表示している。

図 3 より, 共起ヒット件数や, idf に類似した方法で研究キーワードを重み付けした場合には, 類似度が 0.9 だと判定した二人が同じセッションにいる確率が 0.1 にも満たないので, 現実合っているとはいえない。これに対して, 研究キーワードの重みを χ^2 値で計算した場合は, 類似度の高い研究者ほど同じセッ

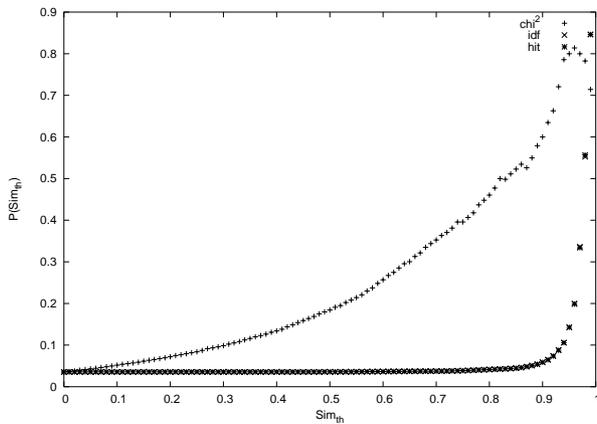


図 3: 類似度の閾値 Sim_{th} と、類似度が閾値以上の二人が同じセッションである確率 $P(Sim_{th})$ の関係

セッションに割り当てられており、 χ^2 値で重み付けした共起行列が、現実における研究分野の類似性をよく表していることが分かる。

研究キーワードを共起ヒット件数で重み付けした場合と、idfに類似した方法で重み付けした場合の系列がほぼ重なっているが、これは、Web 上には様々な情報があるので、ほとんどの研究者が、なんらかの形でほとんどの研究キーワードと 1 件以上共起してしまったことが原因であると考えられる。

2.4 節で、自動要約の研究者である岡崎直観氏を例にとって、「岡崎直観 AND 要約」のヒット件数が 52 件、「岡崎直観 AND 画像」のヒット件数が 43 件であり、あまり差が出ない、ということを示したが、本実験における研究キーワードの重みを χ^2 値で計算した場合には、「岡崎直観 AND 要約」では 194.11、「岡崎直観 AND 画像」では 0 となり、差がはっきりと分かり、岡崎直観氏は要約の研究者であることがよく表れている。

以上のことから、Web から抽出した研究者名と研究キーワードの共起行列に対して、 χ^2 値によって研究キーワードを重み付けすることの有効性が示された。

4.2 クラスタリング結果

研究キーワードの重みを χ^2 値で計算した共起行列を用いて、類似度をコサインで計算し、最大距離法によってクラスタリングを行い、30 個のクラスタを抽出

した結果を、図 4 に示す。ただし、紙面の都合上、全ての結果を示すことができないので、一部のクラスタについてのみ紹介するにとどめる。

図 4 では、いくつかのクラスタの研究者と、そのクラスタの重心ベクトルにおいて、重みの大きな順に上位 10 個の研究キーワードを表示している。すなわち、クラスタが研究分野の研究者を表し、重みの大きな語がその研究分野における主要な研究キーワードを表している。個別のクラスタに対する考察は主観的なものにならざるを得ないが、クラスタ 3 は「マイニング」、クラスタ 8 は「セマンティック Web とオントロジー」、クラスタ 22 は「ロボット」のようにまとめられる。以上のように、研究者のクラスタとして、研究分野が発見できる。

4.3 研究キーワードの選択方法

研究キーワードは、研究者の研究分野を適切に表し、かつ他の研究者と比較が可能な語が望ましい。すなわち、極端な低頻度語は、他の研究者との比較ができないので望ましくないし、研究内容に関係のない一般語も望ましくない。

本稿の評価実験では、論文のタイトルから頻度が 3 回以上の語を研究キーワードとして用いたが、論文のタイトルは、研究内容が一目で分かるように書かれたものなので、研究分野を適切に表していると考えられるし、頻度に 3 回以上、という制限を設けたことで、極端な低頻度語を除いていると言える。ただし、研究内容に関係のない一般語に関しては、図 2 を見れば分かるように「型、法、間、知、時、動」など多くの語が含まれており、これらの語に対しては、ストップワードのリストを作って除くなどの対策を講じる必要があると考えられる。

5 関連研究

我々は、Web 上の情報を用いて研究者の人間関係を抽出する手法について研究を行ってきた [1, 2]。この手法を用いれば、任意の二人の研究者の関係の強さを定義し、その二人の関係が共著関係、同じ研究室のメンバーである関係、同じ研究会などで発表したことのある関係、同じプロジェクトで研究している関係のど

<p>クラスタ 3: 鳥居大祐, 中田豊久, 細井一弘, 相原健郎, 中井隆洋, 豊宏野, 許宰源, 浅田洋平, 久保山哲二, 松尾豊, 友部博教, 藤村滋, 虫明磨毅, 足立史宜, 小野田崇, 猪口明博, 北口真也, 佐藤芳紀, 阿部秀尚, 有村博紀, 浅井達雄, 鹿島久嗣, 茂木明, 鷲尾隆, 元田浩, 村田剛志, 市瀬龍太郎, 山口高平, 久米俊二, 渡邊悠司, 津田宏, 朝日秀真, 許山明弘, 八木龍平, 岩垣守彦, 阿部真美子, 曾我真人, 藪内佳孝, 宮崎和光, 長行康男, 伊藤良彦</p> <p>重みの大きな語: マイニング, 抽出, オントロジー, 発見, 学習, メタ, 知能, 強化, テキスト, コミュニティ</p>
<p>クラスタ 8: コバアレクサンダー, 王凱軍, 小出誠二, 溝口理一郎, 和泉憲明, 武田英明, 岩爪達昭, 小路悠介, 森田武史, 垂見晋也, 來村徳信, 古崎晃司, 見置孝昌, 繁田佳宏, 池田満, 武内雅宇, 渡邊英一, 川村正則, 林雄介, 田中庸平, 金崎弘文, 大森俊秀, 瀧井隆道, 白鳥成彦, 大向一輝, 濱崎雅弘, 谷口智哉, 沼見介, 河合由起子</p> <p>重みの大きな語: オントロジー, コミュニティ, 知能, 共有, 支援, 人工, 構築, 統合, 創造, コンテンツ</p>
<p>クラスタ 22: 加納政芳, 吉田宏徳, 小笠原嘉清, 田島敬士, 嵐山誠, 原口一馬, 大村英史, 今井倫太, 尾形哲也, 辻康孝, 早田浩二, 川野洋, 谷口忠大, 榎木哲夫, 小山慎哉, 香山健太郎, 矢入(江口)郁子, 猪木誠二, 相野聡彦, 村井佐知子, 渋谷恒司, 滝健太, 山本知幸, 今福啓, 近藤英二, 柿倉正義, 岡田昌史, 中村仁彦, 宮城政雄, 工藤康生, 村井哲也, 大崎美穂, 鬼沢武久, 高木友博, 石原一志, 中置一博, 辻野広司, 中谷智広, 山田孝治, 藤吉弘昌, 谷内田正彦, 角所考, 横谷八郎, 神谷祐樹, 藤藤蒙, 奈良博之, 小泉智史, 蔵田武志</p> <p>重みの大きな語: ロボット, 歩行, 物体, 制御, 視覚, センサ, 自律, 端末, クラスタリング, 検出</p>

図 4: JSAI2004 参加者のクラスタリング結果の一部(研究者のクラスと、各クラスターの重心における重みの大きな上位 10 語)

れにあてはまるか(もしくはどれにもあてはまらないか)を判定することができ、研究者の人間関係をネットワークとして図示することができる。現在我々はこの手法に基づく研究者検索システムを開発中であり、本研究の成果も、このシステムにおいて特定の研究分野の研究者を探すときの手がかりとして用いることができるのではないかと考えている。

Web 上の情報を用いてエキスパートを発見する研究はこれまでもいくつか行われてきた。Referral web は social network を用いた情報検索システムである [5, 6, 7]。名前を与えられると、システムはユーザの周辺の social network を表示する。それぞれのユーザはエージェントを持ち、エージェントが social network 上でユーザの興味にあった topic words を持つエキスパートを見つける。Referral web が一人一人のユーザに着目してそれぞれの topic words を抽出しているのに対して、私たちの提案手法では、統計的手法を用いて研究キーワードの重みを計算することで、全研究者の中で各研究者を特徴付けている。この点が異なる。

6 まとめと今後の課題

本稿では、研究者名と研究キーワードの Web 上での共起に基づいて、統計的处理およびクラスタリングを行うことで、研究者のクラスタとして研究分野を発見する手法を提案した。研究キーワードの重みを計算する際に χ^2 を用いることの妥当性を示し、クラスタリング結果について、得られたいくつかのクラスタから、研究分野が発見できることを示した。

今後の課題としては、クラスタリング結果に対する客観的な評価を行うことがあげられる。また、本稿では研究者の分野発見を取り上げているが、研究者以外にも、例えばミュージシャンや小説家の場合には、作

品の紹介文やレビューなど、多くの情報が Web 上に存在しているので、今後は様々な方面に適用できると考えている。

参考文献

- [1] Y. Matsuo, H. Tomobe, K. Hashida and M. Ishizuka: Mining Social Network of Conference Participants from the Web, 12th Int'l WWW Conf., 2003.
- [2] 松尾豊, 友部博教, 橋田浩一, 石塚満: Web 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, 20(1E), pp.46-56, 2005.1.
- [3] I. S. Dhillon: Co-clustering documents and words using Bipartite Spectral Graph Partitioning. Proc. of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(KDD), 2001.
- [4] G. Salton and M. J. McGill: Introduction to Modern Information Retrieval, McGraw-Hill Book Company, 1983.
- [5] H. Kautz, B. Selman, and M. Shah: ReferralWeb: combining social networks and collaborative filtering, Communications of the ACM, 40(3), 1997.
- [6] H. Kautz, B. Selman, and A. Milewski: Agent amplified communication, Proc. of the National Conference on Artificial Intelligence, pp.3-9, 1999.
- [7] H. Kautz, B. Selman, and M. Shah: The Hidden Web, AI Magazine, 18(2), pp.27-36, 1997.

主要発表文献

松村真宏, 三浦麻子, 柴内康文, 大澤幸生, 石塚満: 2ちゃんねるが盛り上がるダイナミズム, 情報処理学会論文誌, Vol.45, No.3, pp.1053-1061 (2004.3).

松尾豊, 友部博教, 橋田浩一, 石塚満: Web 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 (2005.1).

- Y. Matsuo and M. Ishizuka: Keyword Extraction from a Single Document using Word Co-occurrence Statistical Information, *Int'l Journal on Artificial Intelligence Tools*, Vol.13, No.1, pp.157-169 (2004.3)
- Helmut Prendinger, Sylvain Descamps, and Mitsuru Ishizuka: MPML: A Markup Language for Controlling the Behavior of Life-like Characters, *Journal of Visual Languages and Computing*, Vol.15, No.2, pp.183-203 (2004.4).
- Helmut Prendinger, Junichiro Mori, and Mitsuru Ishizuka: Using Human Physiology to Evaluate Subtle Expressivity of a Virtual Quizmaster in a Mathematical Game, *Int'l Journal of Human-Computer Studies*, Vol.62, pp.231-245 (2005).
- Helmut Prendinger and Mitsuru Ishizuka: What Affective Computing and Life-like Character Technology Can Do for Tele-Home Health Care. Workshop on HCI and Homecare: Connecting Families and Clinicians (Online Proceedings), in conj. with CHI-04, Vienna, Austria (2004.4).
- Adam Jatowt, Khoo Khyou Bun, and Mitsuru Ishizuka: Change Summarization in Web Collections, Proc. 17th Int'l Conf. on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems (IEA/AIE 2004), Ottawa, Canada, Lecture Notes in Computer Science, LNCS 3029, Springer, pp.653-662, (2004.5)
- Adam Jatowt and Mitsuru Ishizuka: Web Page Summarization Using Dynamic Content, Poster, Proc. 13th Int'l World Wide Web Conf. (WWW04), New York, USA, pp.344-345, (2004.5)
- Gakuto Kurata and Naoaki Okazaki and Mitsuru Ishizuka: GDQA: Graph Driven Question Answering System - NTCIR-4 QAC2 Experiments -, Working Notes of the Fourth NTCIR Workshop Meeting (NTCIR-4), Tokyo, Japan, pp.338-344, (2004.6)
- Naoaki Okazaki and Yutaka Matsuo and Mitsuru Ishizuka: TISS: An Integrated Summarization System for TSC-3, Working Notes of the Fourth NTCIR Workshop Meeting (NTCIR-4), Tokyo, Japan, pp.436-443, (2004.6)
- Helmut Prendinger, Hiroshi Dohi, Hua Wang, Sonja Mayer, and Mitsuru Ishizuka: Empathic Embodied Interfaces: Addressing Users' Affective State, Tutorial and Research Workshop on Affective Dialogue Systems (ADS-04), Kloster Irsee, Germany, pp.53-64 (2004.6)
- Naiwala P. Chandrasiri, Istvan Barakonyi, Takeshi Naemura, Mitsuru Ishizuka and Hiroshi Harashima: Internet Communication Using Real-Time Facial Expression Analysis and Synthesis, *IEEE Multimedia* (Special issue on Multisensory Communication and Experience through Multimedia), Vol.11, Issue 3, pp.20-29 (July-Sept. 2004)
- Naoaki Okazaki and Yutaka Matsuo and Mitsuru Ishizuka: Coherent Arrangement of Sentences Extracted from Multiple Newspaper Articles, *PRICAI 2004: Trends in Artificial Intelligence* (C. Zhang, H. W. Guesgen, and W. K. Yeap (Eds.)). Lecture Notes in AI, LNAI 3157, Springer, pp.888-891 (Proc. 8th Pacific Rim Int'l Conf. on Artificial Intelligence, Auckland, New Zealand), (2004.8)
- Naoaki Okazaki and Yutaka Matsuo and Mitsuru Ishizuka: Improving Chronological Sentence Ordering by Precedence Relation, Proc. 20th Int'l Conf. on Computational Linguistics (COLING 04), Geneva, Swiss, pp.750-756 (2004.8)
- Yutaka Matsuo, Hironori Tomobe, Koiti Hasida and Mitsuru Ishizuka: Finding Social Network for Trust Calculation, Proc. 16th European Conf. on Artificial Intelligence (ECAI2004), Valencia, Spain, pp.510-514 (2004.8)
- Junichiro Mori, Yutaka Matsuo, Mitsuru Ishizuka and Boi Faltings: Keyword Extraction from the Web for FOAF Metadata, Proc. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web, Galway, Ireland, pp.1-8 (2004.9)
- Yohei Asada, Yutaka Matsuo and Mitsuru Ishizuka: A method to automatically find foaf:Group based on the cooccurrence of people with keywords in the Web, Proc. 1st Workshop on Friend of a Friend, Social Networking and the Semantic Web, Galway, Ireland, pp.34-37 (2004.9)
- Yasubumi Nozawa, Hiroshi Dohi, Hitoshi Iba and Mitsuru Ishizuka: Humanoid Robot Presentation Controlled by Multimodal Presentation Markup Language MPML, Proc. 13th IEEE Int'l Workshop on Robot and Human Interactive Communication (RO-MAN2004), Kurashiki, Japan, No.026 (2004.9)
- Adam Jatowt and Mitsuru Ishizuka: Summarization of Dynamic Content in Web Collections, Proc. 8th European Conf. on Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2004), Pisa, Italy, Lecture Notes in Artificial Intelligence, LNAI 3202, Springer, pp.245-254, (2004.9)
- Junichiro Mori, Yutaka Matsuo, Mitsuru Ishizuka, Boi Faltings: Keyword Extraction from the Web for Creation of Person Metadata, Poster Abstracts 3rd Int'l Semantic Web Conference (ISWC2004), Hiroshima, Japan, pp.45-46, (2004.11)
- Junichiro Mori, Yutaka Matsuo, Mitsuru Ishizuka, Boi Faltings: Keyword Extraction from the Web for Person Metadata Annotation, ISWC Workshop Notes VIII (W8) - 4th International Workshop on Knowledge Markup and Semantic Annotation (Semannot2004) (in conjunction with 3rd Int'l Semantic Web Conference (ISWC2004)), Hiroshima, Japan, pp.51-60, (2004.11)
- Adam Jatowt and Mitsuru Ishizuka: Temporal Summarization of Web Pages, Proc. 5th Web Information Systems Conference (WISE 2004), Brisbane, Australia, Lecture Notes in Computer Science, LNCS 3306, Springer, pp.303-312, (2004.11)