

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 6 月 7 日現在

機関番号：23803

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500133

研究課題名（和文） エゴセントリック情報による全体ネットワーク構造の推定

研究課題名（英文） Network structure estimation based on egocentric information

研究代表者

武藤 伸明 (MUTOH NOBUAKI)

静岡県立大学・経営情報学部・准教授

研究者番号：40275102

研究成果の概要（和文）：本研究は、ソーシャルネットワークから収集可能なエゴセントリック情報より、全体ネットワーク構造を推定する手法の開発を目的とする。このネットワークの構造推定は NP-困難クラスに属する組合せ最適化問題を扱うことになり、その効率的な解法として遅延評価付き貪欲法の応用法を考案した。また、ネットワーク構造推定法の妥当性を評価するために、ネットワークの本質的構造を表す評価尺度の考案や、ネットワークデータを含む各種データの可視化法の考案を行った。

研究成果の概要（英文）：In this research, we aim to develop a method for estimating the network structure from egocentric information of the social network. Since this structural estimation of the network belongs to the NP-hard class, we designed a method based on a greedy algorithm equipped with lazy evaluation. In order to evaluate the validity of the network structure estimation method, we have devised evaluation measure of the essential structure of the network, and have devised visualization method of various data including the network data.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	2,100,000	630,000	2,730,000
2011年度	700,000	210,000	910,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：総合領域

科研費の分科・細目：情報学・知能情報学

キーワード：知識発見、データマイニング、ソーシャルネットワーク、エゴセントリック情報、最適化問題

1. 研究開始当初の背景

(1) 人間などの活動単位をノードとし、それら活動単位の関係をリンクとして表現したソーシャルネットワークにおいて、例えば大規模ネットワーク上での情報拡散に代表される分析などは、多様な経営問題や公共政策問題の解決、さらにはソーシャルイノベーション創出に向けて極めて重要な役割を果た

すと考えられる。ところが SNS (Social Networking Site) のような一部の限定されたインターネット上のコミュニティなどを除けば、現実社会においては、プライバシー等の問題から完全な全体ネットワーク構造を知ることは一般に困難な状況にある。実際、個人名を特定し、その個人の属性および全ての知人数などの情報を収集することには大

きな抵抗が予想される。よって、ソーシャルネットワークに関連した収集可能で多様な断片情報より、本来の全体ネットワーク構造をできるだけ精緻に推定することが強く望まれる。

(2) 本研究に関連する国内・国外の研究動向として、ネットワークの構造予測の観点では、リンク予測 (L. Getoor and C. P. Diehl, *Link mining: a survey, SIGKDD Explorations* 7(2): 3—12 (2005)) が挙げられる。このようなリンク予測研究の主要パラダイムでは、ネットワーク構造のほとんどが既知情報として与えられ、すなわち、エゴセントリック情報のみからは得られないデータを用いて、比較的少数の欠損リンクやネットワーク成長で生成されるリンクなどを予測する。

2. 研究の目的

(1) ソーシャルネットワークから収集可能なエゴセントリック情報より、本来の全体ネットワーク構造をできるだけ精緻に推定する手法の開発を目的とする。このようなネットワークの構造推定においては、必然的に NP-困難クラスに属する組合せ最適化問題を扱うことになり、その効率的な解法を考案する。

(2) 開発したネットワーク構造推定法の妥当性を評価するために、全体ネットワークの本質的構造を推定できているかを検証するための評価法を確立する。これを実現するために、ネットワークの本質的構造を表す評価尺度を考案する。

3. 研究の方法

(1) 社会調査から得られるエゴセントリック情報を詳細に分析し、エゴセントリック情報からの全体ネットワーク構造の推定問題を組合せ最適化問題と捉えて目的関数を構築し、定式化を行う。この組合せ最適化問題を高精度で高速に解くために、全体ネットワーク構造推定問題が有するサブモジュラ性に注目し、遅延評価付き貪欲法を用いた効率のよい解法を探求する。

(2) 推定ネットワークが、真のネットワークの本質的構造を推定できているか検証するための評価法を確立する。具体的には、ネットワークを特徴づける特性量を各種考案し、推定ネットワークが真のネットワークのそれら特性量を保持しているかどうかで、推定の妥当性の評価を行う。

(3) (1), (2) の研究を円滑に行うためのツールとして、各種データやネットワークの特徴を反映した可視化法についての研究を行う。

4. 研究成果

(1) 社会調査から得られるエゴセントリック情報を分析し、組合せ最適問題としての目的関数を構築し、全体ネットワーク構造の推定問題を定式化した。この推定問題は、NP-困難クラスに属し、厳密解を求めることが困難である。これについて貪欲法を用いた近似解法について探求した。特に、推定問題のサブモジュラ性に着目し、遅延評価付き貪欲法が有効な解法となることを示した。遅延評価付き貪欲法の各種問題への有効性を確認するために、ネットワーク上での施設配置問題、類似検索における遅延評価を用いたピボット選択等の各種問題の近似解法を、遅延評価付き貪欲法を用いて実装し、既知構造ネットワークを利用した計算機実験を通して、その有効性を検証した。

① 「乱歩モデルに基づく施設配置問題」においては、乱歩モデルに基づき行動するユーザを想定し、与えられた期間内で、設置された複数施設のどれかにユーザが少なくとも一度は到達する確率を考え、この確率を最大にするような施設集合を配置する問題を扱った。この問題の近似解を貪欲法で求める際に、この問題がサブモジュラ性を有することにより、遅延評価付き貪欲法の適用が可能である。

計算機実験では、静岡市の道路網ネットワークなどに対し、既存の手法と比較して遅延評価付き貪欲法の効果を検証した。図1には、既存の各手法と比較した遅延評価付き貪欲法 (GL法) の解品質が優れていることを示した。図1の横軸は設置施設数、縦軸の目的関数値は、ユーザが一度は施設に到達する確率である。図2においては、GL法が、遅延評価なしの貪欲法 (SG法) や、Kメディアン法 (KM法) と比較して、計算効率において優れていることを示した。

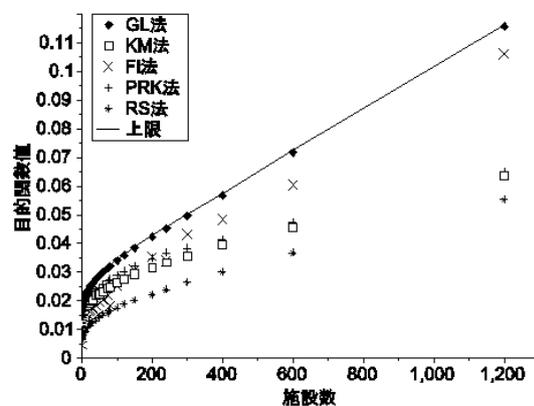


図1. 静岡市道路網についての解品質比較

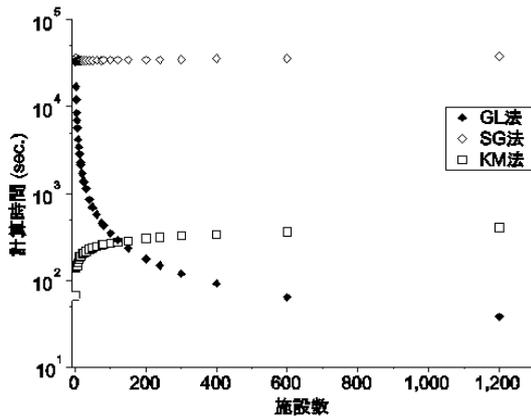


図2. 静岡市道路網についての計算時間比較

②ネットワーク上での施設配置問題についての別の問題設定として、施設が各利用者へ確率的に独立貢献する状況において、施設の開設コストを考慮して、利用者の期待利得の総和を最大化する問題を考える。この問題設定において、貪欲法による2種類の解法を考案した (CG法及びSG法)。これら2つの解法は、それぞれ遅延評価が適用でき (CGLE法およびSGLE法)、図3に示すように、大幅に計算時間を削減することができる。

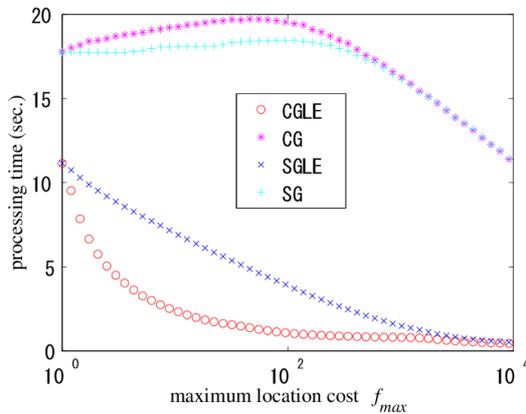


図3. ブログネットワークでの処理効率

③マルチメディアデータについて、与えられたクエリから類似したオブジェクトを検索する類似検索では、一部のオブジェクトをピボット集合として選定し検索を高速化する方法が提案されており、よいピボット集合の集合は検索性能の向上に結び付く。よいピボット集合の尺度として Bustos らの提案した評価関数があるが、この評価関数がサブモジュラ性を有することを示し、遅延評価付き貪欲法を用いて、ピボット集合選択問題を高速化した。図3は、単語データに対するピボット集合選択問題における、単なる貪欲法と遅延評価付き貪欲法の計算時間比較であり、遅延評価の効果が見て取れる。

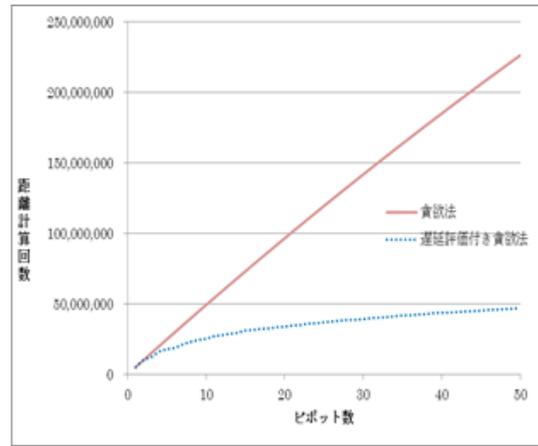


図4. 単語データに対する計算時間比較

(2) エゴセントリック情報からの全体ネットワーク構造推定問題においては、正確なネットワーク構造を推定することは不可能であり、推定結果に、ある種のノイズが含まれることは不可避である。いくらかノイズの含まれるネットワークにおいても、真のネットワークと同様のネットワークの特徴量を得ることができれば、推定したネットワークを用いて研究することが可能である。このためのネットワークの本質的構造を表す特徴量として、複数の評価指標を提案した。提案した評価指標は、ネットワークモチーフ、ノード集合に対する媒介中心性、コリンク構造、隣接数分布、入出隣接数相関、多重度平均、ジニ係数等である。現実のネットワークを用いた実験を通して、これらの評価指標について分析を行い、ネットワークの特徴を把握するための指標としての有効性を示した。

①「ノード集合に対する媒介中心性の提案」では、ネットワークにおけるノード部分集合を対象とし、集合としての中心性値が高くなるノード群の抽出、すなわち既存の中心性をノード集合に拡張定義し、特にノード集合に対する媒介中心性を提案した。提案指標は、個々のノードの媒介度ではなく、集合に含まれるいずれかのノードの媒介度の和を最大にするようにノードを選択するものである。この集合媒介中心性について、貪欲法による近似解によりノード集合を求め、さらに、目的関数を展開することにより導かれる漸化式を用いることによって、効率的な計算を行うことを示した。

②ネットワークにノイズを混入したネットワークモチーフの既存研究として、ノード削除やリンク削除などのノイズに関する研究があるが、本研究では、出次数・入次数を固定したリンク張替というノイズに対するモチーフ解析の頑健性を評価した。実験対象として、ウェブ上での大規模なソーシャルネッ

トワークについてノイズを混入し、真のネットワークにおいて特徴的なモチーフパターンが、ノイズを混入したネットワークでも特徴的であることを検証することによって、モチーフパターンの頑健性を示した。

図 5 は SNS (アメーバブログ) のリンク関係ネットワークにノイズを様々な比率で混入した場合のモチーフパターンの分析例である。この実験の場合にはノイズを $2^1 \sim 2^{10}$ まで比率を変えて混入させた。オリジナルのモチーフパターン (図中の \times) に対して、ノイズを 2^2 の比率で混入させても (図中の \square) モチーフパターンの特徴は保持されており、モチーフパターンの頑健性がわかる。

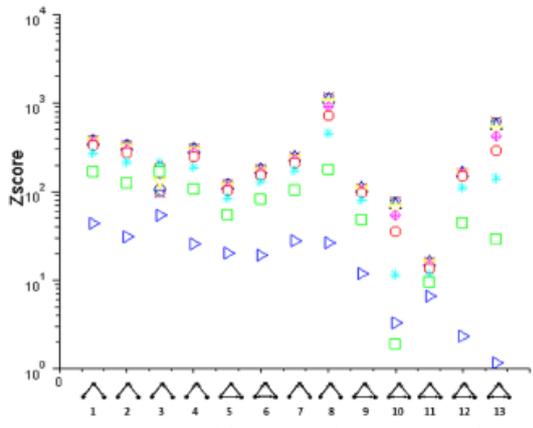


図 5. ノイズを混入した SNS におけるモチーフパターン分析

③ ソーシャルネットワークにおけるユーザ間のメッセージのやりとりは、単純グラフとみなして分析することも可能であるが、メッセージ数に応じた多重グラフとみなすこともできる。ソーシャルネットワークの高次数ノードの相互関係に着目し、多重グラフと単純グラフの分析結果間の差異について調査を行った。このために、コリンク率をベースとした従来の分析手法を多重グラフに適用できるように拡張した手法を提案した。Twitter のフォローネットワークと Mention ネットワークについて、高次数ユーザを対象に分析する場合、多重グラフを用いると、複数のメッセージのやり取りにより形成される高頻度コミュニケーションユーザを検出できることを示した。

図 6(a), (b) はそれぞれ、Twitter データから構築した多重 Mention ネットワークの入次数上位 1000 ノードの平均コリンク率および出次数上位 1000 ノードの平均コリンク率を示したものである。図 6(a) では、ランキングが下位になるにしたがって、平均コリンク率の高いノードが現れる。図 6(b) では、各ノードの平均コリンク率が低コリンクから高コリンクまで分散している。このように Mention ネットワークを多重グラフとして捉

えることによって、入次数上位のユーザと出次数上位のユーザの差異がわかる。

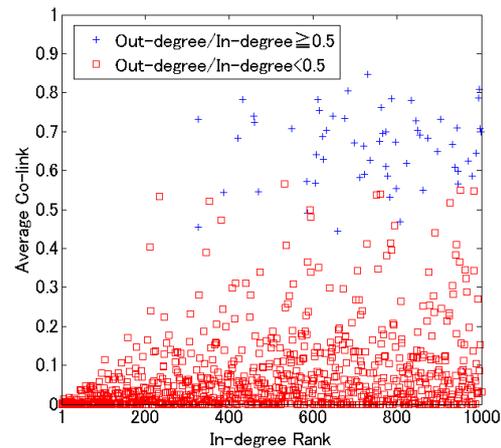


図 6(a) 多重 Mention NW の入次数上位 1000 ノードの平均コリンク率

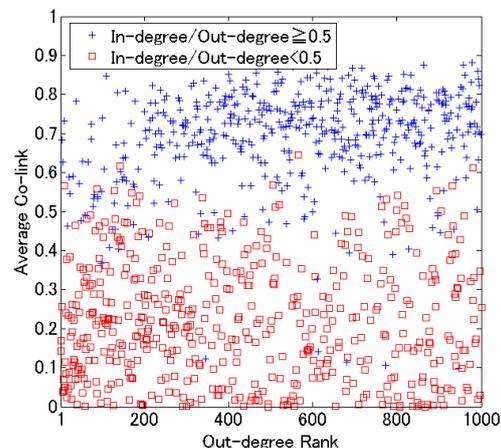


図 6(b) 多重 Mention NW の出次数上位 1000 ノードの平均コリンク率

④ ③と同様にソーシャルネットワークを多重有向グラフとして捉え、ソーシャルネットワークを特徴づける指標として、隣接数分布、入出隣接数相関、多重度平均、ジニ係数を用いて分析を行った。具体的には、Twitter のツイートをお気に入り登録して保管できる Favorites 機能に着目し、これによって形成される Favorites ネットワークについて分析した。同じく Twitter における機能の一つである Mentions 機能についても同様に分析し、結果を比較した。これによって、Favorites ネットワークと Mentions ネットワークの特徴の差異を定量的に把握することが可能となった。

(3) ネットワークやデータの効率的な解析を行うことを目的として、属性データの編集距離に基づくアノテーション付き可視化法、2 部グラフ可視化法の高速度化法、ピボットを

用いた類似度データの可視化法等を確立した。

①編集距離（レーベンシュタイン距離）に基づく可視化に着目し、アノテーション法により、低次元に埋め込まれたオブジェクト集合に対して、どの辺りにどのような共通の特徴を持つオブジェクトが布置されているかを自動的に示す方法を提案した。英単語データを用いた評価実験の結果により、この提案法を用いることで、可視化結果に対し、特定アルファベットを多く含む単語の配置方向として適切にアノテーションが付与できることを示した。図7はアノテーションの付与結果であり、特定アルファベットを含む単語集合に適切にアノテーションを付与することができることが示された。

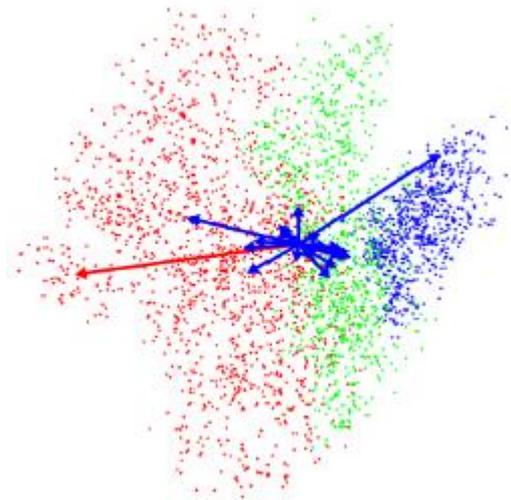


図7. 英単語データに対するアノテーション付与結果

②複数のカテゴリにまたがって所属するオブジェクト間の関係を表すために、2部グラフが頻繁に使われている。大規模かつ複雑な2部グラフの効率的な可視化法としてSpherical Embedding法（既存SE法）がある。しかし、既存SE法では非線形処理を必要とするため、計算量がかかるという問題がある。このことはネットワークの規模が大きく、複雑になるほど顕著である。本研究では、2部グラフ可視化法の高速度化を焦点に当てた新しいSE法（提案SE法）を提案した。図7はYahoo!知恵袋データに対し、質問とベストアンサーに選ばれた回答者から2部グラフを構成し、既存SE法と提案SE法による可視化時間を比較したものである。既存SE法と比較して、平均実行時間は提案SE法の方が高速であることが分かる。

③類似度が定義されたオブジェクト集合を低次元ベクトルとして埋め込むことは、データの隠れた構造やオブジェクト間の関係を

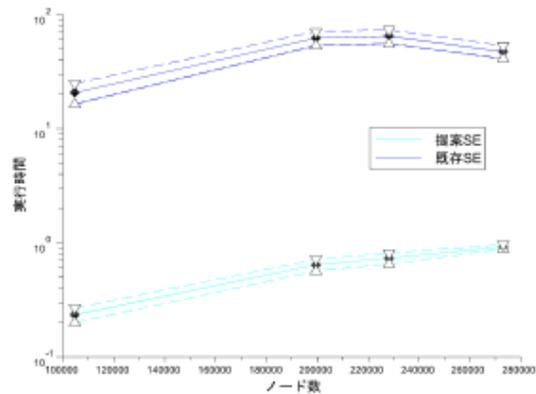


図7. Yahoo!知恵袋データに対する2部グラフ可視化時間の比較

視覚的に把握するために重要である。そのための既存法として、多次元尺度法やスペクトラル法などが利用できるが、本研究ではBustosらが提案した類似検索法を土台にピボット可視化法と呼ぶ新たな手法を提案した。実験では2つのデータ集合を用い、距離を近似させた埋め込みが実現できているかをクラス情報の観点に基づき定性的に評価した。図8はMNIST（0～9の数字の手書き文字データ）に対するピボット可視化法を用いた可視化結果である。実験より、ピボット可視化法を用いれば、代表的な既存可視化法に匹敵しつつ、固有の埋め込みが実現できることを示した。

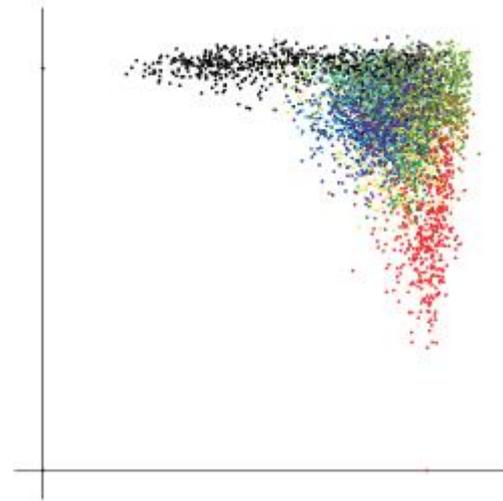


図8. 手書き文字データに対するピボット法を用いた可視化結果

5. 主な発表論文等

〔雑誌論文〕（計4件）

- ①武藤伸明, 斉藤和巳, 大久保誠也, 池田哲夫. 乱歩モデルに基づく施設配置問題. 日本データベース学会論文誌, 査読有, Vol.12 (1), 2013, 採録済.
- ②Shoko Kato, Akihiro Koide, Takayasu Fushimi, Kazumi Saito, Hiroshi Motoda. Network Analysis of Three Twitter

Functions: Favorite, Follow and Mention. Proc. of the 2012 Pacific Rim Knowledge Acquisition Workshop (PKAW2012), 査読有, 2012, 298-312.

- ③小出明弘, 斉藤和巳, 風間一洋, 鳥海不二夫. コリंक構造に着目した多重グラフの特性分析. 日本データベース学会論文誌, 査読有, Vol.11 (2), 2012, 13-18.
- ④武藤伸明, 斉藤和巳, 池田哲夫, 永田大, 伏見卓恭. 非線形施設配置問題における貪欲法の遅延評価による高速化 一複数施設が確率的に貢献する場合. 情報処理学会論文誌 数理モデル化と応用 (TOM), 査読有, Vol.4 (1), 2011, 1-10.

〔学会発表〕 (計 14 件)

- ①加藤翔子, 斉藤和巳. 高多重度有向ネットワーク成長モデルの提案. 第 5 回データ工学と情報マネジメントに関するフォーラム (第 11 回日本データベース学会年次大会) (DEIM2013), 2013 年 3 月 3 日, 福島県, ホテル華の湯.
- ②加藤翔子, 小出明弘, 伏見卓恭, 斉藤和巳. ジニ係数による多重有向グラフとしての Favorites ネットワークの分析. 第 26 回人工知能学会全国大会 (JSAI2012), 2012 年 6 月 15 日, 山口県, 山口県教育会館.
- ③伏見卓恭, 斉藤和巳, 武藤伸明, 池田哲夫. 道路網における看板配置問題のモデル化と効率的解法. 情報処理学会第 74 回全国大会, 2012 年 3 月 8 日, 愛知県, 名古屋工業大学.
- ④伏見卓恭, 斉藤和巳, 武藤伸明, 池田哲夫. ノード集合に対する媒介中心性の提案. 第 4 回データ工学と情報マネジメントに関するフォーラム (第 10 回日本データベース学会年次大会) (DEIM2012), 2012 年 3 月 5 日, 兵庫県, シーサイドホテル舞子ビラ神戸.
- ⑤高橋友美, 小出明弘, 伏見卓恭, 斉藤和巳. Twitter の Hashtag に着目した社会影響分析. DEIM2012, 2012 年 3 月 5 日, 兵庫県, シーサイドホテル舞子ビラ神戸.
- ⑥大森美香, 伏見卓恭, 斉藤和巳. 編集距離に基づくアノテーション付き可視化法の提案. DEIM2012, 2012 年 3 月 5 日, 兵庫県, シーサイドホテル舞子ビラ神戸.
- ⑦伏見卓恭, 斉藤和巳, 武藤伸明, 池田哲夫. 社会ネットワークにおける中心性の定量的な頑健性評価法の提案. 情報処理学会第 73 回全国大会, 2011 年 3 月 4 日, 東京都, 東京工業大学大岡山キャンパス.
- ⑧村松優作, 風間一洋, 斉藤和巳. レビュー類似度によるネットワーク構造分析. 第 3 回データ工学と情報マネジメントに関するフォーラム (第 9 回日本データベ

ス学会年次大会) (DEIM2011), 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.

- ⑨伏見卓恭, 大久保誠也, 武藤伸明, 斉藤和巳. リンク張替に対するモチーフ解析の頑健性評価. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.
- ⑩白賀昌宏, 伏見卓恭, 大久保誠也, 武藤伸明, 斉藤和巳. 大学サイトモチーフの類似ネットワークによる分析法. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.
- ⑪山岸祐己, 風間一洋, 斉藤和巳. ノード属性を考慮した情報拡散影響度の推定. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.
- ⑫鈴木紳吾, 伏見卓恭, 斉藤和巳, 池田哲夫. ピボットを用いた類似度データの可視化法. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.
- ⑬三津山雅規, 武藤伸明, 斉藤和巳, 池田哲夫. 類似検索における遅延評価を用いたピボット選択の高速化. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.
- ⑭久保田大和, 伏見卓恭, 斉藤和巳. 2 部グラフ可視化法の高速化. DEIM2011, 2011 年 2 月 27 日, 静岡県, ラフォーレ修善寺.

〔図書〕 (計 0 件)

〔産業財産権〕

- 出願状況 (計 0 件)
- 取得状況 (計 0 件)

〔その他〕

特になし

6. 研究組織

(1) 研究代表者

武藤 伸明 (MUTOH NOBUAKI)
静岡県立大学・経営情報学部・准教授
研究者番号: 40275102

(2) 研究分担者

斉藤 和巳 (SAITO KAZUMI)
静岡県立大学・経営情報学部・教授
研究者番号: 80379544
池田 哲夫 (IKEDA TETSUO)
静岡県立大学・経営情報学部・教授
研究者番号: 60363727
大久保 誠也 (OKUBO SEIYA)
静岡県立大学・経営情報学部・助手
研究者番号: 90422576

(3) 連携研究者

なし