

科学研究費助成事業 研究成果報告書

平成 29 年 6 月 26 日現在

機関番号：12102

研究種目：基盤研究(B) (一般)

研究期間：2013～2016

課題番号：25280110

研究課題名(和文) コミュニティメディアからの知識抽出に基づくソーシャルキャピタルの変容過程の解明

研究課題名(英文) Dynamic Progresses of Social Capital based on Knowledge Extraction from Community Media

研究代表者

佐藤 哲司 (SATO, Tetsuji)

筑波大学・図書館情報メディア系・教授

研究者番号：70396117

交付決定額(研究期間全体)：(直接経費) 12,200,000円

研究成果の概要(和文)：リアルとサイバー複合空間を対象に、知識創造社会を支える第3の社会資本と言われるソーシャルキャピタルの形成・変容過程を解明する。我々の実生活と不可分な存在となっているツイッターから、リアル空間における生活を支援するツイート抽出・生活の局面ラベルを付与する手法を提案した。コミュニティのノード機能に着目することで、構造的特性と意味的特性を表す中心性指標も提案した。急速に拡大しているテキストコミュニケーションにおける話者の役割や親密さを推定する手法を提案した。また、テキスト投稿時の意図推定や意見分析に有効な特徴量の考察、変化変容を扱うための系列データを対象とする機械学習手法の考案にも取り組んだ。

研究成果の概要(英文)：We will clarify the process of formation and transformation of social capital called third social capital which supports knowledge creation society for real and cyber complex space. We proposed an innovative method to extract the tweets which will support the real life in some aspects of life. By focusing on the node function of the community graph, we also proposed a centrality indicator that represents structural characteristics and semantic characteristics.

According to rapidly expanding text communication, we also proposed a method to estimate the role and intimacy of the speaker. We also devised both a machine learning method for sequence data for dealing with change transformation, and considering the intention estimation at the time of text posting and feature quantities effective for opinion analysis.

研究分野：データ工学

キーワード：社会ネットワーク分析 知識発見 テキストマイニング

1. 研究開始当初の背景

インターネット上の知識・情報共有空間の普及拡大は著しく、我々の実生活に不可分な存在となってきている。ツイッターに代表されるマイクロブログやコミュニティ QA (不特定多数の参加者による質問回答サイト) 料理レシピ投稿共有サイトなどは日常的に利用され、そこから得られる情報に人々の実生活が左右され、またその結果を投稿する情報の再生産が形成されてきている。

2. 研究の目的

本研究は、これまで対面での実生活を前提として議論されてきたソーシャルキャピタル(社会関係資本)をサイバー空間まで拡張し、リアルとサイバー複合空間におけるソーシャルキャピタルの形成過程、並びにその変容過程を解明することを目的とする。社会関係資本といわれるソーシャルキャピタルは、第二次産業を支える物的資本や人的資本に対応する、情報化社会を支える第3の資本と位置づけられ、市民活動を活性化させる重要な要素とされている。

3. 研究の方法

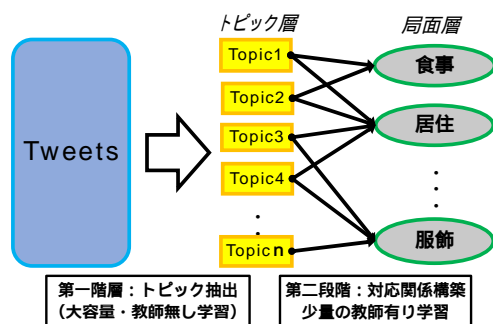
(1) 生活者の日常を投稿しているツイッターや、生活に密着した調理を支援するレシピ投稿共有サイトに登録された大規模なデータを対象に、テキストマイニングや機械学習などの手法を適用して知識発見に取り組む。

(2) コンテンツ間の関係やユーザ間関係をエッジとする社会ネットワークを構築し、発展・進化するネットワークの特徴を捉える新しい手法を考案する。

(3) 様々な意図を持って投稿されている記事集合を対象に、投稿時のコンテキストを考慮した特徴量の提案、意図推定の精度向上などに取り組む。

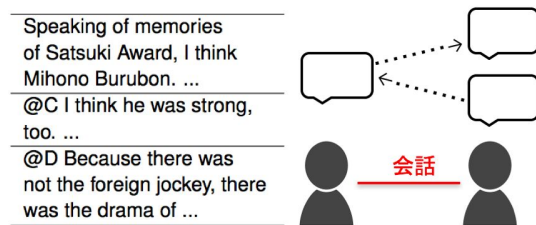
4. 研究成果

(1) 短文かつ文書構造が曖昧なツイートが言及している生活の局面を高精度に推定する手法を提案した。省略が多い短文を扱うこと、更には一つのツイートが複数の局面に言及している場合があることから、単純な機械学習では「食事」や「居住」などの生活の局面を推定することが困難である。このため、

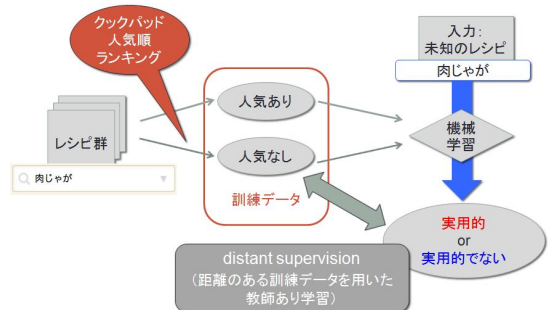


教師無し学習手法である LDA を用いて大量のツイート集合から比較的安定したトピック層を抽出し、トピック層と局面との対応関係にエントロピーフィードバックなどの学習機構を設けることで高精度化を実現した。

(2) ソーシャルメディアに投稿されたテキスト対話情報から、詳細な利用者間の関係の解析を行うための発話役割推定手法に取り組んだ。対話に固有の同時進行性という特徴を考慮した機械学習モデルを提案し、発話役割の推定精度を向上できることを確認した。推定した発話役割を特徴量として、二者間で行われた会話データから両者の親密さを定量化する手法を提案した。また、ソーシャルメディア上のテキストからの教師なしフレーズ抽出手法を提案した。ソーシャルメディア上のテキストには、新たな概念や実態を表す複合語が多数出現するため、従来の辞書に基づいた分析では限界がある。提案手法は、統計的テキストセグメントモデルの教師なし学習でフレーズを抽出するため、情報拡散過程やコミュニティプロフィールの内容解析において有効に機能すると考えられる。

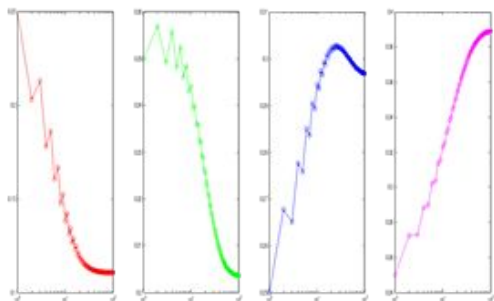


(3) ソーシャルキャピタルに基づく特徴量について、テキスト中に明確には現れない、主に著者による執筆の意図を、詳細な調査に基づき定式化し、同時にその意図を正確に推定できる仕組みを実現するための研究を進めた。具体的には、コミュニティ QA における投稿者と回答者の意見分析、マイクロブログにおける投稿者の地域属性推定、レシピ投稿者による実用的なレシピ手順の推定などについて、その特徴量を明らかにした。



(4) ソーシャルメディアにおけるユーザネットワークでは、多種多様なコミュニティのあり方が観測される。コミュニティ内のノード機能に着目し、下図に示すベクトルで表現したコミュニティの構造的特性に加え、意味的特性を考慮した中心性指標の開発にも着手し、従来の構造のみに着目する手法と一線を

画した手法を提案した。さらに、ソーシャルメディア上でのユーザの投稿内容に着目し、単語の共起ネットワークを構築、分析した。ユーザの投稿内容の視点からも、各コミュニティで明確な差異があることを発見した。



(5) オンラインコミュニティにおける活動の解析のための時系列データ解析の手法の開発に取り組んだ。特に時間軸上でのイベントの発生系列を対象とした正定値カーネルを定義することにより、回帰や低次元表現といった多変量解析をイベント系列に適用することを実現した。さらにコミュニティ活動の分析に適した特徴量抽出を目指して、スパース符号化辞書学習の研究に取り組み、観測データを少数の特徴量の一次結合で表現し、パターン抽出を行う手法を発展させた。

(6) 季節による検索動作への影響の明確化を目的として検討を行った。アンケートと印象評価実験を異なる季節において行い、季節によってどのように検索行動が変化するかを調査した。実験の結果、画像の検索要求においては、男女のスコアに有意差は見られなかったが、夏のエアコンの効いた環境と冬では、秋と異なる検索要求が見られた。また、冬には、人によって大きく差が出ることを明らかにした。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 11 件)

- (1) 加藤翔子, 斉藤和巳, 風間一洋, 佐藤哲司: 多重有向グラフのコア部抽出のための MDSR 法の提案と評価, 日本データベース学会和文論文誌, Vol. 14, Article No. 1, pp. 1 - 6 (Mar. 2016). <査読有>
- (2) 山本修平, 佐藤哲司: 実生活ツイートに対する局面推定の精度向上に関する検討, 情報処理学会論文誌, Vol. 56, No. 6, pp. 1496-1506 (July 2015). <査読有>
- (3) 若林 啓: HHMM 変換を用いた左非循環 PCFG の高速推論. 情報処理学会論文誌データベース (TOD), Vol.8, No.1, pp.45-54 (Mar. 2015). <査読有>
- (4) Yutarō Yamaguchi, Shuhei Yamamoto,

and Tetsuji Satoh: Behavior analysis methods for Twitter users based on transitions in posting activities, Int'l Journal of Web Information Systems, Vol.10, Iss. 4, pp. 363-377, Emerald (Oct. 19, 2014). <査読有> doi: 10.1108/IJWIS-04-2014-0014

- (5) Shuhei Yamamoto and Tetsuji Satoh: Two phase estimation method for multi-classifying real life Tweets, Int'l Journal of Web Information Systems, Vol.10, Iss. 4, pp. 378-393, Emerald (Oct. 19, 2014). <査読有> doi: 10.1108/IJWIS-04-2014-0013
- (6) 若林 啓: 部分統語構造を考慮した階層的確率オートマトンに基づく教師なしチャンキング. 情報処理学会論文誌データベース (TOD), Vol.7, No.2, pp.61-69 (July 2014). <査読有>
- (7) 関洋平: コミュニティ QA における意見分析のためのアノテーションに関する一検討, 自然言語処理, Vol.21, No.2, pp.271-299 (2014). <査読有>
- (8) 堂前友貴, 関洋平: 半教師ありトピックモデルにより選択した地域特徴語を用いた Twitter ユーザの生活に関わる地域の推定, 情報処理学会論文誌データベース, Vol.7, No.3, pp.1-13 (2014). <査読有>
- (9) 酒井紗希, 関洋平: 感性タグを用いて読者間の交流を促進するための一検討, 電子情報通信学会論文誌, Vol. J97-D, No. 1, pp.173-176 (2014). <査読有>
- (10) 山本修平, 佐藤哲司: トピックと局面の対応関係に基づく実生活ツイートのマルチラベル分類, 情報処理学会論文誌データベース (TOD), Vol. 7, No.2, pp. 24-36 (June 2014). <査読有>
- (11) 関洋平: 意見分析コーパスの現状と課題, 情報処理学会論文誌データベース, Vol.6, No.4 (2013), pp.85-103. <査読有>

[学会発表](計 104 件)

- (1) Taro Tezuka, Shun Yasumasa, and Fatemeh Azadi Naghsh: Visualization of n-Gram input patterns for evaluating the level of divergent thinking, Knowledge, Information and Creativity Support Systems: Recent Trends, Advances and Solutions, pp. 229-242, Advances in Intelligent Systems and Computing 364, Springer, Yogyakarta, Indonesia (Mar. 2016). DOI: 10.1007/978-3-319-19090-7_18 <査読有>
- (2) Wajima, K. and Satoh T.: Urgent Question Detection based on the Review Points and Sentiment Words, The 17th

- International Conference on Information Integration and Web-based Applications & Services, pp. 307 - 311, Brussels, Belgium (Dec. 2015). <査読有>
- (3) Yamamoto, S., Wakabayashi, K., Kando, N., and Satoh, T.: BUTE: Bursty Users Tagging Method Estimated by Time Series Data, The 17th International Conference on Information Integration and Web-based Applications & Services, iiWAS2015, pp. 148 - 156, Brussels, Belgium (Dec. 2015). <査読有>
- (4) Fushimi, T., Satoh, T., Saito, K., and Kazama, K.: Comparison of Influence Measures on Structural Changes Focused on Node Functions, The 17th International Conference on Information Integration and Web-based Applications & Services, iiWAS2015, pp. 110 - 119, Brussels, Belgium (Dec. 2015). <査読有>
- (5) Yamamoto, S., Kando, N., and Satoh, T.: LAIM: Life Aspect Inference Method based on Probability Distribution for Real Life Tweets, WI2015, pp. 187 - 194, Singapore (Dec. 6-9 2015). <査読有>
- (6) Zhong Rui and Taro Tezuka, Parametric learning of deep convolutional neural network, Proceedings of the 19th International Database Engineering & Applications Symposium (IDEAS2015), pp. 226-227, Yokohama, Japan, July 13-15, 2015. <査読有>
- (7) Yamamoto, S. and Satoh, T.: Hierarchical Estimation Framework of Multi-Label Classifying: A Case of Tweets Classifying into Real Life Aspects, The 9th International AAAI Conference of Web and Social Media (ICWSM2015), pp. 523-532, Oxford, UK (May 2015). <査読有>
- (8) Shoko Kato, Kazumi Saito, Kazuhiro Kazama, and Tetsuji Satoh: MDSR: An Eigenvector Approach to Core Analysis of Multiple Directed Graphs. The 13th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2014, pp. 447-458, (Dec., 2014). DOI: 10.1007/978-3-319-13560-1_36 <査読有>
- (9) Yohei Seki and Kouta Ono: Discriminating Practical Recipes Based on Content Characteristics in Popular Social Recipes. Proc. of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing Adjunct, Seattle, WA, USA, pp.487-496 (Sept., 2014). <査読有>
- (10) Taro Tezuka, A dictionary learning algorithm for sparse coding by the normalized bilateral projections, Proceedings of the 24th IEEE International Workshop on Machine Learning for Signal Processing (MLSP2014), Reims, France, September 21-24, 2014. <査読有>
- (11) Fumiya Mori, Fatemeh Azadi Naghsh, and Taro Tezuka, The effect of music on the level of mental concentration and its temporal change, Proceedings of the 6th International Conference on Computer Supported Education (CSEDU2014), Barcelona, Spain, April 3-5, 2014. <査読有>
- (12) Mizunuma, Y., Yamamoto, S., Yamaguchi, Y., Ikeuchi, A., Satoh, T., and Shimada, S.: Twitter Bursts: Analysis of their Occurrences and Classifications, 8th International Conference on Digital Society, ICDS 2014, pp. 182-187 (Mar. 2014). <査読有>
- (13) Yamamoto, S. and Satoh, T.: Two Phase Extraction Method for Multi-label Classification of Real Life Tweets, 15th Int'l Conf. Information Integration and Web-based Applications & Services, iiWAS2013, pp. 16-25, (Dec. 2013). <査読有> DOI: 10.1145/2539150.2539197
- (14) Yamaguchi, Y., Yamamoto, S., and Satoh, T.: Behavior Analysis of Microblog Users Based on Transitions in Posting Activities, 15th Int'l Conf. Information Integration and Web-based Applications & Services, iiWAS2013, pp. 63-67, (Dec. 2013). <査読有> DOI: 10.1145/2539150.2539209
- (15) Yamamoto, S. and Satoh, T.: Two Phase Extraction Method for Extracting Real Life Tweets using LDA, The 15th Asia-Pacific Web Conference (APWeb2013), Lecture Notes in Computer Science 7808, pp. 340-347 (Apr. 2013). <査読有>
- (16) Yohei Seki and Kiyoto Miyajima: Finding Impressive Social Content Creators. Proc. of the 36th Ann. Int'l. ACM SIGIR Conf. on Research and Development on Information Retrieval (SIGIR 2013), Dublin, Ireland, pp.1041-1044 (July 2013). <査読有>
- (17) Adam Jatowt, Ee-Peng Lim, Ying Ding, Asako Miura, Taro Tezuka, Social Informatics - Proceedings of the 5th International Conference on Social Informatics, Lecture Notes in Computer Science, Vol. 8238, Springer, 2013, ISBN 978-3-319-03259-7.

- <査読有>
- (18) Yoshie Kubota and Taro Tezuka, Transformed Reality - altering human perceptions by computation -, Proceedings of the International Conference on Culture and Computing 2013, Kyoto, Japan, pp. 39-44(Sept. 2013). <査読有>
- (19) Taro Tezuka, Shun Yasumasa, and Fatemeh Azadi Naghsh, A system using n-grams for visualizing the human tendency to repeat the same patterns and the difficulty of divergent thinking, Proceedings of the 8th International Conference on Knowledge, Information and Creativity Support Systems (KICSS2013), Krakow, Poland, Progress & Business Publishers, pp. 623-634 (Nov. 2013). <査読有>
- (20) 伏見 卓恭, 佐藤 哲司, 齋藤 和巳, 風間 一洋: 距離減衰重みを導入したノード群へのアノテーション付与法, 情報処理学会 他, WebDB Forum 2015, B4-1, 8p., 東京都, 江東区 (Nov. 2015). <査読有>
- (21) 野沢 健人, 若林 啓: トピックモデルに基づく大規模ネットワークの重複コミュニティ発見, 情報処理学会 他, WebDB Forum 2015, pp.88-95, 東京都, 江東区 (Nov. 2015). <査読有>
- (22) 伏見卓恭, 佐藤 哲司, 齋藤 和巳, 風間一洋: 時間減衰付きカテゴリ選択モデルを用いたレビュー傾向分析, 第14回情報科学技術フォーラム, RF-001, 愛媛県, 松山市 (Sep. 15, 2015). <査読有>
- (23) 伏見卓恭, 佐藤 哲司, 齋藤 和巳, 風間一洋: ノード間の隣接関係を考慮したアノテーション法の提案, 日本ソフトウェア科学会, ネットワークが創発する知能研究会(JWEIN2015), 東京都, 千代田区 (Aug. 15, 2015). <査読有>
- (24) 伏見 卓恭, 佐藤 哲司, 齋藤 和巳, 風間 一洋: ノード機能に着目した構造変化に対する影響度指標, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, pp. 1487 - 1495, ホテル安比グランド, 岩手県, 八幡平市 (July 7 - 9, 2015). <査読無>
- (25) 輪島 幸治, 佐藤 哲司: 評価視点と感情表現に基づく質問記事の重要度判定手法の提案, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, 3F-4, pp. 630 - 635, ホテル安比グランド, 岩手県, 八幡平市(July 7 - 9, 2015). <査読無>
- (26) 清野 悠希, 佐藤 哲司: 重心動揺と姿勢計測による学習状態の分析, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, 3H-1, pp. 666 - 671, ホテル安比グランド, 岩手県, 八幡平市 (July 7 - 9, 2015). <査読無>
- (27) 柘植 大, 神門 典子, 佐藤 哲司: Twitterにおける聞き役ユーザ選定手法の提案, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, 6G-1, pp. 1321 - 1327, ホテル安比グランド, 岩手県, 八幡平市 (July 7 - 9, 2015). <査読無>
- (28) 山本 修平, 若林 啓, 神門 典子, 佐藤 哲司: バースト時刻に基づくユーザのタグ付け手法の提案, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, 6G-2, pp. 1328 - 1334, ホテル安比グランド, 岩手県, 八幡平市 (July 7 - 9, 2015). <査読無>
- (29) 伏見 卓恭, 佐藤 哲司, 齋藤 和巳, 風間 一洋: ノード機能に着目した構造変化に対する影響度指標, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02015)シンポジウム, 7D-2, pp. 1487 - 1495, ホテル安比グランド, 岩手県, 八幡平市 (July 7 - 9, 2015). <査読無>
- (30) 山本 修平, 若林 啓, 佐藤 哲司: バースト時刻に基づくフォロー先ユーザ推定手法, WebDB Forum 2014, 芝浦工業大学・豊洲キャンパス, A-5, 東京都, 江東区 (Nov., 2014). <査読有>
- (31) 山本 修平, 佐藤 哲司: 実生活ツイートに対する局面推定の精度向上に関する一検討, 情報処理学会 マルチメディア, 分散, 協調とモバイル(DICOM02014)シンポジウム, 2F-1, pp. 422--429, 月岡温泉, 新潟県, 新発田市 (July, 2014). <査読無>
- (32) 中岡 義貴, 佐藤 哲司: 食材の偏りと調理法に基づくレパートリー拡大のためのレシピ推薦システムの提案, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOM02014)シンポジウム, 7F-1, pp. 1653 - 1660, 月岡温泉, 新潟県, 新発田市 (July, 2014). <査読無>
- (33) 若林 啓: 部分統語構造を考慮した確率オートマトンに基づく教師なしチャタリング. WebDB Forum 2013, 7p., 京都府, 京都市 (Nov., 2013) <査読有>
- (34) 山本 修平, 佐藤 哲司: 二段階抽出法を用いた実生活 Tweet のマルチラベル分類, 情報処理学会, マルチメディア, 分散, 協調とモバイル(DICOM02013)シンポジウム, 1C-2, pp. 64 - 71 (July, 2013). <査読無>
- (35) 山口 裕太郎, 山本 修平, 佐藤 哲

司：投稿活動の変化に着目したマイクロプログユーザの可視化手法の提案，情報処理学会，マルチメディア，分散，協調とモバイル(DICOMO2013)シンポジウム，1C-3，pp. 72 - 79，北海道，河東郡 (July, 2013). <査読無>

- (36) 中岡 義貴，佐藤 哲司：定番度に基づくレシピ推薦システムの提案，情報処理学会，マルチメディア，分散，協調とモバイル(DICOMO2013)シンポジウム，4H-4，pp. 1083 - 1089，北海道，河東郡 (July, 2013). <査読無>

(他 口頭発表 68 件)

〔その他〕

ホームページ等

<http://www.slis.tsukuba.ac.jp/~sato/projects-j.html>

6. 研究組織

(1) 研究代表者

佐藤 哲司 (SATO, Tetsuji)
筑波大学・図書館情報メディア系・教授
研究者番号：70396117

(2) 研究分担者

宝珍 輝尚 (HOCHIN, Teruhisa)
京都工芸繊維大学・情報工学人間科学系・教授
研究者番号：00251984

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

関 洋平 (SEKI, Yohei)
筑波大学・図書館情報メディア系・准教授
研究者番号：00348468

手塚 太郎 (TEZUKA, Taro)
筑波大学・図書館情報メディア系・准教授
研究者番号：40423016

池内 淳 (IKEUCHI, Atsushi)
筑波大学・図書館情報メディア系・准教授
研究者番号：80338607

若林 啓 (WAKABAYASHI, Kei)
筑波大学・図書館情報メディア系・助教
研究者番号：40631908

伏見 卓恭 (FUSHIMI, Takayasu)
筑波大学・図書館情報メディア系・特別
研究員 (PD)
研究者番号：80755702