

令和元年6月25日現在

機関番号：14301

研究種目：基盤研究(C) (一般)

研究期間：2015～2018

課題番号：15K00423

研究課題名(和文)セマンティクスとウェブ・統計データを活用した潜在的関係の知識発見

研究課題名(英文)Discovery of knowledge about implicit relationships utilizing semantics, web data, and statistics

研究代表者

浅野 泰仁 (Asano, Yasuhito)

京都大学・情報学研究科・特定准教授

研究者番号：20361157

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、概念間の潜在的関係の知識を発見・体系化するための基盤技術の開発と、異種の複数のデータに含まれる概念間の潜在的関係を利用して、これらのデータを対応づける応用研究の双方を行った。概念間の潜在的関係とは、文書やデータ上で明示的に示された「AはBの一部である」等の関係とは異なり、文書やデータを総合的に分析して複数の箇所から得られた明示的な関係を結びつけることによって得られる関係である。このような関係を発見するためには、そのためのネットワーク分析技術や文書分析技術が基盤技術として必要であり、また実際のデータに適用するにはそのデータ特有の分析技術が必要となるため、これらの構築を行った。

研究成果の学術的意義や社会的意義

- (1) 本研究におけるネットワーク分析技術の構築の結果、潜在的関係の発見以外にも応用できる手法が開発できた。例えば、ライドシェアリング等で相乗りする際の費用を軽減するのに有効であると思われる、合流経路の最適化手法を開発できた。
- (2) 様々な異種データを実際に結びつける手法を提案することができた。例えば、商品のレビューデータと商品の機能(性能)データから「商品の機能とユーザの使用目的」の関係(デジカメの例を挙げれば、鳥の撮影という目的に必要な機能や、夜景撮影という目的に必要な機能等がある)についての知識を発見する手法を提案できた。

研究成果の概要(英文)：We have studied fundamental techniques for discovering and organizing knowledge about implicit relationships between concepts. We have also studied applications of connecting different kinds of data utilizing implicit relationships of concepts contained in the data. Implicit relationships between concepts are different from explicit relationships such as "A is a part of B" which are described explicitly on documents or data; they are relationships obtained by analyzing documents or data and connecting discovered explicit relationships found on several parts. We have established network analysis methods and text analysis techniques required as fundamental techniques for discovering such relationships. We have also established several analysis methods for particular kinds of data to apply the techniques above.

研究分野：情報学

キーワード：知識発見 知識体系化 エンティティ間の関係

1. 研究開始当初の背景

近年、Wikipedia 等の集合知またはウェブページ集合から、概念間の関係知識を発見するための研究が盛んになってきている。概念間の関係知識とは、「日本と関係の強い国はどの国か」「生活行動と電力消費とはどう関係しているか」などの質問に対する答えとなる知識である。概念間の関係知識はこれらの質問のように社会的に重要なものが多いが、従来の検索エンジンで単純に概念をキーワードとして検索しても見つからないことが多い上、Wikipedia 等の集合知でも概念を説明したページそのものに関係の知識が含まれていることは少ないため、概念間の関係知識を得ようとするとな手でも数多くの検索結果や Wikipedia 等のページを長時間にわたって調査することが必要になってしまっていた。また、異分野の複数のデータを相互に結びつけて利用する研究が LOD(Linked Open Data)や各種の科学分野]で始まりつつあり、そのためにも概念間の関係知識は重要となってきている。

2. 研究の目的

上記のように、Wikipedia 等の集合知またはウェブや統計データから、概念(またはエンティティ、以降は概念で統一)間の関係知識を発見するための研究が盛んになってきている。その研究の多くは「A は B の一種」「A は B の友達」等の比較的単純な顕在的關係の発見にとどまっているが、「日本と関係の強い国はどの国か」等の複雑な質問に答える、あるいは異分野のデータの概念を結びつけて利用する研究には、複数の顕在的關係の組合せからなる「潜在的関係」の発見が必須となる。本研究では、潜在的關係の既存研究で実現できていない、意味付け(セマンティクス)の考慮およびウェブ・統計データへの対応が可能、概念間の潜在的關係の有用な知識の発見手法の基盤となる技術の構築を目的とする。さらに、構築した技術やデータマイニング・機械学習の技術を異分野にまたがる複数の実際のデータ(以降、異種データ)に応用して、その有用性を検証する。

3. 研究の方法

本研究では、潜在的關係の知識を発見・体系化するための基盤技術の開発と、実際の異種データを結びつける応用の双方が重要であると考え、その両面に渡る研究を行うこととした。基盤技術の開発としては、まず、複数のテキスト文書の分析技術、特に同一概念や関連概念について記述されている部分の対応付けが、概念間の関係を意味付けするのに役立つだけでなく、概念に関する知識体系化に必要とされると考え、これを行うことにした。また、概念間の関係発見の基盤技術としては、概念間の関係はネットワーク上の経路で表されることが多くなるため、経路モデルに関する研究を行うことにした。異種データに対する応用としては、関連する複数の概念(またはエンティティ)を内包する複数の異種データが対象となるが、その中から実用上興味深い概念間の潜在的關係が含まれていると考えられる物を選び、上記で開発した技術や既存のテキストマイニング・ネットワーク分析技術を応用して、そのデータに特化した手法を開発することにした。

4. 研究成果

本研究では、上記のように、潜在的關係の知識を発見・体系化するための基盤技術の開発と、実際の異種データを結びつける応用の双方を行った。

基盤技術としては、本研究が発見対象とするエンティティ間の関係はテキストデータやネットワーク構造に含まれていると考えられるため、これらを解析する基本的な手法について研究した。特に、同じトピックに関する複数文書中の対応する部分を整理する手法と、ネットワーク上におけるエンティティを結びつける方法と関連があると考えられる合流経路を計算する手法が挙げられる。これらの結果はそれぞれ査読付き国際会議や査読付き学術論文誌に採録された。

前者としては、(i)クラシック音楽を解説する複数の文書の、解説する楽曲中の場所に基づく対応付けの手法と、(ii)レシピ文書データを分析して同種の料理の複数のレシピ(複数の文書)ごとの対応部分から使用する食材の傾向を発見し、それに基づいて各レシピの古典度を計算する手法と、(iii)教育用スライド等の XML 文書と対話形式教育用脚本の対応付けを行う手法を構築した。

なお、(i)では、要素技術として、段落(文章)に含まれる単語の分散表現の分布でモデル化する手法を考案した。これによって、複数文書中の文章間の類似度を求めることなどができる。これを用いた複数文書中の部分対応付けは、既存の手法である、含まれる単語の分散表現の単純な平均や、文章の分散表現モデルである doc2vec を用いた手法よりも良い性能を示した。

後者としては、複数の移動者が出発地から目的地に向かう際の経路を決定することを考え、(a)これら複数の移動者が異なる出発地から単一目的地へ集合する際の合流経路の最適化手法と、(b)出発地も目的地も移動者ごとに異なる場合の合流経路の最適化手法を提案した。特に(b)に関しては、その取り得るパターン複雑性ゆえに既存の結果がほとんど知られていなかったが、移動者の数が少ない場合には、取り得るパターンに限られている(図 1)ことを理論的に証明し、それに基づく最適化手法を考案することができた。この成果は国際的に高く評価され、データマイニング分野のトップジャーナルである IEEE Transactions on Knowledge and Data Engineering (TKDE)にも採録された。これらの経路最適化技法は、概念間の顕在的關係をネッ

トワークでモデル化することによって、上記の移動者が概念に対応し、上記の経路が概念間の潜在的関係を表すことになるので、複数の概念を結びつける潜在的関係の分析に応用できると期待される他、近年注目されているライドシェアリングやカープーリング(相乗り)等の経路最適化にも応用できると考えられる。

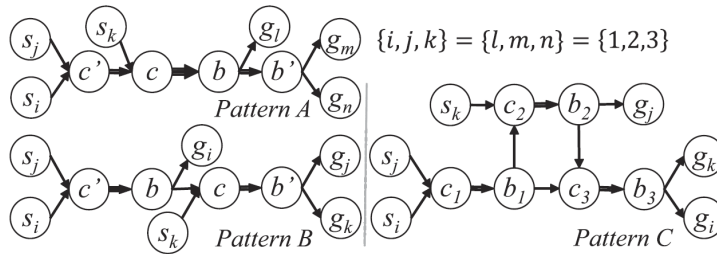


図1 移動者数3の合流経路パターン

実データへの応用としては、(1)商品のレビューデータと商品の機能(性能)データから「商品の機能とユーザの使用目的」の関係についての知識を発見する手法、(2)商品のレビューデータから「特異な評価者と対象の評価サマリ」の関係についての知識を発見する手法、(3)大学

レベルの標準カリキュラムデータとオンラインコースのシラバスデータから、カリキュラムで定義されている知識オントロジーと個々のシラバスとの関係を発見する手法、等を提案した。これらの結果もそれぞれ査読付き国際会議や査読付き学術論文誌に採録された。

(1)の「商品の機能とユーザの使用目的の関係知識」を適用できる例としては、デジタルカメラやPCのように目的に応じて必要とされる機能が変わってくる商品が挙げられる。例えばデジタルカメラで鳥を撮影する際には望遠性能、AF性能、連写性能等の機能が重要になってくる。一方で夜景の撮影をするには、高感度における低ノイズ性能や、手ぶれ補正の性能が重要になってくる。このような関係知識を得るために、商品のレビューデータを収集し、関係知識を表す各種テキストパターンとブートストラップ法を組み合わせた独自の手法を提案している。この手法は、商品機能と目的との関係知識をあまり持ち合わせていない初心者に対して、この知識を直接提示したり、各目的ごとに向いている商品のランキング等を提示する(図2)ことで商品選択を補助できると考えられる。既存の研究として、商品の機能の優劣でランキングする手法は存在したが、提案手法のように目的に応じた機能を見つけることでランキングを行う手法は存在せず、その独自性が高く評価され、The Second International Conference on Data Mining and Big Data (DMBD 2017)において、Best Student Awardを受賞した。

(2)の「特異な評価者と対象の評価サマリの関係」は、Amazonや食べログ等、各種のレビューデータの評価サマリ(Amazonでは各評価対象について、評価者の評価値の平均がサマリとなっている。食べログ等、サマリの算出基準を公開していないところもある)を基に行動を決定している人々が増えている現在、非常に重要と考えられる。本研究では、評価者と評価対象の潜在的関係(レビューのあるなしと、その評価値)を2部グラフで表している。このようなモデル自体は過去にも存在したが、本研究では、評価対象のレビューの数やばらつき具合を定量的に評価する基準と、評価者の各レビューの評価値とサマリが離れている度合いがどれくらい希少かあるいはありふれているかを定量的に評価する基準を提案し、それらを用いて各評価者の特異度と、対象の評価サマリの潜在的関係を定式化する独自の手法を構築することができた。本手法は、特異な評価者が特に影響力を持つと考えられるアーリーレビューにおいて、彼らが結託しているような場合について、特に有効だと考えられる。実際、Amazonのレビューデータを用いてアーリーレビューにおける結託に関する性能を検証したところ、提案手法は最新手法である Fraud Eagle や FRAUDAR よりも良い性能を示した。

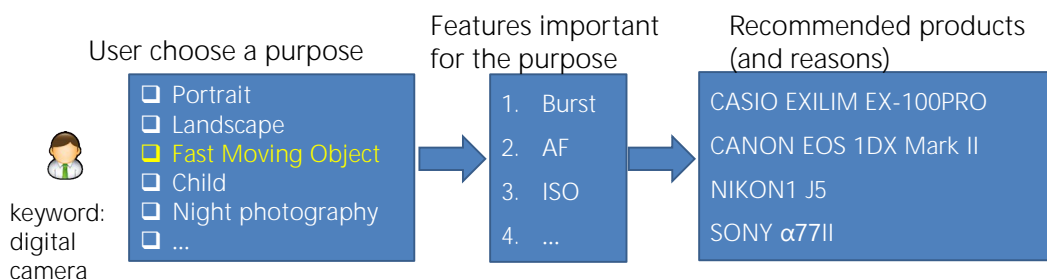


図2 「商品の機能とユーザの使用目的の関係知識」を用いた、目的に重要な機能の提示と、目的に応じた商品ランキング(概念図)

5. 主な発表論文等

[雑誌論文](計 5件)

- (1) Yasuhito Asano, Junpei Kawamoto, "Detecting Anomalous Reviewers and Estimating Summaries from Early Reviews Considering Heterogeneity". IEICE Transactions on Information Systems, E101-D(4), pp.1003-1011, 2018.

- (2) Sopheaktra Yong, Yasuhito Asano, "Purpose-Feature Relationship Mining from Online Reviews towards Purpose-Oriented Recommendation," IEICE Transactions on Information Systems E101-D(4), pp.1021-1029, 2018.
- (3) 伊藤 博典, 浅野 泰仁, 吉川 正俊, "1 教師 1 生徒対話形式教育用脚本の自動生成," 電子情報通信学会和文論文誌, Vol.J101-D, No.3, pp.539-548, 2018.
- (4) 瀧瀬 和樹, 浅野 泰仁, 吉川 正俊, "合流による利益を考慮した単一目的地への集合経路最適化," 日本データベース学会和文論文誌, Vol.15-J, Article No.6, 2017.
- (5) Xinpeng Zhang, Yasuhito Asano, Masatoshi Yoshikawa, "Mutually Beneficial Confluent Routing," IEEE Transactions on Knowledge & Data Engineering, vol.28, no.10, pp. 2681-2696, 2016.

〔学会発表〕(計 8 件)

査読付き国際会議論文

- (1) Sopheaktra Yong, Yasuhito Asano, "Mining Relationship Between User Purposes and Product Features Towards Purpose-Oriented Recommendation," The Second International Conference on Data Mining and Big Data (DMBD2017), pp.9-21, 2017.
- (2) Hironori Ito, Yasuhito Asano, Masatoshi Yoshikawa, "Automatic Generation of Plot for Education by Teacher-Student Dialogue Style," 2017 International Conference on Education and Multimedia Technology (ICEMT 2017), MF1007, 2017.
- (3) Xinpeng Zhang, Yasuhito Asano, Masatoshi Yoshikawa, "Mutually Beneficial Confluent Routing," The 33rd IEEE International Conference on Data Engineering (ICDE 2017) (TKDE Poster Session), pp.43-44, 2017.
- (4) Yiling Dai, Yasuhito Asano, Masatoshi Yoshikawa, "Course Content Analysis: An Initiative Step toward Learning Object Recommendation Systems for MOOC Learners," The 9th International Conference on Educational Data Mining (EDM 2016), pp.347-352, 2016.
- (5) Kazuki Takise, Yasuhito Asano, Masatoshi Yoshikawa, "Multi-user routing to single destination with confluence", The 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (SIGSPATIAL/GIS 2016), pp.72:1-72:4, 2016.
- (6) Yichen Wen, Yasuhito Asano, Masatoshi Yoshikawa, "Ranking Recipes According to Classicality towards Understanding Evolution of Food Preference", IEEE International Conference on Culture and Computing (Culture Computing 2015), pp.99-106, 2015.
- (7) Jiyi Li, Yasuhito Asano, Toshiyuki Shimizu, Masatoshi Yoshikawa, "A Dynamic-Static Approach of Model Fusion for Document Similarity Computation," The 16th International Conference on Web Information Systems Engineering, (WISE 2015), pp.353-368, 2015.
- (8) Taku Kuribayashi, Yasuhito Asano, Masatoshi Yoshikawa, "Towards Support for Understanding Classical Music: Alignment of Content Descriptions on the Web," The 16th International Society for Music Information Retrieval Conference (ISMIR 2015), pp.371-377, 2105.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況 (計 0 件)

名称：
 発明者：
 権利者：
 種類：
 番号：
 出願年：

国内外の別：

取得状況（計 0 件）

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

6 . 研究組織

(1)研究分担者

なし

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2)研究協力者

研究協力者氏名：Xinpeng Zhang

ローマ字氏名：Xinpeng Zhang

研究協力者氏名：清水 敏之

ローマ字氏名：Toshiyuki Shimizu

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。