

平成 30 年 5 月 16 日現在

機関番号：12612

研究種目：基盤研究(C) (一般)

研究期間：2013～2017

課題番号：25330360

研究課題名(和文)セマンティックWebデータに対する推論基盤技術の研究開発

研究課題名(英文)A study of reasoning mechanisms for Semantic Web data

研究代表者

兼岩 憲 (Kaneiwa, Ken)

電気通信大学・大学院情報理工学研究科・教授

研究者番号：00342626

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：本研究では、RDFデータから記述論理のクエリに答える意味的推論システムを開発した。この意味的推論システムとともに、大規模データを推論・検索するためのRDFデータストアを独自に開発した。さらに、ユーザーが簡単に検索できるように、自然言語文による質問とキーワードによる検索方法を実装している。その際、ユーザーの漠然とした関心から検索や質問に答えるために、キーワードや自然言語文からの推論に加えて、機械学習やデータマイニングによってユーザーが求める知識を推定する方法を提案した。

研究成果の概要(英文)：We have developed a query and reasoning system using concept expressions in description logic that returns answers from RDF data. In addition to the system, we have implemented a fast RDF store system for reasoning and searching in large RDF datasets. In order for users to search knowledge, we have proposed a method that translates keywords and Japanese sentences into RDF queries. We have designed algorithms for machine learning and data mining that infer knowledge from users' ambiguous interests.

研究分野：知識表現と推論

キーワード：セマンティックWeb RDFデータ

### 1. 研究開始当初の背景

(1) 近年 検索エンジンを用いて Web から必要な情報を入手することが当たり前になり、知識を覚えることよりも有用な知識をどこからどのように探すかへ関心が移っている。しかし実際に情報を探すとき、目の前の問題に適した知見を膨大な情報から見つけるのは容易ではない。なぜなら、人間が思い描く漠然としたイメージと Web などに存在する文字や画像の情報との間には知識レベルのギャップが存在するからである。

(2) 検索エンジンはユーザーが入力したキーワードに一致する Web ページを検索するが、意味的に近い情報があっても推論を行わない検索ではキーワードがうまくテキスト内に現れなければ探すことができない。こうしたことから Web における知識発見には、人間がはじめに持っている抽象的なイメージをどう具体化するかの問題と情報の検索プロセスをどのように意味的推論へ拡張するかの問題がある。

### 2. 研究の目的

(1) 従来の Web ではキーワードが一致した Web ページを探すだけであったが、セマンティック Web ではデータの意味内容を扱って Web 上の大規模分散データから質問に対する真の答えや推論結果を返すことを目指している。

(2) 本研究では、近年増加しているリンクトデータと呼ばれるセマンティック Web データに対する推論エンジンを開発する。特に、ユーザーの直感的または断片的な入力情報を手がかりにリンクトデータから有益な情報を発見するために、専門的なクエリ(質問)文を入力することなく自動的にクエリ文を生成する機能をもつ意味的推論メカニズムを提案する。そのために、質問が何を問うているかを定めたクエリタイプのオントロジーを新たに構築する。

### 3. 研究の方法

(1) 本研究ではリンクトデータから推論するために、クエリ自動生成を可能にする上位オントロジーを用いた意味的推論メカニズムを設計する。意味的推論メカニズムは、共通項をもつ概念(最大下界の共通概念)を導いたり、共通の特徴をもつインスタンス(共通概念のインスタンス)を導いたりする妥当な推論規則により形式化される。

(2) また、知識構造の基盤としては、研究代表者の過去の科研費成果(決定ルールと構造オントロジーによる意思決定知識ベースシステム)に基づいて、インスタンスや具体的概念を導出するクエリタイプの上位オントロジーを構築していく。研究計画において、意味的推論を実現する3つの重要な項目に

は(i) クエリ生成のための上位オントロジーの構築、(ii) ユーザー入力からクエリタイプを自動生成するメカニズムの設計、および(iii) 意味的推論システムのアルゴリズム設計、が挙げられる。

### 4. 研究成果

(1) 平成25年度では、初年度としてリンクトデータ・プロジェクトやセマンティック Web の研究動向について RDF 推論や論理的推論を備えた既存システムの調査を行った。特に RDF データを格納しクエリを発行できるシステムのうち、世界的に使われている Jena、Virtuoso、Sesame、OWLIM を選定し、実際にベンチマークデータをローカル環境に読み込んで推論エンジンとしての性能を調査している。

既存システムを調査する一方で、本研究の特徴である複雑なクエリの実現のために、上位オントロジーの構築とそれを適用した意味的推論について理論的な分析を行った。特に、事象や合成的な概念のような論理的分析だけでは説明できない対象をオントロジー的な深い分析によってモデル化している。

さらに、意味的推論を使わない従来のキーワード検索(具体的には RDF クエリ)では探すことのできない推論例や複雑なクエリを書かないと導くことのできない推論例を分析した。これを基にして、従来の RDF クエリでは導出できない記述論理の表現に対する変数を RDF クエリ文に記述できるようにしている。その結果、記述論理の表現を含まない RDF データからでも、複雑な記述論理表現を含むクエリ結果を出力できる推論メカニズムを形式化した。

(2) 平成26年度では、リンクトデータからユーザーが求めている情報を得るために表現力の低い RDF データからより表現力の高い記述論理の推論を実現して意味的推論を強化している。Web にあるリンクトデータは RDF で書かれるため OWL オントロジーのような意味記述になっていないことが多い。一方で、ユーザーの関心は表現力が低い RDF では十分に表せず記述論理の表現力が求められる。

本研究では、RDF データから記述論理のクエリに答えるシステムを開発した。これは Web データとユーザーの関心の間に存在するギャップを埋める1つの方法である。実際に、DBpedia の RDF データを使用して本システムで実験を行い、閉世界仮説のもとで高い推論結果を示す実験を行った。

さらに、意味的推論システムと平行して、大規模データを推論・検索するための RDF データストアを独自に開発した。リンクトデータは非常に大きなデータであり、単純な検索でも時間計算量が多くなってしまふ。そのために独自の RDF データストアがインメモリ型で高速な検索を可能にしており、この RDF データストアの上で計算コストの高い意味的

推論を実現できる。

(3) 平成27年度では、クエリ自動生成による推論メカニズムを実現するために次の2つの課題に取り組んだ。

RDF用クエリ言語 SPARQL のクエリ文は SQL にも似た構文をもつが使いこなすのが簡単とは言えない。そこで、ユーザーが簡単に検索できるように、自然言語文による質問とキーワードによる検索方法を実装した。前者は自然言語からの SPARQL クエリへの変換であり、簡単な自然言語文をいくつかのクエリタイプに分類してそれぞれを適切な SPARQL クエリと対応させる。後者は RDF グラフ構造の大規模データから、キーワードを含む部分グラフデータの抽出を実現した。

冗長で断片的なセマンティック Web データからユーザーの質問に答える意味的推論エンジンを実現した。この実現のために、RDF データに対して推論と機械学習を適用する方法論とアルゴリズムを設計した。特に、冗長な RDF データからリソースの意味を特徴づけて重要なデータを抽出する方法を提案している。この特徴抽出を用いて、データ分類器を SVM で学習させる方法を実装し評価実験を行っている。これにより、ユーザーが関心のある対象がどのような概念に分類されるか推定可能になった。

(4) 平成28年度では、ユーザーの漠然とした関心から検索や質問に答えるために、これまでのキーワードや自然言語文からの推論に加えて、機械学習やデータマイニングによってユーザーが求める知識を推定する方法を次のように提案した。

RDF データは URI (Uniform Resource Identifier) やリテラルで記述された Web データのため非常に冗長性が高く、必要な情報がデータ構造内に埋もれてしまうことが多い。そこで SVM の機械学習器を RDF データに適用するために開発したカーネル関数により、RDF データからリソースに関する属性データの推定を可能にしている。従来の SPARQL では、RDF データの中身に精通したユーザーが適切なクエリ文により検索しなければならない。それに対して、本研究ではリソースを指定するだけでクエリ文なしで機械学習によってそのリソースのタイプや属性が推定できる。

データマイニングの手法を RDF データ用に改良することにより、出現頻度の高い RDF データの部分構造を発見できる。こうしたデータマイニング手法は、大量のデータからの頻度計算が必要なため独自開発したインメモリ型 RDF データストアシステムを改良して実現した。これを用いてリンクトデータの意味

的推論を検証するために、DBpedia や YAGO などの大規模リンクトデータを提案手法に適用した。その結果、一般的なリソースに関する概念の推定、リソース間の関係、人物リソースの共通属性などの知識を推定できることが確認された。

(5) 最終年度は、実 RDF データを用いて検索、推論、学習を融合した意味的推論の実用性・応用性の検証実験を次のように行った。

実験で実際に用いた RDF データは Web 上に公開されている DBpedia、YAGO、Wikidata などである。最初の成果として、帰納的学習によって、大規模な RDF データから頻出する部分パターンやルールを発見できることが確認できた。次に、人物、ものなどの属性情報を記述する RDF データを訓練データに用いてカテゴリを学習して、新たな人物やもののデータからカテゴリを推定する評価実験を行った。また、検索、推論と学習を融合した意味的推論の成果として、検索・推論から導けない不完全データを学習器から推定する質問応答システムを実現している。この実験の評価により、ユーザーが求める情報を推定する機能が強化できている。

自然言語文のテキストデータからオントロジーや RDF データを半自動構築して知識ベースを拡充する2つの方法の実験を行った。1つの方法では、国語辞書のような日本語の言い回しが整っているテキストからオントロジーを半自動的に構築できている。もう1つでは、Wikipedia のテキスト文から RDF データを半自動作成して、DBpedia の RDF データを補完する成果を得た。この結果によって RDF データを拡充できるので、推論と学習を融合した意味的推論の性能が強化されるだろう。しかし、自然言語テキストから RDF データの自動構築に関しては、自然言語文の扱いの難しさからさらに精度の向上が必要と考えられる。

## 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6件)

兼岩 憲, 藤原浩司, 「大規模 RDF グラフに対するインデックス付きデータ圧縮と高速検索」, 人工知能学会論文誌, Vol. 33, No. 2, pp. E-H43\_1-10, 査読有, 2018.

荒井大地, 兼岩 憲, 「RDF グラフの冗長な特徴表現に対するカーネル関数とその高速計算」, 人工知能学会論文誌, Vol. 32, No. 1, pp. B-G34\_1-12, 査読有, 2017.

Ken Kaneiwa, Philip H. P. Nguyen, and Riichiro Mizoguchi, A Logical and

Ontological Framework for Compositional Concepts of Objects and Properties, New Generation Computing, Vo. 33, No. 2, pp. 149-172, 査読有, 2015.

藤原浩司, 兼岩 憲, 「大規模 RDF グラフのための効率的なクエリ解決」, 人工知能学会論文誌, Vol. 29, No. 4, pp. 364-374, 査読有, 2014.

長井拓馬, 兼岩 憲, 「RDF データの極小モデル推論に基づく記述論理 ALCH(D) の概念学習」, 人工知能学会論文誌, Vol. 29, No. 3, pp. 343-355, 査読有, 2014.

Philip H. P. Nguyen and Ken Kaneiwa, Event Inference with Relation and Meta-relation Type Hierarchies in Conceptual Structure Theory, Applied Artificial Intelligence, Vol. 28, No. 2, pp. 139-177, 査読有, 2014.

〔学会発表〕(計 10 件)

荒井大地, 兼岩 憲, 「多様な RDF グラフを用いた機械学習のためのカーネル関数」, 第 43 回セマンティックウェブとオントロジー研究会, SIG-SWO-043-\*\*, 査読無, 2017.

廣橋美紀, 兼岩 憲, 「RDF データに対するグラフパターンマイニング」, 第 42 回セマンティック Web とオントロジー研究会, SIG-SWO-042-04, 査読無, 2017.

浜松良樹, 兼岩 憲, 「RDF グラフに対するキーワード検索の高速化と省メモリ化」, 第 42 回セマンティック Web とオントロジー研究会, SIG-SWO-042-05, 査読無, 2017.

香川俊幸, 兼岩 憲, 「FROST におけるデータストアの圧縮と読み込み手法」, 第 40 回セマンティック Web とオントロジー研究会, SIG-SWO-040-06, 査読無, 2016.

荒井大地, 兼岩 憲, 「RDF グラフの冗長な特徴表現に対するカーネル関数」, 第 38 回セマンティックウェブとオントロジー研究会, SIG-SWO-38-\*\*, 査読無, 2015.

兼岩 憲, 藤原浩司, 「大規模 RDF グラフに対する高速検索とデータ圧縮の両立 (優秀賞招待講演)」, 人工知能学会合同研究会, 査読無, 2015.

長井拓馬, 兼岩 憲, 「RDF データの極小モデル推論に基づく記述論理の概念生成」, 第 35 回セマンティックウェブとオントロジー研究会, SIG-SWO-35-\*\*, 査読無, 2014.

藤原浩司, 兼岩 憲, 「大規模 RDF グラフに対する高速検索とデータ圧縮の両立」, 第 35 回セマンティックウェブとオントロジー研究会, SIG-SWO-35-\*\*, 査読無, 2014.

長井拓馬, 兼岩 憲, 「RDF データに対す

る閉世界仮説上の SROIQ クエリ」, 第 4 回情報処理学会東北支部研究会, 査読無, 2013.

藤原浩司, 兼岩 憲, 「高速なクエリ解決を実現する RDF ストアの構築」, 第 4 回情報処理学会東北支部研究会, 査読無, 2013.

〔図書〕(計 2 件)

兼岩 憲, 小項目「記述論理」, 人工知能学事典, 人工知能学会(編), 共立出版, 2017.

兼岩 憲, 「セマンティック Web とリンクトデータ」, コロナ社, 2017.

〔その他〕

ホームページ等

<http://kjk.office.uec.ac.jp/Profiles/66/0006541/profile.html>

6. 研究組織

(1) 研究代表者

兼岩 憲 (KANEIWA KEN)

電気通信大学・大学院情報理工学研究科・教授

研究者番号: 00342626