

平成22年 6月 7日現在

研究種目：若手研究(A)
 研究期間：2007～2009
 課題番号：19680008
 研究課題名（和文） 多戦略学習手法に基づくオントロジー・アライメントに関する研究
 研究課題名（英文） Ontology alignment method based on multiple strategy learning
 研究代表者
 市瀬 龍太郎（ICHISE RYUTARO）
 国立情報学研究所・情報学プリンシプル研究系・准教授
 研究者番号：00332156

研究成果の概要(和文):本研究の目的は、異なるオントロジー間の対応関係を自動的に発見し、複数の異なるオントロジーを連携させるための高度な機械学習技術を開発することにある。そこで、本研究では、複数の機械学習手法を組み合わせることによる高度なオントロジー・アライメント手法の開発と各アプローチを定量的に議論できる実践的なデータセットの開発を行った。

研究成果の概要(英文): We developed high performance ontology alignment method based on multiple strategy learning. In addition to it, we developed practical dataset used for evaluating ontology alignment methods.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	5,100,000	1,530,000	6,630,000
2008年度	4,100,000	1,230,000	5,330,000
2009年度	2,700,000	810,000	3,510,000
年度			
年度			
総計	11,900,000	3,570,000	15,470,000

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：知能情報処理

1. 研究開始当初の背景

大量の情報を人々が容易に入手できるようになり、そのような情報の処理をどのようにするのが大きな課題となっている。セマンティックWebは、入手できる情報にメタデータを付与することで、推論機構などを使って自動的に多くの情報を処理しようという試みであり、次世代のWeb技術として世界的

に注目を集めている。セマンティックWebでは、異なる情報に対して意味の整合性をとるために、人間が持つ概念体系を形式的に整理したオントロジーをメタ情報として用いている。しかし、オントロジーは、作成者の目的、用途などに応じて作られるものであるため、異なるオントロジーを使った場合に、メタデータにつけられた情報同士が相互に利用できないという状況が生まれる。そのた

め、オントロジー同士を精度高く自動的に対応付ける技術（オントロジー・アライメント技術）の開発が喫緊の課題となっている。これまでに様々な取り組みがなされてきているが、客観的に性能を評価するための実践的なデータセットがほとんど整備されていないため、異なる手法間の長所短所の比較が難しく、厳密な議論に基づいて、お互いの短所を補うような多戦略学習に基づく手法の開発が十分に行われていないのが現状である。

一方、国際的には、オントロジーツールの評価研究会（EON）で、オントロジー・アライメントの国際的な性能評価研究会が2004年に開かれ、問題の重要性が認識されてきた。2005年には、AIマガジンに「意味に応じた統合」という特集号が発行され、この問題に対する国際的な注目度がいっそう高まると同時に、オントロジー・アライメント・評価・イニシアティブ（OAEI）が設立され、この問題に対する解決手法を国際的な協調の中で開発して行こうという機運が高まっている。OAEIでは、性能を評価するためのデータセットを研究者に提供することで、手法同士の客観的な評価を進め、コミュニティを通して技術の躍進を狙っている。また、セマンティックWebと情報システム国際ジャーナルでは、オントロジー・マッチングに関する特集号が生まれ、世界中から多数の論文が集まるなど多くの研究者がこの分野に取り組み始めている。

2. 研究の目的

本研究の目的は、異なるオントロジー同士を精度良く自動的に対応付けるためのオントロジー・アライメント手法の開発である。そのためには、単一の学習手法を用いた場合の限界をうち破る手法の開発が必要となると同時に、オントロジー・アライメントに対する各学習手法の特性を明確にすることが欠かせない。そこで、本研究では、次の2つを実施することにより、研究の目的の達成を試みる。

- ・ 複数の機械学習手法を組み合わせることによる、単一の学習アプローチを用いた場合を超える性能のオントロジー・アライメント手法の開発
- ・ 各アプローチを定量的に議論できる実践的なデータセットの開発

目的の達成のためには、これまでに開

発してきた機械学習の手法に基づくものを多戦略学習手法に基づいたものに拡張していくことが、これまでの研究から有効であると考えられる。しかし、そのためには、それぞれの手法の特性を厳密に評価することが欠かせない。OAEIでは、いくつかのデータセットを提供し、国際的な性能評価コンテストを通して、より精度の高い手法の開発を促進している。しかし、OAEIに関わっている研究者の多くは、人間の知識をいかにして記述するかを研究してきた知識工学の研究者であり、機械学習を専門としている研究者は少数であるため、提供されているデータセットは、正答を完全には含まないなど、機械学習の観点からは評価に耐えるようなデータではない。また、データセットは、機械的に生成したものが中心となっており、実践的なデータは提供されていないのが現状である。機械学習の研究では、カリフォルニア大学アーバイン校が実践的なベンチマークデータを整備し、まとめて提供することによって、手法同士の客観的な性能比較が促進され分野が大きく進展した。そこで、本研究では、多戦略学習手法に基づくオントロジー・アライメントの手法の開発と同時に、評価用の実践的なデータセットを同時に開発し、手法の開発と客観的な性能の評価を同時並行に行うことによって、目的の達成を試みる。

3. 研究の方法

3年間の研究を下記のような形式で遂行した。

(1) 2007年度

2007年度は、研究の開始年に当たるため、主に研究に必要な環境の構築を行い、下記の2つに分けて研究開発を実施した。

① データセット作成ツールの開発

本研究では、オントロジー・アライメントを客観的に検証するのに必要な実践的なデータセットの開発を行う。そのため、作成したデータセットを読み書きできるツールを開発し、データセットの作成を容易にできるような環境を構築した。また、各アライメントに対して、正しい(正例)か正しくない(負例)かの判定を人手で行う際に、容易に行えるツールを作成し、次年度以降の研究に向けた環境整備を行った。

② アライメント学習手法の特性

評価

これまでに、研究が行われてきたアライメント学習手法を評価することによって、各手法の特性を明らかにした。そのことによって、学習手法をどのように組み合わせると、有効なアライメント手法が構成できるかの基礎的なデータを得た。

(2) 2008年度

2008年度は、大規模データセットの開発と多戦略アルゴリズムの開発の2つに分けて研究開発を実施した。

① 大規模データセットの構築

本研究では、オントロジー・アライメントを客観的に検証するのに必要な実践的なデータセットの開発を行う。そのため、前年度に開発したパイロットデータを発展させ、データセットの作成を行った。また、それを国際ワークショップに提供し、国際的に多くの参加者が性能評価データとして利用することで、研究の促進を図った。

② データセットを利用した多戦略学習アルゴリズムの開発と評価

データセットを利用して、たくさんのアライメント手法を評価することで、各手法の特性を明らかにした。その結果、さまざまなアライメント手法の中で、有効な手法と効果の薄い手法を判別することができた。また、それらの手法を自動的に組み合わせ、高い性能が得られるような機械学習手法を構築した。

(3) 2009年度

2009年度は、データセットの洗練と多戦略アルゴリズムの高度化の2つに分けて研究開発を実施した。

① データセットの洗練

本研究では、オントロジー・アライメント技術を客観的に検証するのに必要な実践的なデータセットの開発を行った。これらのデータセットなどを用いて、少ない数で高精度な学習をするための方法論を検証した。その結果、大きなデータ

セットに対して、少ないデータ数でも学習が十分にできることが確かめられ、データセットの洗練に十分な正答数などを明らかにした。

② 多戦略学習アルゴリズムの高度化

従来から研究してきた学習アルゴリズムに、ユーザとのインタラクションを導入することで、より高精度な学習ができるような手法の開発をおこなった。その結果、従来の学習手法に対して、ユーザが少し介入するだけで、大きく性能を改善できることが示された。

4. 研究成果

(1) 2007年度

前章で述べた研究で得られた成果に基づき、論文2本の出版、および、学会発表1件を行った。

(2) 2008年度

前章で述べた研究で得られた成果に基づき、論文3本の出版、および、学会発表1件を行った。

(3) 2009年度

前章で述べた研究で得られた成果に基づき、論文3本の出版を行った。

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8件)

1. Raul Ernesto Menendez-Mora, Ryutaro Ichise, The Role of Taxonomy Properties in Information Content Metrics, pp. 22-26, 2010
2. Hoai-Viet To, Ryutaro Ichise, Hoai-Bac Le, An Adaptive Machine Learning Framework with User Interaction for Ontology Matching, Proceedings of IJCAI Workshop on Information Integration on the Web, pp. 35-40, 2009
3. Ryutaro Ichise, Evaluation of Similarity Measures for Ontology Mapping, New Frontiers in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 5447, pp. 15-25, 2009

4. Till Wohlfarth, Ryutaro Ichise, Semantic and Event-Based Approach for Link Prediction, Proceedings of the 7th International Conference on Practical Aspects of Knowledge Management, pp. 50-61, 2008
5. Caterina Caracciolo, Jerome Euzenat, Laura Hollink, Ryutaro Ichise, Antoine Isaac, Veronique Malaise, Christian Melicke, Juan Pane, Pavel Shvaiko, Heiner Stuckenschmidt, Ondrej Svab, Vojtech Svatek, Results of the Ontology Alignment Evaluation Initiative 2008, Proceedings of the 3rd International Workshop on Ontology Matching, pp. 73-119, 2008
6. Ryutaro Ichise, Machine Learning Approach for Ontology Mapping Using Multiple Concept Similarity Measures, Proceedings of the 7th IEEE/ACIS International Conference on Computer and Information Science, pp. 340-346, 2008
7. 市瀬 龍太郎, 計算機による分類概念の構築, 情報の科学と技術, Vol. 58, No. 2, pp. 78-83, 2008
8. 市瀬 龍太郎, 情報の意味的な統合とオントロジー写像, 人工知能学会誌, Vol. 22, No. 6, pp. 818-825, 2007

[学会発表] (計 2 件)

1. Milen Pavlov, Ryutaro Ichise, Finding Experts by Link Prediction in Co-authorship Networks, Proceedings of the 2nd International Workshop on Finding Experts on the Web with Semantics, pp. 42-55, 2007
2. 市瀬 龍太郎, オントロジーマッピングに有効な特徴の抽出, 第 22 回人工知能学会全国大会, 2E1-1, 2008

[図書] (計 0 件)

[産業財産権] (計 0 件)

○出願状況 (計 0 件)

名称 :
 発明者 :
 権利者 :
 種類 :
 番号 :
 出願年月日 :
 国内外の別 :

○取得状況 (計 0 件)

名称 :
 発明者 :
 権利者 :
 種類 :
 番号 :
 取得年月日 :
 国内外の別 :

[その他]

ホームページ等

<http://ri-www.nii.ac.jp/OAEI/2008/index.html>

6. 研究組織

(1) 研究代表者

市瀬 龍太郎 (ICHISE RYUTARO)

国立情報学研究所・情報学プリンシプル研究系・准教授

研究者番号 : 00332156

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし