

Title	Knowledge Discovery about Relationships between Objects on the Web( Abstract_要旨 )
Author(s)	Zhang, Xinpeng
Citation	Kyoto University (京都大学)
Issue Date	2012-03-26
URL	<a href="http://hdl.handle.net/2433/157476">http://hdl.handle.net/2433/157476</a>
Right	
Type	Thesis or Dissertation
Textversion	none

( 続紙 1 )

京都大学	博士 ( 情報学 )	氏名	張 信鵬
論文題目	Knowledge Discovery about Relationships between Objects on the Web ( ウェブ上のオブジェクト間の関係に関する知識の発見 )		
<p>( 論文内容の要旨 )</p> <p>実社会ではオブジェクト間の関係に関する知識を検索する要求が多くある。具体的には、石油と日本との関係、地球温暖化と農業相互の影響に関する関係、人物間の関係などである。既存の関係検索としては、セマンティック関係の検索が挙げられるが、これは予め決められた単純な関係しか表現することができない。本論文はオブジェクト間の多様かつ複雑な暗示的关系に関する知識の発見について、以下の四項目の研究成果を纏めたものである。</p> <p>第一に、オブジェクト間の関係を解析するための「減衰流モデル」の提案である。頂点がオブジェクトを表し、枝がオブジェクト間の明示的关系を表す情報ネットワークにおいては、関係の強さを測るための重要な概念として、距離、連結度、共引用が利用されてきた。しかし、既存研究ではこれら三つの概念とも反映する手法は存在していなかった。本論文では、距離、連結度、共引用の三つの概念全てに基づき、最大減衰流 ( Generalized Maximum Flow ) を用いて関係の強さを測る「減衰流モデル」を開発した。Wikipediaでは、ページをオブジェクトと見なし、ページ間のリンクをオブジェクト間の明示的关系と見なすことができる。Wikipediaのリンク構造により構築したWikipedia情報ネットワークなどの巨大な情報ネットワークには、二つのオブジェクト間の関係にとり重要な情報とそれ以外の情報が存在する。減衰流モデルは、情報ネットワークにおいて関係に重要な枝に低い減衰率を与えることにより、重要な情報を反映した関係の強さを測定することができる。大量データを用いた広範な実験により、減衰流モデルが既存手法と比べてより適切に関係の強さを表現することを確かめた。</p> <p>第二に、Wikipediaにおける関係を理解するための知識の抽出手法の提案である。Wikipediaでは、複数の記事間のリンクをたどることで、記事が表すオブジェクト間の暗示的关系をある程度理解することができる。例えば、「メキシコ湾」が「アメリカ」の重要な「石油」生産地の一つであるという「アメリカ」と「石油」の暗示的关系は、「石油」と「メキシコ湾」間のリンクと「メキシコ湾」と「アメリカ」間のリンクにより表されていると考えられる。しかし、人間がこのような暗示的关系を発見するには多くのページとリンクを調べる必要があるため、現実的な時間では困難である。本論文ではWikipedia情報ネットワークにおいて、オブジェクト間の関係を理解するために、「減衰流モデル」に基づいて、その関係を表す重要な独立した経路を抽出する手法を開発した。実験により、大量の独立した経路が抽出できること、及び、抽出した経路が関係の理解に有用であることを確認した。</p> <p>第三に、関係に関する知識を持つマルチメディア情報の検索手法の提案である。一般のウェブと比較すると、Wikipedia上に存在する知識はまだ限られており、特に、画像などのマルチメディア情報が乏しい。項目二で開発したWikipedia情報ネットワ</p>			

ークから関係に重要な独立した経路を抽出する手法において、経路上にあるオブジェクトがその関係にとって重要な役割を果たす中間オブジェクトとなる。本研究では、中間オブジェクトに着目し、画像の周辺テキストにその関係の中間オブジェクトがたくさん現れていれば、その関係に関する記述があると考え、ウェブから関係に関する知識を含む画像とその周辺テキストを取得する手法を開発した。ユーザ評価実験でこの手法により高い精度で関係に関する知識を含む画像と周辺テキストを取得できることを確認した。また、ウェブから取得したこれらの画像とその周辺テキストは、Wikipedia上に存在しない知識を多く含むことも確認できた。さらに、項目二で説明した手法と本手法を利用し、オブジェクト間の多様な関係を理解するための豊富なマルチメディア情報を提供する関係検索システムを構築した。

第四に、項目一、二、三で説明した技術を利用し、Wikipedia記事に掲載された画像の記事に対する相応しさを評価するシステムを開発した。開発したシステムは、ある記事中の画像が、当該記事オブジェクトとその画像のキャプション中の各リンクに対応するオブジェクトとの関係を表すと仮定し、画像が表す関係を分析することにより、関連度、整合度、典型度の三つの尺度から画像が元の記事とキャプションに対し相応しい度合いを算出する。ユーザ評価実験では、本システムはある程度人間評価に近い画像の相応の判定ができることを確認した。さらに、項目三の技術を利用し、当初の画像より相応しい画像をWebから取得し提示することにより、Wikipediaの編集の支援になる。

注)論文内容の要旨と論文審査の結果の要旨は1頁を38字×36行で作成し、合わせて、3,000字を標準とすること。

論文内容の要旨を英語で記入する場合は、400～1,100 wordsで作成し  
審査結果の要旨は日本語500～2,000字程度で作成すること。

(論文審査の結果の要旨)

本論文は複雑なオブジェクト間の暗黙的關係を理解するために、関係に関する多様性のある知識をWikipediaとウェブから発見する手法を提案している。従来、単純な明示的關係を抽出する手法は存在したが、複雑な暗黙的關係に関する知識を発見する手法が存在しなかったという点で、提案手法の新規性が評価される。

関係を理解するために関係の強さを測定することが有用である。情報ネットワークにおいて関係の強さを測るために、距離、連結度、共引用が三つの重要な概念として利用されてきた。しかし、関係の強さとしてこれら三つの概念を共に反映できる手法は存在しなかった。本論文で提案された情報ネットワークにおける減衰流モデルは、上述の三つの概念全てに基づき、減衰率の設定により情報ネットワークにおいてオブジェクト間の関係に対して重要な情報を反映して関係の強さを測ることができる。また、情報ネットワークでは、次数が高く関係に重要な役割を果たすノードが存在するが、従来の凝集という概念に基づいた手法では、情報ネットワークにおいて次数が高いノードの重要さが低く評価される。それに対し、減衰流モデルでは次数が異なるノードが公平に扱われる。以上に述べるように、減衰流モデルが既存手法と比べて優れた性質を持つことが高く評価できる。また、WikipediaデータセットとDBLPデータセットを用いた広範な実験により、減衰流モデルが既存手法と比べてより適切に関係の強さを表現できることを具体的に示した点も評価できる。

関係の強さの計測だけではなく、関係を理解しやすいように可視化することも重要である。従来、情報ネットワークから関係を説明するための部分グラフを抽出する研究があったが、部分グラフが複雑な場合は理解が困難となる。本論文では、Wikipedia情報ネットワークから関係を表す重要な独立した経路を抽出するための減衰流に基づいた手法を提案している。本手法により抽出された経路は理解が容易である点、冗長な情報を含まない点で評価できる。また、人間評価により、提案手法により抽出された経路が関係の理解に有用であることを確認した点も評価できる。

ウェブの知識は体系化されておらず、関係に関する知識の抽出は難しいとされている。本論文では、Wikipediaから抽出した関係に重要な役割を果たす中間オブジェクトを利用し、関係に関する知識を有する画像とその周辺テキストを取得する手法を提案している。本手法には新規性が認められ、評価実験でこの手法がウェブから関係の理解に有用であり、しかもWikipediaに存在しない知識を取得できることを確認した点も評価できる。さらに、提案手法を利用し開発した「縁」関係検索システムの新規性と有用性、及び、提案した関係の解析手法を利用し、Wikipedia記事の画像の記事に対する相応しさを評価するシステムの新規性も評価できる。

このように、本論文は、Wikipediaまたはウェブにおける関係、特に暗黙的關係に関する知識を発見する手法の提案を行ったものであり、それらの手法を応用した関係検索システムとWikipedia画像の相応しさを判定システムを開発した点で研究の新規性と有用性が高く評価される。

よって、本論文は博士(情報学)の学位論文として価値あるものと認める。

また、平成24年2月10日に実施した論文内容とそれに関連した試問の結果合格と認めた。

注)論文審査の結果の要旨の結句には、学位論文の審査についての認定を明記すること。

更に、試問の結果の要旨(例えば「平成 年 月 日論文内容とそれに関連した口頭試問を行った結果合格と認めた。」)を付け加えること。

Webでの即日公開を希望しない場合は、以下に公開可能とする日付を記入すること。

要旨公開可能日： 年 月 日以降