

深層格とオントロジーを用いた回答抽出による 質問応答システム Metis の回答精度向上

原田 実[†] 西岡 晋太郎[‡]

青山学院大学理工学部情報テクノロジー学科[†]

青山学院大学大学院理工学研究科知能情報コース[‡]

1. はじめに

近年、自然文で書かれた質問文に対して回答を抽出する質問応答システムの研究が盛んに行われており、NTCIR の QAC や CLQA 等、評価型ワークショップなどで研究成果が公表されている。

本研究室で開発している質問応答システム Metis[1]は、意味グラフベースで質問文と知識文を照合し、2つのグラフ間のグラフ類似度で正解を含む確度を判定し、最大の類似度を含む知識グラフから回答を抽出する。

本研究では、主として質問の中心となる述語に質問内容を表す深層格に係る文節あるいはそれを根とする部分木から回答抽出を行うこと、および、数値的な属性に関する単位などをオントロジーから得てそれを補完した質問文を知識文と照合することによる2種類の方法で回答精度を向上させた。

2. Metis の概要

Metis は、図1に示すように、自然言語で入力された質問文と新聞記事や Web 中の文章（知識文）の両方を意味解析システム SAGE[2]で意味グラフに変換する。ここでは、文は文節を頂点、係り受けを辺にしたグラフで表現され、文節の主辞には EDR 辞書の語意が、辺には SAGE が定めた36種の深層格のどれかが割りあてられる。この両グラフの共通部分グラフ（閾値以上の語意類似度を持つ頂点ペアとその間の閾値以上の格類似度を持つ辺ペア）の大きさで文間の類似度（グラフ類似度）を判定する。例えば、図1に示すように Metis に質問文「ペスト菌を発見した細菌学者は誰ですか。」が入力されると意味解析を行った後、質問文解析を行う。この処理では疑問詞が特定され、それに与える意味制約（誰）を決定する。次に、質問文から検索エンジン呼び出すためのキーワード（ペスト菌、発見、細菌学者）を抽出し知識文検索を行う。

Answer accuracy improvement of question answering system Metis by answer extraction using deep case and ontology.

[†]Faculty of Science and Engineering, Department of Integrated Information Technology, Aoyama Gakuin University.

[‡]Graduate School of Science and Engineering, Intelligence and Information Course, Aoyama Gakuin University.

得られた知識文と質問文の意味グラフを照合しグラフ類似度が最上位の知識文から回答（北里柴三郎）を抽出する。

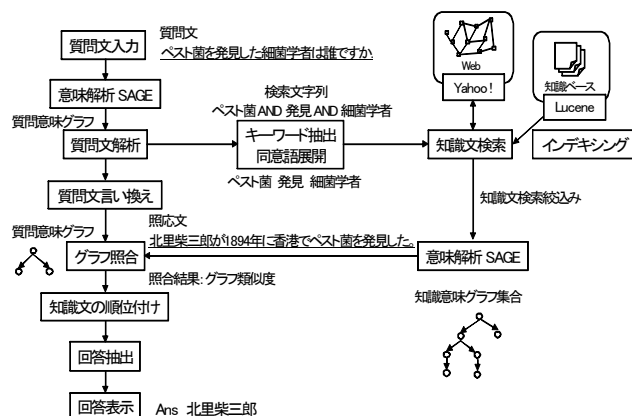


図1 質問応答システム Metis システムの概要

3. 回答抽出方法

回答抽出の基本方針を述べる。質問文とグラフ類似度が閾値以上の知識文を照応文という。質問グラフにおいて疑問詞を持つノードを質問ノードと定め、照応文中で質問ノードと照合したノードあるいはそれを根とする部分木から回答を抽出する。

本研究では、上記の基本方針に加え深層格とオントロジーを利用した回答抽出を行った。以下に、それぞれの回答抽出方法を示す。

3.1 深層格に基づく回答抽出

質問グラフにおいて質問の中心となる述語を表わすノードを主述語ノードと呼ぶ。以下に深層格による回答抽出方法を述べる。

- (1) 主述語ノードから見た質問ノードの深層格を特定する。
- (2) 主述語ノードと照合する照応文のノードに(1)で得た深層格に係るノードを特定する。
- (3) (2)で特定したノードあるいはそれを根とする部分木を回答として抽出する。

例えば、質問文「北里柴三郎はペスト菌をいつ頃発見しましたか？」に対して照合した知識

文が「北里柴三郎は政府より派遣された際、病原菌であるペスト菌を発見するという業績をあげた。」である場合、図 2 に示すように主述語ノード「発見しましたか？」に質問ノード「いつ頃」は time 格で係り、主述語ノードと照合する「発見すると」に time 格で係る部分木「政府より派遣された際」を回答する。

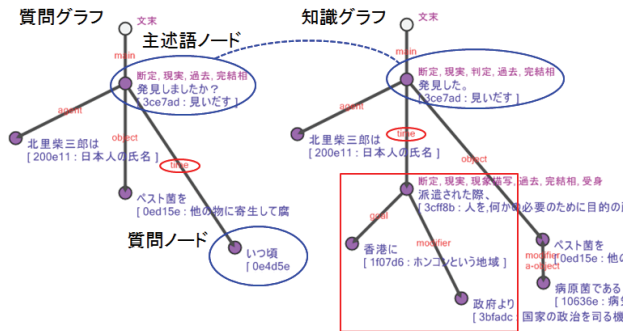


図 2 深層格で係る部分木からの回答抽出

3.2 オントロジーを用いた回答抽出

本システムは照応文より回答を抽出している。しかし、照応文に回答が含まれているとは限らず、その前後の文に回答が含まれている場合がある。例えば、図 3 に示す質問文に対する照応文は知識記事下線部であるが、正解回答はその後の文の二重下線部「178センチ」である。人間の知識や常識をオントロジーに記述しておき、この情報をもとに質問文を補完し、照応文の前後の文から回答を抽出する。

質問文：
モハメド・アリの娘のレイラの身長はどれくらいですか？

知識記事：
(ロサンゼルス 28 日時事) プロボクシングの元世界ヘビー級チャンピオン、モハメド・アリ氏 (57) の末娘レイラさん (21) が 10 月 8 日、ニューヨークでプロボクサーとしてデビューすることが 28 日、発表された。
178センチ、76キロ。

図 3 質問文と知識記事の例

本研究では、事物の属性についてオントロジーを構築し、人間の属性 4 個、組織の属性 1 個、地域の属性 13 個、時間の属性 6 個、数の属性 37 個、競技の属性 80 個用意した。例えば、人間の属性として表 1 のようなオントロジーを構築し、「人間の身長は cm という単位をもつ値である」という知識を補い、照応文の後の文から「178センチ」を回答する。

表 1 人間オントロジー

entity	attribute	value	conceptID
人間	名前	人名	4167196
人間	年齢	歳	3948238
人間	身長	cm	1024589
人間	体重	kg	3932805

4. 評価実験

NTCIR-6 CLQA (factoid 型 200 問) のデータセットに基づき毎日新聞記事 1998 年と 1999 年の 2 年分を知識源として評価実験を行った。

評価方法は、システムが一つの質問に対して 5 位までに正解を回答した割合を求めた回答抽出精度と、システムが一つの質問に対して 5 位までに正解を回答した順位の最高順位の逆数の平均 MRR の二つとした。表 1 と表 2 にそれぞれの実験結果を示す。

表 2 実験結果 (回答抽出精度)

検索	正解数		正解抽出精度		回答抽出精度
	1 位	5 位まで	1 位	5 位まで	
198/200	86	150	43.4% (86/198)	75.7% (150/198)	75.0% (150/200)

表 3 実験結果 (MRR)

回答順位					MRR
1 位	2 位	3 位	4 位	5 位	
86	35	17	7	5	0.559

5. 結論

本研究では、深層格とオントロジーを用いた回答抽出を行った。概念の類似性だけでなく深層格に基づく意味的役割を考慮することにより、照応文に質問で問われている概念がない場合でも回答を抽出することができた。また、事物の属性についてのオントロジーを構築し、照応文の周りから回答を抽出することができ、本研究の有効性を示すことができた。

6. 参考文献

- [1] 西岡晋太郎, 久保田裕章, 坂東晃文, 原田実: "意味グラフ照合による質問応答システム Metis の回答精度向上 - 質問文解析フェーズと検索フェーズの改良を中心として -", 情報処理学会研究報告, Vol.2009-NL-191 No.16, pp.1-8, (2009).
- [2] 原田実, 水野 高宏: "EDR を用いた日本語意味解析システム SAGE", 人工知能学会論文誌, Vol.16, No.1, pp.85-93 (2001).