

質問応答における構文的照合と言い換えの効果

高橋哲朗 縄田浩三 乾健太郎 松本裕治

奈良先端科学技術大学院大学情報科学研究科

{tetsu-ta, kozo-n, inui, matsu}@is.aist-nara.ac.jp

1 質問応答タスク

本稿で対象としている質問応答タスクとは、自然言語で問われた質問に対して、あらかじめ与えられた文書集合を情報源としてその質問の回答となる語句を返すというタスクである。このタスクに対する先行研究では、単語間の類似度や近接性などを使って回答を見付ける手法 (bag-of-words) が主流であった。一方、質問応答タスクを言語理解の問題ととらえると、理想的には、質問と情報源をそれぞれ意味構造に変換し、そのレベルで照合するという方法が考えられる。しかし、こうした深い意味理解に基づく方法は、意味構造を抽出するためのオーバーヘッドやその精度などの問題から、当面は実現困難である。そこでまずは、bag-of-wordsに基づく照合と意味構造間の照合の間でうまく折り合いをつける道を探ることが重要だと考えた。そうした方法の一つに、構文構造間の照合 (以下、構文的照合) を用いる手法が考えられる。質問応答に構文構造を使う研究は、これまでもいくつか見られるが [6, 4, 10], 構文的照合を質問応答に適用することの効果についての詳しい分析はされていない。このような背景から、本稿では質問応答に対する構文的照合の可能性について議論する。

構文的照合による回答探索では、次の2点が議論の焦点となる。

- 構文的照合の方法

質問文と情報源の文書が完全に同一の構文構造で出現することは極めて稀なので、柔軟に照合できる枠組が必要となる。また同時に計算コストも抑える必要がある。

- 言語の多様性

質問文とその正解を含む文書の間には多くの場合表現の差があるので、その差を吸収する必要がある。

これらの問題に対して我々は (a) 柔軟な構文的照合と (b) 言い換えを用いた回答探索、に基づくアプローチを選択した。

以下、2節で (a) について述べ、3節で (b) について述べる。次に4節で実験結果を報告し、5節で誤り分析をおこなう。

2 柔軟な構文的照合

構文的照合には、Collinsらの提案した Tree Kernel [1] を拡張したアルゴリズム [9] を用いた。その拡張は次のようにまとめられる。

- Tree Kernelの対象は順序付きの句構造木だったが、順序なしの木に対して適用可能にした。これにより、同様のアルゴリズムを日本語の表現に適した依存構造木に適用できるようになった。
- ノード間の類似度を導入することにより、形態素間の意味的な類似度を取り入れた。
- Tree Kernelでは与えられたノードをルートとする部分木の数を求めているが、与えられたノードを含む共通部分木の数を求めるようにし、より正確にノード間の対応を取れるようにした。

依存構造木 T_1 と T_2 の間の類似度は式 (1) により求められる。

$$K(T_1, T_2) = \sum_{n_1 \in N_1} \max_{n_2 \in N_2} C(n_1, n_2). \quad (1)$$

ここで N_i は T_i が持つノード集合であり、 $C(n_1, n_2)$ は n_1 と n_2 を含む部分木間の共通の部分木の数を返す。 $C(n_1, n_2)$ は式 (2) で求められる。

$$C(n_1, n_2) = C_{bu}(n_1, n_2) \times C_{td}(n_1, n_2) \quad (2)$$

$$C_{bu}(n_1, n_2) =$$

$$\text{sim}(n_1, n_2) \prod_{k \in \text{ch}(n_1)} \prod_{l \in \text{ch}(n_2)} (1 + C_{bu}(k, l))$$

$$C_{td}(n_1, n_2) =$$

$$\left(C_{td}(p(n_1), p(n_2)) \times \frac{C_{bu}(p(n_1), p(n_2))}{C_{bu}(n_1, n_2)} \right) + 1$$

$\text{ch}(n)$ は n の子ノードの集合を、 $p(n)$ は n の親ノードを表し、 $\text{sim}(n_1, n_2)$ は2つのノード n_1 と n_2 の間の類似度を $[0, 1]$ の範囲で返す関数である。はじめにすべてのノード間の $C_{bu}(n_1, n_2)$ をボトムアップに求め、次に $C_{bu}(n_1, n_2)$ の値を使って $C_{td}(n_1, n_2)$ をトップダウンに求める。そしてそれらの値の積により $C(n_1, n_2)$ が求められる。

$K(T_1, T_2)$ は、bag-of-wordsによる内積の計算量と同じ $O(|N_1||N_2|)$ の時間で計算できる。

3 言い換えを用いた回答の探索

言語表現には冗長性があるため、質問文と情報源の文書が同じ意味を持っていたとしても、多くの場合異なる構文構造で表現されている。例えば (3) において、質問文 Q の「原料にしている」と情報源 D の「原料に作られた」の間の照合は難しい。仮にこれらを同一の表現と見なせたとしても、「バルサミコ酢は」は Q では「している」に係るが、D では「もの」に係るため構文的照合はできない。

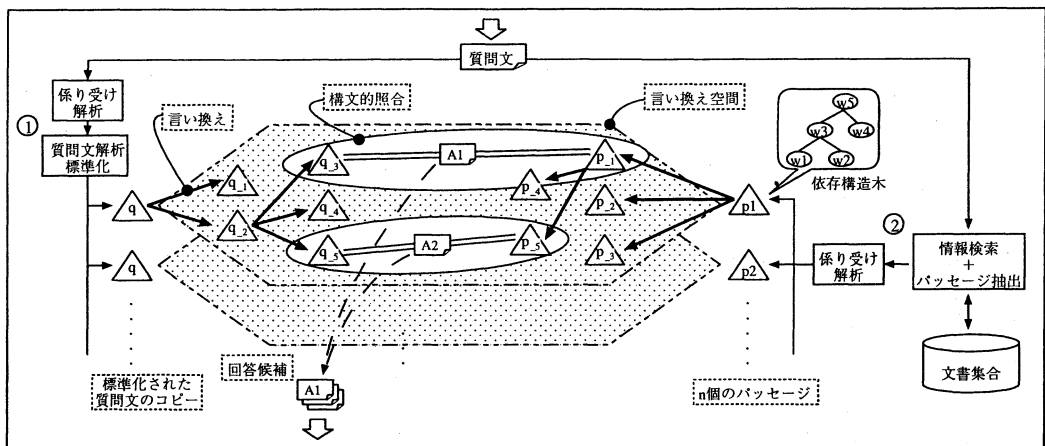


図 1: 言い換え空間における回答の探索

(3) Q. バルサミコ酢は何を原料にしていますか。

D. バルサミコ酢とは、トレヴィアノ種という甘みの強いぶどうを原料に作られたもの。

このような構文構造上の差異を吸収するために、我々は質問応答に言い換えを取り入れた。言い換えには規則ベースの言い換えエンジン KURA [8] を用いた。KURA は与えられた規則に従って、入力テキストに対して適用可能な構文的トランスファをおこない言い換えを生成する。言い換え規則には、これまでに作成されてきた 8 種類 約 15,000 規則¹ に加え、新たに質問応答用に作成した 7 種類 45,000 規則を用いた。これらの規則は名詞句から語釈文への言い換えなどの辞書を用いる言い換えを含む。

言い換え規則は個別の質問に特化したものではなく汎用的であるので、1 回の言い換えで照合が可能になるとは限らない。そこで質問応答を言い換え空間内の探索問題ととらえ、次に示すアルゴリズムを用いた (図 1)。

前処理 (質問文) 質問文の係り受け解析、質問文解析、標準化をおこなう ①(4.1 節参照)

前処理 (情報源) 質問文中の情報を使い照合の候補となる n 個のパッセージを抽出し、各パッセージを係り受け解析することにより依存構造木に変換する ②

探索 類似度が上がらなくなるまで繰り返す

1. 可能な言い換えをすべておこなう
2. 言い換えられた質問文とパッセージのすべての組み合わせにおいて類似度を計算する
3. 最も類似度の高い組み合わせを選択する

回答選択 構文的照合における類似度に基づき回答候補をランキングする。

¹ <http://cf.aist-nara.ac.jp/lab/kura/KuraData/>

4 実験

質問応答タスクは、質問文解析、情報検索、パッセージ作成、回答の検索など複数の段階からなっている。しかし、本稿では構文的照合と言い換えの影響に焦点を当てるために、質問文とその正解を含んでいるパッセージを入力として与え、そこから回答を見付けるというタスク設定で実験をおこなった。

データには、NTCIR-3 の QAC1 [2] で配布されたフォーマルランの 200 問とそれらに対する正解の 1750 セットを元に固有表現抽出とパッセージ抽出をおこない、正しくパッセージを作成することができた 1542 セットを用いた。

4.1 評価モデル

構文的照合と言い換えの効果を調べるために 3 つのモデルを用意し、比較をおこなった。

- ベースライン (BL: Base Line)
- 構文的照合を用いる (SM: Structural Matching)
- 構文的照合と言い換えを用いる (SMP: SM with Paraphrasing)

各モデルでは以下のようなモジュールを用いる。

質問文解析 (BL, SM, SMP) 質問文中のキーワードの抽出とアンサータイプの解析 (「誰」なら人が正解)

質問文の標準化 (BL, SM, SMP) 疑問詞を質問変数に置き換え、人手で記述した規則に従い標準的な形に変換する e.g. 「源頼朝の弟は誰ですか」→「PERSON は源頼朝の弟だ」

対象文書の解析 (BL, SM, SMP) 情報源のドキュメントに対する固有表現抽出と、意味カテゴリの付与

パッセージ抽出 (SM, SMP) キーワードとの近接性と経験則を用い、対象文書からパッセージを作成

回答の選択 (BL) アンサータイプと意味カテゴリが一致する候補の中から、キーワードとの近接性を用い

て回答を選択
 回答の選択 (SM) 構文的照合による質問文とパッセージ間の類似度計算と回答の選択
 回答の選択 (SMP) 言い換えと構文的照合を繰り返し、類似度の上昇が飽和した時点で回答の選択

4.2 結果

各モデルの実験結果を表 1 に示す。“correct”は 1542 問中システムの出力した候補の上位 5 つの中に正解が含まれていた数である。MRR (Mean Reciprocal Rank) は正解が現れた順位の逆数の平均であり、質問応答タスクの評価に広く使われている評価尺度である [2]。

表 1: 各モデルの精度

	BL	SM	SMP
MRR	0.58	0.422	0.423
correct	1132	712	760
wrong	410	830	782

表 1 が示すように、SM や SMP の方が BL よりも精度は悪かった。この原因の分析を次節でおこなう。

5 誤り分析

5.1 構文的照合の効果

BL が正解できなかった 410 事例のうち、SM では 56 事例に正解することができた。SM では BL の使っていない構文情報を使っているため、正解する事例が増えることは予想通りの効果である。しかし、逆に構文的照合がパフォーマンスを下げた事例も多くあった。

そこで、BL では 1 位で正解した 754 事例のうち、SM では答えられなかった 257 事例を調べたところ、主として次の 3 つの問題があることが分かった。

- キーワードのちらばりに対する脆弱性
- 質問文解析誤りに対する脆弱性
- 係り受け解析誤りに対する脆弱性

以下、それぞれについて詳しく述べる。

5.1.1 キーワードのちらばりに対する脆弱性

SM で正解できなかった 257 事例のうち、キーワードのちらばりが原因である事例が 224 事例あった。

次の例を用いて説明する。

(4) Q. 〈PERSON〉は、菅原道真⁽¹⁾と誕生日⁽²⁾が同じ⁽³⁾首相⁽⁴⁾だ。

D. 小淵恵三^(正解)首相⁽⁴⁾へ太宰府から「梅の使節」。天満宮の祭神、菅原道真⁽¹⁾は 6 月 25 日生まれと伝えられ、誕生日⁽²⁾が同じ⁽³⁾という首相⁽⁴⁾は「親しみを感じます」とニコリ。

情報源 D において、正解である「小淵恵三」は、重要なキーワードである「菅原道真⁽¹⁾」や「誕生日⁽²⁾」との間に構文的な関係を持っていない。このような質問に対しても人間は正しい答を見つけられるが、それは、D の中の「誕生日⁽²⁾が同じ⁽³⁾」は「菅原道真⁽¹⁾」と誕生日が同じであるということを理解し、また 2 回目

に現れる「首相⁽⁴⁾」は正解である名詞句「小淵恵三」を指していることが分かるからである。

この例が示しているように、質問文中のキーワードは、しばしば正解とは構文的に離れた位置に、時には文の境界を越えて出現する。そうすると構文的照合の効果はまったくなくなる。現在の SM や SMP では回答を見つけるための情報として近接性を用いていないので、このような場合には SM のように単純な近接性を用いたモデルのパフォーマンスの方がよくなる。

この問題の単純な解決方法としては、近接性に基づくスコアを構文的照合に基づくスコアリングに加えることが考えられる。しかし、1 節で述べたように我々の目標は言語理解としての質問応答であるので、この方法は必ずしも目指すべき方向ではない。この問題に対して我々は照応解析や省略解析を含む深い文脈の解析と統合し評価する方がより重要だと考え、今後の課題の一つとしている。

5.1.2 質問文解析誤りに対する脆弱性

現在の構文的照合に基づくアルゴリズムでは、質問文の構文木中の質問変数 (PERSON, LOCATION など) とマッチしたノードが回答として選ばれる。したがって、質問文解析の誤りにより質問変数を生成し損った場合は回答探索ができない。これに対し SM アルゴリズムでは、質問変数が無い場合でもキーワードの近くの名詞句が選ばれ得るので、回答を見つける可能性がある。このような事例が 257 事例のうち 19 事例あった。

5.1.3 係り受け解析誤りに対する脆弱性

257 事例のうち 14 事例について、不適切な構文木が不正解の原因となっていた。そのうちの半数は係り受け解析誤りによるものであったが、残りの半数は係り受け木の制約の厳しさが原因であった。

図 2 を例に説明する。(Q) は質問文の依存構造であり、(D) は正解である「湯川秀樹」を含む情報源の依存構造である。(D) では、「受けた」という動詞句が「中間子理論」の修飾句として解析されている (a)。しかしそれと同時に「湯川秀樹」の修飾句とも解釈可能である (b)。(b) のように解釈すれば、SM は質問変数「PERSON」に対応する回答として「湯川秀樹」を見付けることができたが、実際には (a) と解釈されたため正解できなかった。このような場合、構文情報を考慮しない BL の方が頑健である。

このような問題に対しては、解析木の代りに解析森を用いて可能性のある解釈を複数与えることにより解析に冗長性を持たせることが考えられるが、その場合どのようにして解析森に対する効率的な照合をおこなうかが新たな問題となる。この問題も今後の課題の一つである。

5.2 言い換えの効果

SM が正解できなかった 830 事例のうち、SMP では 58 事例に正解することができた。これに対し、SM が

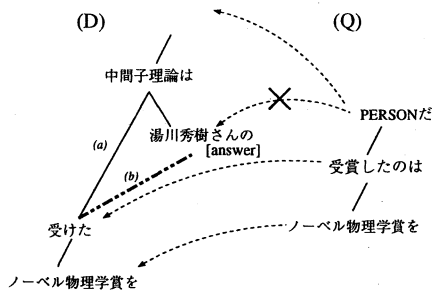


図 2: 本質的な係り受け解析誤りの例

正解した 712 事例にのうち SMP が正解できなかった事例は 10 事例だった。この結果からは言い換えが良い影響を与えているように見えるが、その効果は極めて限定的だった。そこで規則の適用回数を調べたところ、1542 の質問セットに対して生成された言い換えのうち、構文的類似度を上げるために使われた言い換えは 4254 回だった。つまり類似度の上昇に寄与した言い換えは 1 事例あたり 2.76 回であり、この数は我々の予想に反して少なかった。分析の結果その主な原因として次の 3 つがあげられる。

- 依存構造に対して言い換えをおこなっているのに、係り受け解析誤りによって正しい依存構造が作られない場合には言い換えて適用できない。つまり現在の言い換えアルゴリズムが構文解析誤りに対して頑健でない。
- 現在の言い換えは文内に閉じているため、質問文中に現れるキーワードがパッセージ中では異なる文に散らばっていた時に規則を適用できない。この問題の大部分は照応・省略解析により解決される。
- 今回使用した言い換え規則の種類が実際の質問応答タスクに用いるには不十分だった。例えば現在の言い換え規則には次のような規則は含まれていない。“X_[hum] が Y_[movie] を監督する” → “X が Y で知られる”。

Helmjakob ら [3] も我々の手法と同様に言い換えを用いた質問応答タスクをおこなっており、彼らの実験では 1 事例あたりの言い換えの適用回数は 3.14 回だったと報告されている。これと比べると我々の実験結果も少なくはない。しかし、今回の実験では類似度の上昇に寄与していても、正解を見付けるためには役に立たなかった言い換えも多く見られたので、言い換え知識の数を増やすとともに、質を向上させる必要もある。

6 まとめ

本稿では、質問応答タスクに構文的照合と言い換えて適用し、言い換えと構文的照合に基づく類似度計算を繰り返しながら回答を探索する手法を質問応答タスクに適用した。そして、実験を通して構文的照合と言

い換えが質問応答タスクにどのような効果を示すかを調べた。

実験の結果から、現在の実装では我々の予想に反して構文的照合と言い換えの効果が小さいことが分った。また誤り分析を通して、(a) 照応・省略解析、(b) より柔軟なマッチング手法、(c) より大規模な言い換え知識の獲得、の 3 点を重要な課題として認識することができた。

言い換え規則のうち、受動態から能動態への変形のように、構文の変形をおこなう規則については、人手で書き尽くせると考えられ、現状でもある程度多様な規則がすでに実装できている。また語彙レベルの言い換え規則については、照合の部分で吸収することができる。今後必要となるのは、それらの中間にあるフレーズ単位の言い換え知識であるが、それについては Ravichandran ら [7] や Lin ら [5] が提案するように自動的に獲得していく枠組が必要となる。

参考文献

- [1] Michael Collins and Nigel Duffy. Convolution kernels for natural language. In *Neural Information Processing Systems (NIPS)*, 2001.
- [2] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui. Question answering challenge (qac1): Question answering evaluation at ntcir workshop 3. In *Working Notes of the Third NTCIR Workshop Meeting: QAC1*, 2002.
- [3] Ulf Helmjakob, Abd Echibahi, and Daniel Marcu. Natural language based reformulation resource and web exploration for question answering. In *the 2002 edition of the Text REtrieval Conference (TREC)*, 2002.
- [4] A Ittycheriah, M Franz, and S. Roukos. Ibm's statistical question answering system-trec-10. In *the 2001 edition of the Text REtrieval Conference (TREC)*, p. 258, 2001.
- [5] Dekang Lin and Patrick Pantel. Discovery of inference rules for question answering. *Natural Language Engineering*, 2001.
- [6] 村田真樹, 内山将夫, 井佐原均. 類似度に基づく推論を用いた質問応答システム. 情報処理学会自然言語処理研究会, NL-135, pp. 181-188, 2000.
- [7] Deepak Ravichandran and Eduard Hovy. Learning surface text patterns for a question answering system. In *the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, 2002.
- [8] Tetsuro Takahashi, Tomoya Iwakura, Ryu Iida, Atsushi Fujita, and Kentaro Inui. Kura: A revision-based lexico-structural paraphrasing engine. In *the Natural Language Processing Pacific Rim Symposium (NLPRS-2001) Workshop on Automatic Paraphrasing: Theories and Applications*, 2001.
- [9] 高橋哲朗, 乾健太郎, 松本裕治. テキストの構文的類似度の評価方法について. 情報処理学会自然言語処理研究会, NL-150-7, 2002.
- [10] 清田陽司, 黒橋禎夫, 木戸冬子. 大規模テキスト知識ベースに基づく自動質問応答 - ダイアログナビー. 言語処理学会 第 8 回年次大会 発表論文集, 2002.