4ZK-6

Web 文書を情報源とする non-factoid 型質問応答

佐藤 充[†]　石下 円香[†]　森 辰則[‡]
[†]横浜国立大学 大学院 環境情報学府　[‡]横浜国立大学 大学院 環境情報研究院

1 はじめに

質問応答とは自然言語による質問に対して検索文書から回答そのものを抽出する技術である．質問応答のタスクは大きく二つに分けられる．固有名や数量等を問う factoid 型の質問応答と定義や理由等を問う non-factoid 型の質問応答である．日本語における non-factoid 型の質問応答は近年広く研究されるものの，依然難易度の高いタスクである [1].

質問応答の情報源としては，新聞記事等のローカル文書を対象とする場合と Web 文書を対象とする場合とがある．後者は文書の前処理ができない点で不利であるが，新鮮かつ大量の情報が得られる点で実用的である．

本稿では Web 文書を対象とする日本語 non-factoid 型質問応答を扱う．Web 文書を単に検索対象として扱うだけでなく，Q&A サイトの質問・回答事例集合をコーパスとして用い，そこから得た知識を解抽出の際に役立てる．

2 non-factoid 型質問応答

定義や理由等を問うような non-factoid 型質問に対する回答としては，数文にまたがるような比較的長い文章表現が想定される．このような質問の解の適切性は，

【尺度 1】 質問の内容との関連性

【尺度 2】 質問の型に応じた記述スタイルを満たすかの組合せで見積もれると考えられる．記述スタイルとは，例えば定義型質問ならば「…とは〜である」のような表現である．【尺度 1】は簡単には質問と回答の類似度で計算できる．【尺度 2】は人手で作成した語彙統語パタンや機械学習により判定することが多い．

Han ら [2] は英語の定義型質問応答において，これら二つの尺度をコーパスから推定した確率モデルに基づいて計算している．このように質問の型が限定された場合は，型に応じた回答表現のパタンを用意したり，専用のコーパスを学習データとして用いることが有効である．質問の型を限定しない場合でも，入力された質問を予め用意した型に分類し，型ごとに処理を分けることで同様のアプローチを実現できる．

しかし，実際には質問の型が何種類あるかは不明であるし，型ごとに個別の処理方法を用意するのは非効率である．また，質問の型分類が回答精度に大きく影響してしまう．可能ならば質問の型に依らない統一的な手法が望ましい．

水野ら [4] は Q&A サイトの質問・回答事例集合から質問と回答の型の一致を SVM で学習することにより，質問の型の分類を行なわずに【尺度 2】を判定している．この手法では回答の範囲を先に決め，得られた解候補を質問と型が一致するかどうかで分類するため，解候補の範囲を質問に応じて動的に変更できない．

Soricut ら [3] は FAQ サイトの質問・回答事例集合をパラレルコーパスとみなし，回答が質問に「書き換え」られる確率を計算するという，質問の型に依らない手法を提案している．この手法では質問の長さから回答の長さを推定する必要があるのが難点である．

我々の提案手法では，non-factoid 型質問に対する回答として，上記の二つの尺度を兼ね備えたパッセージを解候補として Web 文書から抽出する．【尺度 2】の見積もりには Q&A サイトの質問・回答事例集合をコーパスとして用いるが，入力された質問に適合する回答の特徴表現をコーパスから動的に取得するという点が水野ら [4] や Soricut ら [3] の手法と異なる．先行研究に対する優位点としては，水野ら [4] の手法に対しては回答の範囲を質問に応じて動的に決められること，Soricut ら [3] の手法に対しては回答の長さを推定する必要がないことが挙げられる．

3 質問・回答事例集合の分類

提案手法に先立って試みた手法について説明する．我々は Q&A サイトである Yahoo! 知恵袋[1]の質問・回答事例集合を Q&A コーパスとして利用した．一つの質問には複数の回答が対応するが，回答としてはベストアンサーのみ[2]を使用し，質問とベストアンサーの対を以下では「ペア」と呼ぶことにする．

様々な種類の質問に対応するため，コーパス中のペアを質問の型に応じて分類し，学習データとすることを考えた．質問文が入力された時は学習データ中のいずれかの型に分類し，型ごとに学習データ中の回答文から取得した特徴表現に基づいて【尺度 2】を見積もる．分類する型の種類は「定義」「理由」「方法」というように人手で設定することも考えられるが，種類がいくつあるのかは不明なため，クラスタリングによって分類することにした．その際，質問と回答双方の特徴を考慮してクラスタリングするため，質問の類似度と回答の類似度をそれぞれ別の空間で計算し，両者を合わせたものをペアの類似度とした．

通常の内容語によるクラスタリングでは，質問の型ではなく内容によって分類されてしまう．先に述べた【尺度 2】を実現すべく，記述スタイルによって分類するため，疑問詞や助詞，助動詞等の機能語と一部の内容語[3]を表層表現のまま用い，その他の語は品詞で代表させることにより，質問と回答の双方を一般化した．文中の語の共起をある程度考慮し柔軟に類似度を計算するため，語のスキップ 2-gram(数語の間隔を許した 2-gram) を素性に用いた．

k-means 法により 5 万件のペアを 100 個程度のクラスタに分類した結果を観察すると，質問表現としては「って何」「どうすれば」「違いは何」「どういう意味」「できますか」というように様々な型に応じたクラスタが生成されていた．

各クラスタに属する質問文と回答文から特徴表現を χ^2 値によって取り出した．質問文からはある程度分類に役立つと思われる表現を取得できたが，回答文からは

Web question-answering for non-factoid questions.
Mitsuru Sato, Madoka Ishioroshi and Tatsunori Mori.
Graduate School of Environment and Information Sciences, Yokohama National University

[1] http://chiebukuro.yahoo.co.jp/

[2] さらに，回答文中に URL が表記されているものは除いた．

[3] 質問の焦点になりやすい名詞 (「理由」「方法」「意味」等) やコーパス中で出現頻度が高い動詞と形容詞

あまり有効な表現を得ることができず，この方法を質問応答に適用するのは難しいと判断した．

失敗要因としては，いずれのクラスタ中心からも距離が遠いペアを排除した結果，十分なデータ量が得られなくなった点が挙げられる．記述スタイルが似通ったペアでも適切にマージされないことがあり，素性選択が難しく，類似度計算がうまくいっていないことが窺える．

4 提案手法

提案手法では，入力された質問を型に分類せず，コーパス中から記述スタイルが類似する質問を直接収集する．収集した質問に対応する回答文から特徴表現を取得し，【尺度 2】の見積りに利用する．

こうすることで型分類の失敗を考慮する必要が無くなる上に，コーパスのデータ量の多さ (ペアが 300 万件以上ある) を活かして記述スタイルが良く似た質問でも十分な量を集められる．「質問の型」に応じた回答の特徴表現を用意しておくのではなく，質問が入力された時点で「その質問」に応じた回答の特徴表現を動的に生成するのである．この方式では，入力された質問に，より適合する特徴表現を見つけることができると期待される．

使用するコーパスは 3 章で述べたものと同様である．前処理として，3 章のクラスタリングの時と同様，コーパス中の質問文・回答文双方において機能語と一部の内容語を除く表層表現を品詞に置き換えた．

記述スタイルが類似する質問を収集する際には，質問文同士の「疑問詞を中心とする語の 7-gram」の一致の度合を類似度とする．入力された質問と類似度が高い質問を上位 N 件取得して，対応する回答文の集合から特徴表現を取り出す．特徴表現は，収集した回答文集合とコーパス中の残りの回答文集合の間における χ^2 値の高い 2-gram 上位 M 件とする．例えば，「X が Y を退団した理由は何ですか」という質問を入力すると，「たから」「からです」「ため。」といった表現が得られる．これらの表現とその χ^2 値に基づいて解候補文のスコアリングを行なう．

最後に提案システムの流れを説明する．入力された質問文から内容語を取得し，キーワード集合 K とする．K から数種類のクエリを生成し[4]，Web 検索エンジンに入力して文書を検索するとともに，snippet(検索結果の要約) 中の各語 w_i の snippet 頻度を求め，これを $T(w_i)$ とする．$T(w_i)$ が高い語は質問文の関連語と考えられるので，【尺度 1】の見積りに利用する．また，キーワード $k_i \in K$ に関しては，他の語よりも高い重みを与えるために $T(k_i) = \max_j T(w_j)$ とする．次に上述の手法で回答文の特徴表現である 2-gram b とその $\chi^2(b)$ 値のリストを取得する．そして検索文書中の文 S_i のスコアを次式で見積もる．

$$\mathrm{Score}(S_i) = \frac{\left\{\sum_{j=1}^{n} T(w_{ij})\right\}^{\alpha} \cdot \left\{\sum_{k=1}^{m} \sqrt{\chi^2(b_{ik})}\right\}^{1-\alpha}}{\log(1+|S_i|)} \quad (1)$$

ここで，n は文 S_i 中の語 w_{ij} の異なり数，m は文 S_i 中の 2-gram b_{ik} の異なり数，α はパラメータである．文書中でスコアが高い文が連続した場合，極大値の 1/2 以上のスコアを持つ文をまとめて一つの解候補とし，解候補のスコアは極大値のスコアで代表する．他の文は 1 文で解候補とする．こうして得られた解候補の集合をクラスタリングし，冗長性を制御する．各クラスタ内でスコアが最大の解候補を取り出し，解として出力する．

[4] キーワードから複合語を構成するかどうか等の変化をつける．

[5] 質問の型の分布は，定義型質問 (約 2 割) と理由型質問 (約 3 割) で半数を占め，残りは他の型の non-factoid 型質問である．

[6] http://www.yahoo.co.jp/

5 評価実験

NTCIR6 QAC4 [1] Formal Run のテストセット 100 問[5]を用いて提案システムの性能評価を行なった．式 (1) において $\alpha = 0.5$ とした場合 (提案手法:【尺度 1】と【尺度 2】の組合せ) と $\alpha = 1.0$ とした場合 (ベースライン:【尺度 1】のみ) の結果を比較した．

Web 検索エンジンは Yahoo! JAPAN[6]を利用した．一つの質問に対して 3 種類のクエリを生成し，それぞれ 50 件ずつ文書を検索したが，実際に取得できた文書の異なり数は平均 40 件であった．入力された質問と類似する質問のコーパスからの取得件数は $N = 500$ 件，回答の特徴表現の取得件数は $M = 200$ 件とした．

システムはスコアの降順に解を出力する．今回は上位 5 件までを評価対象とした．正解判定は人手で行ない，回答の一部に正解と思われる表現が含まれていれば正解とした．評価尺度としては MRR(最上位の正解順位の逆数の全質問平均) を用いた．正解を 1 件以上出力できた質問数 (正解質問数) も調査した．結果を表 1に示す．

表 1: 実験結果

手法	MRR	正解質問数
$\alpha = 0.5$ (提案手法)	0.469	66
$\alpha = 1.0$ (ベースライン)	0.318	54

実験結果より，提案手法によって【尺度 2】を見積もることで回答精度が向上することが示せた．コーパス中の回答文から取り出した特徴表現が解抽出の際に有効に働いていると考えられる．しかし，全体のうちおよそ 1/3 の質問については提案手法においても正解を出力できなかった．本稿では【尺度 2】を見積もる方法に焦点を当てたが，文書検索や【尺度 1】の見積りにもまだ改善の余値があるといえる．

6 おわりに

本稿では，Q&A コーパスを用いることで質問の型分類を行なわない non-factoid 型質問応答手法について提案した．評価実験により，提案手法でコーパスから得た知識が解抽出に役立つことを示した．

今後の課題としては，解候補のスコアリング方法の改良や入力された質問文とコーパス中の質問文の類似度計算方法の見直しが挙げられる．

謝辞

本研究の実施にあたり，ヤフー株式会社が国立情報学研究所に提供した Yahoo! 知恵袋データを利用させて頂きました．なお，本研究の一部は文科省科研費特定領域「情報爆発 IT 基盤」(課題番号 19024033) によるものである．

参考文献

[1] Junichi Fukumoto, Tsuneaki Kato, Fumito Masui, and Tatsunori Mori. An Overview of the 4th Question Answering Challenge (QAC-4) at NTCIR Workshop 6. In *Working Notes of the 6th NTCIR Workshop Meeting*, pp. 433–440, 5 2007.

[2] Kyoung-Soo Han, Young-In Song, and Hae-Chang Rim. Probabilistic model for definitional question answering. In *SIGIR*, pp. 212–219, 2006.

[3] Radu Soricut and Eric Brill. Automatic question answering using the web: Beyond the factoid. *Journal of Information Retrieval - Special Issue on Web Information Retrieval*, Vol. 9, pp. 191–206, November 2006.

[4] 水野淳太, 秋葉友良. 任意の回答を対象とする質問応答のための実世界質問の分析と回答タイプ判定法の検討. 言語処理学会 13 回年次大会発表論文集 pp.1002–1005, 言語処理学会, 3 2007.