

Web 質問応答における 複数検索エンジンの組合せによる精度向上

金井 明[†] 佐藤 充[†][†]横浜国立大学 大学院 環境情報学府石下 円香[†] 森 辰則[‡][‡]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 (QA) は、質問に対して関連文書を探して提示するだけでなく、直接質問の答を提示してくれるシステムである。また、近年では、新聞記事集合のような静的かつローカルな文書群ではなく、文書が豊富で日々追加・更新がなされる Web 文書を情報源とした質問応答システム (Web QA) が研究されている。通常、Web 検索エンジンを質問応答用に独自に用意することは現実的でないので、既存の Web 検索エンジンが利用される。

ここで、我々は複数の異なる検索エンジンが Web QA に使用できることに注目している。各検索エンジンの出力する検索結果が同一であることはまずなく、それらを組合せることで情報源の多様性を増す事ができ、Web QA の精度向上に何らかの効果があると期待される。そこで本論文では、QA において異なる Web 検索エンジンを組合せることによる効果について調査を行なった。

2 関連研究

初の Web QA システムの 1 つである START の最近の版では、複数の情報源を活用している [3]。Radev ら [4] は Web QA における確率に基づくアプローチを提案しているが、そこでは 3 つの主要な Web 検索エンジンを組合せて上位 40 文書を得ている。

これらの研究では異なる情報源から得られた文書を使用しているが、文書検索よりも後の段階においては、情報源の違いは考慮していない。これに対して、節 4 で述べる我々の手法では、異なる情報源の間でのデータの冗長性を活用しようとしている。

3 基本となる WebQA システム

本研究で使用する Web QA システムは、Web 検索エンジンの出力を利用し、日本語の factoid 型質問に対し、日本語で答を返すシステムである。また、数百もの Web 文書をダウンロードすることは、非常に時間がかかる処理であるので、応答時間の短縮のために Web 検索エンジンによる短い抜粋出力である snippet を Web QA の情報源として用いている。

図 1 にその Web QA システムの構成を示す。質問文解析部は利用者から質問文を受け取り、キーワードのリストや質問文の型などの情報を抽出する。キーワードのリストが検索質問として Web 検索エンジンに入力され、snippet 群が検索される。文照合部は、文書集合から抽出された文集合をパッセージ抽出部から受けとり、それらを処理する。ここで得られた各文を本論文では**検索文**と呼ぶ。各検索文中の各形態素が一つの解候補として扱われ、それらに対して、次節に述べる方法によりスコアが与えられる。

Effect of combining different Web search engines on Web question-answering for factoid questions.
Akira Kanai, Mitsuru Sato, Madoka Ishioroshi and Tatsunori Mori
Graduate School of Environment and Information Sciences,
Yokohama National University

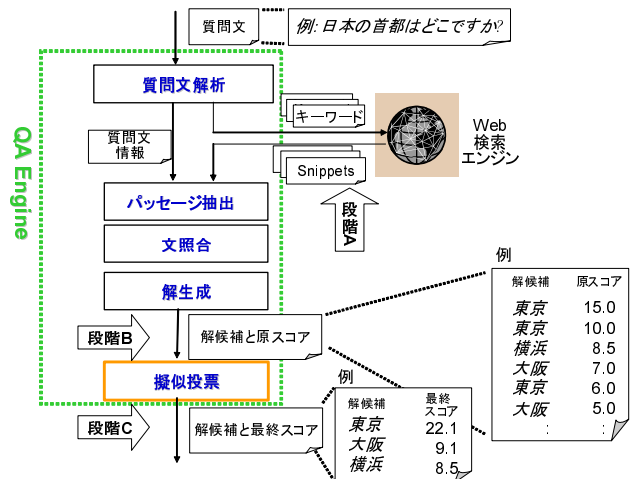


図 1: 基本となる WebQA システム

3.1 解候補のスコア付け

基本となる QA システムでは、解候補に対する複合的な照合スコアを採用している。

$$S(AC, L_i, L_q) = Sb(AC, L_i, L_q) + Sk(AC, L_i, L_q) + Sd(AC, L_i, L_q) + St(AC, L_i, L_q) \quad (1)$$

本論文ではこのスコアのことを**原スコア**と呼ぶ。このスコアは、 i 番目の検索文 L_i にある解候補 AC に対し、質問文 L_q に関する以下の 4 つの部分スコアを計算し、その線形結合を求めたものである。

1. $Sb(AC, L_i, L_q)$: 文字 2-grams の観点で計算した照合スコア
2. $Sk(AC, L_i, L_q)$: キーワードの観点で計算した照合スコア
3. $Sd(AC, L_i, L_q)$: 解候補とキーワードの間の依存構造の観点で計算した照合スコア
4. $St(AC, L_i, L_q)$: 質問文の型の観点で計算した照合スコア

3.2 探索制御の枠組みにおける疑似投票手法

既存の多くの QA システムでは解候補に関する大域的な情報を利用している。特に冗長性は最も基本的であり、かつ重要な情報である。例えば、文書中に複数回出現する解候補に対し、そのスコアを増加させるという投票手法がある [1]。

一方で、我々の QA システムのように解候補の探索に基づく枠組においては投票手法をそのまま利用できない。なぜならば、同枠組では上位 n 件の解候補が見つかり、検索をそこで終了してしまうので、文書集合全体における解候補についての正確な頻度情報が得られないためである。そこで、投票手法の一つの近似として、以下に述べる疑似投票手法が導入されている。まず、上位 n 件の解が必要な場合には、**表層表現の異なる**解候補が n 個見つかるまで探索を継続していることに注意されたい。そのため、探索の過程において、すでにスコア計算の終了状態に至っている解候補と全くおなじ表層表現を

持つが別の解候補が新たに終了状態に至ることが有り得る。よって、探索の過程において終了状態に至ったすべての解候補を記録することで、解候補の頻度情報を部分的に利用できる。本論文では、解候補 AC に対する疑似投票スコア $S^v(AC, L_q)$ を次のように定義する。

$$S^v(AC, L_q) = (\log_{10}(\text{freq}(AC, \text{AnsList})) + 1) \cdot \max_{L_i} S(AC, L_i, L_q) \quad (2)$$

ここで、 AnsList は n -best 探索において終了状態に至った解候補のリスト、 $\text{freq}(x, L)$ を L における x の頻度とする。また、本論文では、この疑似投票スコアを**最終スコア**と呼ぶことにする。

4 異なる Web 検索エンジンの出力結果を組合せる手法

異なる Web 検索エンジンの出力結果を質問応答処理の中で組合せる手法として以下の 3 手法を検討する。

手法 A (従来手法): 各検索エンジンから得られた文書集合 (snippet の集合) を図 1 の段階 A においてそのまま併合し、それを 1 つの QA エンジンに送る。

手法 B: 各検索エンジンから得られた文書集合をそれぞれ個別に解生成までの処理を行ない、図 1 の段階 B において得られた解候補 (原スコアを有する) の各リストを併合する。その後、式 (2) を使って疑似投票を行なう。次の手法 C とは異なり疑似投票は 1 回だけ行なわれる。

手法 C: 各検索エンジンから得られた文書集合をそれぞれ個別の QA エンジンに送り、疑似投票まで行なった結果として得られた解候補 (最終スコアを有する) の各リストを図 1 の段階 C において併合する。併合結果の中においても同じ表層表現を持つ複数の解候補が得られることがあるので、再び式 (2) を使って疑似投票を行なう。

5 評価実験

前節で述べた 3 つの組合せ手法を評価、検討するために、以下のような評価実験を行なった。特に、利用する Web 検索エンジンの数の効果を調査するために、質問応答における各種設定を同一にした状況において、i) 各 Web 検索エンジンを単独で使用した場合、ii) Web 検索エンジンを 2 つ組合せた場合、iii) 3 つ組合せた場合、iv) 4 つ組合せた場合の各々で実験を行なった。

評価に用いる質問文集合としては、NTCIR-3 QAC1 [2] の質問文 200 問のうち、検索 API のタイムアウト等によりいずれかの QA エンジンが一定時間内に回答を返さなかった 8 問を除外した、計 192 問を評価に用いた。Web 検索エンジンとしては、i) AltaVista, ii) goo, iii) Yahoo! Japan, iv) Google, を使用した。

6 実験結果と考察

求解精度の尺度として MRR (正解が最初に現れた順位の逆数の全質問平均) を採用した。その値を図 2 に示す。図 2 によれば、単独の Web 検索エンジンを使ったシステムより Web 検索エンジンを複数組合せるシステムのほうが精度がよくなる傾向にある。特に、質問応答処理を行なう前に文書集合を併合する手法 A に比べ、質問応答処理を個別に行なった後に、解候補を投票処理により併合する手法 B, C のほうが精度がよいことがわかる。一方、手法 B, C の間の差は顕著ではない。以上より、文書集合を先に併合してから質問応答処理を行う方法よ

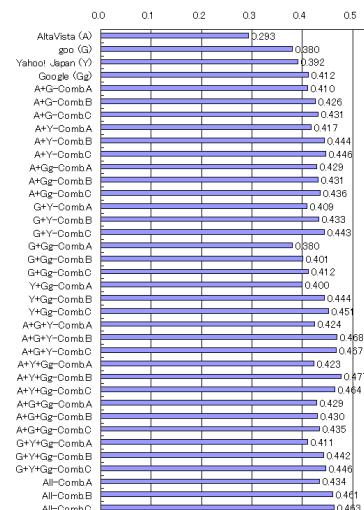


図 2: Web 検索エンジンの各組合せに対する MRR 値。図 2 によれば、単独の Web 検索エンジンを使ったシステムより Web 検索エンジンを複数組合せるシステムのほうが精度がよくなる傾向にある。特に、質問応答処理を行なう前に文書集合を併合する手法 A に比べ、質問応答処理を個別に行なった後に、解候補を投票処理により併合する手法 B, C のほうが精度がよいことがわかる。

また、組合せる Web 検索エンジンの数が増えるに従って、概ね精度が向上することも観察される。一方で、goo と google の組合せにおいて、精度の低下がみられる。調査の結果、google との提携により goo が google と同じ snippet を出力することが数多くあることがわかり、投票における情報源の多様性が保証されないことが原因であった。

7 おわりに

本論文では、Web 質問応答において、異なる Web 検索エンジンを組合せることによる効果について調査した。評価実験より、質問応答処理の前に検索結果を併合する手法よりも、QA エンジンによる個別の質問応答処理の後に、抽出された解候補を組合せるほうが効果があることが示された。

今後は、更に精度の良い質問応答の実現のため、a) Web 検索エンジンの出力のより効果的な組合せ手法、b) snippet における表現の多様性を積極的に引き出すため、Web 検索エンジンに入力する検索要求を生成する手法を複数用意して組合せる方法、などを検討したい。

謝辞

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] Charles L.A. Clarke, Gordon V. Cormack, and Thomas R. Lynam. Exploiting redundancy in question answering. In *Proceedings of SIGIR '01: the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 358–365, 2001.
- [2] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui. Question Answering Challenge (QAC-1) — Question answering evaluation at NTCIR Workshop 3 —. In *Working Notes of the Third NTCIR Workshop meeting – Part IV: Question Answering Challenge (QAC1)*, pp. 1–6, 2002.
- [3] B. Katz, M. Bilotti, S. Felshin, A. Fernandes, W. Hildebrandt, R. Katzir, J. Lin, D. Loreto, G. Marton, F. Mora, and O. Uzuner. Answering multiple questions on a topic from heterogeneous resources. In *Proceedings of TREC 2004*, 2004.
- [4] Dragomir R. Radev, Weiguo Fan, Hong Qi, Harris Wu, and Amardeep Grewal. Probabilistic question answering on the web. *Journal of the American Society for Information Science and Technology*, Vol. 56, No. 3, March 2005.