

non-factoid 型質問応答における まとめの観点からの回答の順位付け手法の提案

長尾 慶一† 渋木 英潔‡ 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
E-mail: {nagao,shib,mori}@forest.eis.ynu.ac.jp

1 はじめに

近年、文書情報に対するアクセス技術として、質問応答が注目されている。質問応答とは、利用者の自然言語による質問に対して情報源となる文書集合から回答そのものを抽出する技術であり、人名や地名、数量などを問う factoid 型質問を対象としたものと、定義・理由・方法等、比較的長い文章表現による記述的な回答が想定される non-factoid 型質問を対象としたものに大別できる。

non-factoid 型質問応答において出力される回答には、正解ではあるが利用者に提示すべき情報の全てが書かれていないものが存在する。例えば、「スケルトンとはどのような競技ですか」という質問に対して、「ウィンタースポーツの一種で、ソリに乗って速さを競う競技です」という回答は正解の一つであるが、他にも「ボブスレーと異なり、競技者がソリに腹這いになって乗り、重心移動によってソリを操作します」と「第2回冬季オリンピックで初めて正式競技となり、第5回冬季オリンピックでも正式競技として採用されている競技です」といった回答も利用者に提示すべき回答であると考えられる。何故ならば、後の2つの回答には、最初の回答に書かれていない「他競技との比較」や「競技の種目」といった観点からの情報が書かれているからである。回答群の中には、こういった各観点からの詳細が書かれている回答が存在する一方、それらの観点を広く網羅しまとめた回答も存在する(図1)。利用者の視点から考えた場合、全体を概観できるまとめとなる回答を最初に読んだ後で、各回答観点の詳細が書かれた回答を読む方が理解しやすいと考えられる。

以上の背景から、本稿では、全体の回答観点を網羅しまとめている順に回答を順位づけする手法を提案する。提案手法は、質問応答システムの存在を仮定して、ある質問に対する回答候補集合が既与とした時に、各回答観点をより網羅しまとめている回答がより上位になるように順位付けを行う。

さて、回答候補集合から各回答観点を網羅しまとめている回答を探すというタスクは、複数文書要約の一種とみなすことができる。この様なタスクにおいて、近年グラフベースの要約手法が提案され、一定の成果を挙げている [1]。この手法では、ある文や文章などのテキスト断片をノードとして、ノード間のリンクに語の重要度に基づく類似度を用いた重み付けを行い、「多くのノードと類似したノード

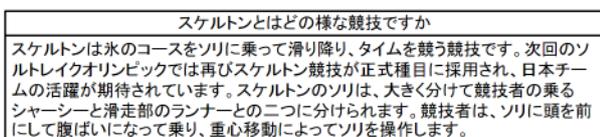


図1 観点を広く網羅しまとめた回答の例

は、まとめノードとして適している」という仮説に基づき、各ノードへ重要度を付与している。

一方で、この様な仮説の基で、回答候補集合に対して重要度計算を行った場合、回答候補の内部構造に言及せず、回答を bag of words として扱うため、たまたまトピック語を要領よく含んだ回答候補に対し高い重要度が付与される可能性がある。そこで、本稿では、各回答観点を網羅しまとめている回答の説明構造を考慮してネットワークを構築することで、上記のタスクに対する精度向上を試みる。

各回答観点を網羅しまとめている回答では、①各観点が端的な表現で記述され、②それらが並んで出現する構造を持っていると想定される。そのため、各観点を表現した部分で、それぞれ同じ観点を共有する(他の)回答と密な内容関連性を持つと考えられる。

この様な特徴から各回答観点を網羅しまとめた回答を検出するには、回答をある単位で分割し、各部分がそれぞれ異なる回答との類似関係に基づくリンク構造を持った回答を探せばよい。そこで、提案手法では、回答候補内のテキスト断片から、他の回答候補へ類似度に基づくリンクを張ることによりネットワークを構築し、このネットワークを用いて各回答候補に対して重要度計算を行う。

重要度計算としては、本稿では HITS アルゴリズム [2] に着目する。HITS アルゴリズムは、Web 文書を対象とした文書検索アルゴリズムで、リンク集に相当するページ (Hub) と、多くの Hub から参照されるページ (Authority) の2種類のページを考慮している。本稿においても、各回答観点を網羅しまとめた回答に高い Hub 値を、ある回答観点を詳細に述べた回答に高い Authority 値を与えるようなネットワークを構築する。

2 関連研究

2.1 回答観点多様性に対する関連研究

石下ら [3]は、回答候補集合に対し、Steinら [4]の要約手法の適用を提案している。即ち、回答候補集合を文書集合と捉え、内容類似性に基づき回答候補同士をクラスタリングした後、各クラスタから代表的な回答候補のみを採用して回答候補のリストを組織化することで、正解提示の精度を保持しつつ、リストの上位に位置する回答候補集合に対し、より多くの回答観点を所持せられるとしている。

また、藤井ら [5]は、Web を事典的に利用するためのシステム Cyclone の提案を行っている。Cyclone では、言葉に関する説明について、あらかじめ「定義」や「略語」といった観点を用意し、収集した回答候補の記述をこれらの観点に分類することで、多様な回答観点を捉えた回答の生成を行っている。

石下ら [3]の手法では、回答の詳細度を考慮した類似度計算を行っていない。各クラスタがある回答観点を表しているとみなすと、各回答観点を網羅しまとめた回答は、複数のクラスタに跨る存在といえるが、実際には、このような回答が各クラスタの回答と均一な内容類似性を持つとは限らず、各観点を表現する語の数などに依存して幅を持つことが多い。この結果、各回答観点を網羅しまとめた回答が、特に内容類似性の高い回答のクラスタに含まれてしまう可能性がある。

そこで、藤井ら [5]の様に回答候補の記述を統合し、回答観点毎に整理するアプローチが考えられるが、例えば「スケルトンとはどのような競技ですか」といった質問では「他競技との比較」などが観点となる様に、回答の観点は質問内容により異なる。そのため、質問の型を限定しない場合、回答のテンプレートをどの様に生成するかが課題となる。本稿では、回答候補同士の内容を比較して観点を捉えることで、質問の型を限定しない手法の提案を試みる。

2.2 HITS アルゴリズム

HITS アルゴリズム [2] は、Kleinberg らにより提案された Web ページのランキング手法のひとつであり、Hub 値と Authority 値の2種類のスコアを重要度に用いる。HITS アルゴリズムは「多くの Hub から参照される Authority は重要である」「多くの重要な Authority を参照する Hub は重要である」という2つの仮定に基づき、繰り返し計算していくことでランキング付けを行うアルゴリズムである。

Mihalcea ら [1]は、Web ページのランキング手法を抜粋型要約に応用する手法を TextRank として提案しており、HITS アルゴリズムによる TextRank も提案している。HITS アルゴリズムで用いられる式(2)、式(3)はそれぞれ Hyperlink によるネットワークを基本としているため、各ノード間の結合度に差異は無く、全てのリンクの重みは等しいものとしていた。しかし、TextRank では、類似度などのノード間の結合の強さに応じてリンクの重みを考慮した計算を行っており、ノード V_i からノード V_j に重み w_{ij} のリンクが張られている場合、式(4)、式(5)に従って、Hub 値と Authority 値を計算する。

$$HITS_A^{t+1}(V_i) = \sum_{V_j \in \text{In}(V_i)} w_{ij} \cdot HITS_H^t(V_j) \quad (1)$$

$$HITS_H^{t+1}(V_i) = \sum_{V_j \in \text{Out}(V_i)} w_{ji} \cdot HITS_A^t(V_j) \quad (2)$$

3 提案手法

3.1 概要

提案手法では、回答候補集合を用いて、回答候補をノードとし、回答候補中の各テキスト断片と、密な内容関連性を持った回答候補とをリンクで結んだネットワーク(図2)を構築し、これに HITS アルゴリズムを適用した重要度計算を行う。この結果、Hub 値の高い回答候補を、各回答観点を網羅しまとめている回答として検出する。

ここで、回答候補同士の比較は、ある回答候補 H の“テキスト断片の内容”と、比較する回答候補 A の“全体の内容”を見て、両者の内容関連性を見積もる。この様な方法をとる理由としては、回答候補 H が Hub としての役割を担っている度合いを見積もるためである。回答候補 H

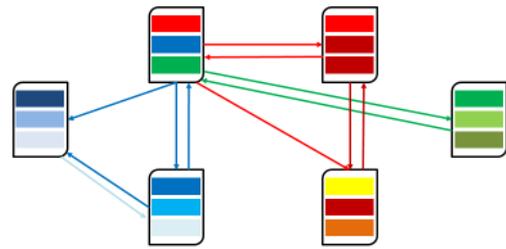


図2 回答候補をノードとしたネットワーク

が Hub としての役割を担っている場合、それが複数の回答観点に対応する記述を持つものに対して、Authority に対応する回答候補 A はそれ全体で少数の回答観点に対する記述となっていることが期待される。そのため、回答候補 H に含まれ、ある一つの回答観点に相当する“テキスト断片”と別の回答候補 A 全体との間にリンクを張ることを考え、内容の関連性を見積もる際には、ある回答候補内のテキスト断片と、別の回答単位全体の間で計算を行うという非対称なものとなる。

提案手法は、大きく分けて「ネットワークの構築」と「重要度計算」の2つの段階に分けられる。

ネットワークの構築では、「①：回答候補の記述をテキスト断片に分割」し、「②：分割したテキスト断片毎に、他の回答候補との内容関連性を比較」したあと、「③：内容関連性に基づいたリンクの生成判定」を行う。また、リンクの生成判定の結果、ある回答候補に含まれるテキスト断片同士が、同じリンク構造を構築している場合や、包含関係にある場合(図3)、これらの断片が同じ観点到属した内容であると考え、「④：テキスト断片同士の統合」を行い、最終的なネットワークを構築する。

重要度計算では、各回答候補は、各テキスト断片が取得した Authority 値の和によって Hub 値を、リンクを張る回答候補の Hub 値の和によって Authority 値を求める。以下では、各処理の詳細な説明を行う。

3.2 回答候補のテキスト断片への分割

1章で述べたように、各回答観点のまとめとなる回答では、各回答観点は、数文節~1文程度の端的な表現の繰り返しによって記述していると想定される。提案手法では、予備調査の結果から、このような端的な表現が一つの述語項構造に対応すると考え、回答候補のテキストを述語の直後で分割する。しかし、Web 文書の様な形式の定まっていな記述に対し、述語項構造を精度よく自動解析することは依然困難な課題である。そのため、現在では解析器への依存を CaboCha¹による文節解析程度に止め、人手で作成した規則を用いた分割を行っている。

3.3 回答候補間のリンク生成

次に、同じ回答観点を持つ回答候補に対してリンクを張ることを目的として、3.2節で分割したテキスト断片毎に、他の回答候補との内容関連性を比較する。

3.1節で述べたように、提案手法では、ある回答候補の“テキスト断片の内容”と、比較する回答候補の“全体の内容”を見て、両者の内容関連性を見積もる。このとき、

¹ <https://code.google.com/p/cabocha/>

両者のテキストの長さに差が生じ、テキスト間での通常の類似度判定はそぐわない。そのため、提案手法では回答候補 A_i の断片 P_{ij} と回答候補 A_k の内容関連性を、次の式(3)による内容包含性により見積もる。

Comprehensibility(P_{ij}, A_k) =

$$\frac{| \{w_n | w_n \in P_{ij} \& w_n \in A_k\} |}{|P_{ij}|} \quad (3)$$

この式は Runs ら [6]が文間の含意関係認識においてベースラインとして用いた尺度と同じである。状況を限定しない汎用的な手法として、本稿においてもこの式を用いる。

各テキスト断片は、自らを含む回答候補以外の全ての回答候補に対して式(3)による評価を行い、閾値 α 以上で最上位の内容関連性を持つ回答候補のみを採用し、リンクを生成する。最上位の候補が複数存在する場合はいずれも採用する。ここで、 α はパラメタであり、予備実験による結果から $\alpha = 0.6$ と設定して処理を行っている。

3.4 テキスト断片の統合

ここまでの処理により、ある回答候補の各テキスト断片が構築したリンク構造は、互いに共通点や相違点を持つ。この中には、全く同じリンク構造を構築している断片同士や、リンク構造が包含関係にある断片同士が存在する。

この様な関係のテキスト断片同士は、実際は同じ回答の観点に属した内容である場合が多い。3.5 節で述べる重要度計算手法の様に、各断片が取得する Authority 値の和によって Hub 値を求めるアプローチでは、同一観点による Authority 値を複数回加算することを認めた場合、異なる回答観点から Authority 値を集めた回答候補と、同一観点の Authority 値を集めた望ましくない回答候補との間に差がなくなり、検出精度を下げる可能性がある。従って、上記の関係のテキスト断片同士を同一とみなして扱う。ある回答候補のテキスト断片集合に対し、断片同士の関係を以下の4つに場合分けし、各場合に対して以下の処理を行う。

- ① それぞれ異なる回答候補へのリンクのみを持つ
 - ② 共通する回答候補へのリンクと、異なる回答候補へのリンクの両方を持つ
 - ③ 一方のリンク構造が他方のリンク構造を包含する
 - ④ どちらも全く同じリンク構造を持つ
- ①と②の場合：それぞれに相違点があるとみなし、そのまま採用する。
- ③の場合：一方の断片が他方の断片に対して、より大きなトピックを表している場合が多い。提案手法では、より密な対応関係で回答候補同士を繋ぐという観点を優先し、両者の断片およびリンク構造を、図3の様に統合する。
- ④の場合：同一観点に属した内容とみなし、これらの断片およびリンク構造を、図3の様に統合する。

3.5 重要度計算

以上の処理により構築されたネットワークに対して、HITS アルゴリズムを適用した重要度計算を行う。具体的には、以下に示す式(4)、(5)を用いて回答候補 A_i に対して Hub 値($HITS_H^i$)と Authority 値($HITS_A^i$)を付与する。

$$HITS_H^{i+1}(A_i) =$$

$$\sum_{P_{ij} \in \text{linked_predicates}(A_i)} \frac{\max(\{HA_k | HA_k = HITS_A^i(A_k) \wedge A_k \in \text{Out}(P_{ij})\})}{1 + \log(\text{total_predicates}(A_i))} \quad (4)$$

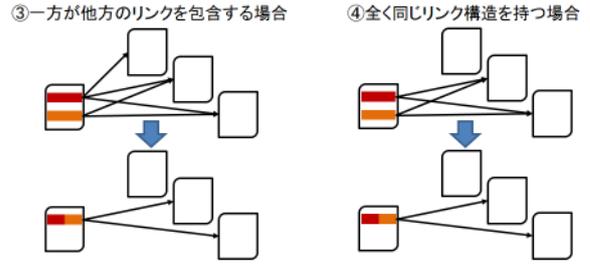


図3 テキスト断片の統合

$$HITS_A^{i+1}(A_i) = \sum_{A_k \in \text{In}(A_i)} \frac{HITS_H^i(A_k)}{1 + \log(\text{linked_predicates}(A_i))} \quad (5)$$

ここで、 P_{ij} は A_i に含まれるテキスト断片、 $\text{Out}(P_{ij})$ は P_{ij} がリンクを生成した回答候補集合、 $\text{In}(A_i)$ は A_i に対してリンクを張る回答候補集合を表す。また、 $\text{total_predicates}(A_i)$ は、3.4 節による統合処理の後、 A_i に含まれているすべてのテキスト断片の数を表し、 $\text{linked_predicates}(A_i)$ は、3.4 節の手法による統合処理の後、 A_i に含まれているすべての断片の内、リンク構造を持つ断片の数を表す。

式(4)は、各テキスト断片から、それぞれリンク先が持つ Authority 値の内、最も良い値をその断片の良さとして取得し、それらの和を回答候補の Hub 値として表現したものである。この時、回答候補が持つ(統合後の)断片の数で除することで、単純に文章量の多い回答候補に対して良い Hub 値が付与されないようにしている。

また、式(5)は、複数の良質な Hub 値を持った回答候補からリンクを張られることで重要性を評価し、複数の回答観点を持つ場合に詳細度のペナルティを与えることで(各項の分母)、重要な回答観点について詳細に述べた回答候補に対して、より良い Authority 値が付与されるようにしている。

なお、実際の HITS アルゴリズムの計算では、各回答候補の Hub 値と Authority 値の計算は、1 サイクル終わるたびに全回答候補の Hub 値および Authority 値の 2 乗和の平方根の大きさで除して、Hub 値と Authority 値が正規化される。提案手法でも同様に 1 サイクルごとに正規化を行う。

式(4)、(5)による計算を、以下の式(6)の条件を満たすまで繰り返す。

$$\sum_{A_i \in N} \left\{ \begin{array}{l} (HITS_H^{i+1}(A_i) - HITS_H^i(A_i)) + \\ (HITS_A^{i+1}(A_i) - HITS_A^i(A_i)) \end{array} \right\} < \epsilon \quad (6)$$

ここで、 N はネットワーク内の回答候補の集合であり、各 Hub 値および Authority 値は正規化後の値である。また、 ϵ は収束条件であり、実験では 0.0001 とした。この値は Mihalcea ら [1]が使った値と同じである。

4 評価実験

4.1 実験設定

提案手法の有効性を検証するための評価実験を行った。回答候補集合を取得する質問応答システムのモデルとして、石下ら [3]による質問応答システムを用いた。石下ら [3]のシステムでは、Web 文書を情報源として、定義や理由といった質問の型に依らない質問文解析を行うことで、幅広い質問に対応した回答を行える点が特徴である。質問文には、評価型ワークショップ NTCIR6 の QAC タ

スクにおける Formal Run テストセットに含まれる質問の内、ID の若い 30 問を利用し、システムによるスコア上位 100 件を回答候補集合として、複数の回答観点をまとめた回答を持つ回答候補が取得された 11 問を評価対象として利用した。

各質問における回答候補集合に対し、以下の 3 つの手法により回答候補集合に対して順位付けを行った。

【手法 A(ベースライン)】

各回答候補をノードとして、Mihalcea ら [1]による HITS アルゴリズムを用いた TextRank をそのまま適用する。

【手法 B】

3.3 節までの処理によりリンク生成を行った後、3.4 節による統合処理を行わず、3.5 節で述べた計算方法に基づき重要度を付与する。

【手法 C】

3 章で述べた手法に従い、回答候補に重要度を付与する。

各手法は、ランキング上位 10 件の回答候補を出力するものとする。各出力に対して、複数の回答観点到分割可能な回答候補を正解として、人手による正解判定を行う。

例えば「スケルトンとはどのような競技ですか」という質問に対して、ある回答候補の記述が「他競技との比較」を記述した部分や、「競技の種目」を記述した部分を持つ様な、複数の回答観点をを持った回答と判断できる場合、この回答候補を正解とみなす。

正解判定の結果に対して、MRR(最上位の正解順位の逆数の、全質問平均)を評価尺度として、各手法の比較評価を行った。

4.2 実験結果と考察

表 1: 各手法の評価実験結果

	MRR
手法 A(ベースライン)	0.168
手法 B	0.283
手法 C	0.575

表 1 の結果より、本稿のアプローチに基づくネットワークが、従来手法によるネットワーク構造と比較して、各回答観点をまとめた回答の検出に有効であることがわかった。また、手法 B と手法 C の結果から、回答候補の断片の統合処理を行うことで、より精度よく検出が行えることがわかった。手法 B では、ある回答候補について、類似度の高い他の回答候補が存在する場合、これらの回答候補同士があらゆるテキスト断片で相互にリンクを張る例が見られた。この結果、これらの回答候補が、まとめとして適切な回答候補と比較してリンクを持つテキスト断片の数に差分がなくなり、誤検出の要因となった(図 4)。

評価対象とした 11 問の内、「北アイルランド和平合意の内容はどのようなものですか」という質問において、手法 A では 10 位以内にまとめとなる回答をランキングできたが、手法 C では 13 位にランクされてしまった。手法 C の 12 位以前のランクの回答候補を見ると、質問に対する回答ではなく、「北アイルランドで行われた選挙に対する、和平合意への支持派・不支持派の得票数」を述べた回答候補が多く、集合全体を見ても、このトピックに関する回答候補が多数派を占めていた。本稿のモデル化によるネットワークでは、回答候補集合は質問に対する回答が多数存在

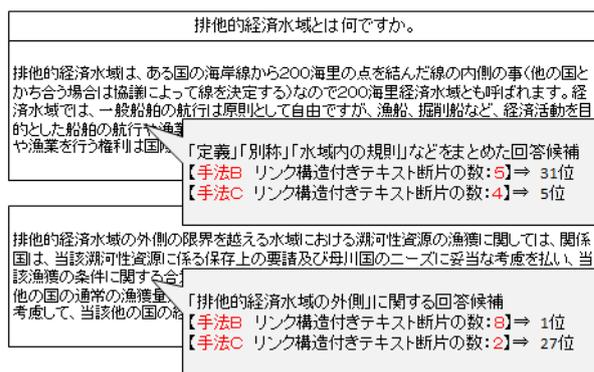


図 4 手法 B, C によるリンク構造の相違点

することが前提であり、上記の様に、質問に対する回答ではないあるトピックの内容が集合の多数派を占める場合など、質問回答システムの回答抽出精度を考慮していない。この様な問題に対処するため、ネットワークを利用したマップ化などインターフェースの工夫を行い、比較的小さな集合の中で情報をまとめている回答候補なども見せられる出力の組織化を検討する必要がある。

5 まとめ

質問応答における様々な回答観点对して、それらのまとめとなる回答候補を検出する手法を提案した。評価実験によって提案手法の有効性を示した。今後、回答候補の分割手法の精度向上を検討するとともに、ランキング以外の出力の組織化手法の検討を行う。

参考文献

[1] Rada Mihalcea, Paul Tarau, “TextRank: Bringing Order into Texts,” In The 2004 Conference on Empirical Methods in Natural Language Processing pp404-411, 2004.

[2] Jon M. Kleinberg, Ravi Kumar, Prabhakar Raghavan, Sridhar Rajagopalan, Andrew S. Tomkins, “The Web as a Graph: Measurements, Models, and Methods,” Computing and Combinatorics, pp1-17, 1999.

[3] 石下円香, 佐藤充, 森辰則, “Web 文書を対象とした質問の型に依らない質問応答手法,” 人工知能学会論文誌, Vol. 24, No. 4, pp. 339-350, 2009.

[4] Gees C. Stein, Tomek Strzalkowski, G. Bowden Wise, “Interactive, Text-Based Summarization of Multiple Documents,” Computational Intelligence, vol.16, pp.606-613, 2000.

[5] 藤井敦, 石川徹也, “World Wide Web を用いた事典知識情報の抽出と組織化,” 電子情報通信学会論文誌. D-II, 情報・システム, II-パターン処理 J85-D-II(2), pp.300-307, 2002.

[6] Masile Runs, Philip M. McCarthy, Danielle S. McNamara, Arthur C. Graesser, “A Study on Textual Entailment,” In Proceedings of the 17th IEEE International Conference on Tools with Artificial Intelligence(ICTAI'05), pp. 326-333, 2005.