解候補スコアの分布を利用したリスト型質問応答

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
E-mail: {ishioroshi,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．質問応答は，利用者が与えた自然言語の質問文に対し，その答を知識源となる大量の文書集合から見つける技術である．質問応答の基本的な仕組みは，

段階 1　知識源となる文書集合の中から，与えられた質問に対する解候補群を見つけること，

段階 2　各解候補に対し，その質問に対する答としての「良さ」を与える数値，すなわち，スコアを付与すること

からなる．利用者が，ある疑問に対する解を知るために質問応答システムを単体で利用する場合には，このスコアに基づき，解候補群を順序づけて上位から提示することが多い．本稿では，この処理を優先順位型質問応答と呼ぶことにする．この場合は解答として採用するか否かは，利用者の判断に委ねられている．

一方，質問応答技術は他の文書処理技術の中で活用されることも期待されている．質問応答の出力を他の文書処理技術の入力として容易に利用可能とするためには，優先順位型質問応答において利用者が行なっていた上記判断を自動的に行なう必要がある．また，「日本三景は何と何と何か」といったように複数の正解が存在する質問が存在することも考慮すべきである．これらのことより，決められた知識源の中から過不足なく与えられた質問の解を見つけ列挙す優る能力も重要であると考えられる．先順位型質問応答の用件に加え，この能力を持つ仕組みをリスト型質問応答と呼ぶ [1]．

本稿では，上記の背景の下，リスト型質問応答を行なうための一手法を提案する．本手法では，優先順位型質問応答により得られた解候補の集合のスコア分布が，正解集合に対するスコア分布と不正解集合に対するスコア分布の混合分布であると仮定し，これら二つの分布を EM アルゴリズムにより分離する．そして，正解側の分布に由来すると推定できる解候補を正解として出力する．質問応答システムには一般に不得意な質問が存在するが，本手法では，二つの分布のパラメタを比較することにより，優先順位型質問応答により解答が適切に見けられているか否かを判断することも可能である．更に，不得意な問題の場合については，得られた解候補群のスコアを再計算し，解候補群を再順位付けすることを提案する．

2　関連研究

リスト型質問に対しては，まだそれほど多くの研究は行なわれていない．秋葉ら [5] は期待効用最大化原理に基づく回答群選択手法を提案している．これは，リスト型質問応答の評価指標である F 値の期待値を求め，期待値を最大化するように回答数を求める手法である．また，スコアの差が最も開いているところよりも上位のものを回答とする手法 [2] や，最大スコアに対する比率に閾値を設けて回答を選択する手法 [4] が提案されている．

3　提案手法

リスト型質問応答の基本は，

段階 3　優先順位型質問応答システムの出力，すなわち，スコア付きの解候補群を正解 (と思しき) 集合と不正解 (と思しき) 集合の二つに分割する

ことである．これは，スコアの値に基づき，上位何件の解候補を正解と判断するかを決定することに等しい．我々は，スコアが適切に付与されているならば，正解集合のスコアの分布と不正解集合のスコア分布が異なる傾向を示すのではないかと考えている．そこで，本手法では，まず，

仮定 1　それぞれの集合のスコアの値が，異なるパラメタの確率分布に従っていること，そして，

仮定 2　解候補群全体のスコアはこの二つの確率分布の混合分布に従っていること

を仮定する．次に，各確率分布のパラメタを **EM** アルゴリズムにより推定し，二つの分布を分離する．最後に，各解候補がいずれの分布に由来するものなのかを推定し，最終的な正解集合とする．

一方，基本となる優先順位型質問応答において，システムの求める解析精度が十分でないこともある．例えば，我々の提案しているシステム [3] では，順位づけにおいても未だ満足のいくものではない．同システムでは，MRR 値[1]が 0.5 程度であり，平均すると 2 位に正解を見つけることができるという評価であるが，実際のところは 1/3 程度の問題について 1 位に正解を返し，2 ~ 5 位に正解を返すのが 1/3 程度，残りの問題については全く正解を得ることが出来ていない．つまり，システムにとって得意な問題と不得意な問題が存在している．

[1] Mean Reciprocal Rank．各問について最上位正解の順位の逆数を評価値とし，それを平均したもの．

我々は，不得意な問題の場合は，各解候補に対するスコアづけがうまくいっておらず，上述の仮定 1，仮定 2 が成立していないと考えた．そこで，推定した確率分布のパラメタに基づき，二つの分布が明瞭に分割できるか否かを調べる．これにより，解答が適切に見つかっているか否かを判断することがある程度可能であると考える．具体的には，二分布が明瞭に分割できると判断できる場合には，正解側の分布に由来すると推定できる解候補を正解として出力する．

一方，明瞭に分割できると判断できない場合には，優先順位型質問応答システムが適切なスコアを付与していない可能性がある．そのような，不得意な問題の場合については，得られた解候補群のスコアを再計算し，解候補群を再順位付けすることを考える．スコアの再計算は，各解候補について質問の疑問詞と置き換えることにより対応する平叙文を生成し，その文と知識源となる文書群のテキストとの間の照合の良さを判断することにより行なわれる．これは，知識源内で質問文中のキーワードと解候補が同じ文脈に出現する可能性を調べることにほぼ等しい．以下ではこの可能性を共起可能性と呼ぶ．

3.1 解候補スコアの分布の計算

解候補のスコアの分布が，正解集合が成す分布と不正解集合のそれとの混合分布であること，並びにその二分布がいずれも正規分布であることを仮定すると，混合分布は式 (1) で与えられる．なお，以下では，各問について各解候補のスコアを最大値を 1 に正規化したものを用いる．

$$p_s(x) = \sum_{i=1}^{2} \xi_i \phi_i(x : \mu_i, \sigma_i^2) \qquad (1)$$

ここで，x はスコアの値に対する確率変数，$\phi_i(x)$ は平均値 μ_i，分散 σ_i^2 の正規分布，ξ_i は二つの確率分布の混合比を表すパラメタである．混合分布であるという仮説が成り立つのであれば，平均値が大きい方の分布が正解集合がなす分布となる．以降では、その分布を ϕ_1 とする。

解候補スコアの観測値の集合が与えられた場合，式 (1) における各パラメタの推定は，EM アルゴリズムにより行なうことができる．

この二分布のパラメタから二つの分布が明確に分割できると判定できる場合には，スコアにより回答候補群を正解の群と不正解の群に分けられると判断し，3.2節の方法により，回答群を求める．二つの分布が明確に判断できるかどうかの判定法は3.4節で述べる．

3.2 解候補の正解判定

前節の方法により解候補のスコア分布が二つの分布に分離可能な場合，ある解候補が正解であるとする判断指標にはいくつか考えられる．スコアの降順で解候補 AC_i が並んでおり，AC_i のスコアを $score(AC_i)$ とすると，例えば，

正解判断指標 1 正解側の正規分布 ϕ_1 において，当該解候補がある閾値 Th_s 以上のスコアである．すなわち，$score(AC_i) \geq Th_s = \mu_1 + \alpha * \sigma_1$．ここで，$\alpha$ は Th_s を設定するためのパラメタである．

正解判断指標 2 二つの正規分布のうち，当該解候補が分布 ϕ_1 に属する確率 $p_{\phi_1}(AC_i)$ がある閾値 Th_p 以上である．すなわち，$p_{\phi_1}(AC_i) > Th_p$．

などが考えられる．ここで，$p_{\phi_1}(AC_i)$ は次式で与えられる．

$$p_{\phi_1}(AC_i) = \frac{\xi_1 \phi_1(score(AC_i))}{\xi_1 \phi_1(score(AC_i)) + \xi_2 \phi_2(score(AC_i))} \qquad (2)$$

なお，指標 1 は，Murata et al[4] などで用いられている，スコアの閾値に基づく手法に似ているが，確率分布の上で閾値を設定している点が異なる．

一方，Fukumoto et al[2] の方法のように，隣接するスコアの差に注目することも有効である．

正解判断指標 3 当該解候補 AC_i が，隣接するスコアの差が最大になる箇所よりも上位に現れている．すなわち，$i \leq argmax_j \{score(AC_j) - score(AC_{j+1})\}$

本稿では上記 3 つの指標を併用することにする．なお，指標 1,2 は個別スコアに対する指標となっているのに対し，指標 3 はスコア間の関係に基づく指標であることに注意されたい．

3.3 解候補スコアの分布の明確さの判定

解候補のスコアが明確に分割できるかどうかの判定は，二つの正規分布の混合比を表すパラメタ ξ_1 の値と隣接解候補のスコアの差の最大値 ($\max_j \{score(AC_j) - score(AC_{j+1})\}$) に基づいて行なう．

ξ_i は二つの分布の混合比であるので，$\xi_1 + \xi_2 = 1$ となる性質がある．予備実験によれば，正解集合に対応する正規分布 ϕ_1 の混合比 ξ_1 が小さい時ほど，図 1(a) に示すようにスコアの分布が明確に分かれており，解候補が適切に見つかっている可能性が高いことが観察された．これは，十分な量の解候補を標本としたときに，その中で正解となる解候補の数はそうでない解候補の数よりも相対的に少なくなっていることによると考えられる．一方で，明確に分布の差が現れていない時に，強制的に二つの分布に分離すると，ξ_1 の値が ξ_2 の値と同等になる (図 1(b))．

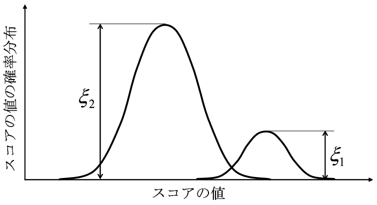
また，解の分布の如何によらず，解候補をスコアにしたがって二群に分けることを考えると，最も素朴な手法は，正解判断指標 3 に示される隣接する解候補のスコアの差が最も大きい箇所で分離するものであろう．この時，その差の最大値が大きい時ほどスコアの分布が明確に二つに分かれることを意味している．

本稿では，以上の二つの指標，すなわち

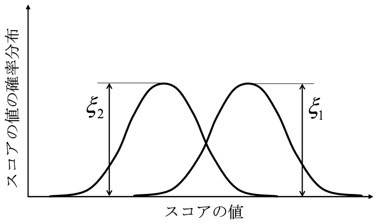
分離指標 1 $\xi_1 <= Th_\xi$

分離指標 2 $\max_j \{score(AC_j) - score(AC_{j+1})\} >= Th_{diff}$

とすることとし，上記分離指標のいずれかを満たす場合を，スコアの分布が明確に分かれていると判断した．



(a) 差が明瞭な場合



(b) 不明瞭な場合

図 1: スコア分布の差と ξ_1 の関係

3.4 共起可能性の計算

3.1節で計算された分布が節の手法により明確に分離されていないと判断された場合には，質問文中のキーワードと解候補の知識源内での共起可能性を計算する．

使用している質問応答システムは質問として平叙文も受け付ける．そのときにシステムから得られる解は，知識源内で平叙文中のキーワードと関連性が高い語である．この語群のスコアの分布を 3.1節と同じように計算すると，平叙文中のキーワードの網羅性が高い文脈が存在する場合，すなわち，共起可能性が高い時ほど二つの分布が明確に分割できるという傾向が見られた．

このことを利用して，質問文中の疑問詞を解候補に置き換えて平叙文を作り，この平叙文を質問応答システムに入力して得られた語群のスコアの分布を見ることにより，質問文中のキーワードと解候補の知識源内での共起可能性を計る．全ての解候補に対して質問文中のキーワードとの共起可能性を算出し，解候補群が，特に共起可能性が高い解候補の群とそうでない群とに明確に分割できる場合には，特に共起可能性が高い解候補の群を回答群とする．

ここでの共起可能性の計算は 3.4節で述べた「ξ_1 の値が小さい時ほど二つの分布が明確に分かれる」という観察に基づいている．その計算過程を図 2に示す．

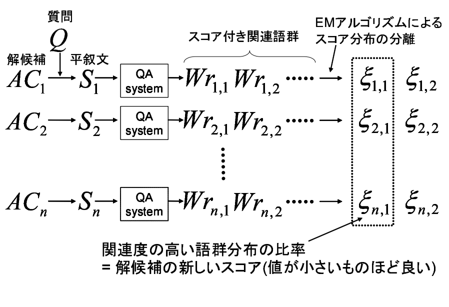


図 2: スコアの再計算

まず，平叙文に対し疑似的な解 (関連語群) を得てその分布を調べる．このとき ξ_1 の値が小さいときほど，共起可能性が高いと判断できる．この時重要になるのは ξ_1 の値の小ささであるので，各平叙文の ξ_1 の値を二つのグループに分けた時の平均値の差が閾値以上であるならば，共起可能性に差が出ていると判断し，ξ_1 の値小さい平叙文を作った解候補の群を正解とする．

共起可能性にあまり差が見られなかった場合には，作られた平叙文中のキーワードのうち，もともと質問文中にあったキーワードを一つ取り除き，その部分を疑問詞にすることで新しい質問文を作成する．各解候補についてそれぞれ作られた質問文について質問応答を行なって得られた解候補群のスコアの分布を計算することにより，新しいスコアを図 2のように再計算する．疑問詞に置き換えるキーワードごとに各解候補に新しいスコアが付けられ，そのスコアの分布が求められる．全てのキーワードに対し各解候補につく新たなスコアの分布を求め，スコアの分布が最も明瞭に分かれるものを採用とする．採用された分布のうち，値が小さいほうの分布に属しているスコアがついている解候補を正解の可能性が高いと判断し，回答群に加える．

もし，全ての場合において，解候補群と質問文中のキーワードの共起可能性に差が見られなかった場合には，3.2節の方法により回答群に加える解候補を選択する．

4 評価実験

QAC1[1] のテストコレクションを用いて評価実験を行なった．以前に述べた各パラメタは，予備実験に基づき $Th_s = \mu_1$(値が大きい方の分布の平均値，すなわち $\alpha = 0$)，$Th_p = 0.5$，$Th_\xi = 0.3$ ，$Th_{diff} = 0.3$ としている．また，優先順位型質問応答システムからの出力は上位 10 位を採用している．

4.1 解答が適切に見つかっている問いの判定

まず，分布の明確さの判定の有効性を見るために，スコアの分布が明瞭な場合とそうでない場合とに分割し，それぞれの場合について 3.2節の三つの正解判断指標を満足するという条件のもと，解候補の正解判定を行なった．平均回答数と平均 F 値 (AFM) を表 1に示す．

表 1: スコアの分布の違いによる平均回答数と平均 F 値 (AFM) の変化

	問いの数	平均回答数	AFM
分布が明瞭	82	1.31	0.555
分布が不明瞭	116	2.44	0.225
合計	200	1.96	0.360

分布が明瞭な場合と不明瞭な場合とで平均 F 値に明確な差がでている．また，分布が明瞭な場合平均回答数を見ると 1.31 と少なにも関わらず平均 F 値が高く，選んだ回答群の精度が良いことが窺える．分布が不明瞭な

場合には平均回答数が多く，回答を多めに選んでも正解が含まれていない問いが多いことが分かる．

このことから，提案手法により解答が適切に見つかっている問いのみを抽出することがある程度可能であることが分かる．

4.2 スコアの再計算

4.1節で分布が不明瞭と判断された問い 116 問に対して以下の 3 つの手法で解候補の正解判定を行なった．手法 1 ではスコアの再計算は行なわず，手法 2,3 は得られた解候補のスコアを再計算し，解の抽出を行なうものである．

手法 1 3.2節の 3 つの条件を全て満たすもののみを回答とする

手法 2 3.4節の手法で新たに平叙文を生成し，スコアの再計算をし，正解判断指標 1 により回答を得る

手法 3 手法 2 に加え，平叙文のみで判定できない場合にはさらに新たな質問文をつくってスコアを再計算し，正解判断指標 1 により回答を得る

手法 2，3 共に新たに求めたスコアの分布が不明瞭と判断される時は，手法 1 の方法で最終的な回答を求めている．それぞれの結果を表 2に示す．表中の手法 1 ～ 3 はそれぞれ上記の手法 1 ～ 3 に対応している．

表 2: 解候補のスコアの分布が不明瞭な問いに対する各手法と精度の違い

	平均回答数	抽出できた正解の数	AFM
手法 1	2.44	50	0.225
手法 2	2.59	51	0.226
手法 3	3.07	55	0.219

上記 3 つの手法にそれほど大きな差は見られないが，回答数が多いほど，抽出できた正解の数が増えているのがわかる．解候補のスコアの分布が明確でない問いでは，正解が解候補群に含まれてはいるが，スコアが上位の方には現れていない場合があるということがわかる．

5 考察

解候補のスコアの分布を求めることにより，解答が適切に見つかっているか否かの判断が可能であることがわかった．また，正解判断指標 1,2,3 により，平均 F 値 0.36 程度の精度が得られることがわかった．

ただ，解答が適切に見つかっている問いを全て抽出できているわけではない．例えば，1 位に正解があるような例でも適切に判断できないこともある．これは，解候補のスコアの分布が明瞭だと判断する条件をある程度厳しくしている為だと考えられる．条件を厳しくしているのは確実なもののみを取り出すためであるが，判断の基準をさらに検討することで精度が向上することが予想される．

問いに関しては，解候補の数を多めに回答に加えることにより，再現率が向上することがわかった．正解と思われる解候補のみをうまく抽出する必要がある．

3.4節でのスコアの再計算は正解と思われる解候補だけをうまく抽出しようというものであったが，あまり精度の向上にはつながらなかった．抽出できた正解の数は増えたが，不正解のものも回答に加えてしまう傾向にある．各解候補とキーワードとの間の共起可能性に明確な差があるかどうかの判定基準や，共起可能性に差があった場合にどこまでの解候補を回答に加えるかなどのパラメタを変えると出力結果が大きく変わるので，今後適切なパラメタを求める必要がある．

また，解候補のスコアの分布が不明瞭な場合では，解候補の中に正解が含まれている場合と，解候補の中に全く正解が含まれていない場合とがある．解候補の中に正解が含まれていないという判定も可能になれば，スコアの再計算の際にかかる時間を短縮できると考えられる．

6 おわりに

本稿では，リスト型質問に対するの二つの手法を提案した．

第 1 に解候補のスコアの分布を求めることにより解答が適切に見つかっている問いの判定をする手法を提案した．スコアの分布を求め，それを利用することはリスト型質問応答に対して有効に働くことが分かった．

第 2 に，質問文中のキーワードと解候補の共起可能性を求め解候補群のスコアを再計算することにより精度を向上させる手法を提案した．スコアの再計算をすることにより抽出できる正解は増えたものの正解を効率良く抽出できているとはいい難く，スコアの再計算の方法，解答の選択の方法などを，見直す必要がある．

参考文献

[1] Jun'ichi Fukumoto, Tsuneaki Kato, and Fumito Masui. Question Answering Challenge (QAC-1) — Question answering evalu ation at NTCIR Workshop 3 —. In *Working Notes of the Third NTCIR Workshop meeting – Part IV: Q uestion Answering Challenge (QAC1)*, pp. 1–6, 2002.

[2] Jun'ichi Fukumoto, Tatushiro Niwa, Makoto Itogawa, and Megumi Matsuda. Rits-QA: List answer detection and Context task with ellipses handling. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pp. 310–314, June 2004.

[3] Tatsunori Mori. Japanese Q/A System using A* Search and Its Improvement: Yokohama National University at QAC2. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pp. 345–352, 6 2004.

[4] Masaki Murata, Masao Utiyama, and Hitoshi Isahara. Japanese Question-Answering system Using Decreased Adding with Multiple Answers. In *Working Notes of the Fourth NTCIR Workshop Meeting*, pp. 353–360, June 2004.

[5] 秋葉友良, 伊藤克亘, 藤井敦. 質問応答における常識的な解の選択と期待効用に基づく回答群の決定. 自然言語処理研究会報告 2004-NL-163, 情報処理学会, 9 月 2004.