

## リスト型質問応答のための回答候補間共起情報を利用したスコアリング改善手法の提案

山本康貴 †

佐藤晴彦 †

小山聡 †

栗原正仁 †

北海道大学大学院情報科学研究科

## 1 はじめに

近年、新しい情報検索技術の一つとして質問応答が注目されている。質問応答とは自然言語で表現された質問に対して適切な回答を提示する技術のことであり、クエリ言語に関する知識を持たないユーザーでも容易に利用可能であるということが大きな特徴である。

質問応答は扱う質問や回答の性質によりいくつかの種類に分類される。代表的なものとしては単純な事物を回答するファクトイド型質問応答や理由、原因を回答する Why 型質問応答などがあるが、基本的な手法としてはどれも知識源となる文書の中から回答候補を列挙してその回答候補に対してスコアリングを行い、最もスコアが高いものを回答とするという点において共通している。

質問応答の一種に「札幌市の姉妹都市はどこですか?」というような回答すべき語が複数存在するタイプの質問を取り扱うリスト型質問応答がある。リスト型質問応答は一見するとファクトイド型質問応答の回答候補のうちスコアが上位のものを回答として提示すればいいように思える。実際、リスト型質問応答というタスクが提案された当初には単純にスコアの上位 10 件の回答候補をそのまま回答とするようなシステム [1] も見られた。ところがこのような手法だけでは過不足無く全ての回答を求めるのは難しく、特に正答の個数は予め与えられないために回答候補からいくつかの回答を正答として抽出すべきかシステムが推測しなければならないという問題がある。そのために回答群抽出の手法としてスコア分布を用いる手法 [2] や階層的クラスタリングを用いる手法 [3] などが提案されているが、これらの手法の多くは回答群抽出を行う前段階のスコアの精度に大きく依存するという問題がある。

そこで本稿ではリスト型質問応答における回答群抽出の精度を向上させることを目的として、その前段階であるスコアリングを回答候補間の共起情報を用いて改善する手法を提案する。

## 2 提案手法

提案手法では従来の手法同様に初期スコアを算出した後、回答群の抽出に入る前に回答候補間の共起強度を計算する。その後初期スコアに対して共起強度に基づいた修正を加える。

## 2.1 初期スコア

本稿では回答候補  $c$  の初期スコア  $S(c)$  はキーワード  $k$  との間の形態素数  $dist(c, k)$  から次の式に従って計算した。

$$S(c) = \sum_i \frac{1}{dist(c, k_i)} \times idf_c \quad (1)$$

$$idf_c = \frac{|D|}{|\{d : d \ni c\}|} \quad (2)$$

ここでキーワードとは質問文中に含まれる特徴語のことであり、本稿においては質問文中に含まれる代名詞と非自立語を除く名詞としている。また、 $|D|$  は知識源となる文書の総数、 $|\{d : d \ni c\}|$  は単語  $c$  を含む文書数で一般に  $idf_c$  は逆文書頻度と呼ばれており、一般語や頻出語のスコアが過剰に高くないよう補正している。

## 2.2 共起強度

共起強度とはある 2 つの単語がどれだけ同時に出現しているかを表す尺度である。共起強度の計算手法はいくつか提案されているが、本稿では次の式のコサイン距離を用いた。

$$\text{コサイン距離} = \frac{|X \cap Y|}{\sqrt{|X||Y|}} \quad (3)$$

## 2.3 共起強度を用いたスコアの修正

リスト型質問応答では回答すべき語が複数存在しており、それら全てを過不足無く回答しなければならない。式 (1) で算出された初期スコアによるランキングでは正答となる回答群の多くは上位に存在しているものの、正答にも関わらず下位になってしまっている回答候補も存在する。10 の質問に対して初期スコアを計算した場合の平均適合率を表 1 に示す。

従来の手法ではこのような初期スコアを元にして回答群の抽出を試みていたが、前述の通り正答となる回答群の一部に低いスコアがついてしまっている場合が多く、そのためにいくら回答群抽出手法を工夫しても

A Scoring Method using Co-occurrence of Candidate Answers for List-type Question Answering

†Kouki YAMAMOTO, Haruhiko SATO, Satoshi OYAMA, Masahito KURIHARA

Graduate School of Information Science and Technology, Hokkaido University

	平均適合率
Q1	0.808
Q2	0.572
Q3	0.469
Q4	0.613
Q5	0.529
Q6	0.204
Q7	0.809
Q8	0.525
Q9	0.71
Q10	0.639

表 1: 初期スコア-平均適合率

	平均適合率
Q1	0.835
Q2	0.925
Q3	0.404
Q4	0.625
Q5	0.493
Q6	0.32
Q7	1.0
Q8	0.331
Q9	0.563
Q10	0.917

表 2: 修正後スコア-平均適合率

全ての正答を抽出することは難しかった。そこで2.2項で述べた共起強度を用いてスコアの修正を行う。

全ての回答候補間の共起強度を計算すると回答候補をノード、共起強度をエッジとしたネットワークが構成できる。これは共起ネットワークと呼ばれる。この共起ネットワーク上でより太いエッジで結ばれている、すなわち強く共起している回答候補同士は互いに関係している可能性が高いと言える。また、リスト型質問の正答となる回答候補同士は同じグループに属する語であるはずなので関係が深く、逆に正答とは異なるグループに属する回答候補は正答とは関係がほとんど無いと考えられる。そこで正答である可能性の高い、つまり高いスコアの付いている回答候補と強く共起している回答候補もまた正答であるという仮説に基づき、エッジの太さに応じた重み付けをして共起している回答候補のスコアに応じた修正を行う。ある2つの回答候補  $c_i$  と  $c_j$  の共起強度を  $C(c_i, c_j)$  とすると回答候補  $c$  の修正後のスコア  $S'(c)$  は式(4)によって求められる。

$$S'(c) = \alpha S(c) + (1 - \alpha) \sum_i (C(c, c_i) \times S(c_i)) \quad (4)$$

ここで  $\alpha$  は  $0 \leq \alpha \leq 1$  の重み係数である。

表1と同様の質問に対してスコア修正を行った場合の平均適合率を表2に示す。

### 3 おわりに

本稿では、リスト型質問応答における回答群抽出を容易にするために回答候補間の共起情報を用いたスコアリングの改善手法を提案した。また、システムの実装を行い回答候補間の共起情報を利用することの有用性を確認した。

提案手法を用いることでスコアが悪化している質問に関しては初期スコアの時点で既に誤答に最も高いスコアが付いてしまっており、それと共起した誤答のスコアまでもが低い値に修正されてしまったことが原因と考えられる。これはリスト型質問応答における初期スコアへの依存性の問題がいまだに残っているということであり今後の課題となっている。現在は初期スコアとは異なる知識を併用して回答群を抽出するための手法を検討中である。

### 参考文献

- [1] Edward W. D. Whittaker, Matthias H. Heie, Josef R. Novak, and Sadaoki Furui. Trec 2007 question answering experiments at tokyo institute of technology. In *TREC*, 2007.
- [2] 石下円香, 森辰則. 優先順位型質問応答の解スコア分布に基づくリスト型質問応答. *Journal of natural language processing*, Vol. 16, No. 1, pp. 67–100, jan 2009.
- [3] Majid Razmara and Leila Kosseim. Answering list questions using co-occurrence and clustering. In Bente Maegaard Joseph Mariani Jan Odijk Stelios Piperidis Daniel Tapias Nicoletta Calzolari (Conference Chair), Khalid Choukri, editor, *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may 2008. European Language Resources Association (ELRA). <http://www.lrec-conf.org/proceedings/lrec2008/>.