

# 学習型質問応答における単語文脈ベクトルの効果の研究

## Research on the Effects of Word Context Vectors in Trainable QA

牧瀬 晃太

三輪 誠

佐々木 裕

Kota Makise

Makoto Miwa

Yutaka Sasaki

豊田工業大学

Toyota Technological Institute

{sd13439, makoto-miwa, yutaka.sasaki}@toyota-ti.ac.jp

### 1 はじめに

キーワードを用いた一般的な検索エンジンはユーザに提示する情報が多く、求める情報にたどり着くのに時間を要する。このユーザの求める情報を得る方法として、短く・簡潔な回答を対象としたファクトイド型質問応答がある。このような質問応答において問題となるのが質問文と回答の文脈の表現の違いである。本研究では、この問題の解決による質問応答の精度向上を目的として、Conditional Random Fields (CRF) を用いた学習型質問応答において単語を表記そのものではなくベクトルとして表現する単語文脈ベクトルの特徴としての利用可能性について研究する。

### 2 学習型質問応答

質問応答システムの多くは以下のような構成になっている。

1. 質問解析：質問タイプの判定と文書検索のための検索語の抽出する。
2. 情報検索：検索エンジンに検索語を入力し上位の文書を抽出する。
3. 回答候補抽出：情報検索で得られた文書から回答候補を抽出する。
4. 回答候補評価：質問の回答としてふさわしいものかどうかを判定して回答候補の順位付けを行う。

構文解析情報を特徴として機械学習を質問応答に適用したものは [5] がある。これは日本語に対して行われた研究で、質問文と文書の単語の組み合わせを特徴

として質問応答システムが構築できることを示した。以下に用いられた特徴を示す。

**QW (Question Word)** :  $n$ -gram ( $1 \leq n \leq N$ ) で質問文の単語を列挙したもの。質問文が “What is cancer?” で  $N=2$  ならば QW:What, QW:is, QW:cancer, QW:What-is, QW:is-cancer となる。

**QI (Question Interrogative)** : 質問内容を表す単語 (who, where, what, how)。

**QP (Question POS)** : 質問文の単語の品詞を列挙したもの。質問文が “What is cancer?” ならば QP:wh, QP:verb, QP:noun となる。

**DW-k,...,DW0,...,DW+k (Document Word)** : 解答に関連する文書から得られた単語。対象の単語が  $w_i$  で  $k=1$  ならば  $w_{i+1}$ ,  $w_{i+0}$ ,  $w_{i+1}$  となる。

**DP-k,...,DP0,...,DP+k (Document POS)** : 解答に関連する文書から得られた単語の品詞。

**CW-k,...,CW0,...,CW+k (Combined Word)** : 解答に関連する文書の単語が質問文に存在するかどうか (true/false)。QW:President, DW-1:President ならば CW-1:true となる

**CP-k,...,CP0,...,CP+k (Combined POS)** : 解答に関連する文書の単語の品詞が質問文に存在するかどうか (true/false)。

**CI-k,...,CI0,...,CI+k (Combined Interrogative)** : 解答に関連する文書の単語と QI を組み合わせたもの。QI:Who, DW-1:President ならば CI-1:President & Who となる。

この研究では  $N=4$ ,  $k=3$  として用いられていた。回答の抽出には最大エントロピーモデル (ME) が使用されている。

### 3 提案手法

本研究では回答候補の出現する文脈と質問文に含まれる情報が関係しているという点に着目する。そこで、英語の生物学文献を対象に ME よりも多くの研究において高い精度を示している CRF を用いて回答候補の抽出モデルを学習し、CRF により出力された条件付き確率をもとに最終的な回答を決定する手法を構築し、単語文脈ベクトルの特徴としての効果を検証する。

#### 3.1 回答候補の抽出

本研究ではスニペットと質問文を構文解析することにより得られる言語解析情報を用いる手法と、それに加えて文脈情報を用いる手法を提案する。言語情報からの素性は関連研究 [5] に従う。スニペットとは解答に関連するとされる部分を文書から抽出したものである。文脈情報として質問文から抽出した Question Focus (質問の対象カテゴリを指定する単語) と対象単語の距離を利用する。距離の算出には単語文脈ベクトル間のコサイン距離を用いる。回答の抽出には閾値を設け、条件付き確率の和が閾値以上の値になるまで回答の抽出を行う。

##### 3.1.1 基本素性

以下の素性を基本素性として用いる。

**B-k, ..., B0, ..., B+k (Base)** : 解答に関連する文書の単語の原形。

**QF (Question Focus)** : 質問文から抽出した Question Focus。

**DWC (Document Word Combination)** : 解答に関連する文書から得られた単語を  $n$ -gram ( $2 \leq n \leq C$ ) で組み合わせたもの。対象の単語が  $w_i$  で  $C=2$  ならば  $DWC:w_{i-1} \& w_i$ ,  $DWC:w_i \& w_{i+1}$  となる。

**DBC (Document Base Combination)** : 解答に関連する文書から得られた単語の原形を  $n$ -gram ( $2 \leq n \leq C$ ) で組み合わせたもの。

**DPC (Document POS Combination)** : 解答に関連する文書から得られた単語の品詞を  $n$ -gram ( $2 \leq n \leq C$ ) で組み合わせたもの。

##### 3.1.2 言語解析情報を用いる手法

この手法では基本素性、関連研究で用いられた素性 (QI, DW, DP, CW, CP), 以下の素性を用いる。

- DW0+CW0+QF
- DW0+CP0+QF
- DW0+CW0+CP0+QF
- B0+CW0+QF
- B0+CP0+QF
- B0+CW0+CP0+QF
- B0+QF
- CW0+CP0+DP0+QI

##### 3.1.3 文脈情報を追加した手法

この手法では基本素性、関連研究で用いられた素性 (QI, DW, DP, CW, CP), 文脈ベクトルを導入した以下の素性を用いる。

**DQF-k, ..., DQF0, ..., DQF+k (Distance QF)** :

QF と DW0 の文脈ベクトルの距離。

- DW0+CW0+QF+DQF0
- DW0+CP0+QF+DQF0
- DW0+CW0+CP0+QF+DQF0
- B0+CW0+QF+DQF0
- B0+CP0+QF+DQF0
- B0+CW0+CP0+QF+DQF0
- B0+QF+DQF0
- CW0+CP0+DP0+QI

距離は 9 つの bin によりブール値で表す。表 1 の例を用いて説明する。ブール値は 0.1 刻みで分ける。distance が 0.09 であった場合 0.1 を超えていないのですべてのブール値が false となる。0.10 の場合は 0.1 を超えるので一つ目のブール値が true になり、他のブール値は false となる。同様に 0.2 の場合は最初の 2 つ目までが true となり、他のブール値は false となる。0.95 の場合は 0.9 を超えているのですべてのブール値が true となる。

表 1: 単語間文脈ベクトルの距離の素性への変換。

distance	ブール値								
	1	2	3	4	5	6	7	8	9
0.09	F	F	F	F	F	F	F	F	F
0.10	T	F	F	F	F	F	F	F	F
0.20	T	T	F	F	F	F	F	F	F
0.95	T	T	T	T	T	T	T	T	T

### 3.2 回答候補の順位付け

各スニペットに対する条件付き確率の高い順に回答候補の抽出を行う。この条件付き確率はスニペット単

位で見た時の回答抽出の確率を表す。その際の条件付き確率の和をそれぞれの回答候補の確度とする。その確度を基に回答候補の順位付けを行う。

## 4 BioASQ チャレンジ

本研究では実験データの取得・システムの評価のため医療分野における英語の質問応答の競技会である BioASQ チャレンジ 2013, 2014[6]に参加した。以下で用いたデータの説明をする。

1. 質問・解答：チャレンジで使用されたもの。解答は人手で付けられている。
2. スニペット：生物・医学文献のデータベースである PubMed[4]に登録されている文献から抽出されたもの。
3. 単語文脈ベクトル：word2vec[3]のスキップグラムモデルを PubMed に登録されている文献の抽象ラクト, 10,876,004 件に適用し文脈ベクトルを作成したたもの。1,701,632 個の単語それぞれに単語文脈ベクトルが付与されている。

Wishart システム [2] は 2013, 2014 年の優勝システムである。Wishart システムは質問文の中から名詞句を抽出。抽出された名詞句を生物学のシソーラスに照合することで曖昧性の解消されたクエリを生成し、そのクエリをもとに類義語や関連する生物学用語を PolySearch[1] というテキストマイニングシステムを用いて検索し、新たなクエリを生成する。それらのクエリを使用し解答に関係する文書を検索し、得られた文書に解答に関連する用語がどれだけ含まれているかという情報をもとに文書に対する順位付けを行っている。そこで 1 位のものを要約したものから回答候補を抽出している。

## 5 実験

実験に用いた手法を以下に示す。

1. Rule-based (Rule)：手書きルールを用いた手法。
2. ベースライン (Baseline)：言語解析情報を用いる手法。
3. +WCV (Word Context Vectors)：文脈情報を追加した手法。

本研究では  $C=3$ , 関連研究に従い  $N=4$ ,  $k=3$  とした。

学習データとして 2013 年のデータを用い(質問正解数:54, スニペット数:1,247), テストデータとして 2014 年のデータ (質問正解数:132, スニペット数:1,985) を

用いた実験と, 2014 年のデータを学習データとして用い, テストデータとして 2013 年のデータを用いた実験を行った。

閾値と CRF に用いたハイパーパラメータは学習データを用いて leave-one-out cross-validation を行うことで決定した。閾値は 0.8 から 1.0 まで 0.01 刻みで行い MRR が一番高いものを採用した。ハイパーパラメータは  $e^{-6}$  から  $e^6$  まで乗数を 1 刻みで行い一番高いものを採用した。交差検定の結果, 2013 年のデータを学習に用いたときはベースラインの閾値は 1, ハイパーパラメータは  $e^0$ 。+WCV の閾値は 1, ハイパーパラメータは  $e^{-1}$ , 2014 年のデータを学習に用いたときはベースラインの閾値は 1, ハイパーパラメータは  $e^{-3}$ , +WCV の閾値は 1, ハイパーパラメータは  $e^{-2}$  として実験を行った。

## 6 結果と考察

2013 年のデータは test1~test3 の 3 つに分かれており, それぞれの結果を表 2~4 に示す。2014 年のデータは test1~test5 の 5 つに分かれており, それぞれの結果を表 5~9 に示す。評価値は以下のものを用いた。

- Strict Accuracy (Strict)  
システムの出した回答候補の中の 1 位のものが正しい回答の割合。
- Lenient Accuracy (Lenient)  
システムの出した上限 100 件の回答候補の中に正しい回答が存在する割合。
- Mean Reciprocal Rank (MRR)  
下記の式で表される数値になる。

$$\text{MRR} = \frac{1}{Q} \sum_{i=1}^Q \frac{1}{\text{rank}_i}$$

$Q$  は問題数,  $\text{rank}_i$  は回答候補内の正しい回答の最上位の順位である。回答候補内に正しい回答が含まれない場合  $\frac{1}{\text{rank}_i}$  は 0 とする。

Whishart システムは 2013, 2014 年 BioASQ チャレンジの優勝システムであり, 2013 年の test3, 2014 年の test5 は不参加であった。CRF を用いたシステムがルールベースで構築したシステムを上回ることがわかった。また, CRF を用いたシステムの中でも本研究の提案手法である質問の文脈ベクトルを素性として加えたものが Lenient Accuracy では Wishart を上回った。以上の結果より文脈ベクトルによる意味情報の付与が精度向上に大きく貢献することを明らかにした。

表 2: 2013 年 test1 の結果.

System	Strict	Lenient	MRR
Rule	0.0000	0.0556	0.0185
Baseline	0.0556	0.2778	0.1194
+WCV	0.0000	<b>0.3333</b>	0.0860
Wishart	<b>0.2222</b>	<b>0.3333</b>	<b>0.3056</b>

表 3: 2013 年 test2 の結果.

System	Strict	Lenient	MRR
Rule	0.0000	0.1500	0.0667
Baseline	0.1500	<b>0.4500</b>	0.1912
+WCV	0.1500	<b>0.4500</b>	0.2057
Wishart	<b>0.2500</b>	0.3000	<b>0.3000</b>

表 4: 2013 年 test3 の結果.

System	Strict	Lenient	MRR
Rule	0.0625	0.1875	0.0906
Baseline	<b>0.1875</b>	0.5625	<b>0.2942</b>
+WCV	<b>0.1875</b>	<b>0.6250</b>	0.2800

表 5: 2014 年 test1 の結果.

System	Strict	Lenient	MRR
Rule	0.0370	0.1481	0.0926
Baseline	0.1852	<b>0.5185</b>	0.2586
+WCV	0.1852	<b>0.5185</b>	0.2655
Wishart	<b>0.4400</b>	0.4800	<b>0.4600</b>

表 6: 2014 年 test2 の結果.

System	Strict	Lenient	MRR
Rule	0.0435	0.1739	0.0942
Baseline	0.0435	<b>0.3478</b>	0.1281
+WCV	<b>0.1304</b>	<b>0.3478</b>	<b>0.1969</b>
Wishart	<b>0.1304</b>	0.1304	0.1304

表 7: 2014 年 test3 の結果.

System	Strict	Lenient	MRR
Rule	0.0000	0.0833	0.0208
Baseline	<b>0.0417</b>	<b>0.4167</b>	0.1088
+WCV	<b>0.0417</b>	<b>0.4167</b>	<b>0.1193</b>
Wishart	<b>0.0417</b>	0.0833	0.0556

表 8: 2014 年 test4 の結果.

System	Strict	Lenient	MRR
Rule	0.0938	0.3438	0.1979
Baseline	0.1875	0.5313	0.2585
+WCV	<b>0.2500</b>	<b>0.5625</b>	<b>0.2875</b>
Wishart	<b>0.2500</b>	0.2813	0.2813

表 9: 2014 年 test5 の結果.

System	Strict	Lenient	MRR
Rule	0.0345	0.1724	0.0787
Baseline	<b>0.1034</b>	0.3103	<b>0.1297</b>
+WCV	<b>0.1034</b>	<b>0.3448</b>	0.1176

## 7 おわりに

文脈ベクトルを CRF の素性として用いることによりファクトイド型質問応答への回答精度が向上することを示した. 今後は文法情報を用いることで更に精度が上がる可能性があると考え.

## 参考文献

- [1] Dean Cheng, Craig Knox, Nelson Young, Paul Stothard, Sambasivarao Damaraju, and David S Wishart. Polysearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites. *Nucleic acids research*, Vol. 36, No. suppl 2, pp. W399–W405, 2008.
- [2] Yifeng Liu. The university of alberta participation in the bioasq challenge: The wishart system. Technical report, 2014.
- [3] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In C.J.C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pp. 3111–3119. Curran Associates, Inc., 2013.
- [4] PubMed. <http://www.ncbi.nlm.nih.gov/pubmed>.
- [5] Yutaka Sasaki. Question answering as question-biased term extraction: a new approach toward multilingual qa. pp. 215–222, 2005.
- [6] George Tsatsaronis, Michael Schroeder, Georgios Paliouras, Yannis Almirantis, Ion Androutsopoulos, Eric Gaussier, Patrick Gallinari, Thierry Artieres, Michael R Alvers, Matthias Zschunke, et al. Bioasq: A challenge on large-scale biomedical semantic indexing and question answering. 2012.