

統計的フレーズ翻訳モデルを用いた言語横断質問応答

有賀美明^{†1} 秋葉友良^{†1}

言語横断質問応答は、質問文と検索対象文書の間で言語が異なる質問応答を扱うタスクである。前処理として翻訳により質問文と検索対象文書の言語を統一することで、単一言語の質問応答の問題に帰着することができる。一方、前処理での翻訳の代わりに質問応答のプロセスに統計的翻訳モデルを組み込む手法が提案され、有効性が示されている。先行研究では、翻訳モデルとして単語翻訳モデルが用いられていたが、統計的機械翻訳の分野ではフレーズベース手法の有効性が示されている。本稿では、翻訳モデルとしてフレーズ翻訳モデルを組み込んだ手法を提案する。評価の結果、提案法は単語翻訳モデルを用いた従来法を上回る性能を示した。

Using Statistical Phrase-Based Translation Model for Cross-Language Question Answering

YOSHIKI ARUGA^{†1} and TOMOYOSHI AKIBA^{†1}

Cross-Language Question Answering (CLQA) is the special case of Question Answering that aims to find the answers written in the language different from that of the query. While a conventional approach for the task is to apply the monolingual QA to the translated query using machine translation, there was proposed the method to incorporate statistical machine translation deeply into the QA process, in which word-based translation models were used. In this work, we propose to use phrase-based translation models to be incorporated in the method. The experimental evaluation using the NTCIR CLQA 1 test collection showed that the performance was improved by using the phrase-based model.

^{†1} 豊橋技術科学大学
Toyohashi University of Technology

1. はじめに

インターネットをはじめとする情報通信網の発展により、個人が利用可能な情報の量は年々増加している。多種多様な言語で記述された大規模な情報の中から求める情報を検索するには多大なコストがかかるため、言語障壁の解消と効率的な検索手法が重要視されている。

質問応答は、組織化されていない検索対象文書から自然言語の質問文に対する直接の回答を求める技術である。言語横断質問応答とは、質問文と検索対象文書の言語が異なる場合の質問応答である。CLEF[1] や NTCIR[2] にて、factoid 型質問応答を対象とした言語横断質問応答システムの評価が行われてきている。

言語横断質問応答システムを実現するためには、翻訳が鍵となる。前処理によって質問文と検索対象文書の言語を統一すれば、単一言語の質問応答の問題に帰着する。一般的な手法としては、質問文を検索対象文書と同一の言語に翻訳してから、単一言語の質問応答システムで回答抽出する手法が主であった。一方、先行研究 [3] では、統計的翻訳モデルをシステムに組み込み、原言語から直接回答抽出を行う言語横断質問応答を提案している。

先行研究では、単語翻訳モデルを用いた手法が提案されていたが、統計翻訳ではフレーズベースに基づいた手法が高い性能となることが知られている。本稿では、フレーズ翻訳モデルを組み込んだ日英の言語横断質問応答システムを実装する。翻訳にフレーズを導入することで単語間の関係を利用でき、パッセージ検索における性能の改善が期待できる。評価実験の結果、フレーズ翻訳モデルを用いた提案法は単語翻訳モデルを用いた従来法を上回る性能を示した。

2. 統計翻訳に基づく言語横断質問応答

日英の言語横断質問応答システムにおいて、前処理として翻訳を用いる一般的な手法と、統計翻訳モデルを組み込んだ手法の構成を図1に示す。両者の違いは、前処理として翻訳モデルを用い、単一言語に統一してから質問応答を行うか、質問応答のプロセスに翻訳モデルを組み込むことで、前処理の翻訳を行わずに、言語の異なるまま質問応答を行うか、である。両者の処理の流れは以下の通りである。

日本語質問文が入力されると、システムはまず質問文解析により質問タイプを得る。同時に、質問文からストップワードを取り除いた後、検索キーとして文書検索を行い、検索スコアの高い英語文書を得る。次に、検索された英語文書から回答候補を抽出する。その後、

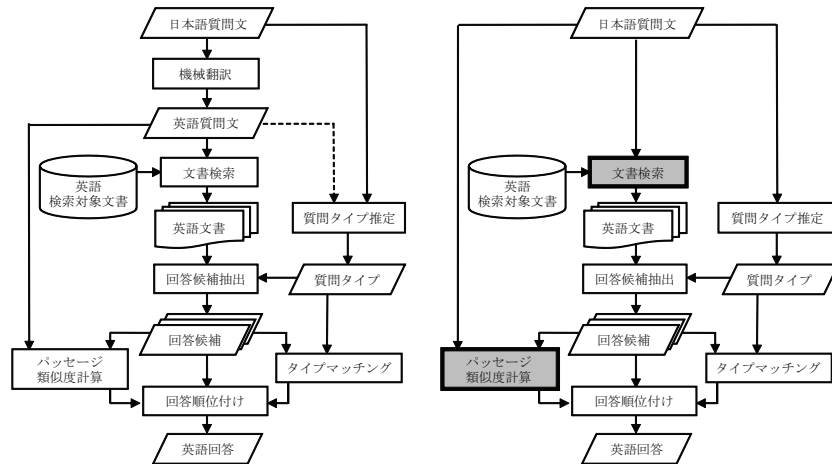


図 1 一般的な手法(左)と先行研究手法(右)のシステム構成
Fig. 1 typically system(left) and previous system(right)

回答候補のスコアリングを行うために、日本語質問文と英語文書における回答周辺の文脈(パッセージ)の類似度の計算と、回答候補の質問タイプとのマッチングを並列に行う。スコアリングの後、スコアに基づき回答候補の順位付けを行い、結果を出力する。

入力質問を翻訳せずに原言語をそのまま用いることによる変更点は、図 1 中の太枠か所の「文書検索」および「パッセージ類似度計算」である。文書検索では、索引付けに翻訳モデルの単語翻訳確率 $t(j|e)$ を用いることで日本語質問文から直接英語文書の検索を行う。パッセージの類似度計算では、質問文と回答周辺の文脈(パッセージ)の類似度を「パッセージから質問文へ翻訳される確率」と見なして計算する。

2.1 文書検索

日本語の質問文から直接英語文書を検索するために、英語の検索対象文書を日本語で索引付けする。その際、単語翻訳確率を用いて英語の索引語頻度から日本語の索引語頻度を求める。

文書 D にて英単語を e 、翻訳先の日本語単語を j とすると、 D を日本語で索引付けした場合の日本語単語 e の出現頻度 $tf(e, D)$ は式 (1) で推定できる。

$$tf(j, D) = \sum_{e \in D} tf(e, D)t(j|e) \quad (1)$$

$tf(e, D)$ は D における e の単語頻度、 $t(j|e)$ は e から j への単語翻訳確率である。単語翻訳確率は統計的機械翻訳の翻訳モデルと同様に対訳コーパスから求める。 $t(j, D)$ を用いれば、英語で索引付けした場合と単語の出現頻度(TF)の点で整合性が保持されるので、ベクトル空間モデルに基づく既存の検索エンジンを用いて日本語質問から英語文書への検索が可能になる。実装には文書検索エンジン GETA を用いた。GETA により抽出された文書のトップ k を質問文と類似した文書として、パッセージ類似度計算の対象とした。

2.2 パッセージ類似度計算

日本語質問文と英語文書における回答周辺の文脈(パッセージ)の類似度を、パッセージが質問文に翻訳される確率として計算する。日本語質問文 J と、検索対象文書中の回答候補 A が含まれるパッセージ E の類似度を、次式で求める。

$$sim(J, E|A) = P(J|E - A)$$

ここで、 $P(J|E - A)$ は単語列 $E - A$ (パッセージ E から回答候補 A を除いた単語列) が J に翻訳される確率である。質問応答の性質から、質問文中には回答に相当する語が現れないので、 E ではなく $E - A$ を用いた。 $P(J|E - A)$ を計算するモデルとして、先行研究では IBM Model 1[3] を用いている。IBM Model 1 は、すべての単語対応が等確率であると仮定し、単語翻訳確率 $P(j|e)$ のみを用いて $P(J|E)$ を計算するモデルであり、式 (2) として定義される。

$$P(J|E) = \prod_{i=1}^m \frac{1}{n+1} \sum_{k=0}^n P(j_i|e_k) \quad (2)$$

式 (2) に基づいて、言語横断質問応答に適用すると次式となる。

$$P(J|E - A) = \prod_{i=1}^m \frac{1}{n - (q - p + 1)} \sum_{k=0,1,2,\dots,q-1,p+1,\dots,n} \{\lambda t(j_i|e_k) + (\lambda - 1)u(j_i)\} \quad (3)$$

$$u(j) = \frac{\sum_k tf(j, D_k)}{\sum_k \sum_i tf(j, D_k)}$$

ここで、 $J = j_1 \dots j_m$ は日本語質問文の単語列、 $E = e_1 \dots e_n$ は英語パッセージ、 $A = e_p \dots e_q$ は E 中の回答候補単語列である。 $E - A = e_1, \dots, e_{p-1}, e_{q+1}, \dots, e_n$ はパッセージから回答候補を除去した単語列である。 $u(j)$ はユニグラム確率で、2.1 節の文書検索の際に付けられた $tf(j, D)$ をもとにした日本語単語の出現数比率(出現数/総単語数)である。ユニグラム

ム確率によるスムージングは、単語翻訳確率の零確率問題に対処するために導入する。

3. 翻訳モデルの改善手法

3.1 クエリ尤度モデルの導入

2.2節で示したように、パッセージ類似度を計算する式(3)はIBM Model 1に基づき、文から文への翻訳確率を求めることで類似度を計算している。一方、言語モデルに基づく情報検索手法としてクエリ尤度モデル(query likelihood model)[14]~[16]が提案されている。そこで、クエリ尤度モデルの観点から式(2)を再考してみる。クエリ尤度モデルでは、文書 D がクエリ Q に適合する確率 $P(D|Q)$ によって文書検索を行う。ベイズの定理を用いて次式を得る。

$$P(D|Q) = \frac{P(Q|D)P(D)}{P(Q)} \propto P(Q|D)P(D) \quad (4)$$

ここで、文書に関する事前分布 $P(D)$ が一様であると仮定すると次式が得られる。

$$P(D|Q) \propto P(Q|D)$$

クエリ $Q = q_1, \dots, q_n$ の各単語が独立に同一の多項分布に従って生成されるとすると、次式のように推定できる。

$$\begin{aligned} P(Q|D) &= \prod_{i=1}^n P(q_i|D) \\ &\approx \prod_{i=1}^n \{\lambda P_{ML}(q_i|D) + (1-\lambda)P_{ML}(q_i|C)\} \\ P_{ML}(q_i|D) &= \frac{n(q_i, D)}{|D|}, \quad P_{ML}(q_i|C) = \frac{\sum_{D \in C} n(q_i, D)}{\sum_{D \in C} |D|} \end{aligned} \quad (5)$$

$n(q_i, D)$ は文書 D における語 q_i の出現頻度、 $|D|$ は文書 D の総単語数、 C は文書コレクションを示す。推定の際に、文書コレクションを用いた補完に基づくスムージングを行うことで、クエリ尤度の計算において頻繁に出現する一般語の影響力を低下させることができ、IDFと同様の効果を与えることができる。クエリ尤度モデルの定義に基づいて、言語横断質問応答に適用すると次式が得られる。

$$P(J|E) = \prod_{i=1}^m P(j_i|E) \quad (6)$$

$$\begin{aligned} &= \prod_{i=1}^m \sum_{k=1}^n P(j_i, e_k|E) \\ &\approx \prod_{i=1}^m \sum_{e \in E} P(j_i|e)P(e|E) \end{aligned} \quad (7)$$

ここで、 $P(e|E) = \frac{n(e, E)}{|E|}$ と推定すると次式となる。

$$\begin{aligned} P(J|E) &= \prod_{i=1}^m \sum_{e \in E} P(j_i|e) \frac{n(e, E)}{|E|} \\ &= \prod_{i=1}^m \frac{1}{|E|} \sum_{k=1}^n P(j_i|e_k) \end{aligned} \quad (8)$$

最終的に、IBMモデルの *null* 単語に相当する $k=0$ を除いて、式(2)と同形の式(8)が導出される。

一方、式(7)の $P(e|E)$ を式(5)で推定すると次式となる。

$$P(J|E) = \prod_{i=1}^m \sum_{e \in E} P(j_i|e) (\lambda P_{ML}(e|E) + (1-\lambda)P_{ML}(e|C)) \quad (9)$$

また、クエリ尤度モデルの定義と式(3)のIBM Model 1と同様に、翻訳確率をユニグラム確率でスムージングすることで、最終的に次式が得られる。

$$P(J|E) = \prod_{i=1}^m \sum_{e \in E} \{\lambda_1 t(j_i|e) + (1-\lambda_1)u(j_i)\} \{\lambda_2 P_{ML}(e|E) + (1-\lambda_2)P_{ML}(e|C)\} \quad (10)$$

IBM Model 1に基づいた式(3)と比較して、クエリ尤度モデルに基づいた式(10)は英語パッセージにおける重要語を考慮することができる。

3.2 統計的フレーズ翻訳モデルの導入

以上の手法は、どれも単語対単語の翻訳確率による類似度を計算している。しかし、統計的機械翻訳ではフレーズベースに基づいた手法が高い性能となることが知られている。そこで、パッセージ類似度の計算に統計的フレーズ翻訳モデルを導入する。

式(7)に基づいて、フレーズ翻訳モデルによる拡張を行う。フレーズ翻訳モデルを導入す

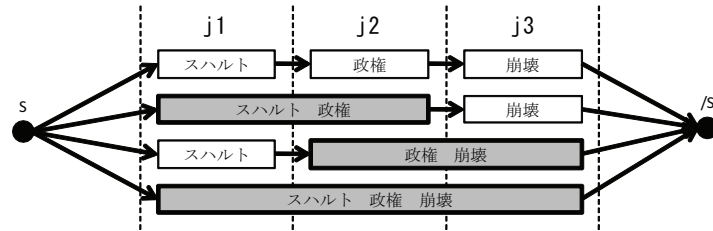


図2 統計的フレーズ翻訳モデルにおけるフレーズ境界

Fig. 2 Phrase border in statistical phrase-based translation models

ることで、フレーズ境界について考えなければならない。図2に、統計的フレーズ翻訳モデルによるフレーズ境界の例を示す。

単語翻訳モデルを用いた場合、翻訳が単語単位のため、境界のパターンは一種類しかなかった(図2中の一番上のパス)。フレーズ翻訳モデルを導入すると、 $2^{|\bar{j}|-1}$ 種類の境界パターンが考えられる。図2の例では、質問文の単語数が3であるため、4パターンのフレーズ境界が考えられる。質問文とより類似したパッセージは、質問文中の日本語フレーズに翻訳される英語フレーズをより多く含んでいる可能性が高い。そこで、多くのフレーズを含むパッセージが重要であると仮定し、各パターンにおける類似度の総和により、質問文とパッセージの類似度を計算する。

$$P(J|E) = \sum_{\bar{j}_1 \dots \bar{j}_m \in S(J)} \prod_{i=1}^m \sum_{\bar{e} \in E} P(\bar{j}_i|\bar{e})P(\bar{e}|E) \quad (11)$$

$S(J)$ はフレーズ境界パターンの集合、 $\bar{j}_1, \dots, \bar{j}_m$ は日本語フレーズ境界パターンの1つ、 \bar{e} は英語フレーズを示す。ここで、 $S(J)$ には最も細かいフレーズ境界パターンとして、単語列から成る境界パターンが必ず含まれる(図2中の一番上のパス)。フレーズ翻訳モデル中に存在しないフレーズは、計算に用いることができないため、 $P(\bar{j}_i|\bar{e})$ の推定には次の2つの場合に分けて計算する。

\bar{j} が1単語のフレーズの場合

$P(\bar{j}_i|\bar{e})$ は式(10)のように単語翻訳モデルから推定する。

\bar{j} が2単語以上のフレーズの場合

$P(\bar{j}_i|\bar{e})$ には、フレーズベース統計翻訳用に学習したフレーズ翻訳確率を用いる。このとき、スムージングは行わない。すなわち、パッセージ E 中のどのフレーズ \bar{e} につい

ても (\bar{j}, \bar{e}) がフレーズテーブルに現れない場合は、 $P(\bar{j}|\bar{e}) = 0$ となり、その境界パターン全体の確率も0となるため、そのようなパターンの計算はスキップされる。

また、 $P(\bar{e}|E)$ の推定には式(9)と同様にスムージングを行う。まとめると、次式のように推定する。

$$P(\bar{j}|\bar{e}) = \begin{cases} \lambda_1 t(\bar{j}|e) + (1 - \lambda_1) u(\bar{j}) & |\bar{j}| = 1 \text{ の場合} \\ T(\bar{j}|\bar{e}) & |\bar{j}| \geq 2 \text{ の場合} \end{cases}$$

$$P(\bar{e}|E) = \begin{cases} \lambda_2 P_{ML}(e|E) + (1 - \lambda_2) P_{ML}(e|C) & |\bar{j}| = 1 \text{ の場合} \\ \lambda_2 P_{ML}(\bar{e}|E) + (1 - \lambda_2) \left\{ P_{ML}(e_1|C) \prod_{i=2}^{|\bar{e}|} P_{ML}(e_i|e_{i-1}, C) \right\} & |\bar{j}| \geq 2 \text{ の場合} \end{cases}$$

$t(*)$ は単語翻訳モデル、 $T(*)$ はフレーズ翻訳モデルを示す。また、2単語以上の英語フレーズ $\bar{e} = e_1, \dots, e_{|\bar{e}|}$ におけるコレクションモデル $P(\bar{e}|C)$ は、検索対象文書の言語モデルから2-gram確率により求める。これを式(11)に代入し、単語列から成る境界パターンを分けて書くと次式のように書き直すことができる。

$$P(J|E) = \prod_{i=1}^m \sum_{e \in E} \{ \lambda_1 t(\bar{j}_i|e) + (1 - \lambda_1) u(\bar{j}_i) \} \{ \lambda_2 P_{ML}(e|E) + (1 - \lambda_2) P_{ML}(e|C) \} + \sum_{\bar{j}_1 \dots \bar{j}_n \in S'(J)} \prod_{i=1}^m \sum_{\bar{e} \in E} P(\bar{j}_i|\bar{e}) P(\bar{e}|E) \quad (12)$$

$S'(J)$ は $S(J)$ から単語列から成る境界パターンを除いたパターン集合を示す。1単語のフレーズは単語翻訳モデルを用いて計算するため、最も細かい分割となるパスが必ず存在することを保証することができ、このパターンの場合における類似度は従来の単語翻訳モデルによる類似度計算と等しくなる。したがって、提案する統計的フレーズ翻訳モデルを用いた類似度計算手法は、単語翻訳モデルの類似度を基本として、フレーズ翻訳モデルによる類似度を加えたものとして考えることができる。

フレーズの分割パターンは $2^{|\bar{j}|-1}$ 通りあるため、それぞれのパターンを独立に計算すると質問文中の単語数により計算時間が指数的に増加する。しかし、左から右へと各フレーズ境界までの総和を漸進的に計算する動的計画法を用いることにより、効率的な計算が可能である。

3.3 事前分布の導入

これまで式(4)に示すように、事前分布 $P(D)$ は一様であるとみなしていた。ここで

は、事前分布を利用することを考える。事前分布を用いることで、クエリに依存しない優先度を、クエリ尤度と自然な形で結合することができる。

$$\text{sim}(J, E|A) = P(J|E)P(E)$$

$$P(E) \propto |E|$$

上式に示すように、事前分布 $P(E)$ としてパッセージ長を用いている。これは、長いパッセージほど回答候補が出現する可能性が高いことを表している。

4. 評価実験

提案したシステムの評価実験を行う。言語横断質問応答は、質問の回答となる部分のみ抽出するタスクであるが、その性能はパッセージ類似度の計算に大きく依存している。そのため、本稿では回答抽出の前段階のパッセージ検索性能による評価を行った。ここで、パッセージは一文と設定した。

4.1 テストコレクション

言語横断質問応答の評価に NTCIR CLQA1 テストコレクション [2] の英日サブタスクを用いた。検索対象文書は英語版の読売新聞である Daily Yomiuri 2年分 (2000-2001) を用いた。テストコレクションは、factoid 型質問 200 問で構成されている。

4.2 翻訳モデルの学習コーパス

翻訳モデルの学習には、以下の文対応の対訳コーパス [10] を用いて学習を行った。

日英新聞記事対応付けデータ 114404 ペア

ロイター日英記事の対応付け 56782 ペア

上記のコーパスのうち、日英新聞記事対応付けデータは読売新聞と Daily Yomiuri の文対応で構成される対訳コーパス [10] であるので、CLQA1 で検索対象となる 2000,2001 年の対訳は取り除かれている。対訳コーパスは日本語、英語それぞれに対して、前処理を行い正規化した。日本語文に対しては日本語の形態素解析器 MeCab を用いて形態素ごとに区切り、活用語の標準化を行った。また英語に対しては品詞タガーを用いて英単語を原型に直し、全ての語を小文字化した。翻訳モデルの学習は GIZA++[11] と Moses[12] を用いて、単語翻訳モデル及びフレーズ翻訳モデルをそれぞれ学習した。

4.3 質問解析と文書検索の性能

質問解析と文書検索の性能をそれぞれ表 1 と表 2 に示す。

疑問詞パターンによる分類により、CLQA1 および CLQA2 それぞれのテストセットにおいて 80% を超える性能となった。

表 1 質問タイプ判定の性能

Table 1 Performance of a question type determination

	CLQA1	CLQA2
LOCATION	0.833	0.903
TIME	1.000	1.000
PERSON	0.889	0.886
ORGANIZATION	0.231	0.400
DATA	0.960	0.935
NUM	0.867	0.800
ARTIFACT	0.778	0.550
PERCENT	0.800	0.867
MONEY	1.000	1.000
ALL	0.805	0.820

表 2 文書検索の性能

Table 2 Performance of a document retrieval

文書数 k	ACC@1	MRR	recall
5	0.45	0.537	0.670
10		0.540	0.695
50		0.548	0.820

文書検索には検索エンジン GETA を用いている。この文書検索結果の範囲においてパッセージ類似度計算を行うため、文書検索における *recall* のスコアが、パッセージ類似度計算における最大スコアとなる。文書数 k に比例して *recall* が増加していき、文書検索の範囲を $k = 50$ とすることで、質問文に対する *recall* が 80% を上回る性能を示した。効率のため、以降の実験では $k = 5$ に設定した。

4.4 ベースライン手法

ベースライン手法として、統計的機械翻訳 (SMT) を外的に用いる一般的な手法を実装した。前処理として日本語質問文を SMT により英語に翻訳し、既存の検索エンジンを用いて類似パッセージを抽出する。SMT の出力において N-best を用いて検索実験を行った。SMT に用いる翻訳モデルは、4.2 節で示したものをを用いた。また、SMT の性能による要因を排除するために、テストコレクションに含まれる質問の英語訳を入力として実験を行った (ideal と記す)。これは、理想的な翻訳を行った場合に相当する。

4.5 評価尺度

性能評価は NTCIR5 CLQA1 評価セットに基づいて評価を行った。評価尺度として回答候補上位 1 位に対する精度、回答候補上位 5 位に対する精度、回答候補上位 5 位に対する MRR を用いた。MRR の計算式を式 (13) に示す。

$$MRR = \frac{1}{N} \sum_{Q=1}^N \max_{r \in \{1, \dots, 5\}} \left(\frac{A_{Q,r}}{r} \right) \quad (13)$$

$A_{Q,r}$ は Q 番目の質問の第 r 位の回答を意味し、正解なら 1、不正解なら 0 となる。 N は質問数 (今回は $N = 200$ 問) である。 順位の逆数で集計するため、多くの質問で高い順位が出るほど高得点となり、最終的に、平均第何位で回答が得られるかを表す指標となる。

CLQA1 テストコレクションとして配布された正解セットにより、文字列の一致のみで正解判定を行った。

4.6 実験結果

システムの性能を見るために検索実験を行った。 実験結果を表 3 に示す。

4.6.1 システムの評価

表 3 中のベースライン手法と翻訳モデルを組み込んだ手法を比較すると、翻訳モデルを組み込んだ手法が高い性能を示している。 この二つの手法において、同じ対訳コーパスで学習した翻訳モデルを用いていることから、前処理として翻訳を行い訳文を一意に決定してしまうより、翻訳モデルを組み込み複数候補を扱うことが効果的であることが考えられる。 また、翻訳モデルを組み込んだ手法は理想的な翻訳を用いた場合よりも高い性能を示した。 これは、類似度計算の際に直訳だけでなく同意語や関連語も利用できたことが理由と考えられる。

次に、IBM Model 1 とクエリ尤度モデルを比較する。 同じ単語翻訳モデルを用いているため、性能の差は文書コレクションによる補間の効果と見なすことができる。 クエリ尤度モデルの性能が高いことから、英語パッセージにおける重要語の効果が見られる。 また、クエリ尤度モデルにおいてフレーズ拡張を行うことで、最も良い性能が見られた。

表 3 中の $+QT+DR$ は、質問タイプ (QT) と文書検索時に得られた文書関連度 (DR) を類似度計算に反映させた場合の検索性能を示している。 全体的に大幅な性能改善が見られた。

4.6.2 事前分布の実装結果

3.3 節に示すように、事前分布 $+PD$ を用いた場合の性能評価を行った。

実験結果は、事前分布を用いることで性能の改善を示した。 事前分布を一様にするだけで、単語あたりの類似度に正規化されることから、質問文に対して類似しているが冗長であるパッセージに高い類似度を割り当てることができない。 パッセージ長に関する事前分布を用いることで、パッセージ中の回答出現確率を考慮することができ、比較的長いパッセージにも高い類似度を与えることができるようになったことが主な改善の理由であると考えられる。

性能の改善が見られた一方、単語翻訳モデルとフレーズ翻訳モデルを比べて性能の差が小さくなってしまった。 フレーズ翻訳モデルは、多くのフレーズを含むパッセージに高い類似度を割り当てるように設定した。 長いパッセージには多くのフレーズが含まれるため高い類

表 3 システムの検索性能
Table 3 System performance

	MRR	ACC@1	ACC@5
ベースライン (1-best)	0.2263	0.160	0.340
ベースライン (50-best)	0.2507	0.195	0.350
ベースライン (ideal)	0.2931	0.260	0.335
IBM Model 1	0.3658	0.310	0.460
クエリ尤度 (単語)	0.3712	0.310	0.470
クエリ尤度 (フレーズ)	0.3749	0.315	0.475
IBM Model 1 + QT + DR	0.4136	0.330	0.540
クエリ尤度 (単語) + QT + DR	0.4194	0.330	0.545
クエリ尤度 (フレーズ) + QT + DR	0.4245	0.340	0.545
クエリ尤度 (単語) + QT + DR + PD	0.4245	0.335	0.555
クエリ尤度 (フレーズ) + QT + DR + PD	0.4266	0.340	0.555

似度を与えることができることから、すでにモデル中に事前分布の効果が含まれていることが原因であると考えられる。

5. おわりに

本稿では、言語横断質問応答において、検索性能を改善させるために統計的フレーズ翻訳モデルを用いた新しいパッセージ類似度計算手法を提案した。 比較実験の結果、フレーズ翻訳モデルを用いる提案法が、単語翻訳モデルを用いる従来法の性能を改善することを確認できた。

本稿の実装では、質問文について質問解析のプロセスでストップワードをすべて除去した。 しかし、"政権が崩壊"のように付属語を含むフレーズにも有効なものが存在すると考えられるので、質問文側についても付属語を含むフレーズの利用を検討したい。 また、提案手法へのクエリ拡張の導入も検討したい。

参考文献

- 1) B.Magnini et al. The Multiple Language Question Answering Track at CLEF 2003. In Working Notes for the CLEF 2003 Workshop, 2003.
- 2) Y.Sasaki et al. Overview of the NTCIR-5 Cross-Lingual Question Answering Task (CLQA1). In Proceedings of the Fifth NTCIR Workshop, pp.175-185, 2005.
- 3) T.Mori et al. A Method of Cross Language Question Answering Based on Machine Translation and Transliteration. - Yokohama National University at NTCIR-5

- CLQA1 - In Proceedings of the Fifth NTCIR Workshop, pp.215-222, 2005.
- 4) H.Isozaki, et al. NTT's Japanese-English Cross-Language Question Answering System. In Proceedings of the Fifth NTCIR Workshop, pp.186-193, 2005
 - 5) P.F.Brown et al. The mathematics of statistical machine translation: Parameter estimation. In Computational Linguistics 19(2), pp.263-311, 1993.
 - 6) A.Berger et al. Information Retrieval as Statistical Translation. In Proceedings of the 22nd Annual Conference on Research and Development in Information Retrieval (ACM SIGIR), pp.222-229, 1999.
 - 7) J.Xu et al. Evaluating a Probabilistic Model for Cross-lingual Information Retrieval. In Proceedings of the 24th Annual International ACM SIGIR Conference, pp.105-110, 2001.
 - 8) V.Murdock et al. Simple translation models for sentence retrieval in factoid question answering. In it-shape Proceedings of the Information Retrieval for Question Answering Workshop SIGIR, pp.31-35, 2004.
 - 9) 秋葉 友良 他. 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. 情報処理学会研究報告, 2004-NL-163, pp.131-138, Sep. 2004.
 - 10) M.Utiyama et al. Reliable Measures for Aligning Japanese-English News Articles and Sentences. In ACL-2003, pp.72-79, 2003.
 - 11) F.J.Och. GIZA++: Training of statistical translation model. <http://www-i6.informatik.rwth-aachen.de/Colleagues/och/software/GIZA++.html>
 - 12) Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. and Herbst, E.: Moses: Open Source Toolkit for Statistical Machine Translation, Proc. 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions, Association for Computational Linguistics, pp.177-180 2007.
 - 13) A.Fujii and T.Ishikawa. Cross-Language Information Retrieval at ULIS. In Proceedings of The First NTCIR Workshop, pp.163-169, 1999.
 - 14) D. Hiemstra, A linguistically motivated probabilistic model of information retrieval, Research and Advanced Technology for Digital Libraries, vol.1513, pp.569-584, Lect. Notes Comput. Sci., Springer-Verlag, 1998.
 - 15) W.B. Croft and J. Lafferty eds, Language Modeling for Information Retrieval, Kluwer Academic Publishers, 2003.
 - 16) 江口 浩二. 情報検索のための確率的言語モデルに関する動向と課題, 電子情報通信学会論文誌 D Vol.J93-D No.3, pp.157-169, 2010. Workshop, pp.163-169, 1999.