# 部分N-gram頻度情報を利用した質問応答定型表現への言語モデル適応

秋葉友良[†]　伊藤克亘[†*]　藤井敦[‡*]　石川徹也[‡]

[†]産業技術総合研究所　[‡]図書館情報大学　[*]科学技術振興事業団 CREST

**概要**　音声入力に対応した質問応答システムのための言語モデルを作成する手法について述べる。検索対象となる新聞記事から作成したN-gramをベースに、人手で与えた質問文定型表現を用いて適応化する2つの手法を提案する。一つは、ベースとなる新聞記事N-gramモデル中の質問文定型表現に対応するN-gram頻度を、部分的なN-gram頻度で強調する手法である。もう一つは、定型表現を記述文法で表し、新聞記事N-gramと統合する手法である。認識実験を行い、N-gram頻度を重み付きで混合する従来法とくらべ、どちらの手法も単語誤り率を減少させることが示された。特に、前者の手法が認識率と頑健性の面でより良い結果を示した。

# Using Extra N-gram Counts for Statistical Language Model Adaptation in Speech-Driven Question Answering

Tomoyosi AKIBA[†]　Katunobu ITOU[†*]　Atsushi FUJII[‡*]　Tetsuya ISHIKAWA[‡]

[†] National Institute of Advanced Industrial Science and Technology (AIST)

[‡] University of Library and Information Science

[*]CREST, Japan Science and Technology Corporation

**Abstract**　Aiming at speech-driven question answering, we propose two methods to produce statistical language models for recognizing spoken questions with a high accuracy. Both methods use a target collection (i.e., a document set from which answers are derived) to extract N-grams, and adapt them to the question answering task by way of frozen patterns typically used in interrogative questions. The first method magnifies N-gram counts corresponding to the frozen patterns in the original N-gram. The second method combines N-grams extracted from the collection and grammars associated with frozen patterns, to produce a single N-gram model. Our experiments showed that the two proposed methods outperformed a conventional language model adaptation method in terms of the recognition accuracy, and that the first method was more accurate and robust than the second method.

## 1　はじめに

質問応答 (QA) は、1999年のTREC-8[8]にタスクとして採択されて以来、次世代の情報検索技術を目指した評価タスクとして注目されている。従来の情報検索タスクも音声入力に対応するように拡張されてきたが[4, 12]、質問応答では入力が質問文という話し言葉に近い表現が使用されることから、より音声入力に適したタスクであると考えられる。我々はこのような、音声入力を前提とした質問応答システムを開発中である。本稿では、音声認識部で利用する、質問文に対応した言語モデルを構築する手法について述べる。

質問応答システムへの入力は、検索のトピックに関する表現、質問文の文末に現れる定型的な表現、の2つの部分から構成される言語表現となる。例えば、次のような質問文が想定される。

「1976年に火星に軟着陸した探査機は何という名前でしたか」

前半の「1976年に火星に軟着陸した探査機は」の部分は、検索のトピックに関する部分で、新聞記事・辞典など質問応答の検索対象となる文書からそのまま学習した言語モデルで対応できる。本稿では、新聞記事を対象とした質問応答[1]を想定し、以降新聞記事モデルと呼ぶことにする。一方、後半の「何という名前でしたか」の部分は質問文に典型的に現れるパターンであり、新聞記事では稀な表現となるので、新聞記事モデルだけで扱うのでは不十分である。

言語モデルのタスク適応については、一般的な大量の学習データと、対象タスク用の比較的少量の学習データから、対象タスク用の言語モデルを作成する方法が一般的である。この枠組みに従えば、QA質問

文用の言語モデル作成には、新聞記事モデルをベースに、定型表現で構成された学習データを用いてタスク適応する方法が考えられる。しかし、本稿で対象とするQA質問文は、以下にあげる特徴を持つと仮定することで、単に漠然と2つのモデルを混合するよりは、より適切な手法が適用できると考えられる。(1) 前半部分だけなら、適応なしの新聞記事モデルで扱える。(2) 後半部分の語彙は新聞記事モデルでカバーできる程度の一般性を持つ。すなわち「語彙」ではなく「言い回し」を獲得したい。(3) 後半部分は「定型的」で、人手で列挙可能な程度の多様度を持つ。(4) 一文は性質の異なる前半と後半の2つの部分から構成される。よって、両者の接続部分のモデル化が重要となる。特に、2つの部分を独立したテキストから学習し、混合したのでは接続部分をうまく学習できないと考えられる。

本稿では、このような特徴を持つQA質問文を扱う2種類の言語モデル適応手法を提案する。一つは、定型表現の部分に対応する部分的なN-gram頻度情報を割増しする手法である (3節)。もう一つは、人手で記述した文法で表される制約をN-gram言語モデルに導入する手法である (4節)。

## 2 N-gram頻度の混合によるタスク適応

言語モデルのタスク適応については、一般的な大量の学習テキストデータと、対象タスク用の比較的少量の学習データから、対象タスク用の言語モデルを獲得する方法が知られている。多くの適応化手法では少量の学習データが利用可能であることを仮定している。しかし、少量のテキストデータでも獲得のコストは決して小さくない。代替手法として、対象タスクの文法を記述してテキストデータを自動生成する方法 [5]、対象タスクの典型的な例文を用いる方法 [11] が提案されている。提案法では、1節で述べた適応タスクの特徴 (3)「人手で列挙可能な程度の多様度を持つ」を考慮して、定型表現部の言語表現を人手で (例文か文法で) 与えることにする。

適応化手法としては、学習データのN-gram頻度を重み付きで足し合わせる方法が知られている [3, 9]。この手法の概念図を図1に示す。ここで混合するN-gram頻度は、共にテキストデータから直接獲得した頻度である。そのため、それぞれのN-gram頻度、および混合後のN-gram頻度は、次のような性質を持つ。

**長さに関する整合性** 任意の長さのN-gram頻度は、より長いN-gram頻度から一意に計算できる。

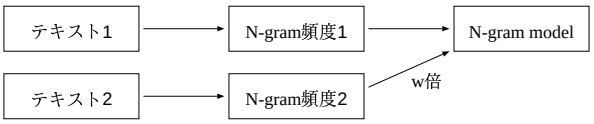


図 1: N-gram頻度の混合によるタスク適応

(例) tri-gram頻度 $C_{(3)}$ が与えられているとする。この時、bi-gram $w_p w_q$ の頻度は、以下のいずれかの方法で求められる。

$$C_{(2)}(w_p w_q) = \sum_{w_i} C_{(3)}(w_i w_p w_q) = \sum_{w_i} C_{(3)}(w_p w_q w_i)$$

同様に、1-gram頻度 $C_{(1)}$ は次のように求められる。

$$C_{(1)}(w_p) = \sum_{w_i} C_{(2)}(w_i w_p) = \sum_{w_i} C_{(2)}(w_p w_i)$$

この性質は、単語の連続としてのテキストの性質が反映された結果である。

このN-gram頻度の混合による適応化手法を、本稿のQA質問文タスクに適用することを考える。前半を新聞記事から、後半を定型表現を記した例文から学習できる。しかし、タスクの特徴 (4)「一文は性質の異なる前半と後半の2つの部分から構成される」を考慮すると、新聞記事と定型表現のどちらの学習データにも接続部分に対するテキストを含んでいない。そのため、接続部分をうまく学習できないと考えられる。

## 3 部分N-gram頻度情報を用いたタスク適応

タスクの特徴 (2)「後半部分の語彙は新聞記事モデルでカバーできる程度の一般性を持つ」を考慮すると、定型表現に対応する語彙はすでに新聞記事モデルに含まれていると仮定できる。また、定型表現に対応する単語の連接情報 (N-gram頻度) もある程度含まれていると考えられる。そこで、新聞記事モデルに含まれる定型表現に関するN-gram頻度をそのまま活用してタスク適応することを考える。

提案する手法の概念図を図2に示す。新聞記事から作成したベースとなるN-gram頻度情報に対し、**部分的なN-gram頻度情報**を与える。ここで「部分的」とは、テキストデータから作成した整合的なN-gram頻度**ではない**ことを意味する。何らかの知識源によって、任意のN-gram頻度を直接与える。その結果、N-gram

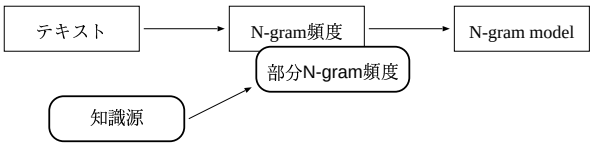


図 2: 部分 N-gram 頻度を用いたタスク適応

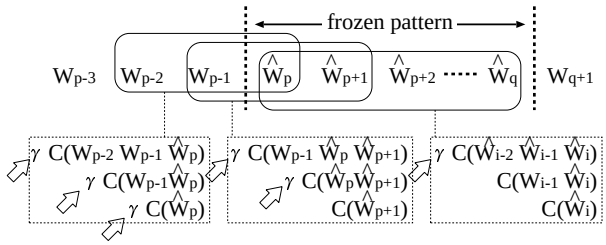


図 3: 部分 N-gram 頻度による定型表現の強調

頻度は、2 節で述べた「長さに関する整合性」が成り立たない。

## 3.1 部分 N-gram 頻度情報を用いた確率計算

整合性のない N-gram 頻度を扱う場合、次に述べるような注意が必要となる。まず、より短い N-gram 頻度は、より長い N-gram 頻度から計算することができない。したがって、N-gram 頻度情報は、長さ $n$ 毎に保持する必要がある。以下では、長さ $n$ 毎に N-gram 頻度を $\mathbf{C}_n$ と記す。

スムージング・モデルの計算式にも注意が必要である。バックオフ・スムージングの一般式は次のように表される。

$$P(w_i|w_{i-n+1}^{i-1}) = \begin{cases} d_{w_{i-n+1}^i} P_{ML}(w_i|w_{i-n+1}^{i-1}) \\ \qquad \cdots C(w_{i-n+1}^i) > 0 \\ \alpha(w_{i-n+1}^{i-1}) P(w_i|w_{i-n+2}^{i-1}) \\ \qquad \cdots C(w_{i-n+1}^i) = 0 \end{cases} \tag{1}$$

ここで $d$, $P_{ML}$, $\alpha$ は、それぞれ、ディスカウント係数、最尤推定による N-gram 確率、確率の総和を 1 とするための正規化係数である。このうち、$\alpha$ は他の値から自動的に求まるので、$d$ と $P_{ML}$ について計算式を再検討する。

最尤推定による N-gram 確率は、通常次の式で求めることができる。

$$P_{ML}(w_{i-n+1}^i) = \frac{C(w_{i-n+1}^i)}{C(w_{i-n+1}^{i-1})}$$

しかし、部分的な N-gram 頻度を用いる場合、$\mathbf{C}_n$ と $\mathbf{C}_{n-1}$ の不整合から、両者を同時に用いることはできず、$\mathbf{C}_n$ だけから次のように計算する必要がある。

$$P_{ML}(w_{i-n+1}^i) = \frac{\mathbf{C}_n(w_{i-n+1}^i)}{\sum_{w_i} \mathbf{C}_n(w_{i-n+1}^i)}$$

同様に、ディスカウント係数 $d$ の計算にも注意が必要である。例えば、Witten-Bell 法 [7] では、次の式を用いる。

$$d_{\mathrm{WB},w_{i-n+1}^i} = \frac{C(w_{i-n+1}^{i-1})}{C(w_{i-n+1}^{i-1}) + r(w_{i-n+1}^{i-1})}$$

ここで $r(w_{i-n+1}^{i-1})$ は、文脈 $w_{i-n+1}^{i-1}$ の次に出現する異なり単語数である。部分的な N-gram 頻度を用いる場合は、長さ毎に $r$ の情報源を特定する必要がある。

$$d_{\mathrm{WB},w_{i-n+1}^i} = \frac{\sum_{w_i} \mathbf{C}_n(w_{i-n+1}^i)}{\{\sum_{w_i} \mathbf{C}_n(w_{i-n+1}^i)\} + \mathbf{r}_n(w_{i-n+1}^{i-1})}$$

ここで $\mathbf{r}_n(w_{i-n+1}^{i-1})$ は、N-gram 頻度情報 $\mathbf{C}_n(w_{i-n+1}^i)$ から求めた、文脈 $w_{i-n+1}^{i-1}$ の次に出現する異なり単語数である。

以上のように、長さ $n$ の確率値は、$\mathbf{C}_n$ だけを使って計算するように注意する。

## 3.2 部分 N-gram 頻度情報による定型表現の強調

前節に述べた確率値計算に注意すれば、任意の部分的な N-gram 頻度を導入できる。特に、頻度情報は長さ毎に与えることができる。このことを利用して、定型表現を含む文だけが相対的に確率値が高くなるように、部分的な N-gram 頻度を与える。

定型表現を表す単語列 $\hat{w}_p^q = \hat{w}_p \hat{w}_{p+1} \cdots \hat{w}_{q-1} \hat{w}_q$ と、その左文脈となる単語列 $w_{p-N+1}^{p-1} = w_{p-N+1} \cdots w_{p-1}$ を考える。この単語列 $\hat{w}_p^q$ を含む文集合が共通して持つ、以下の N-gram 頻度を $\gamma$ 倍して強調する (図 3)。

1. 定型表現内部の単語列に対して、最も長い N-gram 頻度だけを強調する。

$$\mathbf{C}_N(\hat{w}_{i-N+1}^i) = \gamma C(\hat{w}_{i-N+1}^i) \tag{2}$$

例えば tri-gram モデルを構築する場合、tri-gram 頻度だけを $\gamma$ 倍する。

$$\mathbf{C}_3(\hat{w}_{i-2}\hat{w}_{i-1}\hat{w}_i) = \gamma C(\hat{w}_{i-2}\hat{w}_{i-1}\hat{w}_i)$$

2. 定型表現の接頭単語列に対して、接頭単語列長以上の長さの N-gram 頻度を強調する。接頭単語帳を $k$ とすると、すべての文脈単語列 $w_{p-n+k}^{p-1}$ について、

$$\mathbf{C}_n(w_{p-n+k}^{p-1}\hat{w}_p^{p+k-1}) = \gamma C(w_{p-n+k}^{p-1}\hat{w}_p^{p+k-1}) \tag{3}$$

tri-gram モデルの場合、すべての文脈単語列 $w_{p-2}w_{p-1}$ について、次のように各長さの頻度を強調する。

$$\begin{aligned}
\mathbf{C}_3(w_{p-1}\hat{w}_p\hat{w}_{p+1}) &= \gamma C(w_{p-1}\hat{w}_p\hat{w}_{p+1}) \\
\mathbf{C}_2(\hat{w}_p\hat{w}_{p+1}) &= \gamma C(\hat{w}_p\hat{w}_{p+1}) \\
\mathbf{C}_3(w_{p-2}w_{p-1}\hat{w}_p) &= \gamma C(w_{p-2}w_{p-1}\hat{w}_p) \\
\mathbf{C}_2(w_{p-1}\hat{w}_p) &= \gamma C(w_{p-1}\hat{w}_p) \\
\mathbf{C}_1(\hat{w}_p) &= \gamma C(\hat{w}_p)
\end{aligned}$$

3. それ以外の N-gram 頻度情報は、元の N-gram 頻度と同じとする。$n = 1 \cdots N$ について、

$$\mathbf{C}_n(w_{i-n+1}^i) = C(w_{i-n+1}^i) \tag{4}$$

# 4 N-gram モデルへの文法的制約の導入

QA 質問文への言語モデル適応のもう一つの手法を示す。定型表現を記述文法で表し、新聞記事などから学習した N-gram モデル (以降、ベース N-gram と呼ぶ) と統合する手法である [2, 10]。

N-gram でモデル化される単語列は、全ての単語が互いに連接可能な単語ネットワークとして表現することができる。一方、記述文法 (正規言語[1]) で表される単語列は、文法によって部分的な単語連接だけを許した単語ネットワークで表現される。ベース N-gram の任意の位置から記述文法の先頭単語へ、記述文法の末尾単語からベース N-gram の任意の位置へ、両ネットワークを結合することによって、両モデルを統合し、一つの単語ネットワークで表すことができる (図 4)。

統合した単語ネットワークにおいて、連接しない (弧の存在しない) 単語列に対し確率値が 0 となるように、

[1]本手法で扱える記述文法は、(N-gram の表現力の制約の故に) 正規言語までである。自然言語の記述として広く用いられている文脈自由言語は、そのまま埋め込むことはできない。しかし、有限長の文は必ず正規言語で表現できること、文脈自由文法を正規文法に近似するアルゴリズムが知られている [6] こと、などから実用上ほとんど問題はない。

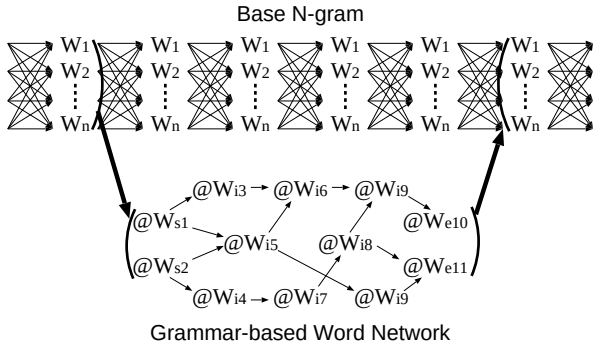


図 4: 単語ネットワークの統合

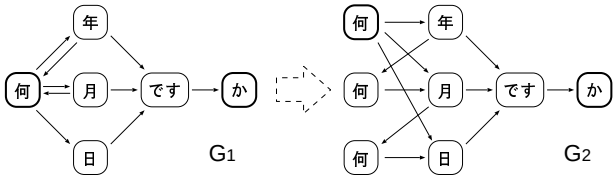


図 5: 単語ネットワーク

N-gram 確率を割り当てれば、汎用 N-gram と記述文法の両方の性質を同時に保持するモデルを獲得できる。

## 4.1 単語ネットワーク

正規言語を、単語を頂点とし可能な単語連接を有効弧で表した単語ネットワークで表現することを考える。このような単語ネットワークは、例文の集合から簡単に獲得可能である。例えば、年月日を尋ねる発話を表した以下の例文から文法を獲得することを考える。

何/年/です/か　　何/年/何/月/です/か

何/月/何/日/です/か

この 3 文から獲得できる連接可能な単語対は以下の通りである。

$A$ = { (何, 年) (何, 月) (何, 日) (年, 何) (月, 何) (年, です) (月, です) (日, です) (です, か) }

この単語対だけが連接可能であると考えると、文法 ($G_1$) は 4 つ組 $(W_a, W_s, W_f, A)$ で表現できる。ここで、$W_a$,$W_s$,$W_f$ は、それぞれ、全単語集合、先頭単語集合、末尾単語集合であり、

$W_a$ = { 何 年 月 日 です か },

$W_s$ = { 何 }, $W_f$ = { か }

となる。$G_1$ のグラフ表現を図 5 左に示す。

文法 $G_1$ は「何年何年ですか」「何月何年ですか」「何年何日ですか」のような、意図されない言語表現まで

受理してしまう。そこで、文法作成者の持つ言語知識を利用して、好ましくない表現を排除し、図 5 右のような文法 $G_2$ に修正することを考える。新たに導入したノード (文脈) 毎に、新たな単語記号を導入して、次のような文法 $(W'_a, W_s, W_f, A')$ として表現する。

$W'_a = W_a \cup \{$ 何 1 何 2$\}$

$A' = \{$(何, 年) (何, 月) (何, 日) (年, 何 1) (何 1, 月) (月, 何 2) (何 2, 日) (年, です) (月, です) (日, です) (です, か) $\}$

文法 $G_2$ は、「何年何月ですか」「何月何日ですか」のような、作成者の意図する表現だけを受理し、それ以外を排除する。このように、単語ネットワーク (正規言語) では、人の持つ言語知識を利用して、N-gram では獲得不可能な、単語間の長距離の依存関係も表現することが可能である。

ここで、ベース N-gram と統合するために、単語ネットワークで表された文法は以下の 2 つの条件を満たす必要がある。第 1 に、ベース N-gram と文法の語彙は区別されている必要がある。これは、文法の語彙に特別な単語記号を割り当てれば良い。本稿では、単語記号の先頭に "@" をつけて表すことにする。ベース N-gram の語彙を $W_U$、文法の語彙を $W_A$ とすると、$W_U \cap W_A = \phi$ である。第 2 に、文法の先頭単語は、先頭以外の個所に現れてはいけない。同様に、文法の末尾単語は、末尾以外に現れてはいけない。すなわち、文法の語彙 $W_A$ は、互いに共通部分のない、先頭単語集合 $W_B$、中間単語集合 $W_I$、末尾単語集合 $W_E$ から構成されるとする $(W_A = W_B \cup W_I \cup W_E \wedge W_B \cap W_I = \phi \wedge W_I \cap W_E = \phi \wedge W_B \cap W_E = \phi)$。

## 4.2 文法部への N-gram 頻度割り当て

文法部の単語ネットワークにも確率値を割り当てるために、頻度情報を与える。ここで与える頻度も「部分的」であるため、3 節で述べた計算方法に従うことに注意する。頻度の与え方は様々な方法が考えられる。例えば、最も簡単な方法として、全ての分岐で等しい頻度を与える方法が考えられる。ここでは、タスクの特徴 (2) から、ベース N-gram の頻度情報を利用する方法を採用した。文法の語彙をベース N-gram と同じ単位で構成すれば、文法内部の単語列 $@w^i_{i-n+1}$ に対して、対応するベース N-gram の単語列 $w^i_{i-n+1}$ が必ず存在することを利用する方法である。

文法内部の単語列 $@w^i_{i-n+1}$ に対する N-gram 頻度 $\mathbf{C}_n(@w^i_{i-n+1})$ を、ベース N-gram での単語頻度 $\mathbf{C}_n(w^i_{i-n+1})$ を利用し、そのまま値をコピーして与える。また、ベース N-gram から文法へ遷移する部分の頻度 $C_n(w^{i-k}_{i-n+1}@w^i_{i-k+1})$ と、文法からベース N-gram へ遷移する部分の頻度 $C_n(@w^{i-k}_{i-n+1}w^i_{i-k+1})$ も、対応するベース N-gram での頻度 $C_n(w^{i-k}_{i-n+1}w^i_{i-k+1})$ を利用する。このうち前者については、頻度を $\gamma$ 倍して強調する。これは、文法でモデル化する定型表現を優先的に扱うためである (図 6)。

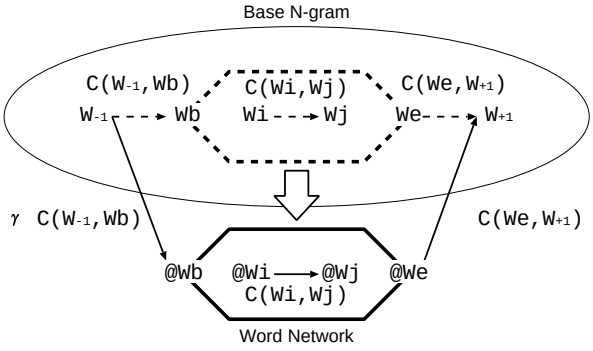


図 6: 頻度情報のコピー (bi-gram の場合)

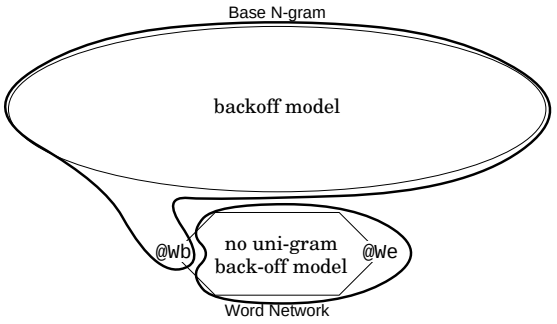


図 7: Selective Back-off Smoothing

## 4.3 選択的なバックオフ・スムージング

獲得した単語ネットワークと頻度情報から、N-gram 確率を計算する。この時問題となるのは、スムージング手法と文法的制約は、両立できないということである。バックオフ・スムージングでは、高次の N-gram が存在しない場合、低次の N-gram で補間する。ネットワーク全体を等しくスムージングすると、文法によって記述された二値的な単語連接の制約が uni-gram で補間され、結局全ての単語間の連接を許すモデルとなってしまう。一方、全くスムージングを行わないモデルを作成することも出来るが、その場合文法部の二値的制約は獲得されるが、N-gram 部にゼロ頻度問題が生じ、精度が落ちてしまう。

そこで、ネットワークの部分によってスムージング手法を切り替える選択的なバックオフ・スムージングを用いる (図 7)。ベース N-gram の単語と文法の先頭単語 ($w_i \in W_U \cup W_B$) を予測する確率値は、式 (1) で表される通常のバックオフ・スムージングで計算する。ただし、uni-gram 確率値は $W_U \cup W_B$ を全単語集合とみなして計算する (これは、次に述べるように、$W_I \cup W_E$ について uni-gram 確率を 0 とするからである)。残りの単語、すなわち文法の中間単語と末尾単語 ($w_i \in W_I \cup W_E$) を予測する確率値は、uni-gram へバックオフしないように計算する。すなわち、長さ $n > 2$ については、式 (1) をそのまま使用するが、$n = 2$ については次の計算式を用いる。

$$P(w_i|w_{i-1}) = \begin{cases} d'_{w_{i-1}^i} P_{ML}(w_i|w_{i-1}) & \mathbf{C}_2(w_{i-1}^i) > 0 \\ 0 & \mathbf{C}_2(w_{i-1}^i) = 0 \end{cases} \quad (5)$$

この計算によって、次の関係が常に成立する。

$$P(w_i) = 0 \quad \text{if} \quad w_i \in W_I \cup W_E \quad (6)$$

$$\alpha_1(w_{i-1}) = 0 \quad \text{if} \quad w_{i-1} \in W_B \cup W_I \quad (7)$$

選択的なバックオフスムージングで獲得した N-gram モデル $P(w_i|w_{i-N+1}^{i-1})$ は、以下のような性質を示す。

- 文法部の中間単語および末尾単語 ($W_I \cup W_E$) を予測する確率値 $P(w_i|w_{i-N+1}^{i-1})$ は、単語列 $w_{i-N+1}^i$ が文法により許されていない場合、必ず 0 となる。

文法で許されない単語列には頻度情報が与えられない。よって、$n > 1$ の N-gram 頻度 $\mathbf{C}_n(w_{i-n+1}^i)$ は 0 となり、$P(w_i|w_{i-N+1}^{i-1})$ は、uni-gram までバックオフされる。また、$w_i \in W_I \cup W_E$ なので、式 (6) より $P(w_i) = 0$。よって、

$$\begin{aligned} P(w_i|w_{i-N+1}^{i-1}) &= \alpha_{N-1} \cdots \alpha_2 \cdot \alpha_1(w_{i-1}) P(w_i) \\ &= 0 \end{aligned}$$

特に、ベース N-gram の単語 $w \in W_U$ から $w_i \in W_I \cup W_E$ を予測する確率は必ず 0 となり、ベース N-gram から文法の途中への単語への遷移は生じないことがわかる。

- 文法部の開始単語および中間単語 ($W_B \cup W_I$) からベース N-gram の単語 ($W_U$) を予測する確率は、必ず 0 となる。

文法の途中からベース N-gram の単語への N-gram 頻度は与えられていない。よって、この場合も $n > 1$ の N-gram 頻度 $\mathbf{C}_n(w_{i-n+1}^i)$ は 0 となり、確率値計算は uni-gram までバックオフする。この時、$w_{i-1} \in W_B \cup W_I$ なので、式 (7) より $\alpha(w_{i-1}) = 0$。よって、

$$\begin{aligned} P(w_i|w_{i-N+1}^{i-1}) &= \alpha_{N-1} \cdots \alpha_2 \cdot \alpha_1(w_{i-1}) P(w_i) \\ &= 0 \end{aligned}$$

以上により、獲得したモデルは文法の二値的な性質と N-gram の汎用性を兼ね備えていることが分る。また、文法で記述した言語表現に対し、uni-gram へバックオフしないで (discount なしで) 計算すること、$\gamma$ で重み付けすること、から相対的に高い確率値を与える。これにより、文法で表現した表現を強調して扱うタスクに適応した言語モデルを得ることができる。また、従来のバックオフ言語モデルと互換性があり、言語モデルを置換えるだけで既存の音声認識デコーダでそのまま利用可能という、既存のシステムとの互換性の面でも優れた特徴を持つ。

## 5 実験

新聞記事 111 か月分から 2 万語の N-gram 頻度情報を抽出した。これをベースに、本稿で述べた種々の適応化手法の比較を行った。スムージング手法は、すべて Witten-Bell 法 [7] を用いた。適応タスクの学習データとして、QA 質問文の定型表現を受理する文法 (単語ネットワーク) (図 8) を作成した。また、この単語ネットワークから全文生成を行い、定型表現のパターンの集合を作成した。表記違いの単語を含めると、172 パターンが得られた。

まず、新聞記事だけから bi-gram および tri-gram を作成した (*BASE* と記す)。従来法として、N-gram 頻度の混合による手法 (2 節) で適応モデルを作成した。定型表現のパターン集合を適応化テキストデータとみなして N-gram 頻度を抽出し、重み $w$ をかけて新聞記事 N-gram 頻度と混合、bi-gram および tri-gram を作成した (*MIX* と記す)。提案法の 1 として、3 節で述べた手法でモデルを作成した。定型表現のパターン集合を用い、新聞記事モデルの N-gram 頻度中の定型表現を (重み $\gamma$ で) 強調し、bi-gram および tri-gram を作成した (*EMP* と記す)。提案法の 2 として、作成した文法を、4 節で述べた手法を用いて新聞記事 N-gram と (重み $\gamma$ で) 統合、bi-gram および tri-gram を作成した (*NET* と記す)。

評価データには、新聞記事 100 文 (*NP*) と QA タスク用質問文 50 文 [13](*QA*) を、男性 2 人女性 2 人によって読み上げた音声データを用いた。作成したネッ

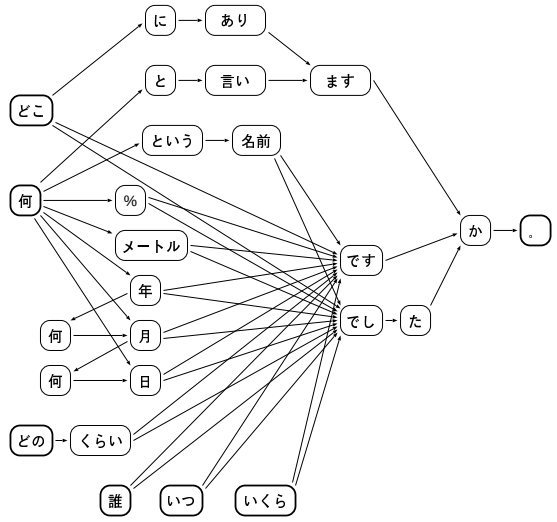


図 8: QA タスク定型表現の文法

表 1: 認識実験結果

| target (# of sent.) | language model | WER (2-gram) | WER (3-gram) |
|---|---|---|---|
| *NP* (100) | *BASE* | 18.0 | 10.7 |
| | *MIX* | 18.0 | 10.6 |
| | *EMP* | 18.0 | 10.3 |
| | *NET* | 18.8 | 10.4 |
| *QA* (50) | *BASE* | 26.3 | 16.9 |
| | *MIX* | 21.2 | 15.4 |
| | *EMP* | 19.6 | 13.8 |
| | *NET* | 23.4 | 14.7 |
| *QA'* (36) | *BASE* | 28.1 | 17.2 |
| | *MIX* | 21.8 | 15.2 |
| | *EMP* | 20.4 | 13.3 |
| | *NET* | 20.8 | 13.6 |

トワーク文法は、29 単語と比較的小規模のものであるが、質問文のうち 72%の 36 文 (*QA'*) が、この文法のモデル化する表現を含んでいた。

デコーダには大語彙音声認識デコーダ julius[14] のバージョン 3.2 を使用し[2]、音響モデルには 2000 状態 16 混合性別非依存 triphone を、言語モデル重みは新聞記事 N-gram(*BASE*) での最適値を用いた。bi-gram の比較には第一パスの結果を、tri-gram の比較には第一パスと第二パスの結果を用いた。

実験結果を表 1、図 9、図 10 に示す。

表 1 は、各手法で重みパラメータを調節して最も良い結果を示したものである。どの適応化手法もベースラインの新聞記事モデル (*BASE*) に比べて、単語誤り率 (WER) を改善している。また、従来法 (*MIX*) と比べ、提案法 (*EMP*,*NET*) は、おおむねより良い結果を示していることがわかる。

図 9、図 10 に、各手法における重みパラメータと単語誤り率との関係を示す。*NET* モデルは、文法で扱える表現だけの発話 (*QA'*) に対しては、重み $\gamma$ の影響なく WER を引き下げるが、文法から少しはずれた発話を含む場合 (*QA*)、重みを上げることで WER が悪化する。文法の表現とは大きく異なる発話 (*NP*) の場合は、この傾向は認められない。これは、類似した発話を無理矢理文法で扱える表現で誤認識しまうことによる弊害と思われる[3]本手法を用いる場合、漏れのない文法記述が必要となることを示唆している。一方、*EMP* モデルではこのような傾向は認められず、より頑健な手法と考えられる。また認識率の面でも、*NET* よりも良い結果を示した。

## 6 まとめ

音声入力に対応した質問応答システムの言語モデルを獲得するため、検索対象となる新聞記事から作成した N-gram をベースに、人手で与えた質問文定型表現を用いて適応化する 2 つの手法を提案した。認識実験の結果、N-gram 頻度を重み付きで混合する従来法とくらべ、どちらの手法も単語誤り率を減少させることが示された。特に、前者の手法が認識率と頑健性の面でより良い結果を示した。提案法は、既存の N-gram 言語モデルを、比較的多様度が小さい (人手で記述できる程度の多様性を持つ) 表現に対応するための適応化手法として、他の分野にも適用可能であろう。

## 参考文献

[1] NTCIR workshop3 質問応答タスク. http://www.nlp.cs.ritsumei.ac.jp/qac, 2001.

[2] T. Akiba, K. Itou, A. Fujii, and T. Ishikawa. Selective back-off smoothing for incorporating grammatical constraints into the n-gram language model. In *Proceedings of International Conference on Spoken Language Processing*, 2002. (to appear).

[2]提案法で用いた部分的な N-gram 頻度から計算した N-gram 確率表は、LR モデルと RL モデルで不整合が生じる (一方には存在するが他方には存在しない N-gram エントリが生じる)。そのため、両者を同一のデータ構造で管理する julius ではそのまま扱うことができない。この点を修正し、LR モデル RL モデルを個別のデータ構造に保持するように変更を行った。

[3]特に、実験に使用したデコーダ (julius) では、第一パスで認識中の仮説を最尤近似するため、影響が大きく現れたと考えられる。

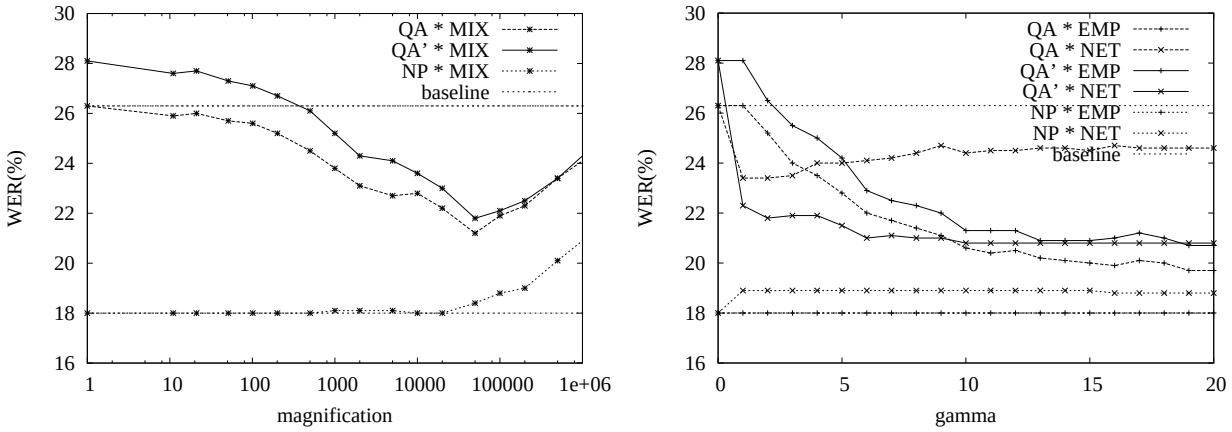


図 9: bi-gram 単語誤り率

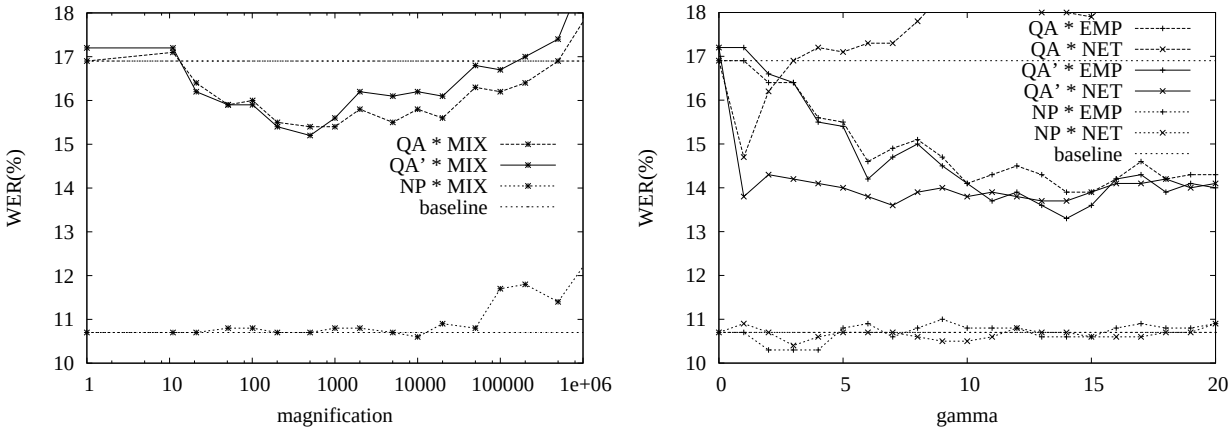


図 10: tri-gram 単語誤り率


[3] M. Federico. Bayesian estimation methods for n-gram language model adaptation. In *Proceedings of International Conference on Spoken Language Processing*, pp. 240–243, 1996.

[4] A. Fujii, K. Itou, and T. Ishikawa. Speech-driven text retrieval: Using target IR collections for statistical language model adaptation in speech recognition. In A. R. Coden, E. W. Brown, and S. Srinivasan eds., *Information Retrieval Techniques for Speech Applications (LNCS 2273)*, pp. 94–104. Springer, 2002.

[5] L. Galescu, E. Ringger, and J. Allen. Rapid language model development for new task domains. In *Proceedings of International Conference on Language Resources and Evaluation*, pp. 807–812, 1998.

[6] F. C. N. Pereira and R. R. Wright. Finite-state approximation of phrase-structure grammars. In *Proceedings of Annual Meeting of the Association for Computational Linguistics*, pp. 246–255, 1991.

[7] P. Placeway, R. Schwartz, P. Fung, and L. Nguyen. The estimation of powerful language models from small and large corpora. In *Proceedings of International Conference on Acoustics Speech and Signal Processing*, Vol. 2, pp. 33–36, 1993.

[8] E. Voorhees and D. Tice. The TREC-8 question answering track evaluation. In *Proceedings of the 8th Text Retrieval Conference*, pp. 83–106, 1999.

[9] 伊藤, 好田. N-gram 出現回数の混合によるタスク適応の性能解析. 信学論, J83-D-II(11):2418–2427, 2000.

[10] 秋葉, 伊藤, 藤井, 石川. 音声入力による質問応答システムのための音声認識用言語モデルの検討. 言語処理学会第 8 回年次大会発表論文集, pp. 244–247, 2002.

[11] 岡登, 石井, 花沢. タスクの例文を用いた自由発話音声認識のための言語モデルの構築. 日本音響学会秋季研究発表会講演論文集, pp. 73–74, Oct 2001.

[12] 伊藤, 秋葉, 藤井, 石川. 音声入力型テキスト検索システムのための音声認識. 日本音響学会秋季研究発表会講演論文集, pp. 193–194, Oct 2001.

[13] 佐々木, 磯崎, 平, 廣田, 賀沢, 中島. 質問応答システムの比較と評価. 信学技法 NLC-24, pp. 17–24, 2000.

[14] 鹿野, 伊藤, 河原, 武田, 山本（編）. 音声認識システム. オーム社, 2001.