特許文に対するクロストリガーモデルを用いた統計的機械翻訳システム

松本 琴美† 山本 幹雄‡ 内山 将夫§

†筑波大学情報学類 ‡筑波大学システム情報工学研究科 §情報通信研究機構

1 はじめに

フレーズ単位で翻訳を行うフレーズに基づく統計的機械翻訳 [5] は単語単位に翻訳を行う方式よりも大きく性能が向上する。これはフレーズの範囲内の局所的な情報をカプセル化することにより、訳語決定および局所的な並び換えの曖昧性の問題を一部回避できていることによる。しかし、これらの曖昧性は局所的な情報だけでは解決できない場合も当然あり得る。本報告では、トリガー言語モデル [1] を言語横断方向に拡張したクロストリガーモデル [3] を用いることによって、フレーズよりももう少し範囲の広い文脈情報を利用して翻訳性能を改善する方法を検討する。

2 トリガーモデルとクロストリガーモデル

2.1 トリガーモデル

トリガーモデル [1] は大域的言語モデルの1つであり、長距離にわたる単語間の共起関係あるいは依存関係を組み入れたモデルである。文脈中のある単語 w_a の出現が後続の別の単語 w_b の出現確率に比較的大きな影響をおよぼすとき、これら2つの単語をトリガーペア (trigger-pair) と呼び、$w_a \rightarrow w_b$ と表す。L 個のトリガーペア $w_{i,1} \rightarrow w_{i,2} (i = 1 \sim L)$ があるとき、これらのトリガーペアの関係を組み入れたトリガー言語モデルでは、履歴 h 後の単語 w の確率を次のように定義する。

$$P_t(w|h) \approx P(w|d_1(w,h), d_2(w,h), ..., d_L(w,h))$$

ここで、$d_i(w,h)$ はL個のトリガーペア中の i 番目のトリガーペア $w_{i,1} \rightarrow w_{i,2}$ について、$w_{i,1} = w$ かつ $w_{i,2} \in h$ のときだけ活性化される2値関数である。

一般にトリガーペアは相互情報量 (MI) を用いて選択され、トリガーモデルの推定には最大エントロピー (ME) 法が用いられる [2]。

2.2 クロストリガーモデル

音声認識のためにトリガーモデルを用いる場合、履歴としては現在認識中の部分以前の認識結果を用いることが多い。この場合、音声認識誤りによる誤った単語からのトリガーによって、残念ながら性能が悪化する場合もありうる。単言語のトリガーモデルを用いる場合は機械翻訳の場合も同じである。しかし幸いなことに、テキスト機械翻訳の場合、一般に原言語文は翻訳開始前に全て与えられていることが多い。このため、もし原言語文中の単語から目的言語の単語にトリガーをかけることができれば、誤認識の可能性のない情報を利用でき、また、現在翻訳している部分やそれ以後の部分の情報も容易に用いることができる。

言語をまたがってトリガーをかけるモデル（以下、クロストリガーモデルと呼ぶ）は、音声認識のために Kim(2004)[3] によって提案されている。本報告では基本的には Kim と同じモデルを統計的機械翻訳に適用するが、上記のように音声認識以上の効用が期待できる。

クロストリガーモデルにおける目的言語文中の単語 e の確率は原言語文文書の一部に含まれる単語集合 $\boldsymbol{f} = f_1, f_2, ..., f_n$ を言語横断履歴として与えられた場合、以下のように定義される。

$$P_t(e|\boldsymbol{f}) \approx P(e|d_1(e,\boldsymbol{f}), d_2(e,\boldsymbol{f}), ..., d_L(e,\boldsymbol{f}))$$

ここで、$d_i(e,\boldsymbol{f})$ は i 番目のクロストリガーペア $f_i \rightarrow e_i$ に対して、$e_i = e$ かつ $f_i \in \boldsymbol{f}$ のときにだけ活性化される2値関数である。

クロストリガーがうまく働く例を図1に示す。フレーズ"played it"は"それで遊んだ"、"それ(楽器)をひいた"、"それ(スポーツ)をした"など多くの訳が存在するが、この例の場合、目的言語(日本語)側では"ギター"と"ひいた"の距離が大きいため、*n*gram モデルによる修正は期待できない。これに対して、クロストリガーモデルでは"guitar"から"ひいた"にトリガーをかけることができ、"それをひいた"という訳の確率を上げることができる。

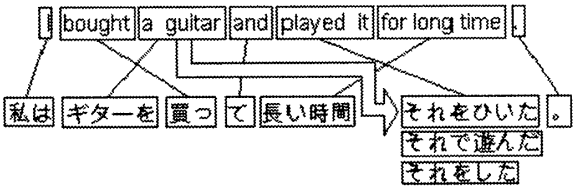

図1:クロストリガーが有効に働く例

2.3 線形補完

クロストリガーモデルにおいても既存のトリガーモデルと同様、*n*gram モデルと線形補完を行うことによって局所的な情報を補う。クロストリガーモデルの確率を P_t、重みを λ_t とし、*n*gram モデルの確率を P_n とすると、最終的な言語モデルの確率 P_{LM} は次のように定義できる。

$$P_{LM} = \lambda_t P_t + (1 - \lambda_t) P_n$$

重み λ_t はデベロップメントデータに対するパープレキシティもしくは BLEU 値が最も良くなるように設定した。

3 実験

3.1 パープレキシティによるトリガーモデルの比較

テストセットパープレキシティ(*PP*)を用いて、トリガーモデルとクロストリガーモデルの性能比較を行った実験を行った。実験条件を以下に示す。

データ:NTICIR7[4] の特許文対訳データ (PSD) の学習データ 1,798,571 文、デベロップメントデータ 927 文、テストデータ 899 文を用いた。デベロップメントデータおよびテストデータには元の対訳特許コーパス (PPC) から文脈を補った。また、これら3つのデータに重なりはない。

Statistical machine translation system using cross trigger model for patent sentences

Kotomi MATSUMOTO†, Mikio YAMAMOTO‡, Masao UCHIYAMA§

†Collage of Information Sciences, University of Tsukuba

‡Graduate School of Systems Information Engineering, University of Tsukuba

§National Institute of Information and Communications Technology

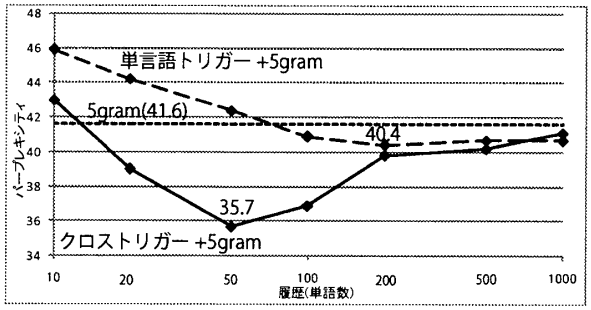

図 2:各トリガーモデルのパープレキシティの比較

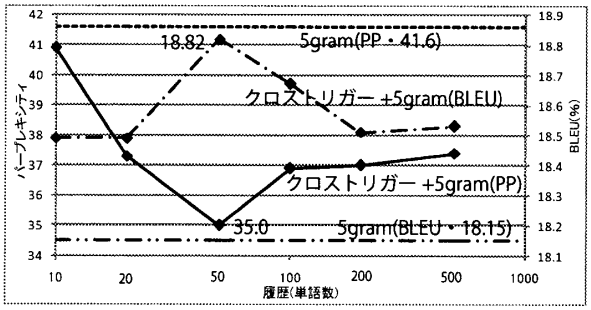

図 3:BLEU 値とパープレキシティ

ngram モデル:SRILM*により、学習データ全文を用いて、線形補完改良 KN ディスカウンティング法を使用した 5gram モデルを作成した。

トリガーペア:全ての単語ペアを考えると膨大になってしまうため、トリガーペアのどちらの単語も機能語でなく、かつ 9 文書以上に現れるトリガーペアのみを対象とし、学習データ全文を用いて各トリガーペア $w_a \rightarrow w_b$ の MI を求め、各 w_b につき上位 3 トリガーペアのみを選択した。機能語は日本語: 助詞・助動詞・感動詞・フィラー・アルファベット以外の記号、英語:記号、とした。この際の履歴として、単言語トリガーモデルではその単語の 5 単語前から文書の最初まで、クロストリガーモデルでは対応する原言語文書全てを用い、最終的なトリガーペア数は単言語トリガーモデルで 58,049 ペア、クロストリガーモデルで 57,126 ペアとなった。

トリガーモデル:トリガーモデルの学習は学習データから約 1% のデータを抜粋した 16,973 文を用いて行い、履歴として、単言語トリガーモデルでは 5 単語前から $5+n$ 単語前までの n 単語を、クロストリガーモデルでは対応する原言語文の文末から n 単語前までの n 単語を用い、履歴の長さ n は 10, 20, 50, 100, 200, 500, 1000 とした。ME モデルの学習には Maximum Entoropy Modeling Package†を用いた。クロストリガーモデルの線形補完重みは 0.2 とした (これはデベロップメントデータに対する *PP* が最も低くなる値である)。

各モデルの *PP* を図 2 に示す。*n*gram モデルのみを用いたベースライン 41.6 と比較して、単言語トリガーモデルで最大 3.8%、クロストリガーモデルで最大 14.2% 減少した。履歴の長さと *PP* の変化に着目すると、履歴が 50 単語の時に最も良い性能を示しており、履歴は長すぎても短すぎても良くないということがわかる。

3.2 統計的機械翻訳への適用

クロストリガーモデルを統計的機械翻訳システムに適用した。フレーズテーブルの学習は特許 B 区分 199,062 文により Pharaoh[5] を用いて行った。クロストリガーモデルの学習は NTICIR7 の学習データから約 10% のデータを抜粋した 172,710 文を用いて実験を行った。また、計算機のメモリの制限のため、履歴の長さは 500 単語までとした。これ以外の条件は前節と同様であり、評価指標としては BLEU 値を使用し、クロストリガーモデルの線形補完重みは 0.09 とした (これはデベロップメントデータに対する BLEU 値が最も高くなる値である)。

BLEU 値と *PP* を図 3 に示す。*n*gram モデルのみを用いたベースラインの BLEU 値 18.15% と比較して、クロストリガーモデルの BLEU 値は最大 0.67% 上昇した。これは有意水準 1% で有意である。また、*PP* と同様に BLEU 値も履歴の長さが 50 単語の時に最も良い性能を示している。

以下に、実際にクロストリガーモデルによって性能が向上したと思われる例を示す。以下の翻訳例を見ると、"excited" が 5gram モデルでは"励磁"と誤って訳されているのに対して、クロストリガーモデルでは"励起"と正しく訳すことができている。これは、"*wavelength*"から"励起"へトリガーがかかることによって、"励起"の確率が高くなったためである。

原文 as a result , rhodamine b contained in the transfer support 12 is excited to release fluorescent light whose peak *wavelength* is 605 nm .

正解文 その結果、転写支持体１２に含まれているＲｈｏｄａｍｉｎｅＢが励起されて、６０５ｎｍの波長にピークを有する蛍光が発せられる。

5gram モデル これにより、搬送支持台１２を励磁してレーザ光の波長はローダミンＢに収容されて解除ピーク６０５ｎｍである。

クロストリガーモデル これにより、搬送支持台１２が励起さ蛍光ピークの波長は６０５を解除光に含まれるローダミンＢｎｍである。

4 おわりに

トリガーモデルの履歴に原言語文を利用し、原言語文から目的言語文にトリガーをかけるクロストリガーモデルの統計的機械翻訳への応用を検討した。クロストリガーモデルの *PP* は 5gram モデルに比べて最大 14.2% 減少し、これを統計的機械翻訳に適用した結果、最大 0.67% の BLEU 値の上昇が見られた。

参考文献

[1] R.Rosenfeld."A Maximum Entropy Approach to Adaptibe Statistical Language Modeling".Computer Speech and Language,10(5),pp187-228,1996.

[2] 北研二. 確率的言語モデル. 東京大学出版会,1999.

[3] Woosung Kim, Sanjeev Khudanpur. "Language Model Adaptation for Automatic Speech Recognition and Statistical Machine Translation".A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy. 2004.

[4] 内山将夫, 山本幹雄, 藤井敦, 宇津呂武仁."特許情報を対象とした機械翻訳: 共通基盤による評価タスクを目指して". 情報処理学会研究報告,No.76(20070724),pp.133-138.2007.

[5] P. Koehn: "Pharaoh: a beam search decoder for phrase-based statistical machine translation models", AMTA, pp. 115-124 (2004).

*http://www.speech.sri.com/projects/srilm/

†http://www2.nict.go.jp/x/x161/members/ mutiyama/software/