論文 / 著書情報
Article / Book Information

論題(和文)	音声理解のための言語モデル自動獲得
Title(English)	
著者(和文)	松岡達雄, Robert Hasson, Michael Barlow, 古井 貞熙
Authors(English)	SADAOKI FURUI
出典(和文)	電子情報通信学会総合全国大会 情報システム1, , , pp. 347-348
Citation(English)	, , , pp. 347-348
発行日 / Pub. date	1996,
URL	http://www.ieice.org/jpn/books/t_g.html
権利情報 / Copyright	本著作物の著作権は電子情報通信学会に帰属します。 Copyright (c) 1996 Institute of Electronics, Information and Communication Engineers.

SD-4-11

音声理解のための言語モデル自動獲得

Language model acquisition from a text corpus for speech understanding

松岡達雄、Robert Hasson†、Michael Barlow、古井貞熙
Tatsuo Matsuoka, Robert Hasson†, Michael Barlow, and Sadaoki Furui

NTTヒューマンインタフェース研究所、†Eurecom Institute
NTT Human Interface Laboratories, †Eurecom Institute

1. まえがき

本報告では、音声理解システムにおいて、音声認識結果である自然言語をシステムを駆動する意味言語に変換するための言語モデルを、コーパスから自動的に獲得する方法について述べる。音声理解の問題は、自然言語を意味言語へ翻訳する問題ととらえることができる。機械翻訳の分野では、二カ国語が一対となったテキストコーパスを用いて、翻訳のための言語モデルを統計的に推定する方法が提案されている[1-3]。Vidalら[4]は、この方法をベースに、自然言語と意味言語が一対となったコーパスを用いて、音声理解のための翻訳言語モデルを推定する方法を提案した。Vidalらの方法では、自然言語の文法規則と意味言語の文法規則の共起確率を翻訳言語モデルとするが、文法規則数が多いと共起確率の推定が困難になる点が問題であった。

以下では、文法規則数が多い場合にも有効な言語モデルが推定できる方法を提案する。提案法は、McCandlessらによる文脈自由文法の推定法[5]を用いて、有限状態オートマトンで表現された文法の状態数を削減することにより、翻訳言語モデルの推定精度を向上する。米国ARPAの音声理解評価タスクである航空旅行情報システム(Air Travel Information System: ATIS)を対象として評価を行い提案法の有効性を示す。

2. 音声理解のための言語処理

図1に提案法の処理の流れを示す。まず、自然言語、意味言語それぞれについて文法ネットワークを推定し、次に、その文法ネットワークと自然言語と意味言語が一対になった学習テキストを用いて翻訳言語モデルを推定する。

2.1 文法ネットワークの生成

翻訳言語モデルを推定するには、まず自然言語（入力言語）、意味言語（出力言語）各々についてECGI(Errror Correcting Grammar Inference)アルゴリズム[6]により有限状態オートマトンで表現される文法を推定する。ECGIアルゴリズムは、学習セット中のテキストを一文ずつパージングし、その文を受理するために必要な状態／遷移を付加していく。入力文と文法ネットワーク中のパスの最適アライメントをDP的手法により求め、その最適パスに沿って必要な状態／遷移を付加する。

ECGIアルゴリズムによって推定される文法の状態数は学習セット中の文の提示順序により異なる。すなわち、単語数の多い文から提示すれば、単語数の少ない文はすでに有限状態オートマトンに存在する状態間を遷移することになり、その結果、状態数の少ない文法が生成される。

2.2 翻訳言語モデルの推定

次に入力言語の文法規則と出力言語の文法規則の共起確率を、入出力文が一対になった学習テキストにおける共起頻度から推定する。

入力言語の文法をG_I、出力言語の文法をG_Oとすると、与えられた問題は入力文xに対して次式を満足する出力文$\hat{y}$を求めることとなる。

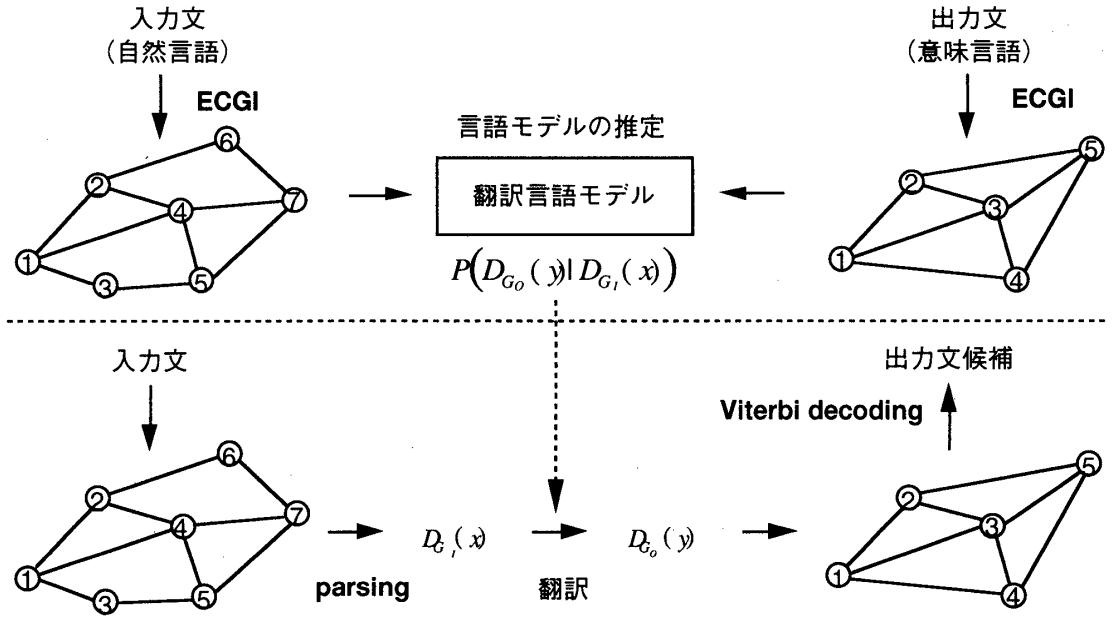

図1　統計的言語モデルによる音声理解の言語処理

$$\hat{y} = \arg\max_{y \in L(G_o)} p(y \mid x) \qquad (1)$$

文章x、yはそれぞれ文法規則の系列$D_{G_I}(x)$、$D_{G_O}(y)$として表現できる。

$$D_{G_I}(x) = \left\{ r_{I1}^x, r_{I2}^x, ..., r_{In}^x \mid r_{Ii}^x \in G_I \right\} \qquad (2)$$

$$D_{G_O}(y) = \left\{ r_{O1}^y, r_{O2}^y, ..., r_{Om}^y \mid r_{Oi}^y \in G_O \right\} \qquad (3)$$

G_I、G_Oが正規文法ならば$D_{G_I}(x)$、$D_{G_O}(y)$は一意に決定できる。文法にあいまい性がある場合にはViterbiアルゴリズムによって近似的に求めることができる。これより、(1)式は(4)式のように書き直せる。

$$\hat{y} = \arg\max_{y \in L(G_o)} p\left(D_{G_O}(y) \mid D_{G_I}(x)\right) \qquad (4)$$

(4)式を、

$$p\left(D_{G_O}(y) \mid D_{G_I}(x)\right) \approx p\left(r_O \mid r_{I1}^x, r_{I2}^x, ..., r_{In}^x\right) \approx \frac{N(r_O, x)}{N(x)} \qquad (5)$$

として推定する。$N(x)$はG_Iが生成可能な文全体の数であり実際には計算不可能なので以下のように近似を行う。

$$\hat{p}\left(r_O \mid r_{I1}^x, r_{I2}^x, ..., r_{In}^x\right) \approx \frac{p^\alpha(r_O) \prod_{r_I \in D_{G_I}(x)} p^\beta(r_O \mid r_I)}{\sum_{\substack{r \text{ with the same} \\ \text{initial state as } r_O}} \left[p^\alpha(r) \prod_{r_I \in D_{G_I}(x)} p^\beta(r \mid r_I) \right]} \qquad (6)$$

以上に述べた翻訳言語モデルの推定では、文法G_I、G_Oの規模が大きいほど、つまり、文法規則数が多いほど$p\left(D_{G_O}(y) \mid D_{G_I}(x)\right)$の推定がスパースになるという問題がある。したがって文法規則数が少ないことが望ましい。そこで、さらに文法規模を縮小するため次節に述べるような方法で有限状態オートマトンの状態数を削減した。

2.3 文脈自由文法の推定による文法状態数の削減

McCandlessらはコーパスから確率的文脈自由文法を推定する方法を提案している[2]。この方法では単語Bigram確率を距離尺度として、ボトムアップに文法規則を生成する。単語や非終端記号間の距離は(7)～(9)式のように定義する。

$$\left\| u_i, u_j \right\| = d(P_i, P_j) + d(P_j, P_i) \qquad (7)$$

$$d(P_i, P_j) = \sum_{C \in Context} P_i(C) \times \log \frac{P_i(C)}{P_j(C)} \qquad (8)$$

$$P_i(C) = P(C \mid u_i) \approx P(u_{left}, u_i \mid u_i) \times P(u_i, u_{right} \mid u_i) \approx \frac{N(u_{left}, u_i)}{N(u_i)} \times \frac{N(u_i, u_{right})}{N(u_i)} \qquad (9)$$

すなわち、同じ文脈で出現する頻度の高い単語同士の距離を近いと判断し、マージして新たな非終端記号を生成する。単語がマージされて非終端記号となることで有限状態オートマトンの状態数を削減することができる。

2.4 前処理

タスク固有の知識を必要としない前処理によっても文法上の状態数の削減が可能である。例えば、Boston、New Yorkなどの単語は「地名」という一つの非終端記号にまとめることができる。同様に、数字、日付、曜日、月、航空会社、空港名、座席クラスなども具体的な単語の代わりに非終端記号で置き換え可能である。文脈自由文法による非終端記号へのマージと、この前処理により、入力言語の文法状態数は約2500から約1500へ、出力言語の状態数は約1000から約600へ削減できた。

3. 実験

ATISタスクにおける意味言語はSQLであるが、ATISコーパスには一意にSQL表現に書き直せるWIN(Wizard Input)文が入力文に対応して与えられているので、WIN文を意味言語として実験を行った。LDC(Linguistic Data Consortium)より頒布されているATIS2のコーパス中で、文脈に独立にWIN文を生成可能なクラスAに分類されたテキストのうち、WIN文に括弧の含まれないテキストを対象として実験を行った。1915文を学習に、213文を評価に用いた。

ECGIアルゴリズムによって推定された文法の評価セットに対するカバー率は約90%であった。また、同じ非終端記号が有限状態オートマトンの異なる場所に複数存在するなど、文法に冗長性があることがわかった。

評価は翻訳結果と正解WIN文との文章単位での正解率により行った。学習セットに対する文正解率は95.3%、評価セットに対しては62.4%であった。ARPRにおける他の評価結果(unweighted error rate: 3.8%～30.6%)と比較するため、置換、脱落、挿入を考慮した単語誤り率を求めたところ、14.2%であった。コーパスからの自動推定を重視してタスク固有のヒューリスティックスなどは用いていないことを考慮すれば、有望な結果と考えられる。

4. まとめ

音声理解のために自然言語を意味言語に翻訳する言語モデルをコーパスから自動的に推定する方法について述べた。Vidalらの方法において文法の規模が大きくなった場合の問題点を明らかにし、アルゴリズムを修正するとともに、文法上の状態数を削減して翻訳言語モデルの推定精度を向上する方法を提案した。ATIS2のコーパスにより評価を行い、コーパスから翻訳言語モデルを自動獲得する可能性を示した。問題として、ECGIアルゴリズムにより推定される文法の冗長性や、文脈自由文法の推定において同一コンテキストでの出現頻度のみで単語をマージしていることなどが考えられる。今後これらの問題点を解決するとともに、タスク固有のヒューリスティックスが導入容易なものについては考慮しながら、音声入力から意味理解までの全体的な評価を行っていきたい。

謝辞

ECGIアルゴリズムのプログラムを提供してくださったUniv. Politecnica de ValenciaのVidal教授と、修士論文を送ってくださったMITのMcCandless氏に感謝します。

参考文献

1. P. F. Brown, et al., Computational Linguistics, Vol. 16, No. 2, pp. 79-85, 1990
2. P. F. Brown, et al., Proc. Int. Conf. on Theoretical Methodological Issues in Machine Translation, pp. 83-100, 1992
3. P. F. Brown, et. al, Computational Linguistics, Vol. 19, No. 2, pp. 263-311, 1993
4. E. Vidal, et al., Proc. EUROSPEECH-93, pp. 1187-1190, 1993
5. M. K. McCandless et al., Proc. EUROSPEECH-93, pp. 981-984, 1993
6. H. Rulot, et al., Proc. ICASSP-89, pp. 643-646, 1989
7. 松岡、R. Hasson、M. Brlow、古井、信学技報、SP95-94、pp. 7-12、1995.12