TR-IT-0102

音声認識のための統計的言語モデルの研究

Stochastic Language Modeling for Speech Recognition

磯谷 亮輔
Ryosuke Isotani

1995.3

概 要

筆者が1992年4月27日から1995年3月31日まで、ATR自動翻訳電話研究所およびATR音声翻訳通信研究所において行なった、音声認識のための統計的言語モデルの研究について報告する。研究内容は、主に以下の2つである。

- 付属語の N-gram、自立語の N-gram を用いた音声認識
- 統計的言語モデルにおける冗長語処理の検討

1 はじめに

1992年4月27日から1995年3月31日まで、ATR自動翻訳電話研究所およびATR音声翻訳通信研究所において、音声認識のための統計的言語モデルの研究を行なった。本報告書では、以下の2つの研究を中心に、筆者がATRで行なった研究の概要を述べる。詳細については、発表論文を参照されたい。

- 付属語の N-gram、自立語の N-gram を用いた音声認識
- 統計的言語モデルにおける冗長語処理の検討

2 研究目的と背景

文音声認識の性能向上のためには、音響モデルの精緻化だけでは限界があり、言語的な制約を用いて候補を絞ることが必要である。そのための言語モデルとして、従来は主に人手で開発した文法が用いられ、語彙や言い回しを限定した比較的狭いタスクに対しては成功を収めてきた。しかし、より広いタスクや自由な言い回しを扱おうとする場合、文法が大規模になり、その開発、管理が困難になるという問題点がある。

一方、近年大規模なテキストデータベースが利用可能になったことを背景に、統計的な手法を用いて言語的な制約を自動的に獲得する手法が盛んに研究されている。統計的手法によれば、エキスパートによる知識や、文法の開発、管理の手間が不要で、容易に言語モデルが構築できる。また、言語現象を確率的な偏りとして扱うことにより、効率的に候補を絞ることができ、かつ低頻度の現象に対する柔軟な処理も可能となる。現在、統計的な言語モデルとして、単語の連鎖統計モデル (単語 N-gram モデル) が広く用いられており、その有効性が示されている。しかし、このモデルで表現できる言語的制約は限られており、特に今後自由発話の認識を扱おうとする場合、より強力でかつ頑健な言語モデルが望まれる。そこで、より効果的で効率的な言語情報の獲得法と、その音声認識への利用法について検討を行なった。

3 研究の概要

3.1 付属語の N-gram、自立語の N-gram を用いた音声認識

従来より、単語の連鎖統計を用いた N-gram モデルが音声認識において有効であることが知られている。単語の N-gram モデルは、日本語の場合の文節内の単語の連接関係などを表現するのには適している。しかし、情報が局所的なものに限られるため、離れた単語間に現れる文節間の構文的あるいは意味的な関係を表現する能力には欠ける。そこで、単語 N-gram を発展させ、より大局的な関係を表現できる新しいモデルとして、文中の付属語のみに注目した連鎖統計 (N-gram)、自立語のみに注目した連鎖統計を用いる新しい統計的言語モデルを提案した。

文中の文節末にあらわれる助詞などの付属語のみに注目してその連鎖を統計的にモデル化することにより、構文的な接続関係が表現できると思われる。同様に、自立語の連鎖により、語と語の意味的な関係を表す情報が抽出できると期待される。本手法はまた、従来の単語 N-gram などと組み合わせて用いることも可能である。

助詞の連鎖統計モデル

最初に、文中の付属語のみに着目した連鎖統計の学習および音声認識の後処理への利用について検討した。これにより、係り受け関係のような、通常の単語連鎖統計モデル (単語 N-gram) とは異なる言語情報の獲得が期待できる。

予備的な検討として、各文節末の助詞に注目したときの連鎖統計を用い、文節認識により得られた文節ラティスから文認識結果を得る際の後処理として利用する実験を行なった。約12,000文からなるテキストデータベースを

用いて文節末の助詞の連鎖統計を学習し、話者2名の各7会話 (137文353文節) の音声を対象に、文節認識結果の文節ラティスからの候補選択実験を行なった。その結果、音響スコアのみを用いる場合に対し、助詞の誤りの約2割、全体の誤りの約1割が修正され、本モデルが助詞の誤りの修正に有効である可能性が示された。以上の実験は、学外実習生の粟津氏 (東北大) に行なってもらった [Awatsu92TR]。また、付属語 bigram (2単語連鎖) による後処理を従来の文節間文法を用いた方法と比較してその有効性を確認し、本手法を ASURA デモシステムに組み込んだ。

自立語と付属語の連鎖統計モデル

つぎに、自立語の連鎖統計についても検討を行なった。これは、付属語の連鎖が構文的な関係を表すのに対し、語の意味的な関係を表現することを期待したものである。付属語の場合と同様に、各文節の最初に現れる自立語のみに注目し、約10,000文のテキストデータからその bigram の確率を学習した。また、付属語としては従来は助詞のみを扱っていたが、助動詞も加えて bigram の確率を学習した。

評価実験は、話者を12名にふやし、付属語 bigram、自立語 bigram、およびそれらを組み合わせた場合について、文節ラティスからの候補選択実験を行なった。その結果、単独で用いた場合および両者を組み合わせた場合のそれぞれについて、音響スコアのみを用いる場合に対して文認識率が向上することを確認した。付属語 bigram と自立語 bigram を組み合わせた場合、従来の文節間文法を用いた場合の認識率を上回り (70.5% → 71.6%)、本モデルの有効性が示された [Isotani93ASJ03]。

候補選択における局所的制約の併用

文節ラティスからの候補選択において、付属語の N-gram、自立語の N-gram と従来の単語 N-gram の併用について検討した。これにより大局的な制約と局所的制約の両方が表現できる。また、パラメータ推定の精度の向上のため、削除補間法を導入した。

局所的制約としては、文節間の関係のみに注目し、文節境界での付属語－自立語の連鎖のみを対象とした。大局的制約として付属語 bigram と自立語 bigram、局所的制約として付属語－自立語 bigram をテキストデータから学習し、12名の話者に対して文節ラティスからの候補選択実験を行なった。局所的な制約を併用することにより、大局的制約のみの場合に比べさらに認識率が向上した (表 1) [Isotani93SP06, Isotani93EUROSPEECH, Isotani93ASR, Isotani95IEICE]。

音声認識処理への組み込み

ここまでは、音声認識への応用は文節ラティスからの候補選択という形で行なってきたが、直接認識処理に組み込んで文音声認識ができるように、従来の単語 bigram による局所的な単語連鎖統計と、付属語の bigram、自立語の bigram による大局的な単語連鎖統計を統一的に扱う言語モデルを提案した。また、局所的制約と大局的制約の独立性を仮定することによりパラメータ数の増加を抑えられることを示した。実際に本言語モデルを組み込んだ認識システムを構築し、特定話者1名の文認識実験により評価を行なった。その結果、提案手法は、テストセットパープレキシティ、認識率で単語 trigram にはやや及ばないものの、単語 bigram の2倍程度のパラメータ数で単語 bigram と trigram の中間程度の認識率を達成し、少ないパラメータで言語的制約を効率的にモデル化できることを確認した (表 2) [Isotani94ARPA, Isotani94ICASSP, Isotani95CSL]。

3.2 統計的言語モデルにおける冗長語処理の検討

自由発話では「あのー」、「えーと」などの冗長語が数多く現れ、音声認識用の言語モデルでもこれに対処する必要がある。そこで、統計的言語モデルにおける冗長語の扱いについて検討を行なった。冗長語は文から取り除い

表 1: 自立語 bigram、付属語 bigram を用いた文節ラティスからの候補選択 (文認識率: %)

Inter-phrase language model

none: no model　　CFG: inter-phrase CFG

t–t: function-word bigram　　h–h: content-word bigram

t–h: function-word-to-content-word bigram

Speaker	none	CFG	1) t–t	2) h–h	3) t–t + h–h	4) t–h	5) t–t + h–h + t–h
MHO	49.8	59.1	57.9	55.2	61.0	59.1	64.5
MIK	80.7	85.7	82.6	81.5	83.4	81.9	82.6
MSH	53.3	59.8	57.1	54.4	58.7	57.1	60.6
MST	70.7	76.4	74.5	73.0	78.8	73.0	77.6
MTK	68.0	74.5	74.5	71.0	76.1	73.0	75.3
MTT	66.0	69.9	70.7	73.0	76.4	76.4	77.6
FAK	86.5	89.2	86.5	84.9	86.9	85.7	87.6
FAS	59.8	66.8	69.9	65.3	72.2	67.6	71.4
FFO	61.0	66.0	62.9	66.0	66.8	69.1	69.1
FKN	46.9	53.5	54.3	51.6	58.5	57.0	62.4
FNY	71.0	77.6	77.2	78.0	82.2	78.0	84.2
FRS	57.5	67.6	62.2	65.3	67.2	68.0	72.6
Average	64.3	70.5	69.2	68.3	72.4	70.5	73.8

表 2: 局所的制約と大局的制約を併用した文音声認識

Language Model	Perplexity	Sentence Recognition Rate	Word Accuracy	Ratio of Number of Parameters
Bigram	41.2	51.3%	69.6%	1.0
Trigram	36.3	54.0%	71.2%	5.4×10^3
Proposed	38.1	52.5%	70.7%	1.9

ても文の文法性や意味に影響しないこと、および冗長語の種類や文中での出現位置には偏りがあること、の2点に着目し、冗長語を含む文の統計的言語モデルを提案した。

予備的な評価として、テスト文に対するパープレキシティの値で評価し、冗長語を通常の単語と同様に扱った場合との比較を行なった。実験では、パープレキシティは若干高くなり、提案手法の効果を確認するには至らなかった [Isotani94ASJ09]。

3.3 その他の成果など

音声認識性能評価用音声データベース

自動翻訳電話プロジェクトにおける認識方式の評価用に用いるための音声データベース (F セット) を設計、作成した [Isotani93TRb]。

CMU 長期出張

米国カーネギー・メロン大学 (CMU) に約4か月間滞在し, Dr. Waibel の指導のもとで spontaneous speech の音声認識のための統計的言語モデルについて研究し, semantic parser の出力を利用した "slot-based language model" [Suhm94AAAI] などについて検討を行なった.

解析済みコーパスを用いた言語モデル

フランスからの学外実習生 Halber 氏に、英語の解析済みコーパスを用いた単語の長距離依存性の獲得について検討してもらった。詳細は、A. Halber: "Capturing Long Distance Dependencies from Parsed Corpora" (TR-IT-0096) を参照。

4 発表論文リスト

論文

[Isotani95IEICE] R. Isotani, S. Matsunaga and S. Sagayama: "Speech Recognition Using Function-Word N-grams and Content-Word N-grams," *IEICE Trans. Inf. & Syst.* (to appear).

[Isotani95CSL] R. Isotani and S. Matsunaga: "Speech Recognition Using a Stochastic Language Model Integrating Local and Global Constraints," *Computer Speech and Language* (投稿中).

学会発表 (国内)

[Isotani93ASJ03] 磯谷、嵯峨山、粟津: "付属語の N-gram、自立語の N-gram を用いた音声認識", 日本音響学会講演論文集, pp. 95–96 (1993-03).

[Isotani93SP06] 磯谷、嵯峨山: "自立語と付属語の連鎖統計モデル用いた音声認識のための候補選択", 電子情報通信学会技術研究報告, Vol. 93, No. 8, pp. 73–78 (1993-06).

[Isotani93ASJ10] 磯谷、松永、嵯峨山: "局所的 / 大局的単語連鎖統計を併用した音声認識", 日本音響学会講演論文集, pp. 179–180 (1993-10).

[Sakamoto93ASJ10] 坂本、磯谷、松永: "対話音声における統計的言語モデルの選択の効果", 日本音響学会講演論文集, pp. 193–194 (1993-10).

[Isotani94ASJ09] 磯谷、松永: "統計的言語モデルにおける冗長語処理の検討", 日本音響学会講演論文集, pp. 145–146 (1994-10–11).

学会発表 (国際会議)

[Isotani93EUROSPEECH] R. Isotani and S. Sagayama: "Speech Recognition Using Particle N-grams and Content-Word N-grams," *Proc. Eurospeech'93*, vol. 3, pp. 1955–1958 (1993-09).

[Isotani93ASR] R. Isotani, S. Matsunaga and S. Sagayama: "Continuous Speech Recognition Using Stochastic Global Language Model," *Proc. IEEE ASR Workshop*, pp. 107–108 (1993-12).

[Isotani94ARPA] R. Isotani and S. Matsunaga: "Speech Recognition Using a Stochastic Language Model Integrating Local and Global Constraints," *Proc. ARPA HLT Workshop*, pp. 88–93 (1994-03).

[Isotani94ICASSP] R. Isotani and S. Matsunaga: "A Stochastic Language Model for Speech Recognition Integrating Local and Global Constraints," *Proc. ICASSP'94*, vol. 2, pp. II-5–8 (1994-04).

[Suhm94AAAI] B. Suhm, L. Levin, N. Coccaro, J. Carbonell, K. Horiguchi, R. Isotani, A. Lavie, L. Mayfield, C. P. Rosé, C. Van Ess-Dykema and A. Waibel: "Speech-Language Integration in a Multi-Lingual Speech Translation System," *Proc. AAAI-94 Workshop*, pp. 92–99 (1994-07–08).

ATR テクニカルレポート

[Awatsu92TR] 粟津、磯谷、嵯峨山: "助詞の連鎖統計を用いた言語モデルとその音声認識への応用", TR-I-0279 (1992-09).

[Isotani93TRa] 磯谷: "付属語の bigram、自立語の bigram を用いた音声認識関連プログラムユーザーズマニュアル", TR-I-0357 (1993-03).

[Isotani93TRb] 磯谷、嵯峨山: "音声認識性能評価用音声データベースの作成", TR-I-0358 (1993-03).

その他

[Isotani95ATRJournal] 磯谷、村上、Lucke: "言葉を覚えて声を聞く — 高精度な音声認識のための言語的知識の獲得", *ATR Journal*, 18 (1995-02).