

ユーザの内部状態に着目した対話音声認識手法についての検討

藤崎 博也¹ 阿部 賢司¹ 黒川 一滋¹ 酒巻 宏¹ 大野 澄雄²

¹ 東京理科大学 ² 東京工科大学

1. はじめに

筆者らは、先に、知的情報検索システムの対話管理モジュールに用いるための新しい対話管理手法として、ユーザおよびシステムをそれらの内部状態に着目して別々の有限状態オートマトンとしてモデル化し、両モデルの状態を考慮して対話を効率的に管理する手法を提案した [1][2]。

この手法では、対話状態をモデル化するのに比べ、それぞれのモデルが簡単になり、状態の推定が容易になる。さらに、状態に特化した言語モデルを作成することにより、ユーザ発話に対する音声認識率の向上が期待される。本報では、このような観点から、対話音声認識への効果を、状態毎の言語モデルの perplexity に着目して検証した結果を述べる。

2. ユーザの内部状態に着目した音声認識

一般の音声認識では、言語系全体を考慮した平均的な言語モデルにもとづいて入力音声を認識する。この場合の言語モデルは、その汎用性は高いが、モデルが複雑になるため、音声認識率は低い。これに対し、話題に特化した言語モデルを用いる場合には、その応用範囲が限定される反面、モデルが簡単になるため、音声認識率は高くなる [3]。さらに、発話者の内部状態に特化したモデルはより簡単になり、音声認識率は、より高くなることが予想される。

本報では、このような観点から、話題を“情報検索における対話”に限定した場合の言語モデルと、その対話におけるユーザの内部状態に特化した言語モデルとを、perplexity に着目して比較し、状態に特化した言語モデルを用いることによる音声認識への効果を検証した結果について述べる。

Recognition of user utterances in a spoken dialogue system based on estimation of internal states of the user
Hiroya Fujisaki¹, Kenji Abe¹, Kazushige Kurokawa¹, Hiroshi Sakamaki¹ and Sumio Ohno²

¹Science University of Tokyo, 2641 Yamazaki, Noda, 278-8510

²Tokyo University of Technology, 1404-1 Katakura, Hachioji, 192-0982

3. 対話モデルの概要

我々が提案した対話管理手法では、ユーザおよびシステムの内部状態を別々の有限状態オートマトンとしてモデル化し、ユーザの内部状態や発話内容を推定・予測しながら対話を効率的に管理する。

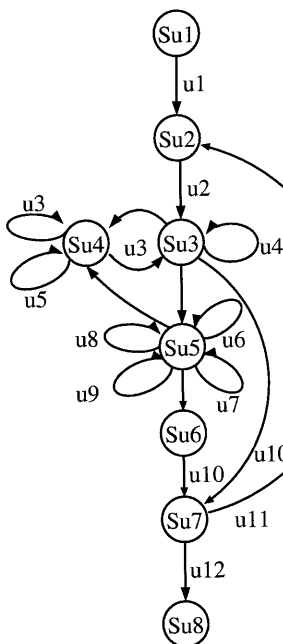
この提案手法を具体化するため、まず、情報検索を目的とした模擬対話 (100 対話 : 3417 発話) を収集・分析し、各発話の意味的なタイプ (以下、本報では、これを“発話タイプ”と呼ぶ)、および、それらの発話が生起する際のユーザ・システムの内部状態を調査した。次に、その結果に基づいて、各発話に発話タイプとユーザ・システムの内部状態に関する情報を付加したタグつき対話コーパスを作成し、そこから求められる状態遷移規則にしたがって、対話管理システムにおいて重要な役割を果たすユーザモデル・システムモデルを構築した。

これらのモデルを構築する際に用いたタグつき対話コーパスの一部を以下に示す。また、構築したユーザモデルを図 1 に示す (システムモデルも同様に記述することができる)。この図において、Su1 ~ Su8 はユーザの内部状態を、また u1 ~ u12 は各状態からの出力 (発話タイプ) 表す。各状態からの出力 (発話タイプ) は比較的限定されているため、状態毎の発話に基づいて構築した言語モデルは、対話コーパスの全発話に基づいて構築した言語モデルよりも簡単になる。

<タグつきコーパスの一例>

U1 論文を探しているんですけど。/Ou1/Su1/
S1 どのような論文をお探ですか。/Os1/Ss1/
U2 自然言語処理関係の論文が欲しいんですけど。
/Ou2/Su2/
S2 自然言語処理というキーワードで検索してよろしいですか。/Os2/Ss3/
U3 はい。/Ou4/Su3/
S3 該当する論文は以下の通りです。ご希望の論文はございますか。また、論文の詳細情報を表示することもできます。/Os4/Ss6/
U4 5 番目の論文が欲しいんですけど。/Ou8/Su6/
S4 5 番目の論文を出力します。他にご希望はありますか。/Os7/Ss7/
U5 いいえ。/Ou12/Su7/
S5 ご利用ありがとうございました。/Os11/Ss8/

(U1~U5: ユーザ役の発話、S1~S5: システム役の発話)



(ユーザの内部状態)
 Su1:初期状態
 Su2:検索要求を伝えたい
 Su3:検索結果待ち
 Su4:検索式を訂正したい
 Su5:検索結果を吟味したい
 Su6:論文を提示して欲しい
 Su7:一検索を終了したい
 Su8:終了状態

(ユーザモデルの出力)
 u1:検索開始の挨拶
 u2:検索要求
 u3:キーワード訂正要求
 u4:検索実行の許可
 u5:検索やり直し要求
 u6:他ページ表示要求
 u7:詳細表示要求
 u8:論文指定
 u9:論文指定の解除
 u10:指定した論文の要求
 u11:新規検索希望
 u12:検索終了要求

図 1. ユーザモデル

4. ユーザの内部状態毎の言語モデルと perplexity

ユーザ状態に特化した言語モデルと、対話コーパス全体から求めた言語モデルとを perplexity に着目して比較する。なお、言語モデルとしては、文字を単位とした文字バイグラム、および、形態素を単位とした形態素バイグラムを用い、それらの perplexity : P を次式 (1) にしたがって求める。

$$P = 2^H \tag{1}$$

この式における H は言語モデルのエントロピーであり、バイグラムモデルの場合には、要素 i と要素 j が連続して生起する確率 P(i, j) と、条件つき確率 P(j|i) を用いて次式 (2) のように表すことができる。

$$H = - \sum_{ij} P(i, j) \log_2 P(j|i) \tag{2}$$

対話コーパス全体から求めたバイグラムモデルの perplexity と、状態毎のバイグラムモデルの perplexity とを比較した結果を表 1、2 に示す。表 1 は文字バイグラムを用いた場合、表 2 は形態素バイグラムを用いた場合の結果である。いずれの場合も状態毎のバイグラムの方がコーパス全体から求めたバイグラムよりも perplexity が小さくなる。すなわち、提案手法にもとづいてユーザの状態を予測し、その状態におけ

る言語モデルを音声認識に利用することにより、対話コーパス全体から求めた言語モデルを用いる方法よりも音声認識率が向上することが予想される。なお、ここでは、ユーザ状態の予測結果が常に正しいことを前提として論じているが、例えば、“ユーザが検索結果に満足するか否か”といった、システムが予測しにくい部分に関しては、状態に特化した言語モデルと平均的な言語モデルとを併用する必要がある。

表 1 文字バイグラムモデルの perplexity

バイグラムの種類	perplexity
1) 平均的な文字バイグラム	3.649
2) 状態毎の文字バイグラム : a) ~ h) の平均	2.087
a) ユーザ状態 : Su1 における文字バイグラム	1.552
b) ユーザ状態 : Su2 における文字バイグラム	2.480
c) ユーザ状態 : Su3 における文字バイグラム	1.550
d) ユーザ状態 : Su4 における文字バイグラム	2.965
f) ユーザ状態 : Su5 における文字バイグラム	2.739
g) ユーザ状態 : Su6 における文字バイグラム	1.701
h) ユーザ状態 : Su7 における文字バイグラム	1.624

表 2 形態素バイグラムモデルの perplexity

バイグラムの種類	perplexity
1) 平均的な形態素バイグラム	4.393
2) 状態毎の形態素バイグラム : a) ~ h) の平均	2.326
a) ユーザ状態 : Su1 における形態素バイグラム	1.601
b) ユーザ状態 : Su2 における形態素バイグラム	2.646
c) ユーザ状態 : Su3 における形態素バイグラム	1.810
d) ユーザ状態 : Su4 における形態素バイグラム	3.267
f) ユーザ状態 : Su5 における形態素バイグラム	3.606
g) ユーザ状態 : Su6 における形態素バイグラム	1.685
h) ユーザ状態 : Su7 における形態素バイグラム	1.665

5. おわりに

すでに提案した対話管理手法にもとづき、ユーザの内部状態に特化した言語モデルを作成し、それを音声認識に利用したときの有効性を、perplexity に着目して検証した。

参考文献

[1] H. Fujisaki, H. Kameda, S. Ohno, T. Ito, K. Tajima and K. Abe: "An intelligent system for information retrieval over the internet through spoken dialogue," *Proceedings of Eurospeech'97*, vol.3, pp.1675-1678 (1997).

[2] 阿部 賢司, 飯島 岐勇, 黒川 一滋, 馬場 一嘉, 藤崎 博也, 大野 澄雄: "情報検索のための対話管理モデルとその評価," 情報処理学会第 60 回全国大会講演論文集, vol.2, pp.7-8 (2000).

[3] 大黒 慶久, 中川 聖一: "音韻認識率・文発声法・Perplexity 及び文認識率との相互関係," 信学技報, SP88-113 (1988-12)