音節連鎖統計情報のタスク適応化

6D−5

松永昭一　山田智一　鹿野清宏
NTT ヒューマンインタフェース研究所

1　はじめに

音声認識は音響情報と言語情報を用いたパターン認識の問題と考えられる [1]。その中で、タスク適応化技術は認識率を向上させるための重要な要素の一つである。従来は、話者適応化手法などが音響的適応化手法として用いられ、その有効性が確認されてきた [2]。本報告では、適応化技術の言語処理への導入を、統計的言語情報の適応化という観点から試みる。具体的には、音節の 3-gram の統計量に対して、学習用テキストを用いて、認識用タスクに適応化させる。適応化の効果はタスクの複雑度 (パープレキシティ、perplexity) を用いて検討する。

2　タスク適応化の問題設定

予め、ユニバーサルな音節連鎖の統計量 μ が与えられると仮定する。これを用いて、発声されたタスク T を認識する (図 1(1-1))。この認識性能はパープレキシティを用いて間接的に評価できる。パープレキシティは、タスクを認識する過程で言語モデルによって、予測される平均音節数であるとみなせる。そのため、パープレキシティが低くなると認識率が向上することが期待される [3]。そのため、従来は μ を作成するために、T に類似したテキストを用いなければならなかった [4]。
本稿では、T とは無関係な μ を学習テキスト M で適応化させ、T を評価する (図 2(1-2))。この時、M と T は類似したタスクを選ぶ。適応化は学習テキストに削除補間を施すことで行う。(1-2) のパープレキシティが (1-1) より低ければ、適応化の効果があるとみなす。

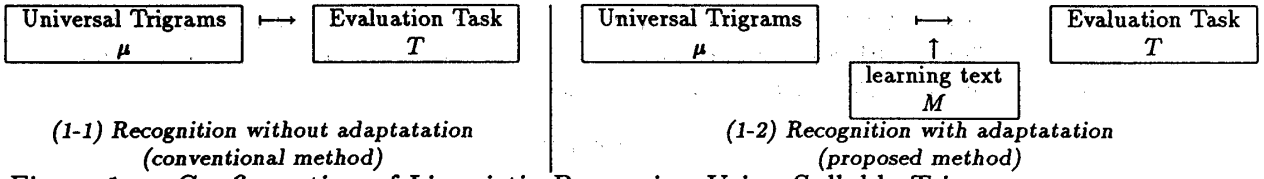

Figure 1:　Configuration of Linguistic Processing Using Syllable Trigrams

3　タスク適応化の実験と検討

一般的な μ を得るためには、非常に大規模なテキストデータを必要とする。現段階では、これは不可能なため、ある一定量のデータを削除補間法で平滑化した値で代用する [5]。すなわち、

$$\mu = \lambda_0 P_0 + \lambda_1 P(w_n) + \lambda_2 P(w_n \mid w_{n-1}) + \lambda_3 P(w_n \mid w_{n-2}, w_{n-1}) \qquad (1)$$

を基準言語モデルとする。ここで w_n は音節を表し、$P(w_n \mid w_{n-2}, w_{n-1})$ は音節列 $w_{n-2}w_{n-1}$ の後に、w_n が出現する確率。λ は重み係数であり、$\sum_{i=0}^{3} \lambda_i = 1$ である。ここでは μ_1(データ数約 54K、国際会議登録) と μ_2(約 943K、会議登録と旅行手続き) を用いる。
評価データ T は、新聞の声の欄より抜粋した 100 文節である。学習データ M は同じ文献の他の箇所より作成する。ここで適応化法として

Syllable Trigram Adaptation for Statistical Language Modeling
Sho-ichi MATSUNAGA, Tomokazu YAMADA and Kiyohiro SHIKANO
NTT Human Interface Laboratories

(3-1) M の 1-gram のみで適応 $\tilde{\mu} = \eta_0\mu + \eta_1 P(w_n)$
(3-2) M の 2-gram のみで適応 $\tilde{\mu} = \eta_0\mu + \eta_1 P(w_n \mid w_{n-1})$
(3-3) M の 3-gram のみで適応 $\tilde{\mu} = \eta_0\mu + \eta_1 P(w_n \mid w_{n-2}, w_{n-1})$
(3-4) M を平滑化した値で適応 $\tilde{\mu} = \eta_0\mu + \eta_1 m$ 、(m は、M の中で式 (1) の削除補間法で求まる 3-gram の平滑化した値)
(3-5) M の各 gram で適応 $\tilde{\mu} = \eta_0\mu + \eta_1 P(w_n) + \eta_2 P(w_n \mid w_{n-1}) + \eta_3 P(w_n \mid w_{n-2}, w_{n-1})$
を用いた (η は重み係数,P は M に関する値)。結果を表 1 に示す。この結果、(3-4),(3-5) が効果があることがわかる。

表 1. 適応化手法の違いによる perplexity (M は 2000 文節、$\mu = \mu_1$)

適応手法	なし μ(1-1)	(3-1)	(3-2)	(3-3)	(3-4)	(3-5)	M(1-1)
perplexity(T)	32.1	25.2	22.9	24.2	19.4	19.3	22.0

次に、学習データの個数を変化させて μ_2 に対して (3-4) を用いて学習を試みた。結果を表 2 上段に示す。この結果を、M を基準モデルとして作成した m により評価した結果 (下段) と比較すると、学習データが多い場合にはそれのみで基準モデルを作った場合のパープレキシティに近づくことがわかる。

表 2. 適応化学習量の違いによる perplexity ($\mu = \mu_2$)

学習データ M の文節数	0	1k	2k	5k	9k
μ を M で適応化した評価	25.2	20.9	17.8	14.9	14.1
m(M が基準情報) による評価	-	28.1	22.0	16.1	14.6

次に、3-gram の頻度別、あるいは確率別に重み係数を決定する削除補間法の効果を検討する。頻度別では、学習データでの頻度 (出現回数) がある閾値 (F) 以上のものはクラス 1 に、以下のものはクラス 2 として補間する。確率別では μ を用いた確率値が、M の 3-gram の値 ($P(w_n \mid w_{n-2}, w_{n-1})$) より高いものはクラス 1 で、それ以外をクラス 2 とする。補間は (3-4) を用いる。結果を表 3 に示す。この結果、各手法とも効果があることがわかる。確率別の手法の重み係数の値は、「学習データは評価データに類似しているため、学習データの確率値が高いものは学習データの統計量に重みを置き (M の重みが 1 に近い)、その他は基準データの統計量に重みを置いて (μ の重みが 1 に近い)、評価すると評価データのパープレキシティが下がる」という直感に合致した適応化の指針を示している。

表 3. クラス別適応化法の効果 ($\mu = \mu_1$,M は 2000 文節)

削除補間法	perplexity	クラス	学習データの割合 (%)	μ の重み	M の重み
頻度別 ($F = 1$)	16.9	1	30.6	0.000	1.000
		2	69.4	0.715	0.285
確率別	14.1	1	33.8	1.000	0.000
		2	66.2	0.011	0.989
単一	19.4	-	100	0.209	0.791

4 むすび

音節連鎖の統計情報量を削除補間法で適応化する手法について述べ、その効果を評価データのパープレキシティを用いて検討した。今後は、音声認識実験を通して適応化法を評価する。

謝辞 日頃御指導頂く古井部長、相川主任研究員を初めとする皆様に感謝致します。

< 文献 > [1]L. R. Bahl, et al:"A maximum likelihood approach to continuous speech recognition",IEEE PAMI(1983) [2]K. Shikano, et al:"Speaker adaptation through vector quantization", Proc. ICASSP(1986) [3] 村瀬、中川:" 音韻ラティス, 単語ラティス,Perplexity, 平均文長および文認識率との相互関係 " 信学技報 SP(1989) [4]T. Kawabata, et al:"Japanese phonetic typewriter using HMM Phone units and syllable trigrams", Proc. ICSLP(1990) [5]F. Jelinek, et al:"Interporated estimation of Markov source parameters from speech data", Pattern Recognition in Practice (1980)