音声言語情報処理 16－5
（1997. 5. 26）

文長を考慮した言語モデルの検討

小川 厚徳　武田 一哉　板倉 文忠
名古屋大学 大学院 工学研究科
〒464-01 名古屋市千種区不老町1
ogawa@itakura.nuee.nagoya-u.ac.jp

あらまし 連続音声認識システムの言語モデルとしてはN-gramモデルが一般に用いられている。しかし、マルコフ性を仮定したN-gramモデルでは言語確率は文長が長くなるほど小さくなる。そのため、音響モデルに対する言語モデルの重みを大きく設定するほど認識結果文章の文長が短くなることを実験で確認した。本研究では一般化したベルヌーイ試行のモデルを用いて文長を考慮した言語モデルを提案し、文認識実験によりN-gramモデル、N-gramモデルによる対数言語確率を単語数で割ることで正規化する方法との比較を行なった。その結果、提案モデルでは文長を考慮することにより単語正解率、単語正解精度ともに改善され、また、重みによらず正解に近い文長で認識結果文章を得ることができた。

キーワード N-gramモデル、文長、言語モデルの重み、一般化ベルヌーイ試行、正規化

A Language Model Based on Generalized Bernoulli Trials

Atsunori OGAWA　Kazuya TAKEDA　Fumitada ITAKURA
Graduate School of Engineering, Nagoya University
Furo-cho 1, Chikusa-ku, Nagoya 464-01 JAPAN
ogawa@itakura.nuee.nagoya-u.ac.jp

Abstract Recently the N-gram language model has been generally used in continuous speech recognition systems. But, in general, the N-gram based on Markov modeling gives lower probabilities to longer sentences. Thus when the language model is weighted heavier than the acoustic model, the recognition results tend to be shorter. In this paper we propose a language model based on generalized Bernoulli trials. Since this model considers sentence length, it shows high word correct and accuracy rates in comparison with the usual N-gram model and its normalized version. Moreover, independent of the weight for the language model, we managed to realize sentences of almost the correct length.

key words N-gram, sentence length, weight for the language model, generalized Bernoulli trials, normalize

1 はじめに

現在、連続音声認識システムにおいては、音響モデルとしてHMM、言語モデルとしてN-gramモデルが一般に用いられている。しかし、マルコフ性に基づくN-gramモデルでは、単語間遷移が行なわれるたびに確率値がかけられる。そのため、いくら自然な単語間遷移を繰り返しても、文長(本研究では1文中の単語数)が長くなると、その文章に与えられる確率は小さくなってしまう。また逆に不自然な単語間遷移を繰り返しても、文長が短い場合には、その文章には比較的大きな言語確率が与えられる。このため、音響確率が同程度の場合、音響モデルに対する言語モデルの重みを大きく設定するほど文長が短い文章が認識結果として選ばれる傾向があると予想される。

また、この重みパラメータには予備実験で求めた最適値を与えるのが一般的である。しかし、入力される音声の長さは様々であるのと、N-gramモデルでは認識結果文章の文長が重みに依存することを考えると、実験的に求めた値が常に最適値であるとは考えられない[1]。さらに、重みが最適値からずれた場合に十分な認識性能が保たれるかも保証されない。

言語モデルで文長を考慮する方法としては、最も簡単には、N-gramモデルによる対数言語確率を単語数で割ることにより正規化する方法が考えられる。これに対し、本研究では、一般化したベルヌーイ試行のモデルを用いて文長を考慮した言語モデルを提案する。そして文認識実験による検討を行なった結果、提案モデルでは、文長と過去に出現した単語(品詞)の頻度を考慮することにより、単語正解率、単語正解精度ともに改善されることが確認できた。また、重みに依存せずに正解に近い文長で認識結果文章が得られることが明らかになった。

2 N-gramモデルの問題

入力音声の音響パラメータの時系列パターン$\boldsymbol{A}$が与えられたとき、連続音声認識システムは次式で与えられる対数同時確率$\ln P_{TOTAL}(\boldsymbol{A},\boldsymbol{W})$を最大にするような単語列$\boldsymbol{W}=w_1,w_2,\cdots,w_n=\boldsymbol{w}_1^n$を認識結果文章とする。

$$\ln P_{TOTAL}(\boldsymbol{A},\boldsymbol{W}) = \ln P_{AM}(\boldsymbol{A}|\boldsymbol{W}) + \alpha \cdot \ln P_{LM}(\boldsymbol{W}) \quad (1)$$

ここで、$P_{AM}(\boldsymbol{A}|\boldsymbol{W})$は、単語列$\boldsymbol{W}$が発声された場合に音響パラメータの時系列パターン$\boldsymbol{A}$が観測される条件付き確率で、音響モデル(HMM)によって求められる。$P_{LM}(\boldsymbol{W})$は、単語列$\boldsymbol{W}$が発声される事前確率で、言語モデルによって求められる。また、αは音響モデルの重みを1としたときの言語モデルの重みである($\alpha \geq 0$)。言語モデルとしてN-gramモデルを用いた場合、言語確率は次式で与えられる。

$$P_{NG}(\boldsymbol{W}) = \prod_{i=1}^{n} P(w_i|\boldsymbol{w}_{i-N+1}^{i-1}) \quad (2)$$

(2)式からも明らかなように、N-gramモデルでは文長nが大きくなるほど言語確率$P_{NG}(\boldsymbol{W})$が小さくなる傾向にある。このため、(1)式において音響モデルの出力確率が同程度であれば、言語モデルの重みαを大きく設定するほど文長nが小さい単語列$\boldsymbol{W}$が認識結果文章に選ばれる傾向があると予想される。

3 文長を考慮した言語モデル

本節では前節で述べたマルコフ過程に基づく言語モデルの問題に対処する方法として、対数言語確率を正規化する方法と、一般化したベルヌーイ試行によりモデル化する方法について述べる。

3.1 正規化対数言語確率

正規化対数言語確率はN-gramモデルにより単語列$\boldsymbol{W}$に対して(2)式で与えられる言語確率$P_{NG}(\boldsymbol{W})$を文長nを考慮して$1/n$乗したものである。

$$P_{NG\text{-}NOR}(\boldsymbol{W}) = \left\{ \prod_{i=1}^{n} P(w_i|\boldsymbol{w}_{i-N+1}^{i-1}) \right\}^{1/n} \quad (3)$$

(3)式の両辺の対数をとると、N-gramモデルによる対数言語確率$\ln P_{NG}(\boldsymbol{W})$が文長$n$で正規化されることになる。正規化対数言語確率では単語列$\boldsymbol{W}$がどの程度自然であるかを1単語あたりでとらえることができる。しかし、文長nがいくら大きくなっても1単語あたりで評価されるので、助詞、助動詞などの頻繁に出現する品詞による挿入誤りが多発する可能性があると思われる。

3.2 ベルヌーイ試行モデル

3.2.1 一般化したベルヌーイ試行のモデル

互いに排反なr個の事象$\mathcal{A}=\{\mathcal{A}_1,\mathcal{A}_2,\cdots,\mathcal{A}_r\}$を考え($\mathcal{A}$:全事象)、事象$\mathcal{A}_j$が生起する確率を$P(\mathcal{A}_j)$とする。このとき、$n$回の独立試行において、順序は考慮せず、「$\mathcal{A}_1$が$k_1$回、$\mathcal{A}_2$が$k_2$回、$\cdots$、$\mathcal{A}_r$が$k_r$回生起」する確率は次式で与えられる。

$$P_n(k_1,k_2,\cdots,k_r)=\frac{n!}{\prod_{j=1}^{r}k_j!}\prod_{j=1}^{r}\{P(\mathcal{A}_j)\}^{k_j} \quad (4)$$

ただし、次式が成り立つ。

$$\sum_{j=1}^{r}P(\mathcal{A}_j)=1, \quad \sum_{j=1}^{r}k_j=n \quad (5)$$

(4) 式が一般化したベルヌーイ試行のモデルである。以下でこのモデルを文長を考慮した言語モデルに適用する。

3.2.2 一般化ベルヌーイ試行に基づく言語モデル

まず、unigram モデルに適用する。単語 unigram モデルは単語の生起は以前に生起した単語には依存しないと考えるモデルであるので、単語の出現を独立試行の繰り返しでモデル化可能である。すなわち、(4) 式において事象 $\mathcal{A}_j$ に単語 w_j の出現を考える。これより、単語 unigram に基づくベルヌーイ試行モデルは単語列 $\boldsymbol{W}$ に対して次式の言語確率を与える。

$$P_{UG\text{-}BER}(\boldsymbol{W})=\frac{n!}{\prod_{j=1}^{L}k_j(\boldsymbol{W})!}\prod_{j=1}^{L}\{P(w_j)\}^{k_j(\boldsymbol{W})} \quad (6)$$

ここで、$P(w_j)$ は単語 w_j の生起確率 (単語 unigram 確率)、$k_j(\boldsymbol{W})$ は単語列 $\boldsymbol{W}$ 中での単語 w_j の生起回数、L は語彙数である。また、次式が成り立つ。

$$\sum_{j=1}^{L}P(w_j)=1, \quad \sum_{j=1}^{L}k_j(\boldsymbol{W})=n \quad (7)$$

次に bigram モデルに適用する。この場合も (4) 式において事象 $\mathcal{A}_j$ に単語 w_j の生起を考える。ただし、試行間は独立でなく直前に生起した 1 単語に応じて単語 w_j の生起確率 $P(w_j)$ の値が異なると考える (この時点で独立試行とは呼べなくなる)。単語 bigram ベルヌーイ試行モデルは単語列 $\boldsymbol{W}$ に対して次式の言語確率を与える。

$$P_{BG\text{-}BER}(\boldsymbol{W})=\frac{n!}{\prod_{j=1}^{L}k_j(\boldsymbol{W})!}\prod_{i=1}^{n}P(w_i|w_{i-1}) \quad (8)$$

ここで、$P(w_i|w_{i-1})$ は $i-1$ 番目の単語として w_{i-1} が生起した場合に i 番目の単語として w_i が生起する確率、すなわち単語 bigram 確率である。$k_j(\boldsymbol{W})$ は単語列 $\boldsymbol{W}$ 中での単語 w_j の生起回数、L は語彙数である。また、次式が成り立つ。

$$\sum_{j=1}^{L}P(w_j|w_i)=1, \quad \sum_{j=1}^{L}k_j(\boldsymbol{W})=n \quad (9)$$

単語 N-gram ベルヌーイ試行モデルは通常用いられている単語 N-gam 確率に加え、$n!/\prod_{j=1}^{L}k_j(\boldsymbol{W})!$ の項により、単語列の長さと各単語の出現頻度を考慮に入れた単語 N-gram モデルの拡張モデルであると考えることができる。従って、通常の単語 N-gram が直前の $(n-1)$ 単語のみを考慮して確率値を与えるのに対し、文頭から現在までの広いコンテキストを考慮して確率値を与えることができる。また、単語 N-gram モデルを用いる場合と異なり、文長により確率値のレンジが大きく変わることはない。よって、音響・言語モデルの重みパラメータを文長に依存せずに決定することができる。

3.3 品詞カテゴリーベースの言語モデル

本研究で用いる言語モデルは品詞カテゴリーをベースとしたものである。以下に各言語モデルが単語列 $\boldsymbol{W}$ とそれに対応する品詞列 $\boldsymbol{H}=h_1,h_2,\cdots,h_n$ に対して与える言語確率を示す。なお、( ) 内は以降で用いる略表記である。

品詞 N-gram モデル

- 品詞 **unigram** モデル (Unigram)

$$P_{UG}(\boldsymbol{W},\boldsymbol{H})=\prod_{i=1}^{n}P(h_i)P(w_i|h_i) \quad (10)$$

- 品詞 **bigram** モデル (Bigram)

$$P_{BG}(\boldsymbol{W},\boldsymbol{H})=\prod_{i=1}^{n}P(h_i|h_{i-1})P(w_i|h_i) \quad (11)$$

- 品詞 **trigram** モデル (Trigram)

$$P_{TG}(\boldsymbol{W},\boldsymbol{H})=\prod_{i=1}^{n}P(h_i|h_{i-2},h_{i-1})P(w_i|h_i) \quad (12)$$

品詞 N-gram 正規化対数言語確率

- 品詞 **unigram** 正規化対数言語確率 (UG-NOR)

$$P_{UG\text{-}NOR}(\boldsymbol{W},\boldsymbol{H})=\left\{\prod_{i=1}^{n}P(h_i)P(w_i|h_i)\right\}^{1/n} \quad (13)$$

- 品詞 **bigram** 正規化対数言語確率 (BG-NOR)

$$P_{BG\text{-}NOR}(\boldsymbol{W},\boldsymbol{H})=\left\{\prod_{i=1}^{n}P(h_i|h_{i-1})P(w_i|h_i)\right\}^{1/n} \quad (14)$$

品詞ベルヌーイ試行モデル

- 品詞 unigram ベルヌーイ試行モデル (UG-BER)

$$P_{UG-BER}(\boldsymbol{W},\boldsymbol{H}) = \frac{n!}{\prod_{j=1}^{M} k_j(\boldsymbol{H})!} \prod_{j=1}^{M} \{P(h_j)\}^{k_j(\boldsymbol{H})} \prod_{i=1}^{n} P(w_i|h_i) \quad (15)$$

- 品詞 bigram ベルヌーイ試行モデル (BG-BER)

$$P_{BG-BER}(\boldsymbol{W},\boldsymbol{H}) = \frac{n!}{\prod_{j=1}^{M} k_j(\boldsymbol{H})!} \prod_{i=1}^{n} P(h_i|h_{i-1}) \prod_{i=1}^{n} P(w_i|h_i) \quad (16)$$

以上において $P(h_i), P(h_i|h_{i-1}), P(h_i|h_{i-2},h_{i-1})$ はそれぞれ品詞unigram,bigram,trigram確率、$P(w_i|h_i)$ は品詞 h_i 中での単語 w_i の生起確率、M は品詞カテゴリー数、$k_j(\boldsymbol{H})$ は品詞列 $\boldsymbol{H}$ 中での品詞 h_j の生起回数である。また、ベルヌーイ試行モデルについては (7),(9) 式に対応して次式が成り立つことを示しておく。

$$\sum_{j=1}^{M} P(h_j) = \sum_{j=1}^{M} P(h_j|h_i) = 1, \quad \sum_{j=1}^{M} k_j(\boldsymbol{H}) = n \quad (17)$$

4 文認識実験

本節では前節で述べた各言語モデルを認識システム [2] に組み込み、文認識実験により比較検討した。

4.1 実験条件

各言語モデルの基本となる確率値は品詞 N-gram モデルの確率値である。これを ATR の旅行に関する旅行会社と客の対話テキストデータベース [3] を用いて推定した。ただし、「えーと」などの間投詞や言い淀みなどの不要語はあらかじめ除いた。総文数は 7740 文、語彙数は 4784 語、品詞カテゴリー数は 27 種類である。また、品詞 bigram、trigram モデルの確率値は Katz の back-off スムーズィング法 [4] でスムーズィングした。品詞 unigram モデルについては確率値のスムーズィングは行なっていない。

音響モデルには状態数 3, 混合数 4 の文脈依存 HMM を用いた。総状態数は 12000 状態である。学習には日本音響学会の研究用連続音声データベース [5] の話者 54 人分、8128 文を用いた。表 1 に分析条件を示す。

表 1: 分析条件

標本化周波数	16kHz
量子化ビット数	16bit
分析窓	ハミング窓
フレーム長	25ms
フレーム周期	10ms
プリエンファシス	0.97
MFCC	12 次
ΔMFCC	12 次
Δパワー	1 次

4.2 実験

前述の言語モデルの学習データから、文長が 9,12,15, 18,21 である文章を各 30 文、計 150 文選び出し (平均文長 15) 音声を収録し評価セットとした。よって、この実験は言語モデルに対しては closed な文認識実験である。発声者は男性話者 1 名で、スタンドマイク (SONY ECM-23FII) を用いて防音室で収録した。そしてこの評価セットに対して、(1) 式のように音響モデルの重みを 1.0 で固定し、言語モデルの重み α を N-gram モデル、ベルヌーイ試行モデルの場合は 0.0 から 30.0 まで 2.0 刻みで、正規化対数言語確率の場合は 0.0 から 300.0 まで 20.0 刻みで変化させて文認識実験を行ない、文認識実験における単語認識率と認識結果文章の平均文長を求めた。N-gram モデル、ベルヌーイ試行モデルと正規化対数言語確率とで重みの設定のレンジが異なるのは、言語確率のレンジが異なるためである。

4.3 実験結果

4.3.1 単語認識率

まず、単語認識率を図示する。単語認識率には、挿入誤りを許す単語正解率 (%Correct) と挿入誤りを許さない単語正解精度 (Accuracy) を示す。以下の図 1,2 に品詞 N-gram モデル、品詞ベルヌーイ試行モデルについて言語モデルの重みと単語正解率、単語正解精度の関係をそれぞれ図示する。また、図 3,4 に品詞 N-gram 正規化対数言語確率について同様の関係を図示する。図 1,3 より、単語正解率ではベルヌーイ試行モデル、正規化対数言語確率ともに N-gram モデルを上回っている。しかし、図 2,4 より、単語正解精度ではベルヌーイ試行モデルが N-gram モデルを上回っているのに対し、正規化対数言語確率は下回っている。これは 3.1 節で述べたように正規化対数言語確率では挿入誤りが多発していることを示している。表

表 2: N-gram モデルからの改善 [%]

	%Corr (UG,BG)	Acc (UG,BG)
BER	+3.90, +6.38	+2.14, +3.53
NOR	+4.14, +3.67	−18.62, −15.85

表 3: (最大値−5[%]) 以上である重みの範囲

	%Corr (UG,BG)	Acc (UG,BG)
N-gram	4.0~16.0, 6.0~20.0	6.0~16.0, 6.0~20.0
BER	6.0~24.0, 8.0~26.0	6.0~26.0, 8.0~26.0

2 に N-gram モデルを基準として、ベルヌーイ試行モデル、正規化対数言語確率で単語認識率が最大値でどの程度改善されたかを示す。

4.3.2 認識結果文章の平均文長

次に、認識結果文章の平均文長を図示する。図 5 に品詞 N-gram モデル、品詞ベルヌーイ試行モデルについて言語モデルの重みと認識結果文章の平均文長の関係を示す。また、図 6 に品詞正規化モデルについて同様の関係を示す。図 5 より、N-gram モデルでは重みを大きくするほど認識結果文章の文長は短くなる傾向があることが確認できる。しかし、ベルヌーイ試行モデル (特に BG-BER) では N-gram モデルのような傾向は見られず、重みによらずに正解に近い文長で認識結果文章を得られることが分かる。この性質により、ベルヌーイ試行モデルでは、図 1,2 のように比較的広い範囲の重みで安定して高い単語認識率を得ることができ、N-gram モデルのように重みが最適値からずれた場合に急激に認識率の低下が起こるということを避けることができる。例として表 3 に N-gram モデル、ベルヌーイ試行モデルにおいて単語認識率が (最大値−5[%]) 以上である重みの範囲を示す。

また、図 6 より、正規化対数言語確率の場合も N-gram モデルとは違い、認識結果文章の平均文長は重みによらず広い範囲でほぼ一定値をとるといえるが、その値は正解の平均文長よりも大きい。このことからも、正規化モデルでは挿入誤りが多発していることが分かる。

なお、図 1~6 は評価セット全 150 文 (平均文長 15) に対して示したものであるが、評価セット中の各文長 (9,12,15,18,21) ごとに図示しても同様の傾向が得られた。

5 まとめ

一般化したベルヌーイ試行のモデルを用いて文長及び過去に出現した単語 (品詞) の頻度を考慮した言語モデルを提案し、文認識実験により、N-gram モデル、N-gram モデルによる対数言語確率を文長で正規化する方法との比較を行なった。その結果、提案モデルでは言語モデルの重みを大きくするほど認識結果文章の文長が短くなるという N-gram モデルの傾向を抑えることができ、重みに依存せずに正解に近い文長で認識結果文章を得ることができた。また、この性質により、比較的広い重みの範囲で安定して N-gram モデルより高い単語認識率を得ることができた。

これらの結果から、1) 一般化ベルヌーイ試行に基づく言語モデルはより広い範囲のコンテキストを考慮することで高い言語制約を果すことができること、2) 文長を考慮することで、音響・言語モデルの重みパラメータを文長に対して頑健に決定できることが確認された。

今回の実験では品詞カテゴリー数が 27 種類とかなり粗いクラスタリングの下で実験を行なった。そのため認識率も非常に低いレベルとなっている。今後はもう少しクラスタリングの精度を上げて、より高い認識率のレベルで検討を行なう予定である。

参考文献

[1] 小川, 武田, 板倉: 「文長を考慮した言語モデルの検討」, 秋季音講論集,pp.39-40 (1996).

[2] S.J.Young,P.C.Woodland,W.J.Byrne:HTK V1.5 User Manual,Entropic Research Laboratory, Inc. (Dec.1993).

[3] 江原, 井ノ上, 幸山, 長谷川, 庄山, 森元: 「ATR 対話データベースの内容」, ATR Technical Report,ATR 自動翻訳電話研究所 (1990).

[4] S.M.Katz:"Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer", IEEE Trans.,ASSP-35,No.3,pp.400-401 (Mar.1987).

[5] 小林, 板橋, 速水, 竹沢: 「日本音響学会研究用連続音声データベース」, 音響学会誌,Vol.48,No.12, pp.888-893(1992).

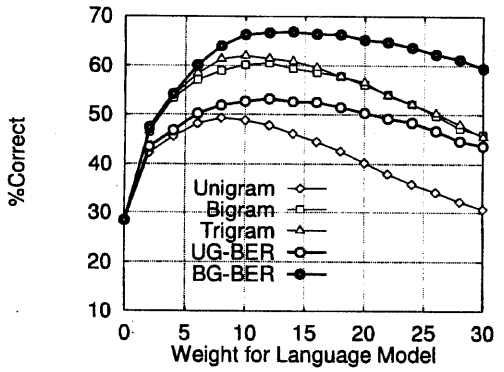

図 1: 言語モデルの重みと単語正解率の関係 (品詞 N-gram モデル, 品詞ベルヌーイ試行モデル)

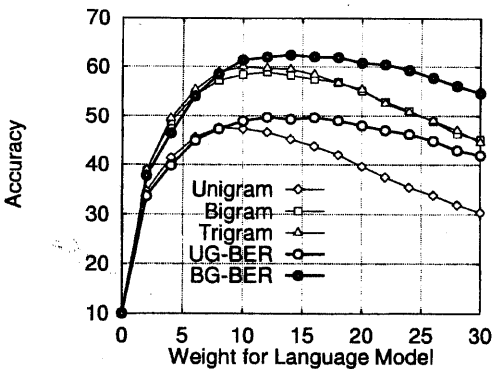

図 2: 言語モデルの重みと単語正解精度の関係 (品詞 N-gram モデル, 品詞ベルヌーイ試行モデル)

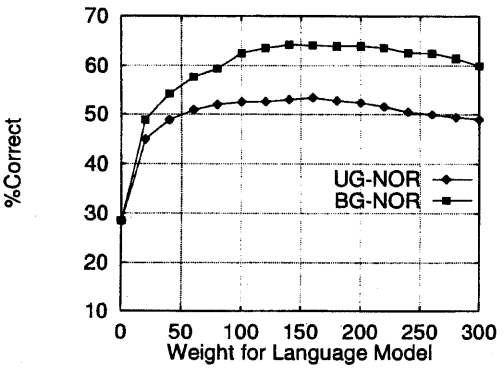

図 3: 言語モデルの重みと単語正解率の関係 (品詞 N-gram 正規化対数言語確率)

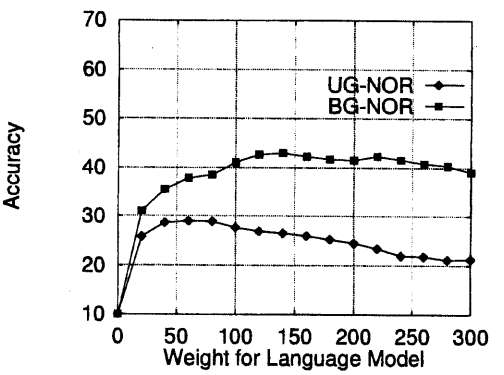

図 4: 言語モデルの重みと単語正解精度の関係 (品詞 N-gram 正規化対数言語確率)

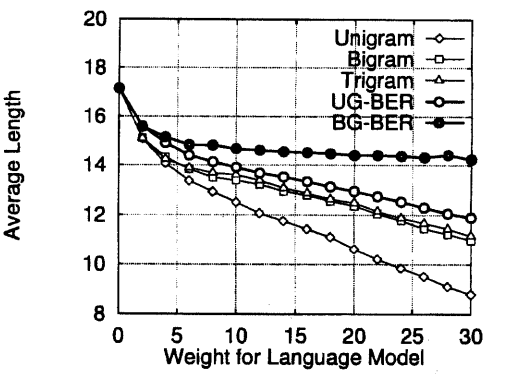

図 5: 言語モデルの重みと認識結果文章の平均文長の関係 (品詞 N-gram モデル, 品詞ベルヌーイ試行モデル, 平均文長 15)

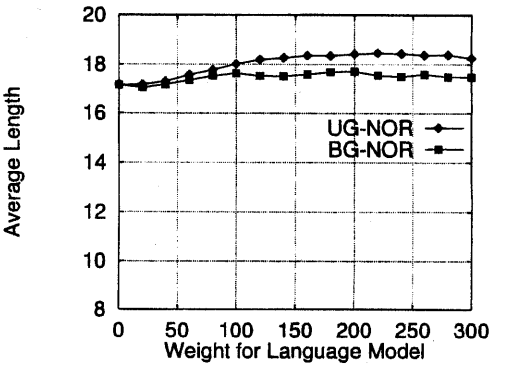

図 6: 言語モデルの重みと認識結果文章の平均文長の関係 (品詞 N-gram 正規化対数言語確率, 平均文長 15)