

認識文字候補からの高速予備選択法

2H-6

菅原一秀, 野崎広志

日本アイ・ビー・エム株式会社 東京基礎研究所

1 はじめに

オンライン文字認識装置において、単語の部分的な文字の入力からいくつかの候補単語を推定、表示し、それらの中からユーザーが選択できるようにしたい。本稿ではこの候補単語の推定のことを予備選択と呼ぶ。この予備選択により、高速に候補を部分的入力からも推定できるようになるので、ユーザーが速くペン入力を使用できるようになる。

予備選択のためには、頻度情報を持った語彙集と、一文字ずつの文字認識のスコア付の認識結果を使う。認識装置には誤認識はつきものであるが、選択すべきリスト中に正解単語が無いのは非常に不都合である。そこでこの予備選択時にはたとえ認識結果からの組み合わせからは出てこない単語でも、それが一文字の置換程度で得られ、かつ、頻度の高いあるいは語彙集に載っているものならば、結果のリスト中に残るような、認識の耐雑音性ともいべき仕組みが必要である。また、予備選択は軽い処理である必要がある。これに時間がかかるようでは次の文字を入力した方が早いことになり意味がなくなる。

本稿では、この2つの点、認識の耐雑音性と、処理の軽さを実現する手法として、単語ごとの文字の出現行列を使い、認識装置からの一文字ごとの認識結果に対しその文字が各単語に含まれているかどうかを調べるものを提案する。さらにインプリメンテーション、予備実験の結果を示す。

2 予備選択の方式と従来技術との比較

ユーザーが文字を入力するにしたがって、文字認識装置からはスコア付の認識候補列が送り出される。予め各単語についてどの文字が出現するかの行列を作成しておき、文字認識のスコア付候補が入力されるたびにその行列を参照してその文字が出現する単語に対しスコアを積算する。この場合、各単語に対して語彙集から得た頻度情報に基づいた初期値を与えることもできる。アプリケーションプログラムは適当な時に候補列選択の要求を出し、上記積算スコアの上位の単語から要求され

た個数だけのものを出力する。従来、1) 認識対象文字を類別し、認識対象単語もそれらの類名で表し探索の範囲を狭める[1]手法が提案されている。また、音声認識の分野では2) 認識対象単語について時間情報を無視した音素の分布を作成し、高速に予備選択を行う方法がある。[2] 手法1)では類名からさらにそこに含まれる単語を導くことが必要であり、その際類別の具合により一つの類に含まれる単語数がリストとしてユーザーに提示するのに望ましい程度に小さくはならない恐れがある。手法2)はあるフレーム(音響分析のための時間枠)では候補が1個のみ与えられることを仮定しており、複数の候補がスコア付で与えられることに対応していない。本稿で提案する方法は[2]の方法を拡張して、複数のスコア付の候補を受け入れることができるようにしたものである。

3 処理方式

3.1 出現行列の作成

文字認識装置で入力が不完全だったり、(後処理を含まない)認識結果に誤りがあるときでも語彙集に基づきもっともらしい認識候補単語を高速に提示するために語彙集中の各単語 i と文字(ここでは仮名) j について出現行列 W

$$W = w(i, j) = \begin{cases} 1 & i \text{ が } j \text{ を含む} \\ 0 & i \text{ が } j \text{ を含まない} \end{cases}$$

を予め作成しておく。

3.2 スコアの初期化と積算

一連の入力を受け入れる前に各単語 i のスコア $s(i)$ を初期化する。初期化スコアは、語彙の頻度情報を使う場合はそれを認識装置のスコアに見合う程度の範囲の値を持つ確率値に直しそれを対数に変換したものを、頻度情報を使わない場合はゼロを使用する。各入力文字に対して認識装置はスコア付きの認識候補列

$$(j(t), k(t)), t = 0, \dots, N-1 \text{ (順位)}$$

$j(t)$ は t 番目の認識候補の文字

$k(t)$ は t 番目の認識候補のスコア

を出力する。これに対して各単語 i のスコアの増分を次式で計算する。

$$\sum_{t=0}^{N-1} w(i, j(t))k(t)$$

3.3 候補単語の出力

アプリケーションプログラムより候補単語を出力せよとの要求が来たら、各単語をスコアでソートし、求められた数の上位の単語リストを出力する。アプリケーションプログラムはこのリストをユーザに提示し、選択させる。

3.4 位置情報の利用

ここまでは出現行列を、単語内の文字の出現位置を無視して一つだけ作るとして説明してきたが、より精度の高い予備選択を行なうために、位置情報も採り入れた出現行列 W_p ($1 \leq p \leq P$ 、 p は単語中での文字の出現位置、 P は打ち切り単語長) を作成し、使うこともできる。

$$W_p = w_p(i, j) = \begin{cases} 1 & i \text{ の } p \text{ 番目の文字が } j \\ 0 & i \text{ の } p \text{ 番目の文字が } j \text{ でない} \end{cases}$$

$p = P$ の時は P 番目以降の文字すべてを対象とする。

この場合スコアの積算は対応する位置の出現行列を使って行なう。位置情報をそのまま使うと出現行列に必要な記憶域が、位置情報を使用しない場合に比べて P 倍になってしまう。折衷案としては単語を二分割して、先頭二文字とその他に分け、出現行列を2つ作成して使用することもできる。

3.5 例

出現行列を2つ使用した場合の例を示す。認識結果が表1のようだったとする。但し括弧内はスコアである。このときに上位幾つかの単語のスコアの例(「ヤサモト」

表 1: 認識結果とスコア

位置 順位	1	2	3	4
1	ヤ(10)	マ(10)	モ(10)	ト(10)
2	ア(5)	サ(5)		
3	カ(3)			

は辞書に無いとする) を表2に示す。

表 2: 上位幾つかの単語のスコア

ヤマモト	40
ヤマトモ	40
アマモト	35
カマモト	33
アサモト	30
アサトモ	30
イマモト	30
:	
カサモト	28
:	
ハシモト	20
:	

この例で見られるように本稿で提案した方式は認識候補文字に含まれていない文字を持つ単語、「イマモト」や「ハシモト」等もある程度のスコアを持たせて出力することができる「耐雑音性」を持っている。上の例では頻度情報を使用していなかったが、使用した場合には「ハシモト」等はずっと上位に来ることが考えられる。本方式は処理が簡単なので大きな認識対象語いに対しても高速に実行することができる。約30,000単語からなる認識対象語いに対しP S/V (CPU: 486DX2, 66MHz) で実験を行ったところ、1候補(とスコア)当たり全単語に対してのスコアの積算が6ms、上位100候補の選出が50msでありこの手法が実用上問題ないほど高速であることが確かめられた。

4 まとめ

オンライン文字認識装置において、単語の部分的な文字の入力からいくつかの候補単語を推定、表示し、それらの中からユーザーが選択できるようにすることを考えた。認識の耐雑音性と、処理の軽さを考慮して、単語ごとの文字の出現行列を使い、認識結果の文字が各単語に含まれているかどうかを調べる手法を提案した。予備実験の結果実用上問題ない速度で実行できることがわかった。今後はこの手法をサブワード単位の音声認識に使用することを検討していく。

参考文献

- [1] 田中栄一、川井さち子、外山孝典: “誤りのある音素記号列訂正のための計算手数の軽減法”, 情報処理学会論文誌, Vol. 26 No.6, pp. 1156-1158, Nov. 1985
- [2] 菅原一秀: “分割ラベルヒストグラムによる単語の予備選択法”, 日本音響学会昭和63年度春季研究発表会講演論文集, pp.235-236, March, 1988