

カナ姓名のペン入力後処理における予測処理について

2H-1

野崎広志 伊東伸泰

日本アイ・ビー・エム(株) 東京基礎研究所

1. はじめに

テキスト入力手段としてペン入力インターフェースに対する需要は高いと考えられているが、普及を妨げている理由の一つとして、手書き文字を意図通りに認識しない認識精度(正解率)の低さがあり、この結果ペン入力の使い勝手を悪くする要因ともなっている。正解率向上のために、認識処理と言語処理を組み合わせ、認識装置からの結果(候補)を言語知識と照合することで有意味な候補を作成する処理(後処理)が必要とされるが、OCR入力処理において後処理部は言語知識との照合を行うためだけに使われており、入力を推測(予測)するためには使われていない。

本論文では帳票の姓名記入欄にペンでカナ入力する際の後処理を想定し、後処理部を、単語としての照合以外にも文字の予測および単語の予測を行う機能を持った言語知識処理部として作成したので報告したい。この予測機能により、部分的な手書き入力から予測結果を選び、予測結果にないときにだけ手書き入力するというインターフェースが可能になる。

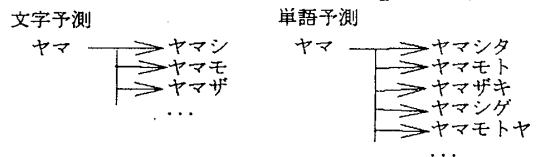
2. 予測処理について

予測処理により、ペン入力で全部の文字を書かずとも、部分的な入力の後は次々に予測結果から正解を選択するだけの操作が可能になるので、ペン入力インターフェースの向上が期待できる。しかし単に多数の予測結果を全て提示すれば良いというわけではなく、例えば、候補数が

非常に多い中から正解を選ぶ操作はかえって負担を強いる(スクロール等)ことになる。そこで、ペン入力における予測処理として以下の点に留意して設計した。(1) 確からしい順に予測結果を返せること、(2) 提示可能な数(その他の基準を満たすもの)に絞れること、(3) 状況に応じて予測範囲を拡大・縮小できること、(4) ペン入力速度に追従できること。

また、予測する文字列範囲の違いから、文字予測と単語予測という用語を使うこととする(図1)。

図1. 予測例(「ヤマ」から)



3. 予測処理の構成

予測処理を行う場合、既に入力された左部分列に対する認識結果 T_j ($n-1$ 文字)から、次の文字 C_n を含めた文字列 $C_1C_2 \dots C_{n-1}C_n$ の確率 $P(T_j, C_n)$ を求めて、この順に候補を提示すればよい。

$$P(T_j, C_n) = P(T_j) \times P(C_n | T_j) \quad (1)$$

ただし、確率 P は認識装置からの確率 P_R と言語情報からの確率 P_L を統合したものであり、またベイズの定理から

$$P(T_j) = P(T_j | x) = P_R(x | T_j) \times P_L(T_j) / V$$

$$(V = \sum P_R(x | T_j) \times P_L(T_j))$$

ただし、 x は認識系における入力の特徴ベクトル。 $P(C_n | T_j)$ は認識に依存しないので

$$P(C_n | T_j) = P_L(C_n | T_j)$$

よって式(1)は

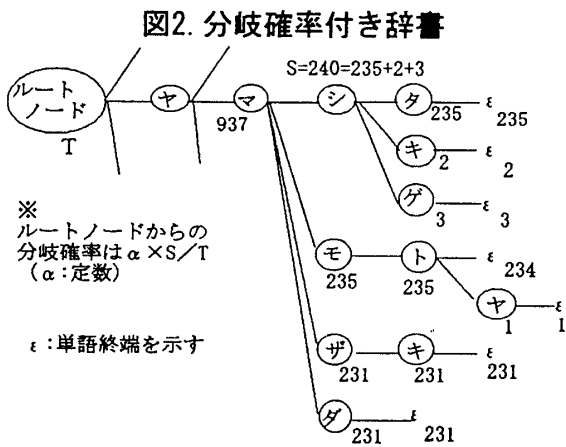
$$P_L(C_n | T_j) \times P_L(T_j) \times P_R(x | T_j) / V$$

したがって、辞書上には $P_L(C_n | T_j)$ と $P_L(T_j)$ の確率値を計算できる値を持っていけばよい。また $P_R(x | T_j)$ は式(2)で近似する。

$$P_R(x | T_j) = \beta \sum \text{Score}(C_i) \quad (2)$$

(Scoreは認識系が出力するゆう度)

このため、分岐確率付辞書を作成した。これは図2に示すようなトライ構造型辞書である。分岐確率は、正規化した単語の出現頻度をルートノードに向かって各ノードにおける総和(スコア)Sを求め、ルートノードでの総和をTとしたときに、 $\alpha \times S/T$ (α :定数)。辞書作成には姓・名の各々が約17万件(異なり読みは約3万、1.6万)のデータを用いた。



4. 予測処理の流れ

1. 認識装置からの候補文字列 T_j をビームサーチにより絞り込む。
2. 候補文字列 T_j をキーとして分岐確率付辞書を検索し、到達したノードでの分岐確率($P(T_j)$)とそのノードからの各子ノード C_n の分岐確率($P(C_n|T_j)$)を得る。
3. 候補数がある個数 (N_{max} : 提示可能な最大候補数)を超えるなら絞り込み、最低でもある基準 S_{min} に達しない候補または第1位の候補からある基準 S_{dif} を超えている候補は落とす。
4. 各 T_j に対する予測結果をソート。
5. 単語予測を行う場合には、単語へ展開した場合にある総数 N_{chd} を超えるなら単語予測を取り消す(単語予測を取り消す場合は、文字予測のみとする)。

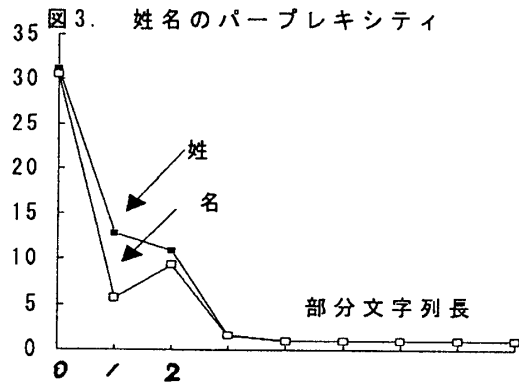
また単語へ展開しながら N_{max} 個以内からこぼれた候補のスコアとその時点での候補のスコアの比率がある値 S_{rat} より小さくなるなら単語予測を取り消す。ただし、

単語に展開しても候補数が N_{max} 個以内に収まるなら展開する。

5. 評価実験と考察

5-1. 予測率の検証

左部分列($n-1$ 文字)が与えられたときの次の1文字を含めた部分単語の数がどれくらいか、情報論的な意味での平均分岐数(Perplexity) $2^{H(n)}$ を求めたので図3に示す。 $H(n) = -\sum P(T_k, C_n) \log P(C_n|T_k)$ 。何も書かない場合を除いて $n-1$ 文字の左部分列が確定していれば10程度でありリストとして提示可能な範囲であった。



6. おわりに

帳票でのカナ姓名入力を例として、ペン入力後処理において予測処理の実現および予測率の検証による有効性を示した。今後の拡張として、カナ以外の文字種への拡張、単語から文節への拡張などの他に、予測結果を利用した文字認識そのものの精度向上等が考えられ、使い勝手の向上以外にも予測処理の重要性が増すものと思われる。

参考文献

[1] 西野文人: "文字認識における自然言語処理", 情報処理, pp. 1274-1280 (1993-10)
 [2] 古和田孝之他: "名前に対する文字認識後処理手法", 情報処理学会第48回全国大会, 2V-5 (1994)