統計的言語モデルに基づく電子カルテ入力支援システムの開発

中村明[1)]　川尻博光[1)]　金川誠[1)]　松本忠博[2)]　池田尚志[2)]　速水悟[2)]　紀ノ定保臣[3)]

1）三洋電機㈱ ヒューマンエコロジー研究所
2）岐阜大学 工学部 応用情報学科
3）岐阜大学大学院 医学系研究科 医療情報学分野

1. はじめに

近年、医療機関において電子カルテシステムの導入が進みつつある。カルテを電子化することにより、診療情報の迅速な共有や症例の検索・分類などの二次利用が可能となり、医療の質向上につながることが期待されている。しかし一方で、操作性に関しては改良が必要との指摘もある。

医療現場では、限られた時間内で多くの患者を診察してカルテを入力する必要があり、できる限り迅速に入力したいというニーズが極めて高い。一般的には汎用のかな漢字変換システムに医療用語辞書を追加して利用されているが、病名・薬品名等の専門用語の中には長い単語も多いため、すべての読みを正確に入力することの負担は大きく、すでに電子カルテを導入したユーザからは入力システムに対する改善の要望が根強い。一方、電子カルテ導入を検討中の医療機関の中には、電子化によりかえって入力操作が煩雑になることを懸念し、導入を躊躇するところもある。電子カルテ導入後の入力負荷を考察した文献[1]では、課題のひとつとして日本語入力システムの改善を挙げている。

長年、紙カルテでの入力・保存に慣れ親しんできた経緯もあり(1999 年までは紙媒体での保管が義務付けられていた)、特に中小規模の診療所では依然としてキーボード入力を苦手と感じるユーザも少なくない。このため音声認識や手書き文字認識を用いたカルテ入力インタフェースも提案されているが、幅広く利用されるには至っていない。すなわち、カルテ入力の負担を軽減することは、電子カルテシステムの普及を促進する上で重要な課題となっている。

テキスト入力の負荷を軽減する技術として予測入力(predictive text entry)があり(文献[2])、特にキーの数が少ない携帯電話などで広く利用されている(文献[3])。手書き文字認識と予測入力を組み合わせた入力方式に関する研究(文献[4])や、重度身障者の言語入力を支援するユニバーサル技術として予測入力を用いた研究(文献[5])も報告されている。筆者らの一部による文献[6]では、電子カルテに予測入力を用いて入力負荷を削減している。

本稿では、電子カルテシステムにおけるテキスト入力効率の向上を目的として、言語モデル(*N*-gram モデルおよびキャッシュモデル)に基づく予測入力システムを開発し、カルテ入力を想定したシミュレーション実験により評価を行った内容を報告する。以下、2 章で予測入力の定式化を行い、3 章でシステムの概要、4 章で言語モデルの構築について述べる。そして 5 章で実験結果を示し考察を行う。

2. 予測入力の定式化

入力文字列から単語を予測する問題は、音声認識・機械翻訳等と同様、観測された入力パターン Y から元の情報源 W を推定する復号問題として定式化できる。

$$\hat{W} = \arg\max_{W} P(W \mid Y) \tag{1}$$

すなわち、事後確率 $P(W|Y)$を最大化する語 W が最良の候補となる。ベイズの定理より

$$P(W \mid Y) = \frac{P(Y \mid W)P(W)}{P(Y)} \tag{2}$$

であり、$P(Y)$は W に依存しないため $P(W|Y)$を最大化することは $P(Y|W)$ $P(W)$の最大化と等価である。

$$\hat{W} = \arg\max_{W} P(Y \mid W)P(W) \tag{3}$$

音声認識では Y は音声パターン、かな漢字変換では Y は読み文字列となる。予測入力では一般に Y は語 W の部分文字列となる。予測入力における部分文字列 Y としては、語 W の表記の前方部分列、語 W の読みの前方部分列、語 W の子音列などさまざまな種類が考えられるが、本稿では通常のキーボードによる日本語入力を想定し、Y が語 W の読みの前方部分列である場合のみを扱う。

式(3)において $P(W)$は言語モデルであり、一方、認識誤りや入力誤りがないと仮定すると $P(Y|W)$は 1.0 となるため、予測入力では原理的には言語モデルのみによって精度が決まる(ただし実際には、後述するような候補リストの最適化など、操作系を含めたシステム全体の設計の良し悪しがユーザの使い勝手に大きく影響する)。

3. 提案システムの概要

3.1. システム構成

図 1 にシステム構成を示す。本システムは PC 上で MS-IME や ATOK など既存のかな漢字変換システムとともに動作する。動作 OS は WindowsXP および Windows2000 である。

I/F 制御部はユーザのキー入力を常に監視しており、かな漢字変換が ON の場合に限り、キー入力のたびに入力文字列を予測エンジンに送る。予測エンジンは言語モデルに基づいて入力文字列から予測候補としてふさわしい単語のリストを生成し画面に表示する。本システムでは、入力文字列は単語の読みの前方部分列であるため、読みが入力文字列に前方一致する語が予測候補となる。

ユーザは予測候補リスト中に入力したい単語があれば上下矢印キーにより選択し、なければ次の読み文字を入力する。予測候補リストは 1 文字追加入力されるたびに更新される。すべての読みを入力しても候補リスト中に入力したい単語が現れない場合には、変換キー(スペースキー等)の押下によりかな漢字変換候補が表示され、予測候補リストは非表示となる。すなわち、本予測入力システムは既存のかな漢字変換システムとともに協調して動作する。

本システムでは、予測候補の選択結果とともに、かな漢字変換候補の選択結果も入力履歴として記憶する。入力履歴は次節で述べるキャッシュ確率を算出する際に参照される。したがって言語モデルにない語句でも、一度かな漢字変換

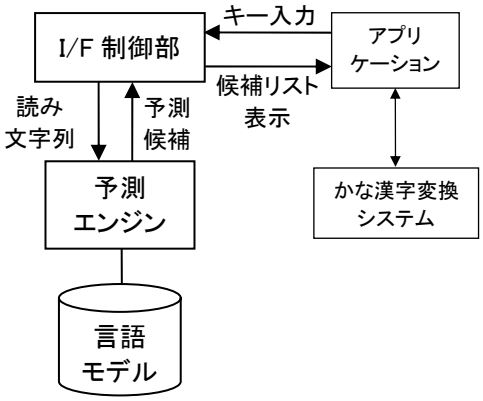

図1. システム構成

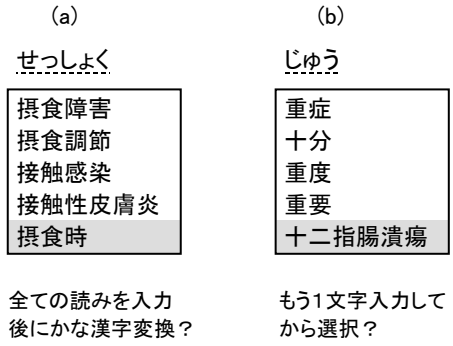

図2. 入力効率向上に寄与しない予測候補の例

により入力すれば、次回以降、読みの一部から予測することが可能となる。

3.2. 言語モデル

本システムで用いる言語モデルは、次式のように *N*-gram モデルとキャッシュモデルを線形結合したモデルである。

$$P(w_i \mid w_1^{i-1}) = \lambda P_M(w_i \mid w_{i-N+1}^{i-1}) + (1-\lambda) P_C(w_i \mid w_{i-L}^{i-1}) \quad (4)$$

$P_M(\cdot)$、$P_C(\cdot)$はそれぞれ *N*-gram 確率、キャッシュ確率であり、*N* は *N*-gram の次数、*L* はキャッシュ長、λ は *N*-gram とキャッシュの混合比を表す。

N-gram 確率は直前の(*N*-1)単語列 w_{i-N+1}^{i-1} に続いて単語 w_i が現れる条件付確率であり、4 章で後述するように実際のカルテから抽出したテキスト、および Web より収集した医療分野のテキストから *N*-gram を構築した。キャッシュ確率は直前の *L* 単語列における単語 w_i の出現確率であり、一般に次式で与えられる($\delta(\cdot)$はクロネッカーの δ 関数)。

$$P_C(w_i \mid w_{i-L}^{i-1}) = \frac{1}{L}\sum_{l=1}^{L}\delta(w_i, w_{i-l}) \quad (5)$$

ただし本システムでは、キャッシュ中に2単語列 w_{i-1}^{i} が出現する場合はキャッシュ内で w_{i-1} に続いて w_i が現れる条件付確率を、またキャッシュ中に3単語列 w_{i-2}^{i} が出現する場合はキャッシュ内で w_{i-2}^{i-1} に続いて w_i が現れる条件付確率をキャッシュ確率とする。

$$P_C(w_i \mid w_{i-L}^{i-1}) = \begin{cases} \dfrac{1}{L-2}\left\{\sum_{l=1}^{L-2}\delta(w_{i-2}^{i}, w_{i-l-2}^{i-l})\Big/\sum_{l=1}^{L-2}\delta(w_{i-2}^{i-1}, w_{i-l-2}^{i-l-1})\right\} & if\ P_C(w_{i-2}^{i}) > 0 \\ \dfrac{1}{L-1}\left\{\sum_{l=1}^{L-1}\delta(w_{i-1}^{i}, w_{i-l-1}^{i-l})\Big/\sum_{l=1}^{L-1}\delta(w_{i-1}, w_{i-l-1})\right\} & else\ if\ P_C(w_{i-1}^{i}) > 0 \\ \dfrac{1}{L}\sum_{l=1}^{L}\delta(w_i, w_{i-l}) & else \end{cases} \quad (5')$$

3.3. 予測候補リストの最適化

予測入力システムでは、単に言語モデルに基づいて生起確率の高い順に予測候補を提示するだけでは、実際の入力効率改善には結びつかないケースが多々、発生する。たとえば、図2(a)の例では入力文字列「**せっしょく**」に対して1位から5位までの予測候補「**摂食障害**／**摂食調節**／**接触感染**／**接触性皮膚炎**／**摂食時**」が得られている。「**摂食障害**」を入力したい場合であればここで1位の「**摂食障害**」を選択すればよい。しかし「**摂食時**」と入力したい場合、5位の「**摂食時**」を選択するよりも、残りの読み「**じ**」を入力後にかな漢字変換を行ったほうが効率よい可能性がある。

実際にどちらが効率がよいかは、予測候補選択の操作仕様、かな漢字変換の精度等に左右されるが、入力効率改善に結びつかないと予想される予測候補は提示しないほうが望ましい。そこで、現在の入力読み文字列を y とし、予測候補 w を今選択するのに必要な操作コストを $t_s(w|y)$、残りの読みをすべて入力してからかな漢字変換により w を直接入力するのに必要な操作コストを $t_d(w|y)$とする。そして、

$$t_d(w|y) \leqq t_s(w|y) \quad (6)$$

が成り立つ場合には候補リストから w を削除する。このように、予測を用いないほうが望ましい語を候補から除外する考え方は文献[5]に示されているが、本研究ではこれを日本語入力に適用している。なお、コスト t_s および t_d は、候補選択や読み入力に必要なキータッチ数、あるはこれらの選択・入力操作に必要となる所要時間を見積もることによって算出する。

次に図2(b)の例では、入力文字列「**じゅう**」に対して1位から5位までの予測候補「**重症**／**十分**／**重度**／**重要**／**十二指腸潰瘍**」が得られている。「**十二指腸潰瘍**」と入力したい場合、通常、この時点で5位の「**十二指腸潰瘍**」を選択する。しかし、ここでは選択せずに次の読み「**に**」(あるいはローマ字入力であれば「n」)を入力して予測候補を絞り込めば「**十二指腸潰瘍**」は1位となるため、この時点で選択したほうがトータルの操作コストが少なく済む可能性がある。

そこで式(6)のときと同様、現在の入力読み文字列を y、予測候補 w を今選択するのに必要な操作コストを $t_s(w|y)$とし、さらに1文字追加入力後の読み文字列を y' として、

$$t_s(w|y') + \triangle t_i(y, y') \leqq t_s(w|y) \quad (7)$$

が成り立つ場合には候補リストから w を削除する。$\triangle t_i(y, y')$ は読み文字列 y' の入力コストと y の入力コストとの差分を表し、y' の最後の1文字を入力するのに必要なコストを意味する。したがって式(7)の右辺は入力読み文字列が y の時点で w を選択するコスト、左辺は1文字追加入力してから w を選択するコストを表す。このように1文字追加入力後に選択する場合とのコスト比較を行って不適切な候補を削除しておくことにより、ユーザが効率の悪い手順で候補選択を行うのを未然に防ぐことができる。また、式(6)による候補削減と同様、入

力効率改善に寄与しない候補を削除した分、他の有効な候補を提示することができる。

本提案システムでは、式(6)および式(7)の2つの条件に基づいて各予測候補の妥当性を評価し、不適切な候補を削除することにより予測候補リストの最適化を行う。これによって入力効率の改善が期待できる。

4. *N*-gram モデルの構築

カルテ文書の入力支援を目的とするため、実際のカルテ文書、および Web より収集した医療文書をコーパスとして *N*-gram モデル(*N*=2,3)を構築した。以下に概要を示す。

［コーパス 1：カルテ文書］

岐阜大学医学部付属病院総合診療部の内科系外来カルテ 3559 件から、SOAP[1]区分が S の項目（主訴・現病歴など主観的情報を自由書式で記述した項目）を抽出（のべ形態素数 898635、語彙数 23572）

［コーパス 2：Web 文書］

内科系の病名 1518 個をキーワードとして Web 検索を行い、各キーワードに対し最大100位までの Web 文書を収集した後、各キーワードとその共起語の出現傾向に基づき段落単位で不要部分を除去（のべ形態素数約 1462 万、語彙数 137466）

コーパス1・2とも、まず日本語文解析システム ibukiC（文献[7]）に医療分野の専門用語など約11万 5000 語を追加し、これを用いて形態素解析を行った。その後、*N* 単語列(*N*=2,3)の頻度をカウントし、カッツ・スムージングにより平滑化を行って bi-gram および tri-gram を構築した。メモリ容量削減のため、tri-gram では頻度1の要素を捨ててから平滑化を行っている。

なお、個人情報保護に配慮し、カルテ文書を実験に用いる際には、まず医師の手により個人を特定し得る項目・内容を全て削除した後、病院内に設置した実験用 PC を用いて、医師の監督の下で作業を行った。

5. 評価実験

5.1. 評価指標

以下の評価指標により、本提案システムでの入力効率改善効果を評価する。

［相対入力文字数］

予測入力を用いた場合の入力読み文字数を、予測入力なしの場合を1として相対的に表した値。提示する候補数 *k* によって結果は異なる。

$R_C(k)$ = (入力読み文字数)／(全読み文字数)　　(8)

［相対キータッチ数］

予測候補の選択・確定や、かな漢字変換に必要なキー操作を含めた総キータッチ数を、予測入力なしの場合を1として相対的に表した値。相対入力文字数同様、提示する候補数 *k* によって異なる。

$R_K(k)$ =
(キータッチ数)／(予測なしでのキータッチ数)　　(9)

本システムでは上下矢印キーで予測候補を選択し、Enter キーにより確定するため、*n* 位候補の選択に必要なキータッチ数は（*n*+1）である。一方、かな漢字変換候補の選択に必要なキータッチ数は実際には変換精度に依存するが、ここでは変換キー1回と Enter キー1回の計二回のキータッチで選択できるものと仮定する。また読み文字の入力に関しては、ひらがな1文字につき二回のキータッチとして計算する。

［テストセットパープレキシティ］

一般的な言語処理システムと同様、テストデータに対する言語モデル自体の性能を表す。

$$TPP(D|M) = \quad 2^{H(D|M)} \qquad (10)$$

（*M* は言語モデル、$D=(w_1 \ldots w_Q)$はテストデータであり
$H(D|M) = -(1/Q)\log P_M(w_1 \ldots w_Q)$）

5.2. 実験結果

カルテ入力を想定したシミュレーションにより、本提案システムの評価実験を行った。実験に用いたテストデータは岐阜大学病院総合診療部の外来カルテ 890 件であり、すべて *N*-gram モデル構築に用いたカルテ文書（コーパス1）とは別のデータである（のべ形態素数 204254、語彙数 11812）。*N*-gram 構築時と同様、医師の手により個人を特定し得る内容・項目を全て削除した後、病院内の実験用 PC 上で医師の監督の下、実験を行った。

表1に結果を示す。(a)はカルテから構築した *N*-gram を用いた場合、(b)は Web から構築した *N*-gram を用いた場合の結果である。R_C、R_K、*TPP* はそれぞれ相対入力文字数、相対キータッチ数、テストセットパープレキシティを表し、「最適化あり／なし」は 3.3 節で述べた不適切な予測候補の削除を「行う／行わない」を表す。*N*-gram とキャッシュの混合比 λ =0.5、キャッシュ長 *L*=10000、提示する予測候補数 *k*=10 とした。

(a)(b)いずれの場合も、候補リスト最適化により不適切な候補を削除することにより、入力文字数は増加するものの、キータッチ数は減少する。これは、候補リスト最適化によってトータルでの入力効率を改善できることを表している。

bi-gram と tri-gram とを比較すると、テストセットパープレキシティは約 15～20％減少するが、キータッチ数・入力文字数の減少はわずかである。(a)と(b)とを比較すると、テストセットパープレキシティは2倍近い差があるにも関わらず、キータッチ数・入力文字数ではそれほどの差はない。

表1．実験結果

(a) カルテから構築した *N*-gram を使用

	bi-gram（最適化なし）	tri-gram（最適化なし）	tri-gram（最適化あり）
R_C	0.301	0.285	0.430
R_K	0.791	0.755	0.593
TPP	58.88	45.99	44.41

(b) Web から構築した *N*-gram を使用

	bi-gram（最適化なし）	tri-gram（最適化なし）	tri-gram（最適化あり）
R_C	0.343	0.330	0.477
R_K	0.830	0.805	0.620
TPP	97.81	83.01	81.21

以上のことから予測入力においては、言語モデル自体の性能だけでなく、操作コストを考慮した予測候補リストの最適化が入力効率改善効果を大きく左右する。また、Web から構

[1] SOAP: カルテを記載する代表的な形式。項目を S(Subjective), O(Objective), A(Assessment), P(Plan)の４種類に分類する。

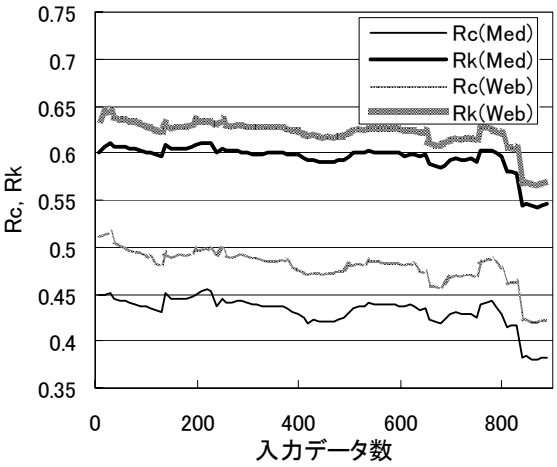

図3. 相対キータッチ数、相対入力文字数の変化

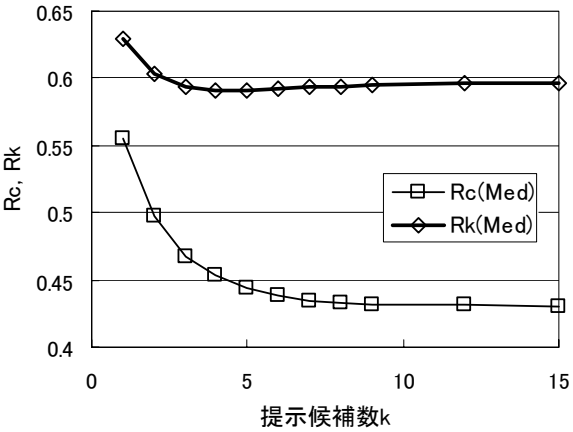

図4. 提示候補数の違いによる入力効率の変化

築したモデルでもカルテから構築した場合に近い入力効率改善効果が得られる可能性がある。

シミュレーションの過程で相対キータッチ数と相対入力文字数が変動した経過を図3に示す。tri-gram を用いて候補リストの最適化を行った場合について、直近の 100 データ(入力データ数が100に達するまでは全入力データ)に対する相対キータッチ数と相対入力文字数をプロットした。データ数が増えるに従ってキャッシュによる適応が進み、キータッチ数・入力文字数とも減少している。全期間に渡ってカルテから構築した *N*-gram のほうが良い結果であるが、適応の進行に伴い Web から構築したモデルとの差はやや縮小する傾向がみられる。

次に、提示する予測候補数 *k* を変化させた場合の結果を示す。*k* が大きいほど *k* 位までに正解候補が含まれる可能性が増すため入力文字数は減少するが、下位の候補を選択する回数が増えるため候補選択に必要なキータッチ数は増加する。そのため、総キータッチ数は *k* がある値のとき最小となる。図4に示すとおり、本実験では *k*=5 のときキータッチ数が最小となった。*k* の最適値はシステムの設計とテストデータによって変動すると考えられるが、筆者らの一部が本稿とは別のシステム・別のテストデータで行った実験(文献[6])でも候補数は4～6個前後が最適との結果が得られている。また文献[5]の結果でも候補数が5個前後のとき最も入力効率がよい傾向がみられる。

候補数を最適値より増やしてもキータッチ数の増加はわずかであるが、実際には提示候補数を多くするとユーザが候補リストの内容を確認する負担が増大する。したがって候補の選びやすさの観点からは *k* はできる限り小さく設計することが望ましい。"The Magical Number Seven"(文献[8])として知られるように、人間が短期記憶で判断できる項目の数は7±2といわれている。この意味では、提示候補数5個というのは無理のない数字であり、逆に 10 個あるいはそれ以上の候補を提示することは適切でない。予測入力はユーザが対話的に操作を進めていくものであるため、実際の使い勝手まで考慮してシステムを設計することが重要である。

6. まとめ

電子カルテシステムにおけるテキスト入力効率の向上を目的として、言語モデル(*N*-gram モデルおよびキャッシュモデル)に基づく予測入力システムを開発した。

実際のカルテ文書、および Web より収集した医療文書から *N*-gram モデルを構築し、カルテ入力を想定したシミュレーション実験により評価を行った結果、言語モデル自体の性能だけでなく、操作コストを考慮した予測候補リストの最適化が入力効率に大きく影響すること、tri-gram を用いた場合に bi-gram より高性能であること、Web 文書を学習テキストに用いてもカルテ文書に近い入力効率向上の効果が得られること等が確かめられた。また提示する予測候補数に関しては、5 個前後が最適である結果が得られた。

今後は、実際に文章を入力しての性能評価や入力所要時間の計測、体感的な使いやすさの評価等を通してさらなる改良を図り、実システムでの運用を目指す予定である。

文　献

[1] 下村淳一, 作野周介, "電子カルテにおける入力負荷の考察", 第 26 回医療情報学連合大会, 3-B-1-7, 2006.

[2] 田中久美子, "少数キーによる入力-ユニバーサルな言語コミュニケーションを目指して-", 情報処理学会誌, Vol.46, No.6, pp.691-696, 2005.

[3] 増井俊之, "携帯端末のテキスト入力手法", ヒューマンインタフェース学会誌, Vol.4, No.3, pp.131-144, 2002.

[4] 福島俊一, 山田洋志, "予測ペン入力インタフェースとその手書き操作削減効果", 情報処理学会論文誌, Vol.37, No.1, pp. 23-30, 1996.

[5] 田中久美子, "重度身障者のための 1 ボタン自然言語入力システム", 言語処理学会第 10 回年次大会, pp.544-547, 2004.

[6] 川尻博光, 中村明, 金川誠, 松本忠博, 池田尚志, "予測入力による電子カルテ入力支援", 第 26 回医療情報学連合大会, 3-B-1-3, 2006.

[7] 山田佳裕, 高松大地, 石原吉晃, 水野智美, 大口智也, 佐藤芳秀, 松本忠博, 池田尚志, "日本語文解析システム ibukiC/S について", 言語処理学会第 12 回年次大会, pp.185-187, 2006.

[8] George A. Miller, "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information", The Psychologocal Review, Vol.63, pp.81-97, 1956.