

サジェスト選択とタイピングの時間推定のためのモデル

荒牧英治* ** 宮部真衣* 篠原 (山田) 恵美子*** Tam Wailok****

橋本泰一***** 鷹合基行***** 宇於崎宏***** 橋田浩一*****

*東京大学 知の構造化センター **科学技術振興機構 さきがけ

*** 東京大学医学部附属病院企画情報運営部 ****産業総合研究所社会知能技術研究ラボ

*****東京工業大学総合プロジェクト支援センター

*****富士ゼロックス株式会社研究技術開発本部 *****東京大学医学部人体病理学・病理診断学分野

eiji.aramaki@gmail.com

1 はじめに

人間はミススペルを行ってしまう。Grudin (1983) はタイピングにおいて 3.2% の割合でタイプミスがあることを報告している。ミススペルは日常的現象であるといえよう。

ミススペルを減らす代表的な方法はスペル訂正 (Spell Correction) を行うことである。これは、通常、文章入力後に単語を走査することで行う。例えば、Word の校正機能や Google の「もしかして」機能がこれに相当する。スペル訂正については、Speller Challenge¹ などワークショップも開かれており評価型の研究が進んでいる。

一方で、入力前にサジェストを行い、事前にミススペルを防ぐアプローチもある。IME 予測変換や統合プログラミング環境 (visual studio, eclipse, Xcode など) における入力支援、Google Suggest など検索クエリの表示がこのアプローチとなる。このサジェストはユーザとのインタラクションが必須となるため、評価型の研究は乏しく、どのようなサジェストでどれくらい入力が効率化されるかはよく知られていない。

本研究では、サジェスト機能を L 文字の文字列を入力するかわりに、 M 個中 N 番目の候補に正解が期待できるという装置とみなす (図 1)。この仮定のもとでは、各変数について人間の反応時間を推定することができれば、どの程度サジェストが時間を節約できたかを計算できる。本研究は、実験データをもとに、この時間ゲインを測る指標を提案する。

提案する指標は、複数のサジェスト機能の比較に使えるのみならず、単一のサジェスト機能を設計する際のパラメータ調整にも有効である。例えば、サジェストの精度の向上は困難であっ

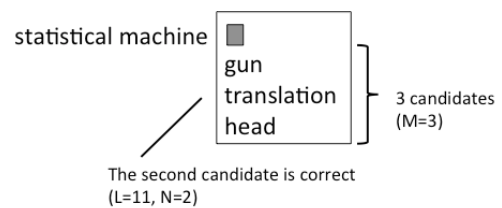


図 1: サジェスト機能をパラメータ化する。

ても、出力する文字数 L などは比較的容易に変更できるパラメータであることが多い。提案する指標を用いれば、どのようなパラメータが最適であるかを、被験者を用いずにただちに決定することができる。

2 関連研究

2.1 サジェスト／補完

前章にて述べたように、サジェスト機能を直接研究したものは少ない。また、それらの多くはキーストロークの回数をベースにサジェストの効率を測っている [1-4]。本研究と同様に時間を扱っているものとしては、牧野ら [5] が数字選択課題にて、リストでの選択時間を測定している。また、今村等 [6] が候補選択における諸条件 (背景色など) について調べている。しかし、本研究のように、ダイナミックに候補が変わっていく環境での実験はなされていない。

2.2 スペル訂正

これまで、スペル訂正 [7,8] では主に文字の類似度や文脈 (周辺の語列) [9-11] に頼っていた。web 検索におけるクエリ訂正 [12] も同様である。

一方、サジェストは次の 2 つの制約がある。まず、対象となる文が完結していないため、後方文脈が使えず、前方文脈と入力中の文字のみしか用いることができない。また、入力をサポートするという観点からは、できるかぎり短い時間で結果を表示する必要がある。このような制約があるものの正解候補を複数表示させ、ユ

¹ <http://web-ngram.research.microsoft.com/spellerchallenge/>

ーザの選択に任せることもできるという敷居の低さも併せ持つ。しかし、この利点を活かすためには、候補をどこまで表示するのが効率的なのかを知る必要があり、本研究の動機となっている。

2.3 ユーザ・インタフェース

ユーザ・インタフェースの分野ではフィッツの法則 (Fitts's law) [13]が、ポインティングデバイスの使いやすさをモデル化している。このモデルは、ボタンなどの対象部までの移動距離と大きさをを用いた関数となっており、マウスやタッチパネル入力にも応用可能な汎用的な指標となっている。本指標も Fitts と同様に、キーボード操作のみならずタッチパネルでの選択にも応用可能な一般的なモデルであり、拡張性に配慮している。

3 サジェストの評価モデル

3.1 基本アイデア

自動評価の基本アイデアは、直接入力 (TYPE) した場合の速度とサジェストを用いた選択入力 (SUGEST) の速度の比を自動的に行う点である。前者の TYPING の時間は、入力文字数 (以下 L) に相関するとみなし、L の関数 $F_{type}(L)$ として表現する。

後者の SELECTING の時間は、L の他にサジェスト機能の性能に依存する。性能を表現する方法は様々であるが、本研究ではこれを「平均 M 個の候補を表示し、平均 N 番目に正しい候補が存在する」というパラメータとして表現する。すなわち、 $F_{select}(L, M, N)$ として表現する。

以上2つの関数を用いると、サジェストの効率率は以下のように定義できる： $F_{type}(L)/F_{suggest}(L, M, N)$ 。上式において、値が大きければ大きいほど、よいサジェストシステムであるといえる。また、逆に、1を下回る場合はサジェストを行わず直接入力の方がよいことになる。

3.2 TYPE 関数

TYPE の時間は入力語の文字数に比例するとみなし、L の線形関数として以下とする：

$$F_{TYPE}(L) = L \times C_{keystroke}$$

ここで、 $C_{keystroke}$ は1回のキー押下に必要な時間とする。

3.3 SUGGEST 関数

SUGGEST の時間は、実際にはタイピングからサジェストにスイッチするための切り替えの時間が必要となる (この時間を C_{switch} と表記する)。また、タイピングが単に文字長 L にのみ依存する一方、候補を選択する時間 (この時間を $F_{suggest}$ と表記する) は M 個中から N 個目の正解を選ぶため、L, M, N に依存すると考えられる： $F_{suggest}(L, M, N) = C_{switch} + F_{select}(L, M, N)$

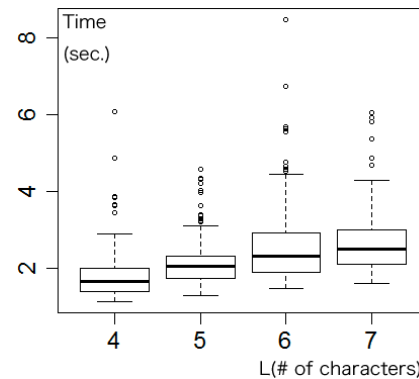


図 2: $F_{type}(L)$ のサンプリング結果。X 軸は L を示す、Y 軸は時間を示す。太線は中央値を示す。箱は第 1 四分位点と第 3 四分位点を示す。その他の記法も BOX PLOT の記法に準ずる。

which
their
other
would
there
which
first
these

	M	N	F_{select}
L	-0.003	0.002	0.024
M	1	0.663	0.625
N		1	0.733
F_{select}			1

図 3: 実験 UI。
(L, M, N) = (5, 7, 5)

表 1: 各変数間の相関係数

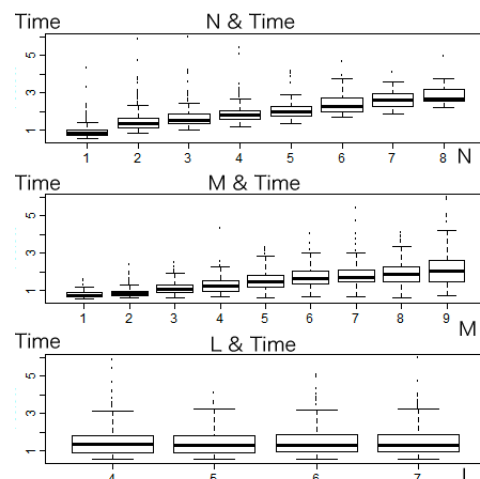


図 4: L, M, N と時間。

ここで、サジェストの候補を選ぶ方式は大別して、カーソルやスペースなどでリストを移動し選ぶ方式と番号や頭文字を入力する方式があるが、本稿では前者のみを扱い、次章にてサンプリングを行う。

4 パラメータの推定

4.1 TYPE 関数の推定

単語刺激を提示するユーザ・インタフェースを用いて、タイピング時間を計測した。実験では、まず、画面上に L 文字の単語刺激が提示され、それを被験者がタイピングする。

【単語刺激】単語刺激は、Google n-gram コーパスの 1gram のうち、英小文字で構成されている L 文字 (L=4...7) の語を頻出しているものから 100 語ずつ用いた。

【被験者】被験者は、共著者を含む 6 人とし、それぞれに 100 刺激を提示した。また、タイプミスがあった場合は、これを結果からのぞいた（脱落とみなした）。

【モデル】本実験によって得られた時間には単語認識の時間が含まれる。実際の入力においては入力すべき単語はユーザの頭の中にあるため、この時間は余分なものであり、取り除きたい。この値を $C_{\text{recognize}}$ とみなすと、実験によって得られる値（時間）は以下のモデルにより得られたとみなせる： $\text{Time} = C_{\text{recognize}} + L \times C_{\text{keystroke}}$

【結果】図 2 に結果を示す。回帰により、 $C_{\text{recognize}}=0.67$ および $C_{\text{keystroke}}=0.30$ が得られた ($R^2=0.14, n=582$)。

4.2 F_{SELECT} 関数の推定

模擬サジェスト機能を実装したユーザ・インタフェース(図 3)を用いて、単語選択の時間を計測した。被験者は画面に L 文字の単語刺激が提示された後、M 個の選択肢から、単語刺激と同じ語を選ぶ (N 番目に正解がある)。選択は、下カーソルキーを押下し単語を選択した後、リターンキーの押下によってなされる。

【単語刺激】4.1 節と同じくする。

【被験者】被験者は共著者を含む 8 人に 200 刺激を提示した。候補数(M)と正解の位置(N)はランダムに生成した ($M < 9, N < 9$)。選択ミスがあった場合は、これを結果から除いた。

【モデル】先の実験と同様に得られた時間には単語認識時間 ($C_{\text{recognize}}$) が含まれるため、以下となる： $\text{Time} = C_{\text{recognize}} + F_{\text{select}}(L, M, N)$ 。

【結果】得られた時間(Time)を各変数の線形とみなすと以下の回帰結果が得られた ($R^2=0.52, n=715$)。

$$\text{Time} = 0.01L + 0.07M + 0.23N + 0.50 \dots \textcircled{1}$$

各変数と時間との関係を図 4 に示す。文字数 L はほとんど Time と相関しない。よって、L は Time の引数から除くこととする。また、各変数間の相関を表 1 に示す。M と N はともに Time と強く相関するものの、M と N 同士も互いに相関している。このような場合、M と N のどちらかのみを回帰として用いるべきであり、本研究ではより強い相関をもつ N のみを用い、再度回帰を行った：

$$\text{Time} = 0.69 + 0.30N \dots \textcircled{2}$$

②式の R^2 は 0.49 ($n=715$)と高く、①式と比べてわずかしこ測精度度を下げていない。また、単語認識時間に相当する時間 (0.69 秒) が、タイピングの単語認識時間の 0.67 秒とほぼ同じであること、および、1 回のボタン押下に必要な

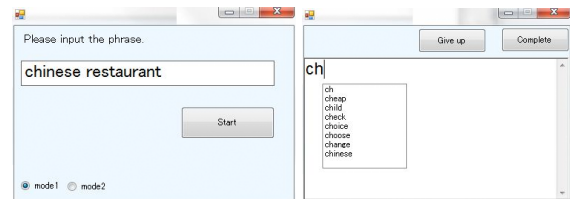


図 5: 模擬アプリ(左図: 入力記憶画面；右図: 入力画面
入力している最中にインクリメンタルにサジェストが行われる)。

L	予測値	実験値の平均	実験値の 95%信頼区間
1	<u>0.30</u>	0.171 (n=14)	0.02-0.32
2	<u>0.60</u>	0.385 (n=35)	0.26-0.51
3	<u>0.90</u>	0.559 (n=47)	0.37-0.75
4	<u>1.20</u>	0.99 (n=47)	0.73-1.24
5	<u>1.50</u>	0.95 (n=23)	0.70-1.21
6	<u>1.80</u>	1.66 (n=16)	1.19-2.14
7	<u>2.10</u>	1.62 (n=12)	1.11-2.13
8	<u>2.40</u>	2.17 (n=4)	0.09-4.25

表 2: TYPE 関数の提案手法による理論値と実験値。太字は予測値が実験値の 95%信頼区間に収まったものを示す。下線はサンプル数が 10 以上あったものを示す。

N	予測値	実験値	実験値の 95%信頼区間
1	1.48	1.67 (n=58)	1.47-1.87
2	1.78	2.20 (n=5)	1.25-3.14

表 3: SUGGEST 関数の提案手法による理論値と実験値。表記は表 2 に従う。

時間 (0.30 秒) もタイピングのそれと同じであることから変数の妥当性を示唆している。すなわち $F_{\text{select}}(N)=0.30N$ とみなす。

4.3 定数 SWITCH の推定

タイピングの途中でサジェストにスイッチする時間 C_{switch} を調査した。4.1 節のユーザ・インタフェースを用い、50%の確率でタイピングの途中 (1 秒以内のランダムな時間) にサジェストが割り込む。それ以外の条件は、4.2 節と同様である。サジェスト割り込んだ場合の入力時間は、以下と考えられる：

$$\text{Time} = C_{\text{switch}} + C_{\text{recognize}} + F_{\text{select}}(N).$$

回帰の結果、 $\text{Time} = 1.85 + 0.263N$ が得られた。 ($R^2=0.96, n=246$)。ここで、4.1 節と 4.2 節で既知となったパラメータから $C_{\text{switch}} = 1.18 (=1.85 - C_{\text{recognize}})$ が得られる。

4.4 サジェストが有効となる条件

実験の結果、以下の 3 つの式が得られた：

$$F_{\text{type}}(L) = 0.30L,$$

$$F_{\text{suggest}}(N) = 1.18 + 0.30N.$$

ここから、サジェストが有効となる条件は $L > 3.9 + N$ とみなせる。

5 実験

実環境で提案する指標による入力時間の予測精度を検証した。実環境ではサジェストとタイピングの両方を選択可能な状態であり、さらに、その選択は単語入力中においても可能である。すなわち、 L や N の両方が変化していく。このような状態での提案指標の有効性を模擬アプリケーションにて調査した。

5.1 実験設定

模擬アプリ上で以下の手順の実験を行った。まず、画面上に 2~4 単語が表示される。被験者は、それらを記憶終了したら、ボタンを押し、入力を開始する。入力は直接入力、または、サジェストにより入力が可能とする。サジェストのアルゴリズムは、現在編集集中の単語ともっとも編集距離が類似した単語を最大 8 個表示する(図 5)。5 人の被験者により、20 の試験句を入力することで、サジェストを採用した場合とそうでない場合の入力の時間を集計した。

5.2 評価

評価として、提案するモデルによる入力時間と、実際の単語の入力時間がどの程度一致するかを調査した。これは、実験結果によりえられた時間の分布から信頼区間を算出し、その範囲に予測時間が収まるかどうかで行った。

5.3 結果

結果を表 2 および表 3 に示す。サンプルが 10 以上あるデータ区間は 10 つある。このうち予測値が実験値の 95%信頼区間におさまったものが 8 区間あり、手法の妥当性を示している。今後、より大規模な実験での検討が望まれる。

5.4 考察

実験でサジェストを利用する割合は 23% (=63 語/265 語)と低く、その内の 92%(=5 語/58 語)が対象となる候補が先頭にきた場合 ($N=1$) に採用している。このように、人間は先頭の候補を採用する傾向が強いのであれば、サジェストが有効である条件 ($L>3.9+N$) から 5 文字以上の候補を提示すれば時間利得が得られる可能性がある(逆に 4 文字以下は常にタイプした方がよい)。

あと何文字打たないといけないかを L として、サジェストが利用された時の L の頻度を図 6 に示す。ほとんど時間利得がないはずの、残り 1 文字や 2 文字でサジェストを使っている頻度が高く、今後の研究課題となる。

6 おわりに

本論文では、サジェスト機能の評価を行う指標を提案した。また、サジェストが有効となる条件を示した。技術的限界として、本実験は非英語話者による英単語での実験であり、他の設定

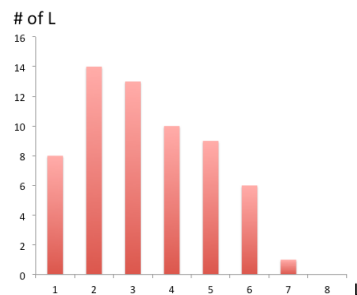


図 6: サジェストによって補完された文字数 (L) とその頻度。

での一般性を持たない。しかし、言語の問題はパラメータの変更で対応可能と予想され、本稿の提案の主眼であるサジェストの効率を自動的に計算する意義は揺るがない。

参考文献

- [1] Nestor Garay-Vitoria and Julio Abascal. 2006. Text prediction systems: a survey. *Univ Access Inf Soc*, 4:183–203.
- [2] Jianhua Li and Graeme Hirst. 2005. Semantic knowledge in word completion. In *ASSETS*, pages 121–128. Alan Newell, Stefan Langer, and Marianne Hickey. 1998.
- [3] Keith Trnka, Debra Yarrington, John McCaw, Kathleen F. McCoy, and Christopher Pennington. 2007. The Effects of Word Prediction on Communication Rate for AAC. In *NAACL-HLT: Companion Volume: Short Papers*, pages 173–176.
- [4] Alice Carlberger, John Carlberger, Tina Mag-nuson, M. Sharon Hunnicutt, Sira Palazue-los-Cagigas, and Santiago Aguilera Navarro. 1997.
- [5] 牧野純. 連続整数リストからの選択時間の測定, 経営情報研究: 摂南大学経営情報学部論集, Vol.9, No.2, pp.39-49 (2002).
- [6] 今村浩一郎, 宗森純, 長澤庸二. 次候補表示方法の入力効率に及ぼす影響について, 情報処理学会研究報告, ヒューマンインタフェース研究会報告, Vol.93, No.35, pp.57-64, 1993
- [7] Brill and R.C. Moore. An improved error model for noisy channel spelling correction. In *Proc. 38th Annual Meeting of Association for Computational Linguistics (ACL2000)*, pp. 286–293, 2000.
- [8] G. Navarro. *ACM computing surveys. A guided tour to approximate string matching*, Vol. 33, No. 1, 2001.
- [9] E. Mays, Fred J. Damerau, and R.L. Mercer, Context based spelling correction, *Information Processing and Management: an International Journal*, Vol. 27 No. 5, pp. 517–522, 1991.
- [10] D. Yarowsky, Decision lists for lexical ambiguity resolution: application to accent restoration in Spanish and French. In *Proc. 32nd Annual Meeting of Assoc. for Computational Linguistics (ACL1994)*, pp. 88–95, 1994.
- [11] AR. Golding and Y. Schabes, Combining Trigram-based and feature-based methods for context-sensitive spelling correction. In *Proc. 34th Annual Meeting of Assoc. for Computational Linguistics (ACL1996)*, pp. 71–78, 1996.
- [12] Q. Chen, M. Li, and M. Zhou. Improving query spelling correction using web search results. In *Proc. 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, pp. 181–189, 2007.
- [13] PM. Fitts (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, volume 47, number 6, June 1954, pp. 381–391.