

# アイデアを逃さず記録するための 操作ログ認識システムの開発

有賀美明 秋葉友良

豊橋技術科学大学

## 1 はじめに

ある瞬間の思考(特に, アイデア)を後から思い出すのは非常に困難なことである。そのため, 思考を逃さず記録した上で, 後から必要なときに, 記録した思考の閲覧・検索を可能とする, 思考のロギングシステムが望まれる。

現在利用可能な思考を記録する手段としては, 手書きのメモ, パソコンや携帯電話を使ったキー入力による記録, レコーダによる音声の録音などが考えられる。しかし, 手書きメモでは手元を見るために視覚が奪われ, また書いた文字を確認することで思考が中断されてしまう。また, 計算機を使って効率的に検索するためには, 手書き文字を後から OCR で認識して電子化する手間が必要となり, また手書き文字認識の精度が問題となる。パソコンや携帯電話を使ったキー入力による記録では, 最初から記録が電子化されるという利点はあるものの, やはり入力された文字の確認が必要であり, 特に日本語入力では変換候補の確定作業が必要となり, 思考が中断されてしまう。

一方, IC レコーダなどを用いた音声の録音は, 入力の確認・確定の負荷がないために, 思考を逃さず記録する手段としては適している。しかし, 他人が周りに密集している場所(公共交通機関の中やオフィス, など)では使いづらいという大きな問題があり, いつでも使える手段ではない。また, 電子的に記録した音声データは, 音声認識により文字化して検索や閲覧に備える必要があるが, 雑音の多い場所での録音は音声認識の精度に影響を与えてしまい, 記録時の環境に依存してしまう。

そこで, 記録時の環境に依存しないキー入力の利点と, 逃さずに記録する音声の利点を兼ね備えた, 思考のロギングシステムを提案する。提案システムでは, パソコンのキーボード, 携帯電話のテンキー, iPhone 等のタッチパネル, などの操作そのもの(操作ログ)を, 確認・修正を行うことなく記録し続ける。これにより, 確認や修正といった思考の妨げとなる行動を排除して, 漏れなく記録することを可能にする。記録・蓄積された操

作ログは, 記録者の閲覧・検索の要求に応じて, 文字列へと変換されて提示される。操作ログから文字列への変換(以降, 操作ログ認識と呼ぶ。)は, 音声認識と同様の手法を利用するが, 音声データの代わりに操作ログシーケンスを入力とする点が異なる。完全な変換が一度では難しい場合でも, 記録者が変換結果を見て誤りを指摘したり, 記録時の手がかりをキーワードとして与えたりするなど, 記録者とのインタラクションを通して, 漸進的に認識精度を上げていくことが期待できる。

本研究では, 提案する思考のロギングシステムのうち, 操作ログから日本語文字列を復元する, 操作ログ認識システムの開発を行った。システムは, 重み付き有限状態トランスデューサ(Weighted Finite State Transducer)で実装を行った。また, 構築したシステムについて, 操作ログ認識の評価実験を行った。操作誤り対策モデルを導入することで, 既存の IME の性能を越える性能を得ることが確認できた。

## 2 システムの概要

提案する思考のロギングシステムの全体像を図 1 に示す。

記録する操作ログは, 使用するデバイスにより様々な形式が考えられる。ノートパソコン等のキーボードを使う場合は, 各キーがアルファベット(あるいは, かな入力の場合はカナ文字)に対応しているので, アルファベットのシーケンスを時間情報とともにそのまま記録する。携帯電話に附属するテンキーを使用する場合は, 携帯電話用の様々な入力メソッド<sup>1</sup>から一つを選択する。iphone 等のタッチパネル上の操作を記録する場合は, パネル操作の絶対位置のシーケンスを, 時間情報とともに記録する。<sup>2</sup>

<sup>1</sup>ボケベル打ち, ニコタッチ, T9, などが利用できる。

<sup>2</sup>記録者は, パネル上にキーボードやテンキーをイメージして操作を行う。あるいは, 手書き文字入力を想定し, ストロークのシーケンスを記録することも考えられる。

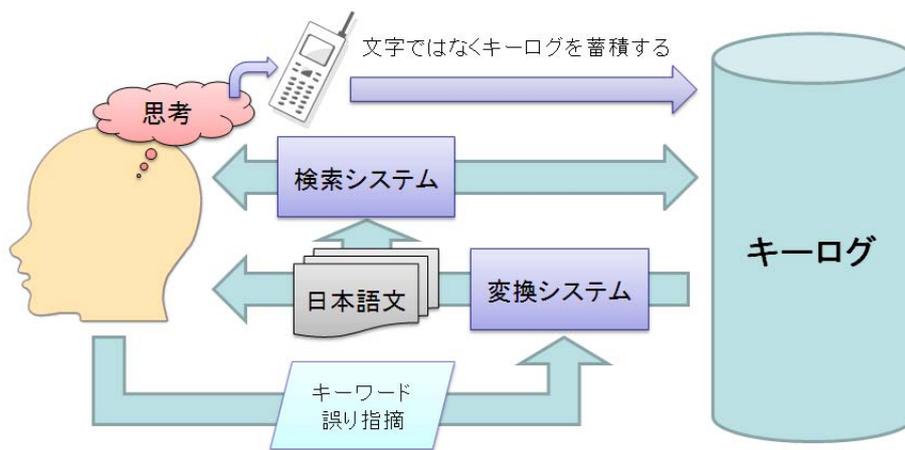


図 1: システムの全体図

蓄積された操作ログに対し、認識処理によってテキストに変換したり、検索により記録時刻を特定することで、過去の思考を後から参照できるようにする。このシステムにおいて、ユーザは記録者であり、同時に参照者でもある。したがって、完全な変換が難しい場合でも、ユーザとのインタラクションを通して、漸進的に認識精度を上げていく仕組みを導入することを考えている。例えば、記録者が変換結果を見て誤りを指摘したり、記録時の手がかりをキーワードとして与えたりするなどの方法が考えられる。

本研究では 操作ログから日本語テキストを復元する操作ログ認識システムの開発に焦点を当てる。統計的手法により、与えられた操作ログシーケンス  $y$  に対する条件付き確率  $P(x|y)$  が最大となる変換候補  $x$  を出力することで、操作ログから日本語文を復元する。この問題はベイズの定理を用いることで、以下の式に示すように言語モデルといくつかの変換モデルの積として推定することができる。

$$\arg \max_x P(x|y) \approx \arg \max_x \max_{z_1 \dots z_n} P(x)P(y|z_1)P(z_1|z_2) \dots P(z_n|x)$$

各モデルは重み付き有限状態トランスデューサ (WFST) により記述し、全てのモデルを統合し最小コスト法により操作ログから日本語文を復元する。

### 3 WFST を用いた変換モデル

重み付き有限状態トランスデューサ (WFST)[1] とは、入力、出力、重みにより構成される状態遷移モデルである。直観的に理解しやすい上に拡張性が高いことから、

近年、音声認識 [2]、音声合成、機械翻訳、OCR、情報抽出、など種々の音声言語処理分野で利用されている。

WFST は Compose と呼ばれる操作によって、一方の WFST の出力を、もう一方の WFST の出力にあてがうことで、2 つの WFST を合成した WFST を生成することができる。本研究では、入力操作ログシーケンスと各変換モデルを WFST で記述し、Compose 操作によって、出力のテキストを求めた。

システムは、WFST で記述した以下の 5 つのモデルにより構成する。

#### 3.1 入力モデル

入力モデルは、操作ログを入力として一文字ずつに分解し、一直線上にならべた WFST で構成する。

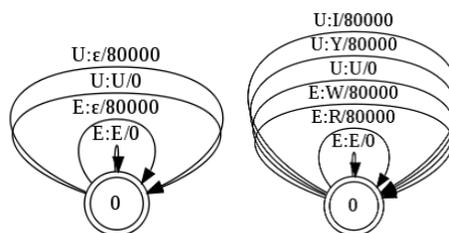


図 2: 挿入誤り対策モデル (左) と押し間違い対策モデル (右)

#### 3.2 誤り対策モデル

誤り対策モデルは、入力に対する誤りを除去するモデルである。被験者により実際に得られた操作ログを

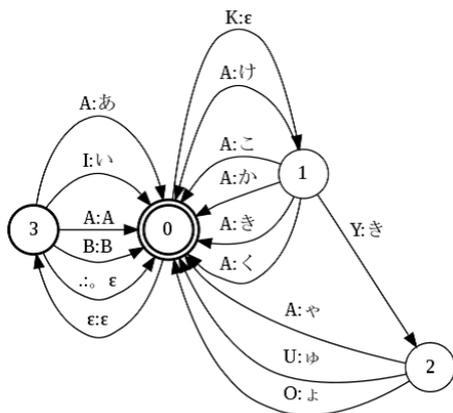


図 3: かな変換モデルの一部

分析したところ、特に多かった「挿入誤り」と「近いキーの押し間違い」に対して対策を講じた。

- 挿入誤り  
各入力に対して、スキップするパスを高いコストで追加することで対策する。
- 近いキーの押し間違い  
各入力に対して、入力されたキーの付近にあるキーのパスを高いコストで追加する。

誤り対策モデルの例を図 2 に示す。

### 3.3 かな変換モデル

かな変換モデルは、入力モデルの出力と辞書モデルの入力の整合性を図るためのモデルである。本研究では、操作ログはキーボードによるローマ字入力としたため、ローマ字列の入力を、ひらがな・アルファベットが複合した出力に変換するモデルとし、ローマ字・ひらがなの対応関係から作成した。ローマ字を全てひらがなに変換すると、アルファベットで構成される単語を示すことができなくなってしまふ。そのため、元となるアルファベット列を残しておくパスが必要となる。かな変換モデルの例を図 3 に示す。

### 3.4 辞書モデル

辞書モデルは、ひらがなを漢字に変換するモデルである。初めに、言語モデルの各語彙の読みを Chasen により調べ、一文字ごとに分解し一直線のモデルを語彙ごとに作成する。語彙の中に CD といったアルファベットで構成される単語が含まれている場合、アルファベット一文字ずつに分解しモデルを作成する。辞書モデル

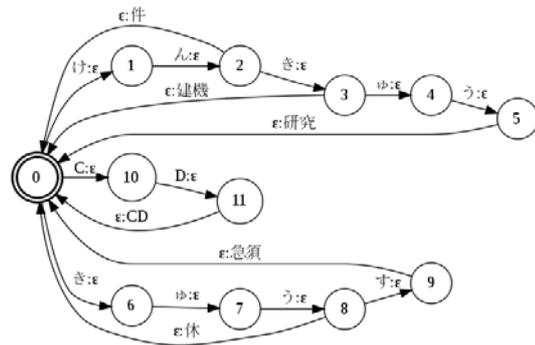


図 4: 辞書モデルの一部

75 (k)	0.875	1.016
69 (e)	1.125	1.281
78 (n)	3.016	3.094
78 (n)	3.203	3.328
75 (k)	3.547	3.688
89 (y)	3.891	3.985
85 (u)	4.125	4.235
85 (u)	4.360	4.469
83 (s)	4.594	4.766
85 (u)	4.766	4.875
82 (r)	4.953	5.094
85 (u)	5.141	5.250

図 5: 操作ログの例

は木構造化単語辞書とし、これによりプレフィックスを共有することで状態数を削減させた。辞書モデルの例を図 4 に示す。

### 3.5 言語モデル

言語モデルは、出力候補に生起確率を付与するモデルである。言語モデルは 3-gram とし、毎日新聞 2000-2005 年の記事、約 400 万文を学習データとして語彙を出現頻度上位 2 万語に制限して作成した。

## 4 操作ログ認識システムの実装

以上の変換モデルを WFST で表現し、認識システムを構築した。WFST のツールには OpenFST[4] を利用した。

操作ログ取得プログラムにより、操作ログを取得する。例として、「研究する」という日本語文では、図 5 の操作ログが得られる。操作ログは左から、押したキーのキーコード、キーを押した時間、キーを離れた時間である。なお、時間は操作ログ取得プログラムの起動時を基準とした時間とした。この操作ログから入力モデルを作成する。

表 1: 日本語テキストへの変換結果の単語誤り率 (%)

	完全	被験者 1	被験者 2	被験者 3	被験者 4
誤り対策なし	6.20	10.25	10.53	18.85	45.98
押し間違い対策	6.20	9.96	8.37	18.27	44.54
挿入誤り対策	6.20	7.36	9.52	11.65	34.63
IME2007	5.92	7.94	10.39	17.41	40.52

先に示した入力モデル，誤り対策モデル，かな変換モデル，辞書モデル，言語モデルの 5 つのモデルを順番に Compose し，出来上がったモデルからコストが最小となるパスを抽出することで日本語文を復元する．

## 5 評価実験

構築した操作ログ認識システムの性能評価実験を行った．評価実験のため，今回は自由な思考の記録ではなく，被験者にあらかじめ用意した原稿を見せて，それを入力した操作ログからどれだけの精度で元の文書が復元できるかを調査した．原稿には，Web から取得した論説文書を用いた．また，記録する操作ログは，ノートパソコンのキーボードを用いたローマ字入力とした．被験者は，手元のキーボードを見たり変換候補の確認をすることなく，原稿の文書を入力させた．

被験者 4 人から実際に得られた操作ログと，原稿のローマ字列を再現した完全な操作ログに対し，操作ログ認識システムにより日本語文を復元し，正解と比較することで単語誤り率 (Word Error Rate) を算出した．認識手法は，誤り対策なし，押し間違い対策，挿入誤り対策の各条件で行った．また，比較として，操作ログと同様の操作を文字として記録し，Microsoft 社の Office 附属の日本語 IME である IME2007 によるかな漢字変換によって得られた日本語文書を求め，同様に単語誤り率を算出した．実験結果を図 1 に示す．

完全な操作ログについては，誤り対策を講じても結果は変わらず，WER は IME より少し高くなった．復元結果を確認したところ，「時」と「とき」や「更に」と「さらに」など，人によって表記が揺れてしまう部分の違いが多く，意味として見ると大きな差は見られなかった．

被験者による操作ログは，誤り対策を講じることで全てにおいて性能改善が見られた．各対策における改善率は被験者によって異なることから，人によって誤りに偏りがあることが確認できた．

誤り対策により最も改善された結果において，IME を上回る精度で日本語に復元することができた．これにより，誤り対策が有効に働いていることが確認できた．

## 6 まとめ

デバイスの操作を操作ログとして記録する，思考のログインシステムを提案した．このシステムの内，操作ログから日本語文を復元する認識システムの実装を行い，実際に復元が可能であることを確認した．特に，操作間違いを想定した誤り訂正を行うことで，既存の日本語かな漢字変換 IME を上回る精度で日本語文を復元することができた．

今回は，近いキーの押し間違いと挿入誤りについての対策をそれぞれ講じた．これら 2 つの対策を同時に適用することも考えられるが，WFST の Compose 操作だけでは状態数が多くなり過ぎて計算不可能であった．今後より多くの WFST モデルを考慮するためには，効率よく探索する手法 (デコーダ) が必要である．また，今回は誤り対策におけるコストをヒューリスティックに定めたが，今後は適切なコストの設定方法についても検討する．また，今回の認識システムでは，操作ログに含まれている時間の情報を扱っていない．時間情報を利用することで復元精度をより向上させることができると考えられる．

## 参考文献

- [1] Emmanuel Roche, Yves Schabes. "FINITE-STATE LANGUAGE PROCESSING", The MIT Press, 1997.
- [2] 大西 翼, ディクソン・ポール, 岩野 公司, 古井 貞熙, "WFST 音声認識デコーダにおける on-the-fly 合成の最適化処理", 電子情報通信学会論文誌, D Vol.J92-D, No.7 pp.1026-1035 (2009).
- [3] Bowen Zhou, Stanley F. Chen and Yuqing Gao, "Constrained Phrase-Based Translation using Weighted Finite-State Transducers", In Proceedings of ICASSP, pp.1017-1020, 2005.
- [4] OpenFst Library, <http://www.openfst.org/>