E-029

音声誤認識された日本国内住所の事後訂正
Correcting Japanese Addresses on Failed Voice Recognition

米持 幸寿†
Yukihisa Yonemochi

1. はじめに

音声認識で日本国内の住所を入力する，というニーズがある．残念ながら一般用語を学習した機械学習音声認識では誤った文字列が出力されることがしばしば起こる．

今日の音声認識は機械学習を用いたものがほとんどであり，正しい結果を得るには認識モデルを正しく訓練することが効果的である．しかし，実際のシステム開発における現場では音声認識モデルを再訓練するには知識，コストなどの面で多くの障壁がある．その障壁を避けるために、現存する音声認識サービスが出力した結果が「間違っている」ことを識別し，それを適宜訂正することも重要である．

本研究では，音声認識により誤認識された日本国内の住所文字列を，音声認識の処理後に訂正して正しい結果を得る方法を提案し，その性能を測定する．

2. 先行研究

吉岡ら[1] の研究ではマルチモーダルの価値の研究であり古く本研究とはスコープが異なる．北岡ら[2]の研究で誤認識の検出が検討されているが訂正に関する研究ではない．近年では住所の音声認識の訂正研究は見当たらない．

3. 背景

本研究では，日本国内の住所表記の入力において，実際のシステム開発におけるいくつかの制約を受け入れつつ，音声認識結果が「間違っている」ことを識別し，それらを簡易な手段で訂正する方法を検討する．

3.1 研究の制約

本研究では「完全なデータを入手し」「音声認識モデルを訓練する」というアプローチは行わず，既存する無償あるいは安価な音声認識サービスが返す結果を事後訂正する方法を検討する．実行環境はスマートフォンとし，できる限りネット上に新サービスなどを起動しないこととする．

3.2 日本国内住所と郵便番号データ

日本のすべての住所が登録されているものは戸籍だが入手困難である．地図企業が提供する POI（Point Of Interest）でも住所が入手可能だが一覧で入手するには高額である．

本研究では日本郵便の郵便番号データに掲載されている範囲で日本国内の住所として取り扱うこととする．日本郵便のサイトからダウンロードできる．地名と事業所の二種類が提供されているが，本研究では全国の地名データ「KEN_ALL.CSV」を入力データとして使う．データは郵便番号に対して「都道府県」「地区町村」「町域名」とそれらの読み（半角カナ）が 3 段の木構造で収録されている．

3.3 構造上の例外

町域名の右にカッコ（）書きでさらに細かい地区名が掲載されていることがあり，最大 4 段木構造である．特定の屋号がその後ろについていることがあるが特定の区町名に接続されているため，それは 4 段目の一部として取り扱う．「以下に含まれないもの」といった行があり，全てが掲載されておらず不完全であることを前提に利用する．

4. 音声認識の性能試験

本研究では Apple iOS の標準 API（SFSpeechRecognizer）を使い，日本国内の住所を入力した際の性能を計測し，その中で失敗しているものに対する訂正を試みる．

4.1 音声誤認識の検証システム

以下の仕様の実験アプリケーションを作成した．

1. スマートフォンで音声認識エンジンを起動する
2. 郵便番号データから各地点の住所の読みをカタカナで取得し，音声合成で発話し，その音を音声認識する
3. 音声認識された文字列を取得し，入力された元データの住所表記とペアで保存し，比較する

本研究では先頭 4,063 件の住所データを検証する．そのうち 1,358 件（約 33%）が想定した文字列を出力しない．ただし，目検査で「口で話しかけたら認識されそう」なものを選択して手動で再実行したところ，1,214 件まで減らすことができる．この他「1 条，2 条」表現が全角・半角の違い，「2 線，5 線」が「2000，5000」となる問題があり，それらを自動変換すると誤認識として 1,194 件が残る．

4.2 誤認識している地名の特徴

誤認識している地名は表 1 に示すように分類できる．

表 1　誤認識している地名

	同じ読み	一部異なる読み
一部異なる出力	網走市 ○**嘉多山**（カタヤマ） ×片山（カタヤマ）	札幌市南区 ○**小金湯**（コガネユ） ×小金井湯（コガネイユ）
全く異なる出力	網走市 ○**能取**（ノトリ） ×の鳥（ノトリ）	札幌市厚別区厚別町 ○**小野幌**（コノッポロ） ×この頃（コノゴロ）

5. 提案手法

郵便番号に紐づく住所データは木構造であるため，幹から枝・葉に向かって検索し，一つに辿り着くように検索することで正解を導くことができると考えられる．表記完全一致で見つかる時は音声認識が正解しており，見つからない時は訂正を試みる．間違った音声入力と比較したとき，

† Pandrbox 合同会社　Pandrbox LLC

表記と読みに少しずつ類似点があることが表 1 からわかる．このことから，表記と読みを総合して類似度を測り，曖昧検索する方法を提案する．

5.1 木構造からの選択

郵便番号に紐づく住所データの概念構造と音声認識された文字列との関係を図 1 に示す．都道府県，市区町村，町域名などを個々のノードとし，それぞれに正式な表記と読みが与えられる．

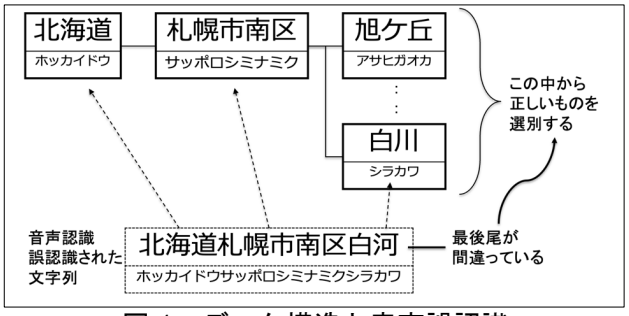

図 1　データ構造と音声誤認識

これを木構造のデータとしてプログラム内部に保持することにより音声誤認識された文字列を正しく訂正する．

5.2 曖昧検索のための比較スコア

図 1 に示された木構造を文字列から検索するにあたり，完全一致で見つかるものは「正しく認識された」文字列である．認識が失敗しているものは完全一致では見つからないため，そこから類似度を測ることで「最も近い」ものを選択する必要がある．文字列の類似度を比較する方法として，本研究では編集距離「ダメラウ・レーベンシュタイン距離」[3][4]の使用を提案する．また表記だけでなく音が重要と考え，文字列の読みも利用する．郵便番号データではカタカナ読みが提供されているが，子音，母音が検索の参考になるため，個別に扱いたい．そこで，カタカナをローマ字に変換し，ローマ字同士で比較に利用する．

注意しなくてはいけないのは，グローバル企業が提供する音声認識 API のほとんどが「読み」を出力しないことである．本研究では、国内メーカーの音声認識 API が読みを返すこと，表記文字に読み仮名を振るサービスがあること、といった状況から，音声認識結果文字列に読みデータが存在することを前提に検証を行う．検証には、手作業で準備した読みデータを利用する。

保持する木構造データと，音声入力された文字列および読みを同じ長さ切り出して比較し類似度が高いものを選択する．本提案では，比較スコアを以下の式で求める．

s: 比較スコア（相違度）
t1: 音声入力された表記テキスト
t2: 比較するノードの表記テキスト
r1: 音声入力された表記の読み（ローマ字）
r2: 比較するノードの表記の読み（ローマ字）
length: 文字列長さ
dl: ダメラウ・レーベンシュタイン距離

$$s = \frac{dl(t1,t2)\ +1}{t1.length} \times \frac{dl(r1,r2)+1}{r1.length}$$

長い町名と短い町名が前方一致で短い同一の文字列であることがある．例として「江部乙町東」に対する「江部乙町」がある．この場合どちらもスコアが 0 となり比較できない．0 になることを避けるため分子に+1 を置いている．また，長いものがより優先されるように長さで割る．

提案手法で選択されたノードより下位のノードの表記と読みを使って該当するスコアを計算し，もっとも成績のよい（スコアの値が小さい）ものを選択する．

6. 検証

最終的に選択された住所を入力の住所表記と比較することで訂正性能を確認する．検証結果を表 2 に示す

表 2　実験結果

状況		件	率
対象データ（先頭）		**4063**	
音声合成から音声認識で失敗		1358	33%
人間がリトライ後の失敗		1214	30%
パターン（1 条，1 線）編集後		1194	29%
提案手法で訂正できた	最後尾を訂正	-927	-23%
	途中から訂正	-174	-4%
訂正できない（残り）		**93**	**2%**

郵便番号データから抽出した 4,063 件のデータに対してパターン調整処理後の 1,194 件が音声認識で正しい住所が入力できない．同データの表記と読みを木構造で保持し，前方から編集距離を使用してあいまい検索することで 1,101 件を正しく訂正することができる．

93 件（2%）は訂正できなかった．音声認識結果が，入力しようとした住所と一文字しか合致しないほどかけ離れている，酷似した地名が多い，などが原因である．

7. おわりに

対象となった 4,063 件に対して音声認識による正解率 71%を訂正することで 98%まで改善できることを確認した．

今後，次のような追加検証を検討している．残りの 93 件に対してさらに訂正方法を検討する．今回の全体の 3% 程度に対して検証したが，全体に対して検証する．音声認識の読みデータと形態素解析等で生成する読みデータ間でどの程度性能差があるか検証する．

提案手法は木構造を持った語彙に対して有効と考えられるので住所以外のものにも応用できることが期待できる．

参考文献

[1] 吉岡理, et al. “音声認識機能を含むマルチモーダルインタフェースをもつ住所入力システムの開発と評価.” 電子情報通信学会論文誌 D 80.5 (1997): 1007-1015.
[2] 北岡教英, 角谷直子, and 中川聖一. “音声対話システムの誤認識に対するユーザの繰返し訂正発話の検出と認識.” 電子情報通信学会論文誌 D 87.7 (2004): 1441-1450.
[3] Damerau, F. “A technique for computer detection and correction of spelling errors.” Communications of the ACM 7(3) (1964): 659-664.
[4] Levenshtein, V. “Binary codes capable of correcting deletions, insertions and reversals.” Soviet Physice -- Doklady 10(1966): 707-710.