

圧力センサから読み込まれた点字のスペル訂正アルゴリズム

下平 啓太† 伊藤 祥一‡

長野工業高等専門学校 電子情報工学科†‡

1 はじめに

視覚障害者の情報収集手段の一つに点字があり、現在の社会においては、様々な場所で点字が併記されることが増えてきている。しかし、実際に視覚障害者が点字を利用するためには、点字独自の文法を学習し、指先で点の微妙な突起を読み取らなければならない。これは特に中途視覚障害者にとって困難なことであり、厚生労働省による調査結果¹⁾によれば、視覚障害者のうち「点字ができる」者は12.7%に留まっている。

そこで筆者らは、点字を指でなぞることによって翻訳し、その結果を音声として出力できるウェアラブル点字リーダの開発を行っている。点字の読み取りには感圧センサを使用し、これを指先に取り付けて使用する。このデバイスを使用することで、点字を利用できない視覚障害者が簡単に点字を読むことができるようになることを目指している。

2 誤入力訂正システム

現在のウェアラブル点字リーダでは、感圧センサが正しく点字を読み取れなかったり、圧力データを処理する際に点が欠損してしまったりすることがある。図1では、感圧センサが「コーセン(高専)」という点字を読み取ったのにもかかわらず、圧力データを処理する際に一部の点が欠損し「コーエン」という間違った語に翻訳されてしまっている。

先行研究²⁾では、そうした問題を解決すべく、翻訳処理部において誤入力を検出し訂正するシステムの開発が行われた。本研究ではそれらを引き継いでシステムの開発・改良を行い、実際に点字を感圧センサが読み取った際の誤入力がどの程度訂正可能なのか検証し、システムの評価を行う。

2.1 誤入力訂正システムの処理手順

本システムの処理は、以下に示す5つの手順で構成されている。

1. 1つの点字を2進数6ビットの数値と見立てた任意の個数の点字が、本システムに入力される

Spell Correction Algorithm for Braille Image with a Pressure Sensor

† Keita Shimodaira: National Institute of Technology, Nagano College

‡ Shoichi Ito: National Institute of Technology, Nagano College

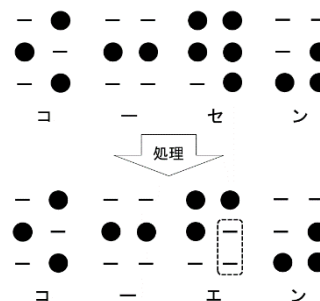


図1 圧力データの処理による誤入力
表1 各辞書の比較

名称	収録語数	最終更新
IPAdic	約39万語	2007年7月
NAIST-jdic	約48万語	2011年10月
NEologd	約317万語	最低月2回

2. 入力された向きの点字(以降、正点字と呼ぶ)と、正点字を上下左右逆さまにした向きの点字(以降、逆点字と呼ぶ)を生成する
3. 正点字と逆点字のそれぞれについて、一致する語が辞書にあるかどうか検索する
4. 一致する語がなかった場合、辞書の中から最も近いと思われる語を決定する
5. 正点字と逆点字それぞれについて決定された語を比較し、より近いものを訂正結果として出力する

2.2 固有名詞への対応

先行研究では、検索用の辞書データにIPAdic³⁾が用いられていた。IPAdicは約39万語を収録した形態素解析用の辞書であるものの、2007年に更新が終了している。そのため、2007年以降の新しい固有名詞や固有表現が収録されていない。

そこで、辞書としてIPAdicの後継であるNAIST-jdic⁴⁾と、IPAdicに収録されていない語が収録され、新しい語が毎週追加され続けているmecab-ipadic-NEologd⁵⁾(以降、NEologdと呼ぶ)を新たに採用した。各辞書の情報を比較したものを表1に示す。

本研究ではこれらの辞書を統合し、重複している語を削除することで生成された約217万語の辞書を用いることで、IPAdicでは対応できなかった「国土交通省」のような固有名詞にも対応可能となった。



図 2 円い容器に記されている点字

2.3 分かち書きへの対応

日本の点字表記について定めている日本点字表記法⁶⁾によると、点字の主な表記法として分かち書きがある。分かち書きは、文章の意味がおかしくならない程度に細かく言葉を区切るというもので、例えば「私は伊藤研究室の所属です」という文章は「私は_伊藤_研究室の_所属です」というように分かち書きできる。

また、「伊藤研究室」というような複合名詞については、3拍以上の独立性の強い意味のまとまりが2つ以上あれば境目で区切ると定められているため「伊藤_研究室」というように分かち書きできる。対して「夏休み」というような複合名詞は、意味のまとまりが2拍以下であるため分かち書きしない。

本システムでは、分かち書きされた文章が入力された場合でも、その文節ごとに誤入力訂正を行うことで分かち書きに対応している。

2.4 助詞への対応

2.3節で述べたように、点字には分かち書きという文法がある。しかし、本システムで用いている辞書データには「自立語」しか収録されておらず、「自立語+助詞」は収録されていない。そのため自立語に助詞が付属している語が入力された際に、正しい入力であるにもかかわらず、誤って訂正されてしまう可能性がある。

そこで、分かち書きされた文節の末尾の文字が助詞であった場合は、助詞とそれ以外の語をシステムの内部で分割するようにした。分割後に誤入力訂正を行うことで、「私は」「研究室の」「所属です」というような、助詞を含む文節の分かち書きを正しく処理できるように改善された。

2.5 正点字と逆点字の比較

点字の中には、図2のような円い容器に記されているものもある。ウェアラブル点字リーダの利用者がどちらの向きからなぞっても正しく出力されるよう、システムは逆点字に対応する必要がある。

先行研究では入力された点字と異なっている点の少なさを基準とし、システム内部で算出されたコストを正点字と逆点字で比較することで出力を決定していた。しかし、この方法で図2に記されている「ソース」を逆点字にした「ヒーチ」を入力した場

表 2 誤入力訂正システムの動作結果

	入力	出力	異なる文字数	優先出力
正点字	ヒーチ	キーチ	1文字	—
逆点字	ソース	ソース	0文字	○

合、「ヒーチ」を訂正した「ヒータ」の方が逆点字よりも優先されてしまうため、正しい訂正が行われなかった。

そこで本システムでは、正点字と逆点字それぞれの入力について、訂正後に出力された語が入力と何文字異なっているかを比較して出力を決定するように改良した。

図2に記されている点字について、本システムの実際の動作をまとめたものを表2に示す。「ソース」を逆点字にした「ヒーチ」が誤入力として入力されているが、出力された語と異なる文字数が逆点字の方が少ないため、逆点字である「ソース」の方が正しい出力であると処理されていることが分かる。

3 今後の展望

現段階では、2.4節で述べた助詞への対応において、対応可能な助詞が「は」「を」「に」「の」「が」「です」の5つのみとなっている。今後は対応可能な助詞を増やし、さらなる誤入力訂正の精度向上を図る。

また、本システムを実際にウェアラブル点字リーダと統合し、感圧センサが読み取った点字に対して期待される処理結果が得られるかどうかを検証することで、このシステムの評価を行う。

参考文献

- [1] 厚生労働省社会・援護局障害保健福祉部企画課：平成18年身体障害児・者実態調査結果、<http://www.mhlw.go.jp/toukei/saikin/hw/shintai/06/index.html>
- [2] 小出優真, 伊藤祥一, 藤澤義範: センサから読み取られた点字データの誤り訂正, 信学技報 IEICE Technical Report WIT201802 (2018-06), pp.7-10, 2018.
- [3] 浅原正幸, 松本裕治: ipadic version 2.7.0 ユーザーズマニュアル, <https://ja.osdn.net/projects/ipadic/>, 2003.
- [4] 浅原正幸, 松本裕治: NAIST Japanese Dictionary version 0.4.0 ユーザーズマニュアル, <https://ja.osdn.net/projects/naist-jdic/>, 2008.
- [5] 佐藤敏紀, 橋本泰一, 奥村学: 単語分かち書き辞書 mecab-ipadic-NEologd の実装と情報検索における効果的な使用方法の検討, 言語処理学会第23回年次大会発表論文集, pp.875-878, 2017.
- [6] 日本点字委員会: 日本点字表記法 2018年度版, 日本点字委員会, 2019.