話し言葉に対する音声認識誤りの分析

An Analysis of Speech Recognition for Error-patterns in Spoken Language

黒澤 義明 Yoshiaki KUROSAWA
目良 和也 Kazuya MERA
市村 匠 Takumi ICHIMURA
広島市立大学 Hiroshima City University

Abstract: The purpose of this paper is to analyze some errors of speech recognition and propose an approach to correct them using rules automatically captured without any correct/incorrect decisions. We noted the lists of morphemes and calculated their plausibility as a nonsentense collected from various corpora. As a result of experimental evaluation, we found over new 100 rules, which increased accuracies of our experiment. Therefore, we came to the conclusion that our approach is effective.

1. はじめに

近年，コンピュータ技術の発展に伴い，多くの人々が家庭でコンピュータを使うようになってきている．しかし，依然としてコンピュータを使用する上で，問題に直面している人々がいる．例えば，キーボードによるタイピングなどの，インタフェースに直結したスキルの問題である．

こうしたスキルに精通していないことによって生じる問題を解消するためには，キーボードやマウスを入力装置として使うのではなく，我々が通常行っているような，音声を用いた入力装置を使う必要がある．

このような必要性にもかかわらず，音声認識装置の精度はそれほど上がっていない．まったく予想もしない出力が生じることがあるかと思えば，既に経験した誤りと同一の誤りが生じることがある．このため，多くの研究者は，この誤りの問題に関心を持ち，音素等の特徴から，誤りを修正する規則を学習することを目的とした，機械学習の研究を行っている（Litman et al. 1999; Hirschberg 1999）．機械学習の立場では，誤りを修正するため，様々な特徴を持った，そしてその特徴に基づいて正事例/負事例とラベル付けを行った，より多くの事例が必要となる．しかもユーザ個人の特徴記述が不可欠である．例えば，単語と単語の時間間隔，単語の繰り返しの頻度など様々な特徴を加味しなければならない．しかし，コンピュータに不慣れなユーザが，こうした事例を用意し，しかもユーザの特徴に基づいた規則を学習させることは，簡単なことではない．

そこで本研究は，事例を自動的に取り扱い，特に日々の生活～例えば，質問応答～に関連した認識誤りを自動的に修正する方法を考察する．特殊な操作法を必要とする方法ではなく，コーパスから採取した統計情報に基づき，さらに音声入力による自然な応答に基づいた修正判断を行う方法を提案する．この方法は家庭での使用を前提としており，学習に際してはコンピュータについての知識が必要ないため，これまでの方法より扱いやすいという利点がある．

こうした家庭で使うことを主眼とし，また取り扱いが容易であるという観点から，Linux 上で動作する音声認識システム（Julius）を使用する（Kawahara et al. 2000; Lee et al. 2001; 鹿野 他 2001）．そして，Julius の結果をもとに，後処理として修正を行う．

なお，後処理として行う利点としては，音声認識システムにおいて考慮されていない特徴を容易に追加できる点にある（Ringger and Allen 1996）．さらに高精度な形態素解析器を誤りの修正に利用できる点も利点であると考えられる．

2. 音声認識誤り

Julius は Linux を含む UNIX 環境において，多くの研究者が使用している，高精度の音声認識エンジンである（Kawahara et al. 2000; Lee et al. 2001; 鹿野 他 2001）．また，Linux は近年家庭でも使用されるようになっており，その上で動く Julius もまた広く使用することができると考えられる．したがって，我々は Julius を用いて，自動誤り訂正についての分析を行う．

•実験

まず最初に，認識の際の誤りを分析するため，以下の手順に基づいて実験を行う．

文： 「はい，そうです」「いいえ，そうではないです」など Q&A システムで使用されるような 26 の文を使用した．このうち，10 文は「はい」から始まり，10 文は「いいえ」から始まる文であり，指定通りに発声することが求められた．一方，残り 6 文は「かまいます」等，「はい」「いいえ」のどちらかを用いて発声するように求められた．

被験者： 大学生または大学院生 5 名（うち，女性 1 名）．全員，日本語が母国語であった．

手続き： 被験者は個々に机の前に座り，紙に書かれた文を，10 回繰り返して，声に出して読み上げるよう教示が行われた．なお，音量調整・雑音除去等，音声入力に関する最適化は行わなかった．ただし，明らかな音量不足等，システムに正しく入力が行われなかった場合には，必要に応じて，10 回以上の読み上げが行われた．

音声認識評価： Julius は 2 回走査を行う解析器であり，最初に 2-gram をもとにした計算，そして 2 回目に 3-gram をもとにしたより正確な計算を行う．また，2 回目の走査の際には最適解以外の候補を表示することができるため，この候補を使用して，不明瞭な入力だったかどうかの指標とする．すなわち，明瞭な入力であれば，複数の候補が共通の文として認識されるのに対し，不明瞭な入力であれば，一見おかしな候補が表示される．そこで，2 回目の走査結果として，全 5 文の候補を表示させることにする．

結果

Table 1.に個々の走査結果を示す．「1st Pass」は最初の走査結果を示し，「2nd Pass Best」は 2 回目の走査中の最適解を示す．「All Candidates」は，全ての候補を示す．それぞれ，260 個，260 個，そして 1560 個のデータを含んでいる．

表 1　解析結果

		Pass		
		1st Pass	2nd Pass Best	All Candidates
被験者	S1	19.2 % (50/260)	32.7 % (85/260)	27.0 % (421/1560)
	S2	27.3 % (71/260)	32.3 % (84/260)	25.8 % (402/1560)
	S3	27.7 % (72/260)	43.5 % (113/260)	35.6 % (556/1560)
	S4	33.1 % (86/260)	50.0 % (130/260)	39.9 % (623/1560)
	S5	43.7 % (118/260)	60.0 % (162/260)	51.9 % (841/1560)

表 1 から明らかなように，「2nd Pass Best」が最も成績がよい．しかし，Julius はこれらの文に最適化されておらず，不正確な結果になっている．

主な誤りの原因を，以下に示す．

a. 不明瞭な音節に伴う誤った語彙化
b. 不明瞭なポーズや雑音による過語彙化
c. 子音落ち（母音前）
d. 子音落ち（母音無し）

例えば，「いいえ，かまいます」を意図した発声により，「いいえかマイナス」と誤認識が起こる．これが a.である．「はい，そう思います」を意図したときに，「配送」と認識される誤りが b.である．さらに，「はい」が「I」と認識される誤りが c.である．最後の d.は.「ありませ ん」の「ん」が認識されず，「ありませ」または「あります」と解釈されてしまう誤りである．

一般的には，こうした誤りを含む多くの文に，非文としての解釈が成立する．例えば「いいえ，市が今さん」は文字通り非文であり，こうした文を構成する品詞列（感動詞/助詞-終助詞 /名詞－一般）は文法的に出現しにくいと考えられる．

我々は，こうした品詞列がおかしいということにすぐ気がつく．しかしながら，コンピュータに認識させることは容易ではない．そこで，こうした品詞列の自動処理法が必要となることは明らかである．

3. 認識誤り訂正の自動化

現在までに，形態素誤りの修正に関する研究が行われている（Kurosawa et al. 2003; Mera et al. 2004）．しかしながら，これらの研究は知識ベースを自動で構築せず，手動での記述を行うことで修正を行っている．

この手法に従えば，「いいえかマイナス」のような誤りを発見する際に，誤りを含んだ品詞列（感動詞/助詞-終助詞 /名詞一一般）に着目し，規則を作成する．図 1 に修正規則の例を示す．

```
感動詞/助詞－終助詞/名詞一一 般¥
    0/(いいえ)/1/(か)/2/(マイナス)          ¥
        0/1/2->{... 形態素列 “いいえ，かまいません”... }
```

図 1　修正規則例

規則は，3 行から成り立っており，主規則部，副規則部，効果部と呼ぶ．主規則部には品詞列を記載し，副規則部にはさらに詳細な条件を記載する．そして，効果部に出力内容を記載する．記載内容については，今回の研究の範囲ではないため，詳述を避ける．

この表記法を用いることにより，数多くの形態素誤りを修正することができる．しかしながら，音声認識誤りに対して同様の手続きを行うと，規則数が爆発的に増えるという問題がある．すなわち，形態素解析誤りは意図した表記と解析結果が一対一対応になるため，主規則の内容が固定となり，したがって規則数が限られるが，音声認識誤りにおいては，意図した表記が得られるかどうかがわからない．誤り分類（a.－d.）に示したように，複数の可能性が生じ，しかもその並びはまったく予想できないため，どのような規則を記述すればよいかわからない．したがって，形態素解析誤りと同様に，手動で規則を記載することは困難であると考えられる．

すなわち，一般的な手法，機械学習を用いた手法が必要となる（cf. Hirschberg et al. 1999; Litman et al. 1999; 内山 1999）．ただし，この主の方法の問題点としては，事例を用意しなければならないところにある．先にも述べたように，音声認識誤りは，様々なタイプがあり，事例を用意する場合には，非常にたくさんの事例が必要となる．この必要な事例を，研究者が個々のユーザに合わせて用意するのは困難である．また，個々のユーザが事例を収集するのも一層困難である．

以上の理由により，こうした誤りを扱う際には，品詞列が正しいか誤っているかを自動的に判定する手法が必要となる．そこで，本研究では，以下のふたつの手続きを用いて，自動的に非文であることを認識し，自動的に非文を修正する規則を生成する．

・非文の認識

コーパスを利用して，値を計算する．次のふたつのルールを有する．(F1) 品詞列に含まれる全ての 2-gram または 3-gram の値，そして，(F2) 品詞列全体（すなわち，N-gram：N は品詞数）の頻度である．非文の場合には，これらの値が 0 になることが予想されるため，積を計算し，非文の尺度とする．

なお，修正規則を作成するために必要な手続きを述べる際には，上のように片仮名で「ルール」と述べる．これは，修正規則とは別である．

・修正規則候補の作成

Julius の出力に着目する．2 章に述べたように，我々は 6 種類の出力を利用する（「1st Pass」「2nd Pass Best」そして，4 種類の「All Candidates」）．一般的には，「1st Pass」が非文となる可能性が高い．そして，2 章の「手続き」の箇所で述べたように，明瞭な入力であれば，2 回目の走査結果に共通項目が増えるという特性があった．そこで，ふたつの条件を同時に満たすときに，前者の「1st Pass」品詞列を誤り（すなわち，主規則部に相当）と考え，後者の「2nd Pass Best」を正解（効果部に相当）と考えることとした．例えば，「2nd Pass Best」と，4 種類の「All Candidates」が全て同じだった場合を(G1)，さらに，「2nd Pass Best」と，4 種類の「All Candidates」のうち 3 種類が同じだった場合を(G2)等と定義し，(G1)–(G5) で表される 5 種類のルールを用意した．

・修正規則の生成

また，より正確な判定を行うため，修正規則に含まれる主規則と効果部の品詞列の値を(F1)または(F2)の判定ルールに基づいて計算することにより，規則候補として正しいかどうかを判定することとする．例えば，図 1 の場合には主規則部「感動詞/助詞-終助詞/ 名詞一般」の値と，効果部「感動詞/動詞-自立/ 助動詞/助動詞（いいえ，かまいません）」の値とを比較して，規則として適切かどうかを判断することになる．

4. 実験

前章までに述べた非文かどうかの判断を行うため，コーパスを用意した．本研究では，青空文庫（青空文庫 2004）から，およそ 100 万文を含む 2012 作品を抽出した．

品詞数を計数するため，一旦形態素解析を行った．本研究で使用する形態素解析器は茶筅である（Matsumoto 他 2000）．さらに形態素解析の誤りを修正するため，誤り修正法を使用し，より正確な値の計測を行う（Kurosawa et al. 2003; Mera et al. 2004）．

なお，計測の結果，非文認識ルール(F1)は有効でないことが明らかとなった．何故なら，2-gram の計算では，2 章で採取した品詞列の中には，ふたつの組み合わせを除いて，ほとんど全ての値が 0 ではなかったからである．したがって，積を計算しても 0 とならないことが明らかとなった．形態素解析の誤りが依然含まれているため，誤った品詞列を計数している可能性もある．また，今回の採取データ中の品詞組み合わせがたまたま 0 にならなかった可能性もある．詳しい検討が必要であるが，今回は，仮に設定する別の基準をもとに考察を進める．新たに作成した基準は次の通りである．

(F1a) 主規則部*10 の N 乗＜効果部
　　N：品詞数
(F1b) 主規則部*100＜効果部

2 章で採取したデータを用いて，クローズドテストして評価を行う．

得られた修正規則候補は，303 個であった．このうち，判定ルールによる出力結果，及び精度を以下に示す（表 2）．

表 2　規則判定結果

		規則			精度
		総数	正解/不正解		
判定ルール	F1a	58	42	16	72.4
	F1b	79	39	40	49.4
	F2	123	87	36	70.7

表 2 の結果，F1a と F2 の判定ルールに基づく判定結果が良好であることがわかる．ただし，あくまで F1a は仮の値であり，詳細な議論のためには最適化が必要であると考えられる．したがって，以降の分析の対象とはしない．

(F2)ルールを適用し，規則の判定を行い，その

後，規則を適用して，修正を行った．その結果を次の表に示す(表 3)．表中,「b」は基準値(baseline を示しており，表 1 に示した値と同一である．したがって，どのように値が変わったかを比べることが可能である．たとえば被験者 S1 の結果は，当初正解が 85 個であったが，新たに 17 個が正しく修正されたことを示す（O:17）．その結果，精度が 6.5%アップし，39.2%になったことがわかる．

表 3　規則適用後の精度

		Pass	
		2nd Pass Best	All Candidates
被験者	S1	39.2% (102/260) 6.5%↑ b:85, O:17, X:0	32.1 % (500/1560) 5.1%↑ b:421, O:123, X:44
	S2	31.6 % (82/260) 0.7%↓ b:84, O:6, X:8	26.5 % (413/1560) 0.7%↑ b:402, O:72, X:61
	S3	48.9 % (127/260) 5.4%↑ b:113, O:15, X:1	44.1 % (688/1560) 8.5%↑ b:556, O:143, X:11
	S4	56.2 % (146/260) 6.2%↑ b:130, O:18, X:2	48.3 % (754/1560) 8.4%↑ b:623, O:176, X:45
	S5	65.8 % (171/260) 5.8%↑ b:162, O:9, X:0	60.1 % (856/1560) 8.2%↑ b:841, O:115, X:8

この表から明らかなように，被験者 S2 において，精度の逆転現象が見られるという問題もある．しかしながら，これらの問題は，共起関係などの重要な言語知識を持たないことによって起こると考えられる．例えば，否定の助動詞「ん」が感動詞「いいえ」と共起しやすい等という知識はまったく持っていない．このため，このような修正規則が得られないとは言え，本研究の有効性が否定されたわけではない．本研究の手法は，通常の出力として用いる「2nd Pass Best」の出力を修正することが可能であり，さらに過適用が少ないため，有効な手法であると言える．

5. おわりに

音声認識処理によって得られた誤りを分析し，品詞列に注目することにより，学習のためのトレーニングデータを必要とせず，また個人に対する認識訓練を必要としないで，修正に必要な規則を学習し，適切に修正を行えることがわかった．

今後の課題は，入力を一文に制限することなく，自由な発話場面から同様の規則を発見できるかどうか検討することである．その際，単語の繰り返しなどの誤りを判断するために使用される尺度（Krahmer et al. 1999; Krahmer et al. 2001）と，本研究の計算との比較を行うことが重要になると考えられる．

参考文献

青空文庫．2004. WWW page, Available from < http://www.aozora.gr.jp/>.

Hirschberg, J., Litman, D., and Swerts, M. 1999. Prosodic cues to recognition errors. Proceedings of the 1999 International Workshop on Automatic Speech Recognition and Understanding, pp. 349-352.

Kawahara, T., Lee, A., Kobayashi, T., Takeda, K., Minematsu, N., Sagayama, S., Itou, K., Ito, A., Yamamoto, M., Yamada, A., Utsuro, T., and Shikano, K. 2000. Free software toolkit for Japanese large vocabulary continuous speech recognition, Proceedings of the 6th International Conference on Spoken Language Processing, pp. 476-479.

Krahmer, E., Swerts, M., Theune, M., and Weegels, M. 1999. Problem spotting in human-machine iteraction, Proceedings of EUROSPEECH, pp.1423-1426

Krahmer, E., Swerts, M., Theune, M., and Weegels, M. 2001. Error detection in spoken human-machine interaction, International Journal of Speech Technology 4:19-30

Kurosawa, Y., Ichimura, T., and Aizawa, T. 2003. A description method of syntactic rules on Japanese filmscript. Proceedings of the 7th International Conference on Knowledge-Based Intelligent Engineering Systems & Allied Technologies, pp.446-453.

Lee, A., Kawahara, T., and Shikano, K. 2001. Julius --- an open source real-time large vocabulary recognition engine, Proceedings of the 7th European Conference on Speech Communication and Technology, pp.1691-1694.

Litman, D., Walker, M., and Kearns, M. 1999. Automatic detection of poor speech recognition at the dialogue level. Proceedings of the 37th Annual Meeting of the Association of Computational Linguistics, pp.309-316.

Matsumoto, Y., Kitauchi, A., Yamashita, T., Hirano, Y., Matsuda, H., Takaoka, K., and Asahara, M. 2000. Morphological analysis system ChaSen version 2.2.1 manual. http://chasen.aist-nara.ac.jp/.

Mera, K., Kurosawa, Y., and Ichimura, T. to appear in 2004. Emotion oriented interaction system for elderly people. Knowledge Based Intelligent Systems for Health Care (T. Ichimura and K. Yoshida Eds.), Advanced Knowledge International.

Ringger, E. and Allen, J. 1996. Error correction via a post-processor for continuous speech. Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.427-430.

鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄 2001. 音声認識システム，オーム社.

内山将夫．1999．形態素解析結果から過分割を検出する統計的尺度，自然言語処理，6(7)，pp.3-28.

連絡先：
黒澤 義明
広島市立大学 情報科学部
〒731-3194 広島市安佐南区大塚東 3-4-1
Phone: 082-830-1581 Fax: 082-830-1792
E-mail: kurosawa@its.hiroshima-cu.ac.jp