G-003

キーワード抽出に基づく統計的音声発話理解
Statistical Speech Understanding based on Keyword Extraction

内田 佳孝 * Yoshitaka Uchida　**嶋田 和孝** † Kazutaka Shimada　**峯脇 さやか** † Sayaka Minewaki　**遠藤 勉** † Tsutomu Endo

1. はじめに

近年の音声認識技術の向上により，音声入力を用いた実用的な対話システムの実現を目指した研究が進められている．実用的な音声対話システムの構築には，音声認識誤りへの対処が不可欠である．しかしながら，音声認識誤りは，システム内の語彙や文法の範囲を広げたとしても，本質的に避けられない問題である．実用的な音声入力を用いた対話システムを構築するためには，音声認識の精度を高めるとともに，音声認識誤りに対して頑健な枠組みが必要である．

音声認識誤りの問題に関しては，その影響を軽減するためにキーワードもしくはキーフレーズの抽出に基づく発話理解法の研究が行われてきた [6, 8, 11, 13]．これらの手法は，音声認識誤りを減少させ，ユーザからの比較的自由な発話を可能にする．また，Bouwman ら [2] や駒谷ら [5] は音声認識エンジンからの N-best 出力による内容語に対して信頼度を計算し，それを利用することによって音声認識誤りの影響を軽減させる手法を提案している．

本稿では，これらの先行研究での議論を踏まえ，音声認識誤りに対して頑健な発話理解を実現するために，キーワード抽出に基づく発話理解法を提案する．図 1 に提案手法の概要を示す．提案手法では，音声認識エンジンから得られた N-best 出力からキーワード抽出を行う．続いて，得られたキーワードを，対話コーパスからの情報を利用して文に復元する．その復元された文が妥当かどうかの評価を行い，最終的に依存構造木を出力する．

2. キーワード抽出

まず，音声認識エンジンから出力される N-best 出力からキーワード列を抽出する手法について述べる．抽出には，(1) 内容語の信頼度，(2) 連想確率，という 2 つの尺度を用いる．音声認識エンジンには記述文法方式の音声認識エンジン **Julian**‡ を使用した．

2.1 内容語の信頼度

ここでは，Bouwman ら [2] や駒谷ら [5] が提案している内容語に関する信頼度の計算方法を拡張したものを用いる．内容語とは，品詞が，動詞，名詞，指示詞，形容詞，連体詞，副詞であると定義する．音声認識結果の N-best 出力には対数スケールでスコアが付与されている．このスコアより，

$$p_i = \frac{e^{\alpha \cdot score_i}}{\sum_{j=1}^{N} e^{\alpha \cdot score_j}}$$

を計算する．ここで，$1 \leq i \leq N$ であり，α はスムージング係数である．ある内容語 w が i 番目の文に含まれるときを $\delta_{w,i} = 1$，含まれないときを $\delta_{w,i} = 0$ とすると，入力音声に w が含まれていた確率 p_w は，

$$p_w = \sum_{i=1}^{N} p_i \cdot \delta_{w,i}$$

となり，Bouwman および駒谷らは，これを内容語 w の信頼度としていた．

提案手法では，これを形態素単位で計算していく．まず，語の発話時間や品詞，文法情報などを用いて，形態素長の正規化を行う．その後，以下の式で形態素毎に全ての語に対して計算を行う．

$$P_c(w_j) = \sum_{i=1}^{N} p_i \cdot \delta_{w_j,i}$$

ここで，j は形態素位置であり，$\delta_{w_j,i}$ はある語 w_j が i 番目の文の j 番目の形態素に含まれるとき 1，含まれないとき 0 とする．この処理によってキーワードになり得る語を抽出する．キーワード以外の語の方が高い信頼度を持つ場合は，そのキーワードは抽出しない．

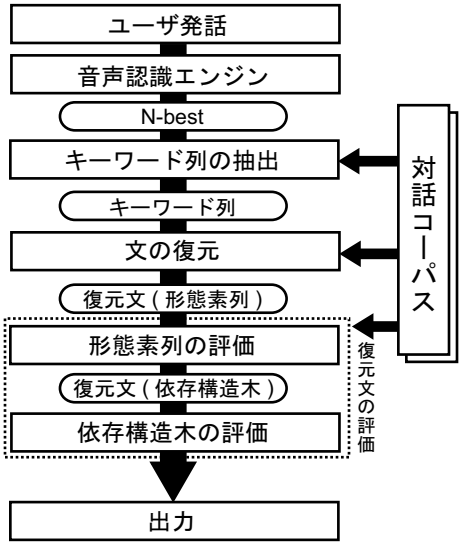

図 1: 処理の概要

2.2 連想確率

2.1 の内容語の信頼度は，音声認識結果に対する信頼度である．次に対話ドメインでの信頼度を考える．ここでは，単語間の意味的なつながりの強さ計算するために持橋らが提案している**連想確率** [7] を導入する．

持橋らは，語の意味とは他の語との連想関係により定義され，その手がかりは言語的な束縛＝共起関係であるという考えに基づき，共起確率分布から連想確率を定義している．単語の意味的な重みを表す指標として，単語

*九州工業大学 情報工学部 知能情報工学科．現在は，ジャストシステム．
†九州工業大学 情報工学部 知能情報工学科
‡http://julius.sourceforge.jp

の共起確率分布の情報量から計算される連想情報量を提案しており，それと共起確率との組み合わせにより連想確率は計算される．連想情報量 $ap(x)$ とは 1 語当たりの平均共起確率であり，様々な語と平均的に共起するような語 (機能語など) はこの値が小さくなる．

$$ap(x) = \exp(\sum_{w \in L} p(w|x) \log p(w|x))$$

ここで，w は意味空間上での語であり，L は意味空間上での語の集合であることを意味し，単語 $x \in L$ である．$p(w|x)$ は語 x, w の共起確率である．

語 w_i から語 w_j への連想確率 $a(w_j|w_i)$ とは，共起確率 $p(w_j|w_i)$ に比例し，連想情報量により重み付けを行ったものであり，以下のように定義される値である．

$$a(w_j|w_i) = \frac{p(w_j|w_i)ap(w_j)}{\sum_{w_j} p(w_j|w_i)ap(w_j)}$$

2.3 2 つの尺度によるキーワードの抽出

2.2 で説明した連想確率を 2.1 で述べた手法により抽出されたキーワード群に対して適用していく．まず，抽出されたキーワード群から考えられる n 個のキーワード列 $KP = \{kp_1, kp_2, \cdots, kp_i, \cdots, kp_n\}$，$kp_i = \{k_1, k_2, \cdots, k_j, \cdots, k_m\}$ を作成する．kp_i の要素 k_j が持つ連想確率 $P_a(k_j)$ は以下の式で計算される．

$$P_a(k_j) = \sum_{\substack{l=1 \\ l \neq j}}^{m} a(k_l|k_j)$$

キーワード列 kp_i に対する内容語の信頼度 $P_c(kp_i)$ と連想確率 $P_a(kp_i)$ は，

$$P_c(kp_i) = \prod_{j=1}^{m} P_c(k_j), \quad P_a(kp_i) = \prod_{j=1}^{m} P_a(k_j)$$

となる§．この $P_c(kp_i)$ と $P_a(kp_i)$ を $\sum_{i=1}^{n} P_c(kp_i)$ および $\sum_{i=1}^{n} P_a(kp_i)$ で正規化したのち，2 つの尺度の調和平均をとり，その値をキーワード列 kp_i に対するスコア $P(kp_i)$ とする．

$$P(kp_i) = \frac{2}{\frac{1}{P_c(kp_i)} + \frac{1}{P_a(kp_i)}}$$

これを，KP 全てに対して計算し，スコア順にソートした N-best 解をここでの出力とする．図 2 にキーワード列の抽出の例を示す.

3. 文の復元

2. で得られたキーワード列のみでは，音声発話を理解したとは言い難い．キーワード間には依存関係が存在しており，それらの関係を同定する必要がある．ここでは，キーワード間の依存関係を同定するためにコーパスから機能語などを補完し，キーワード列を文に復元する手法をとる．

キーワードからの文の復元には，内元ら [12] の先行研究がある．内元らの文の生成部での出力は，依存構造木だが，提案手法では，評価を (1) 形態素列の評価，(2) 依存構造木の評価，の 2 段階に分けて行う．その理由としては，依存構造を付与しない場合でさえも，膨大な数の文がキーワードから復元されることにある．さらに，そこから依存構造を考えるとキーワード 3 つの場合は 2 倍，4 つの場合は 5 倍の文を評価しなければならない．提案手法ではこのコストを考慮し，まず，形態素列での評価を行い，その後，正しいと評価された形態素列のみから作成される依存構造木を用いて評価するという方法をとる．まず，キーワード列を入力とし，そのキーワードを含む全ての文節をコーパスから獲得し，文節の候補を復元する．この際，フィラーや冗長表現などを含む文節は抽出しないなどの簡単な制約を加えている．次に，得られた文節の全ての組み合わせを考え，文を復元する．文として復元された形態素列に対して，頻出系列パターンを素性とした分類器を作成し，評価を行う．続いて，正例と判断された形態素列に対して依存構造木を作成し，それらを素性とする分類器によって評価を行う．分類器には Support Vector Machine (SVM)[3]¶を用いた．

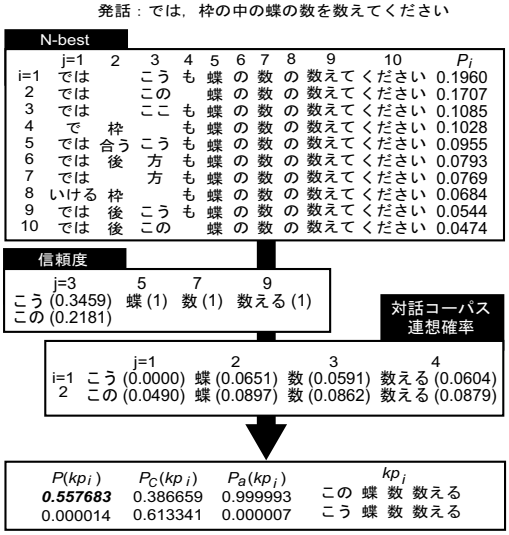

図 2: キーワード抽出の例

3.1 形態素列の評価

まず，復元された文の形態素列の評価のための分類器について説明する．日本語は語順が比較的自由なため，文の評価では，本質的には依存関係を考慮しなければ正しい評価がくだせないものも多い．しかし，ここでの評価とは，ドメイン内において受理可能か否かの評価である．そのため，形態素列だけから評価できるものも少なくないと考えられる．

例えば,「蝶」,「数」および「数える」というキーワード列から，コーパスの情報によって「蝶が数を数える」という形態素列が復元されたと仮定する．ここで,「蝶」「が」「数」という系列パターンを含む文は，それだけで

§ $P_a(k_j) = 0$ になる場合もあるが，その場合は小さな値を与えている．

¶実装には工藤氏の TinySVM を使用した．http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM

は，正例か負例かの判断ができないが，同時にこの文が「数」「を」「数える」という系列パターンを含んでいれば，明らかに負例であると判断できる．このように，系列パターンの組み合わせにより文の妥当性を判断できると考えた．

頻出系列パターンの抽出には PrefixSpan アルゴリズム [9] を使用する．また，素性には形態素の集合 (Bag-of-Words：BOW) も加える．素性の重みは 0 か 1，つまり文中に形態素やパターンが出現すると 1，しない場合は 0 として表現する．

3.2 依存構造木の評価

形態素列の評価において，正例と判断されたものに対して依存構造木を作成し，さらに，それらに対する評価を行う．依存構造木は日本語の係り受けに関する 3 つの制約を満たすもの全てを作成する‖．

ここでの評価基準も 3.1 と同様で，あるドメイン内で受理可能か否かの 2 値分類である．素性には頻出部分木を用い，部分木の組み合わせにより，その文の構造が正しいかどうかの評価を行う．依存構造木は形態素単位の依存構造木を用いる．文節単位の依存構造から形態素単位の依存構造へは以下の手続きで変換を行う．

1. 文節内の形態素は直後の形態素にかける．
2. 文節末尾の形態素は，その文節が係っている文節内の最後尾の自立語 (主辞単語) にかける．

頻出部分木の抽出には，FREQT アルゴリズム [1] を使用する．ここでも，BOW を素性に加え，素性の重みは 0 か 1 で表現する．

4. 実験・考察

提案手法について実験および考察を行う．本稿では，ドメインとして，我々の研究グループが構築を進めている，小学校 1 年生の算数のドリルテキストを対象とした対話支援型問題解決システムの世界を扱う [4]．これは，ドリルテキストならびに人間との対話を相互参照しながら問題解決を行うシステムである．使用する対話コーパスは，そのドメインに対する模擬対話であり，人間対人間の 27 対話と Wizard fo OZ 法によって収集した人間対機械の 45 対話から構成されている．総発話数は 2211 発話である．

4.1 キーワード列の抽出

まず，音声認識結果からのキーワード列の抽出処理について述べる．内容語に対する信頼度を求めるために必要なパラメータである N は 10，α は 0.05 とした．連想確率のモデルには，我々が収集した対話コーパス (72 対話・2211 発話) と，PASD コーパス (51 対話・1084 発話)[10] から作成されたものを使用した．実験には 300 発話 (被験者 6 名 ×50 発話) に対して音声認識エンジン **Julian** から出力された N-best 候補を，人手で形態素長の正規化したものを用いた．使用した発話の平均キーワード数は 4.42 である．評価には，スコアが最も高かった候補が正しいキーワード列であるかどうか (キーワード列正解率)，N-best 解中に正しいキーワード列が含まれいるかどうか (N-best 解正解率) で行った．さらに，抽出されたキーワード全てに対する適合率 (P) および再現率 (R) で評価した．それぞれの評価基準に対して，(1) 音声認識結果の第 1 候補中のキーワード，(2) 内容語の信頼度のみを用いた場合，(3) 提案手法，について比較した．結果を表 1 に示す．

‖語順は固定して作成している．

表 1: キーワード列抽出実験の結果

	(a) 第 1 候補	(b) 信頼度のみ	(c) 提案手法
キーワード列正解率	43.67	34.33	46.00
N-best 解正解率	——–	——–	52.33
適合率	80.64	78.21	81.86
再現率	85.14	82.58	86.43

表 2: 系列パターン＋ BOW 実験結果 (F 値)

	C=0.05	0.1	0.5	1
m =10	53.12	61.57	65.60	64.04
20	62.18	70.17	73.20	72.57
30	65.41	73.36	**76.29**	76.01

提案手法は，第 1 候補のみ，内容語の信頼度のみを用いる手法 (先行研究を拡張した手法) と比べて，全ての尺度においてポイントが向上している．これにより，内容語の信頼度と連想確率によるキーワード抽出手法が有効であることが確認された．全体的に精度が低いのは，基準となる音声認識結果の文正解率**が低いためであり，N-best 出力中に抽出すべきキーワードが含まれていない場合があったためである．この問題については，例えば，文の復元の段階で連想確率を用いることにより，抽出洩れもしくは音声認識誤りにより得られなかった内容語の復元も可能だと考えており，今後の課題の一つである．

4.2 復元文の評価

まず，形態素列の評価を行う．評価には，キーワード列 60 組 (4 つのキーワード列 50 組，3 つのキーワード列 10 組) から復元した 321126 文からサンプリングされた訓練データ 20000 文とテストデータ 10000 文を用いた．

PrefixSpan アルゴリズムは工藤氏の実装を使用した††．系列パターンの長さは，2 から 5 とした．SVM のカーネル関数は線形関数を使用した．実験では，系列パターンの最小サポート数および SVM のソフトマージンパラメータ C をいくつか変化させて行った．評価は F 値によって行った．

$$F \text{ 値} = \frac{2}{\frac{1}{P} + \frac{1}{R}}$$

適合率 (P) と再現率 (R) は，分類器が正例として正しく判断した文の数に基づき算出される．実験結果を表 2 に示す．表中で，C はソフトマージンを，m は最小サポート数を表す．太字は全体での最高値である．使用した素性数は，m =10 の場合が 85312，20 の場合が 38742，30 の場合が 24499 であった．

続いて，依存構造木の評価を行う．評価には，形態素列の評価で用いた訓練データおよびテストデータから，正例を 1140 文，負例を 2111 文サンプリングし，それら

**第 1 候補の文正解率は 22.33%であった．

††http://cl.aist-nara.ac.jp/~taku-ku/software/prefixspan

表 3: 部分木＋ BOW 実験結果 (F 値)

	0.05	0.1	0.5	1
m =5	81.94	87.59	**90.91**	90.31
10	72.30	83.64	88.89	88.58
15	61.98	77.96	86.27	86.98

に対して依存構造木を作成した 12500 文を用いた．訓練データは 10000 文，テストデータは 2500 文とした．

FREQT アルゴリズムには，工藤氏の実装を使用した[‡‡]．部分木の長さは 3 から 6 とした．ここでも，SVM のカーネルは線形関数を使用し，頻出部分木の最小サポート数と SVM のソフトマージンパラメータ C をいくつか変化させて実験した．評価には，形態素列による評価と同じ尺度を用いる．実験結果を表 3 に示す．ここで使用した素性数は，m =5 の場合で 14109，10 の場合は 5460，15 の場合では 4342 であった．

双方の実験結果から，提案手法の有効性は示せた．キーワードからの文生成により，キーワード間の依存関係を扱えるようになるため，提案手法はキーワード抽出に基づく音声認識手法の精度向上に繋がると考えられる．しかしながら，他のドメインで同程度の精度が得られるかどうかは今後の課題となる．内元らの先行研究は，生成文の評価を形態素解析および係り受け解析の応用として扱っていることや，その目的が自然な文の生成であり，提案手法とは問題設定が異なるため，単純な比較は難しい．今回の実験は比較的小規模であり，今後さらに大規模な実験を行う必要がある．その際の計算的負荷についても議論の対象となるだろう．解決策としては String Kernel や Tree Kernel などのカーネル関数の導入などが挙げられる．また，現在は対話の文脈やその発話の意図などを全く考慮していない．そのような情報をどのように提案手法に反映させ，精度を向上させるかも今後の課題の一つである．

5. おわりに

本稿では，音声認識誤りに頑健な音声発話理解を実現するためにキーワード抽出に基づく音声発話理解手法について述べた．提案手法の特徴は，(1) コーパスの情報を用いたキーワード抽出，(2) キーワードから復元された文による音声発話理解の実現，である．キーワード抽出では，内容語の信頼度と連想確率の 2 つの尺度を用いることにより，内容語の信頼度のみの場合に比べて，キーワード列の正解率で約 11%の精度の向上が見られた．キーワードからの文の復元に関しては，形態素列と依存構造木による 2 段階の評価を行った．限られたドメインであるが，一定の有効性を示せた．キーワードからの文生成は，キーワード抽出に基づく音声認識手法の精度向上に繋がると考えられる．

今後の課題としては，より大規模な実験や他ドメインでの提案手法の有効性の確認，各プロセスのさらなる精度向上や一般性の向上などが挙げられる．

[‡‡] http://cl.aist-nara.ac.jp/~taku-ku/software/freqt

参考文献

[1] K. Abe, S. Kawasoe, T. Asai, H. Arimura, S. Arikawa : Optimized Substructure Discovery For Semi-structured Data, Proc. 6th European Conference on Principles and Practice of Knowledege Discovery in Databases (PKDD), LNAI 2431, pp.1-14 (2002).

[2] C. Bouwman, J. Sturm, L. Boves : Incorporating Confidence Measures in the Dutch Train Timetable Informataion System Developed in the ARICE project, Proc. ICASSP (1999).

[3] C. Cortes, V. Vapnilk : Support Vector Networks, Machine Learning, Vol.20, pp.273-297 (1995).

[4] T. Endo, T. Kagawa : Cooperative Understanding of Utterances and Gestures in a Dialogue-based Problem Solving System, Computational Intelligence, Vol.5, No.2, pp.152-169 (1999).

[5] 駒谷和範，河原達也：音声認識結果の信頼度を用いた効率的な確認・誘導を行う対話管理，情報処理学会論文誌，Vol.43 No.10，pp.3078-3086 (2002).

[6] 宮崎昇，中野幹生，相川清明：n-best 音声認識と逐次理解法によるロバストな音声発話理解，情報処理学会研究報告，2002-SLP-40-32，pp.121-126 (2002).

[7] 持橋大地，松本裕治：連想としての意味，情報処理学会研究報告，99-NL-134，pp.155-162 (1999).

[8] M. Nakano, N. Miyazaki, J. Hirasawa : Under-Standing Unsegmented User Utterances in Real-Time Spoken Dialogue Systems, Proc. 37th Annual Meeting of the Assosiation for Computational Linguistics (ACL), pp.200-207 (1999).

[9] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal, M. Hsu : PrefixSpna :Mining Sequential Efficiently by Prefix-Projected Pattern Growth, Proc. International Conference of Data Engineeringm (ICDE), pp.215-224 (2001).

[10] 重点領域研究「音声対話」音声対話コーパス WG 編集：平成６年度文部省科学研究費補助金重点領域研究「音声・言語・統合の統合的処理による対話の理解と生成に関する研究」対話音声コーパス，CD-ROM Vol.1-4 (1995).

[11] 竹林洋一：音声自由対話システム TOSBURGII — ユーザ中心のマルチモーダルインタフェースの実現に向けて—，電子情報通信学会論文誌，Vol.J77-D-II, No.8，pp.1417-1428 (1994).

[12] 内元清貴，関根聡，井佐原均：キーワードからのテキスト生成，言語処理学会第８回年次大会，pp.375-378 (2002).

[13] 屋野武秀，笹島宗彦，河野恭之：音声対話タスクのための高速なキーワードラティスパーザ BTH，人工知能学会論文誌，Vol.17，No.6，pp.658-666 (2002).