Vol.2009-NL-191 No.12
Vol.2009-SLP-76 No.12
2009/5/21

【パネル討論】音声インタフェースにおけるWebテキスト処理技術の利用

中野 幹生（司会）[†1]　緒方 淳[†2]　清田 陽司[†3]
東中 竜一郎[†4]　翠 輝久[†5]

音声インタフェースの性能の向上と構築コストの低減のために，Web テキスト処理技術を用いる方法に関し，研究事例や課題を共有し，今後の方向性に関して議論する．

[Panel Discussion] Using Web Text Processing Technologies in Building Speech Interfaces

Mikio Nakano (Coordinator),[†1] Jun Ogata,[†2]
Yoji Kiyota,[†3] Ryuichiro Higashinaka[†4]
and Teruhisa Misu[†5]

This panel discussion aims at sharing information on the current state of research and issues in using Web text processing technologies for improving speech interfaces and reducing the cost of building them. We also discuss future directions.

†1 (株) ホンダ・リサーチ・インスティチュート・ジャパン
Honda Research Institute Japan Co., Ltd.
†2 (独) 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology
†3 東京大学 情報基盤センター
Information Technology Center, University of Tokyo
†4 日本電信電話 (株) NTT コミュニケーション科学基礎研究所
NTT Communication Sience Laboratories, NTT Corporation
†5 (独) 情報通信研究機構
National Institute of Information and Communications Technology

1. 趣旨（中野）

音声インタフェースは，音声認識・言語理解・対話管理・言語生成・音声合成など様々な技術の統合システムであり，それらの個々のモジュールが，音響モデル・言語モデル・対話モデル・対話コンテンツなどの知識を用いて動作している．これらの知識の中には，音声インタフェースのタスクドメイン毎に用意しなくてはならないものもあり，その構築には専門家による作業が必要である．たとえば，言語理解の規則を記述したり，統計モデルの構築のための大量のタグ付けデータや書き起こしを用意したりすることなどが必要である．このことが様々な音声インタフェースが構築されて一般に普及するのを妨げる一要因になっていると考えられる．

この問題に対し，従来，様々な研究が行われてきた．ひとつには，専門的な知識がなくても知識を記述できるようにする，対話記述言語とそのインタプリタの研究がある（例えば文献 1)–4) ）．インタプリタは，対話記述言語から言語モデル，言語理解モデル，応答選択規則などを自動生成して対話システムを動作させる．また，対話の実行中や対話終了後にシステムが知識を自動的に構築する研究も行われている（例えば文献 5)，6)）．

近年，この課題に対し，Web 上のリソースを用いる研究が盛んになってきている．Web テキストからテキスト処理技術を用いて，発音・構文・意味的な知識や，常識・最新のニュースなどの幅広い知識を自動的に取り出し，それらを使って言語モデルや対話コンテンツを構築する研究が進んできた．このような手法は，テキスト処理技術と音声処理技術の新たな形のコラボレーションを生んでいる．

本パネル討論は，このような Web テキスト処理技術を音声インタフェースの構築に用いる手法について，その現状と課題に関して情報を共有し，将来課題について議論することを目的としている．パネリストは，Web テキスト処理を用いた音声インタフェースの研究を進められている緒方氏（産総研），東中氏（NTT），翠氏（NICT）および，Web テキスト処理に詳しい清田氏（東大）にお願いした．

以下の各節では，パネリストの方々に研究事例と課題を簡単にまとめていただく．

2. 様々な Web リソースを活用した音声認識技術（緒方）

我々は，音声認識の新たなアプリケーションとして，Web 上の膨大な音声データを検索可能なシステム「PodCastle」の開発を行っている[7)–10)]．ポッドキャスト等の Web 上の音声データは，その内容や録音環境が多様であるため，従来の音声認識システムでは高精度

Vol.2009-NL-191 No.12
Vol.2009-SLP-76 No.12
2009/5/21

な認識を行うことは困難であった．これに対し我々は，従来のように研究技術の枠組みだけで解決するのではなく，Web を通じて得られる知識やリソースを積極的に利用することで，音声認識性能を改善するアプローチについて検討してきた．

2.1 日々成長する音声認識システム

ポッドキャストは幅広いトピックを含み，しかもそれらが日々更新されるという特徴を持っているため，従来のように事前に用意されたテキストで学習された言語モデルのみで認識を行うことは困難である．そこで我々は，最新の幅広いトピックを網羅している Web ニュースサイト (Yahoo! ニュース，Google ニュース) のテキストを言語モデル学習に利用して，常に音声認識システムが最新の話題に対応できる仕組みを構築した[9]．しかし，最新ニュースに出現する新出語の中には，読み (発音) がわからない単語も少なくない．このような単語を音声認識で利用可能とするために，集合知によるキーワード辞書サービス (はてなダイアリーキーワード) を活用した新出語の読み獲得機能も備えている[9]．

2.2 ユーザの協力で育つ音声認識システム

PodCastle では特徴的な機能として，独自のインタフェース (音声訂正[11]) を通じて不特定多数のユーザが音声認識誤りを訂正することを可能にしている．これにより検索サービスとしての質を向上させるだけでなく，ユーザの協力を得ることで，音声認識技術の底上げをはかることも狙っている[8]．このような,「不特定多数のユーザによる音声認識誤りの訂正情報」という新たな Web リソースを活用した，音声認識システムの協調的学習 (collaborative training) の試みを行っており，実際にポッドキャスト音声認識の性能を大きく改善できることを確認している[10]．

ポッドキャスト，動画共有サイトの普及により，今後 Web 上では音声データが増加の一途を辿ることが予測される．そのようなデータを適切に扱うためにも,「日々成長し」,「ユーザの協力で育つ」という 2 つのアプローチは，今後，音声言語処理技術において必要不可欠になると考えられる．

3. 会話を楽しむ音声インタフェースを目指して（東中）

音声インタフェースの研究はタスク指向型のものが多い．例えば，フライト予約や天気情報案内などを行うシステムが代表的である．ユーザはシステムに要求を伝え，システムがユーザの意図を理解し，特定のタスクを遂行する．しかし，長年の研究にも拘らず，これらのシステムはいまだ広く使われていない．これは，音声認識の性能が実環境下では厳しいこともさることながら，結局のところ，タスクを遂行するために音声が最も効率の良い手段ではないからであろう．筆者らは，音声インタフェースが最も求められる場面は，雑談のような楽しめる会話だと考え，ユーザが親しみやすく話したくなる対話システムの研究を行っている．

筆者らが構築した「クイズ型対話システム」[12),13)] は,「この人は誰？」という人名クイズを作成し，ユーザが人名を当てるまでヒントを順次出し続けるものである．例えば，正解が「織田信長」の場合,「戦国武将」や「1534 年生まれ」,「本能寺の変で殺された」などがヒントとなり，ユーザはなるべく少ないヒントで人名を当てなければならない．ヒントは Wikipedia から自動的に作成され，正解の人名との共起にしたがって，難しい順番に並び替えられユーザに提示される．ユーザが間違った人名を答えるとシステムは「惜しい」や「ぜんぜん違う」といった応答を行う．この応答の生成には，Wikipedia から獲得した人名間の距離を用いている．クイズ型対話システムは，クイズというタスクを扱うが，タスクの達成を目的とするものではない．むしろ，答えを考えるという過程を楽しむものである．

筆者らは「共感喚起型対話システム」[14)] の研究も行っている．このテキスト対話システムは，システム自身の親しみやすさや対話の満足度の向上を目的とし，ユーザに対し自己開示と共感表出を行う．例えば，ユーザが「私はホタルが好き」と発話すると，システムは「私もホタルが好きなんです．儚いですよね.」といった応答を行う.「ホタル」の「儚い」といった属性は，連想概念辞書[15)] から得ている．現在は，動物の好き嫌いについての対話のみを対象としているが，実験により，共感を多く行うシステムは，ユーザの共感を誘引することができ，システムの親しみやすさや対話の満足度を向上させることが分かった．感情を持たないシステムの共感表出であっても，ユーザの心理状態に良い影響を与えるという結果は,「楽しみ」を目的とした対話システムの可能性を示すものである．

上記システムの構築により，筆者らは「楽しめる会話」を一部実現できたのではないかと考えているが課題も多い．例えば，クイズ型対話システムは Wikipedia の見出し語を用いた「○○当てクイズ」しか作成できない．これは，Wikipedia では見出し語と本文に，対象とその説明という関係が成り立っていることを利用してクイズを作成しているからであるが，一般の Web テキストからより多くの関係性を抽出できれば，クイズの幅が広がり，会話がより楽しめるものになるだろう．共感喚起型対話システムについては，システムのドメイン知識の作成コストが大きな問題となっている．動物の知識も，連想概念辞書だけでは十分でなく，手作業で網羅することも困難である．できれば Web テキストなどの大規模なデータから，自動的に獲得することが望ましい．より豊富な話題で楽しめる会話を実現するために，今後の関係抽出技術と属性抽出技術の進展に期待する．

Vol.2009-NL-191 No.12
Vol.2009-SLP-76 No.12
2009/5/21

4. 文書知識ベースを利用した音声対話システム（翠）

4.1 研究事例

Web テキスト処理技術の利用に関する討論ではあるが，もう少し広く，文書を知識ベースとして利用するタスク指向型の音声対話システムを取り上げる[⋆1]．従来の研究・開発では，バックエンド知識ベースに関係データベース (RDB) 検索を用いるものが大半であったが，近年の Web 検索に代表される情報検索技術の発展に伴い，新聞記事やマニュアルなど，より一般的な文書検索を対象を用いる研究が盛んに行われている[16),17)]．このようなシステムでは一般的に音声認識結果から作成するクエリと，バックエンド知識のマッチングを行いユーザに提示する．

音声により文書知識ベースを提示する際の課題として (特に情報提示にディスプレイを利用することが困難な場合に) 一度に提示できる情報量が制限されることが挙げられる．そのため，該当するすべての候補を Web 検索エンジンのように提示するのではなく，十分に絞り込んだ上で提示する必要がある．我々はこれまでに，ソフトウェアサポート知識ベースの検索をタスクとして，係り受け情報などのバックエンド知識を利用した検索クエリ明確化のための対話戦略を提案している[18)]．また，該当する文書が一意に絞られた場合であっても，長々と説明を音声により読み上げるのは適切ではない．そのため，要約，質問応答や情報推薦技術を利用して，対話的に情報を提示する枠組みを提案している[19)]．

4.2 Web テキストを音声対話に利用する際の課題

音声インタフェースでは，ユーザが必要としている情報を簡潔に答える必要がある．そのため，テキストインタフェースでの情報検索，質問応答においては，N-best 候補中に正解があれば十分な場合も多いが，音声インタフェースにおいては基本的に第 1 候補で提示する必要がある．そのために，検索精度を上げる (ドメインやタスクに合致した情報の検索・ゴミの棄却) ことが課題であると考える．

Web テキストの利用により，音声認識モジュールや，対話コンテンツの自動構築はある程度の成功を収めているが，ドメイン知識 (=対話の遂行にはどのような情報が重要かなど) は，基本的に人手で与える必要がある．このような知識はドメインやタスクに大きく依存するため，新しい対話システム構築の障害になっている．この問題に対しては，Web の大量のテキストからボトムアップに知識を獲得する技術 (例えば文献 20)) が，対話システムへの適用が可能であると期待される．

⋆1 ここに挙げる研究事例の多くは，翠が京都大学河原研究室にて行った研究成果である．

5. Web テキストの「緩い組織化」（清田）

Web テキスト処理を音声インタフェースに応用する際には，Web テキストに対して「語彙集合の獲得」「ドメイン分類」「コーパスへのタグ付け」など，さまざまなレベルの「組織化」を行う必要がある．従来の音声インタフェース研究は，厳密な基準にもとづいた精緻な組織化を前提としていた．しかし，厳密な基準による組織化を行うためには専門家による高コストの作業を必要とするため，Web テキストの長所である膨大なデータ量を活かすことが難しい．この問題を解決するには，ある程度の誤りを許容し，専門家以外でもできる「緩い組織化」がひとつのポイントになると考えている．

Web テキストの世界では「緩い組織化」の考え方が普及しつつある．代表例としては Flickr，delicious，YouTube などで採用されているフォークソノミーが挙げられる．フォークソノミーでは，(専門家ではない) 個々の参加者がそれぞれの観点に基づいたタグを自由に付与することができる．中には的外れとも思えるタグも付与されることもあるが，システム全体としてみれば多様な観点を反映した分類が形成されているように見える．「緩い組織化」の考え方は Wikipedia の中でも見られる．Wikipedia 日本語版に登録されている 50 万以上の記事は，Wikipedia カテゴリと呼ばれる仕組みを用いて整理されていて，全体として緩い分類体系を構成している．筆者らが開発している図書館ナビゲーションシステム[21)]では，Wikipedia カテゴリを利用して「Winny」というキーワードから「著作権法」「ソフトウェア」「通信」などの図書館分類のキーワードを自動的に導出している．

音声認識において欠かせない語彙抽出でも,「緩い組織化」の考え方は利用されている．田中ら[22)] はサーチエンジンで得られるスニペットから文字単位のトライ木を構築し，分岐数の変化を利用してセグメンテーションを行うことで，言語に依存しない緩い定型表現の抽出を実現している．

タグ付けにおいても「緩い組織化」を利用できる可能性が生まれている．Amazon が提供している Mechanical Turk というサービスでは，多数の参加者にタグを付与させることで，一カ所あたり数セントという低コストでタグ付きデータを構築することができる．その成果が EMNLP 2008 などで報告されている[23)]．

上記で述べた「緩い組織化」の考え方を推し進めていくことで，音声インタフェースにおいて以下のような応用が可能ではないかと考えている．

- ドメイン辞書の構築

Vol.2009-NL-191 No.12
Vol.2009-SLP-76 No.12
2009/5/21

Wikipedia カテゴリを利用すると,「テレビ番組名」「菓子の商品名」など，これまで存在しなかった分野の用語辞書を獲得することができる．また，田中らの考え方を応用することでドメイン毎の未知語を獲得できる可能性がある．

- 対話シナリオの構築
  Web 上のコンテンツ間の関係を示すメタデータを Mechanical Turk のような仕組みで付与することで，Web コンテンツを活用した対話システムが実現できるかもしれない．
- エンターテインメント性のあるインタフェース
  音声対話システムの利用にあたっての心理的な障壁を乗り越えるには，エンターテインメントの要素を持たせることも重要だと考える．フォークソノミーでは,「もっと評価されるべき」「これはひどい」など，コンテンツの面白さを基準としたタグ付けがなされている例が多くみられる．これらのタグを対話の媒体として利用できないだろうか?

6. おわりに

以上，パネリストの方々に研究事例と課題を簡単にまとめていただいた．パネル討論では，共通の課題と方向性に関してパネリスト以外の参加者の方々とともに深く議論する．

参考文献

1) Araki, M., Kouzawa, A. and Tachibana, K.: Proposal of a multimodal interaction description language for various interactive agents., *IEICE Transactions on Information and Systems*, Vol.E88-D, No.11, pp.2469–2476 (2005).
2) 桂田浩一，中村有作，山田　真，山田博文，小林　聡，新田恒雄：MMI 記述言語 XISL の提案，情報処理学会論文誌，Vol.44, No.11, pp.2681–2689 (2003).
3) Nishimura, Y., Minotsu, S., Dohi, H., Ishizuka, M., Nakano, M., Funakoshi, K., Takeuchi, J., Hasegawa, Y. and Tsujino, H.: A Markup Language for Describing Interactive Humanoid Robot Presentations, *Proc. IUI'07* (2007).
4) W3C: Voice Extensible Markup Language (VoiceXML) Version 2.0, W3C Recommendation (2004).
5) Holzapfel,H., Neubig,D., and Waibel,A.: A dialogue approach to learning object descriptions and semantic categories, *Robotics and Autonomous Systems*, Vol.56, No.11, pp. 1004–1013 (2008).
6) Sudoh, K. and Nakano, M.: Post-dialogue confidence scoring for unsupervised statistical language model training, *Speech Communication*, Vol.45, No.4, pp.387–400 (2005).
7) 緒方　淳，後藤真孝，江渡浩一郎：PodCastle: ポッドキャストをテキストで検索，閲覧，編集できるソーシャルアノテーションシステム，WISS 2006 論文集，pp.53–58 (2006).
8) Goto, M., Ogata, J. and Eto, K.: PodCastle: A Web 2.0 Approach to Speech Recognition Research, *Proc. of Interspeech 2007* (2007).
9) Ogata, J., Goto, M. and Eto, K.: Automatic Transcription for a Web 2.0 Service to Search Podcasts, *Proc. of Interspeech 2007* (2007).
10) 緒方　淳，後藤真孝：PodCastle:ポッドキャスト音声認識のための集合知を活用した音響モデル学習，第 3 回音声ドキュメント処理ワークショップ (2009).
11) 緒方　淳，後藤真孝：音声訂正：選択操作による効率的な誤り訂正が可能な音声入力インタフェース，情処学論，Vol.48, No.1, pp.375–385 (2007).
12) Higashinaka, R., Dohsaka, K., Amano, S. and Isozaki, H.: Effects of Quiz-style Information Presentation on User Understanding, *Proc. Interspeech*, pp.2725–2728 (2007).
13) Sawaki, M., Minami, Y., Higashinaka, R., Dohsaka, K. and Maeda, E.: "Who is this" Quiz Dialogue System and Users' Evaluation, *Proc. 2008 IEEE Workshop on Spoken Language Technology (SLT 2008)*, pp.149–152 (2008).
14) Higashinaka, R., Dohsaka, K. and Isozaki, H.: Effects of Self-Disclosure and Empathy in Human-Computer Dialogue, *Proc. 2008 IEEE Workshop on Spoken Language Technology (SLT 2008)*, pp.109–112 (2008).
15) 岡本　潤，石崎　俊：概念間距離の定式化と既存電子化辞書との比較，自然言語処理, Vol.8, No.4, pp.37–54 (2001).
16) 伊藤克亘，藤井　敦：NTCIR-3 ワークショップにおける音声入力 web 検索タスク，情処研報, SLP-43-5 (2002).
17) Akiba, T. and Abe, H.: Exploiting Passage Retrieval for N-Best Rescoring of Spoken Questions, pp.65–68 (2005).
18) Misu, T. and Kawahara, T.: Dialogue Strategy to Clarify User's Queries for Document Retrieval System with Speech Interface, *Speech Communication*, Vol.48, No.9, pp.1137–1150 (2006).
19) 翠　輝久，河原達也：質問応答・情報推薦機能を備えた音声による情報案内システム，情報処理学会論文誌，Vol.48, No.11, pp.3078–3086 (2007).
20) 鳥澤健太郎，隅田飛鳥，野口大輔，柿澤康範，風間淳一，De Saeger, Stijn，村田真樹，山田一郎，塚脇幸代，太田公子：ウェブ検索ディレクトリの自動構築とその改良 - 鳥式改-.，言語処理学会第 15 回年次大会発表論文集 (2009).
21) Kiyota, Y., Tamura, N., Sakai, S., Nakagawa, H. and Masuda, H.: Automated Subject Induction from Query Keywords through Wikipedia Categories and Subject Headings, *The Sixth International Conference on Language Resource and Evaluation (LREC 2008)* (2008).
22) Tanaka-Ishii, K. and Nakagawa, H.: A multilingual usage consultation tool based

Vol.2009-NL-191 No.12
Vol.2009-SLP-76 No.12
2009/5/21

on internet searching: more than a search engine, less than QA, *Proceedings of the 14th international conference on World Wide Web (WWW 2005)*, pp.363–371 (2005).

23) RionSnow, BrendanO'Connor, D. J. A. Y.N.: Cheap and Fast — But is it Good? Evaluating Non-Expert Annotations for Natural Language Tasks, *Proceedings of EMNLP 2008*, pp.254–263 (2008).