

人工知能と対話技術

(株) ホンダ・リサーチ・インスティテュート・ジャパン
船越 孝太郎

2016/10/21

北九州学術研究都市産学連携フェア
ひびきのAI社会実装研究会

Honda Research Institute **JP**

講師所属紹介

- ・ HRI-JP：ホンダ・リサーチ・インスティテュート・ジャパン
 - ・ 2003年設立
 - ・ HONDA（本田技術研究所）の研究専門子会社
 - ・ 先端技術研究（ソフトウェア）
 - ・ <http://www.jp.honda-ri.com/>
- ・ 人工知能学会誌 31(5) 2016/9に
研究拠点紹介記事が掲載されています
（「AI書庫」より無料でアクセス）



Honda Research Institute **JP**

概要

人と話すコンピュータ／ロボット

何ができるのか，どうつくればいいのか，今後どうなるのか

- ・ 人工知能と対話システム
- ・ 対話システムの作り方
- ・ まとめと展望

人工知能と 対話システム

人工知能とは

人工知能Q&A（一般向け解説）

<http://www.jp.honda-ri.com/ai/faq/01.html>

- ・ “知的な”コンピュータソフトウェア

- ・ 人・生物が行っている様々な問題解決を自動的に行えるようにしたプログラム

[自律] 例：人の助けを借りず自分だけで部屋を掃除する

[認識] 例：いろいろなものが写っている風景写真の中で人の顔だけを見つける

[分類] 例：顔写真に写っている人が誰かを当てる

[予測] 例：サッカーで、どこにボールを蹴ったらパスが通るか判断する

[学習] 例：教えられたり試行錯誤したりして、新しいことができるようになる

[計画] 例：夕飯までに済ませられるように買い物に行く店の順番を決める

[推論] 例：テスト問題やクイズを解いたり、ゲームに勝つ手を見つけたりする

[言語] 例：話しかけられるとそれを理解して適切な返事をする

[創造] 例：新しい道具やゲームを作ったり、作曲したりする

上記のいずれかに該当すると人工知能とよばれる（ことがある）

- ・ 「強いAI」：全部できるAI, 「弱いAI」：いずれか, かつ, その一部だけ

対話システムとは

「人間と自然言語でコミュニケーションし、
情報を授受するソフトウェア」

⇒ 人工知能（弱いAI）の一種

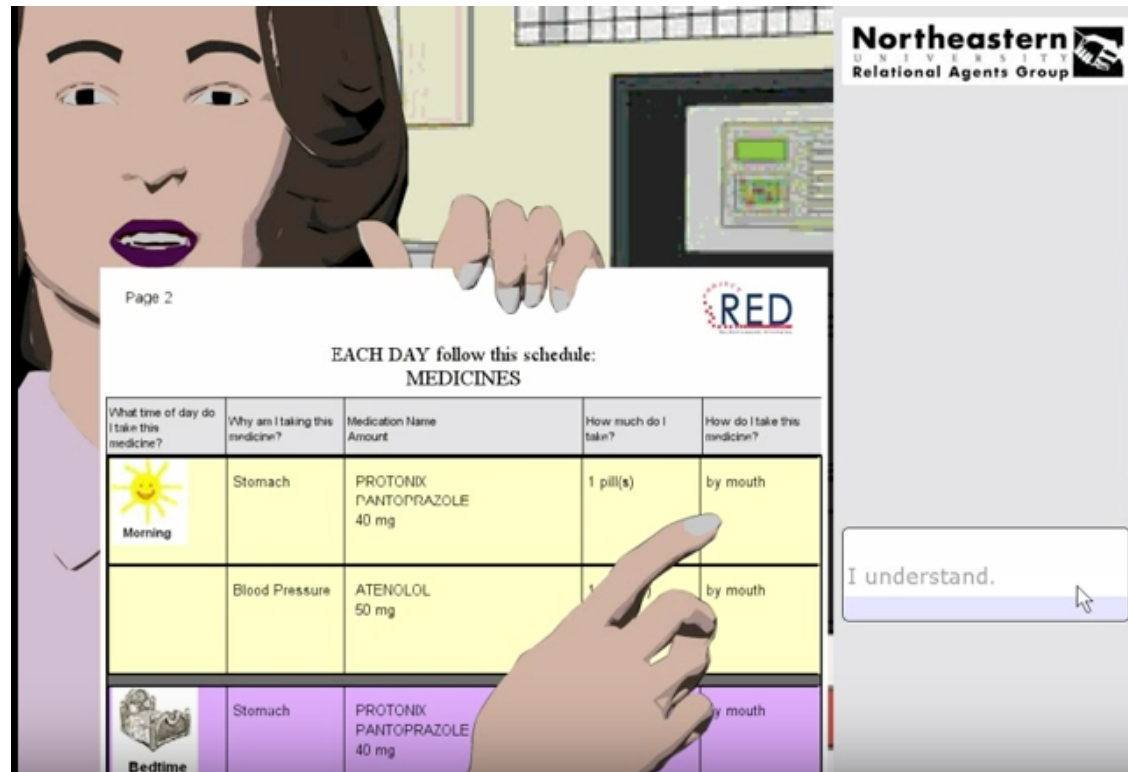
会話ロボット = マルチモーダル対話システム

- ・ マルチモーダル： 音声言語 + α の情報授受手段を持つ
 - ・ 入力が多モーダル：画像を中心とした非言語情報処理
 - ・ 出力が多モーダル：ディスプレイ, LED, アクチュエータ

対話システムの価値

- ・ インターフェースとしての価値
 - ・ 複雑な情報を素早く、簡便に入力できる
(任意の量の情報を任意の単位で伝えられる)
 - ・ 両手がふさがっていても、遠方からでも操作できる
(音声の場合)
 - ・ 使い方を学ばなくても使える
(理想的には、現実はずしもそうではない)
- ・ アプリケーションとしての価値
 - ・ 会話訓練
(語学学習者、自閉症患者など)
 - ・ コミュニケーションパートナー
(独居老人など：孤独感解消、ストレス発散、認知症予防)
 - ・ より効果的な情報伝達メディア
(場合によっては人が伝えるよりもよい)

より効果的な情報伝達メディア



<https://www.youtube.com/watch?v=TGUQkWQRLuU>

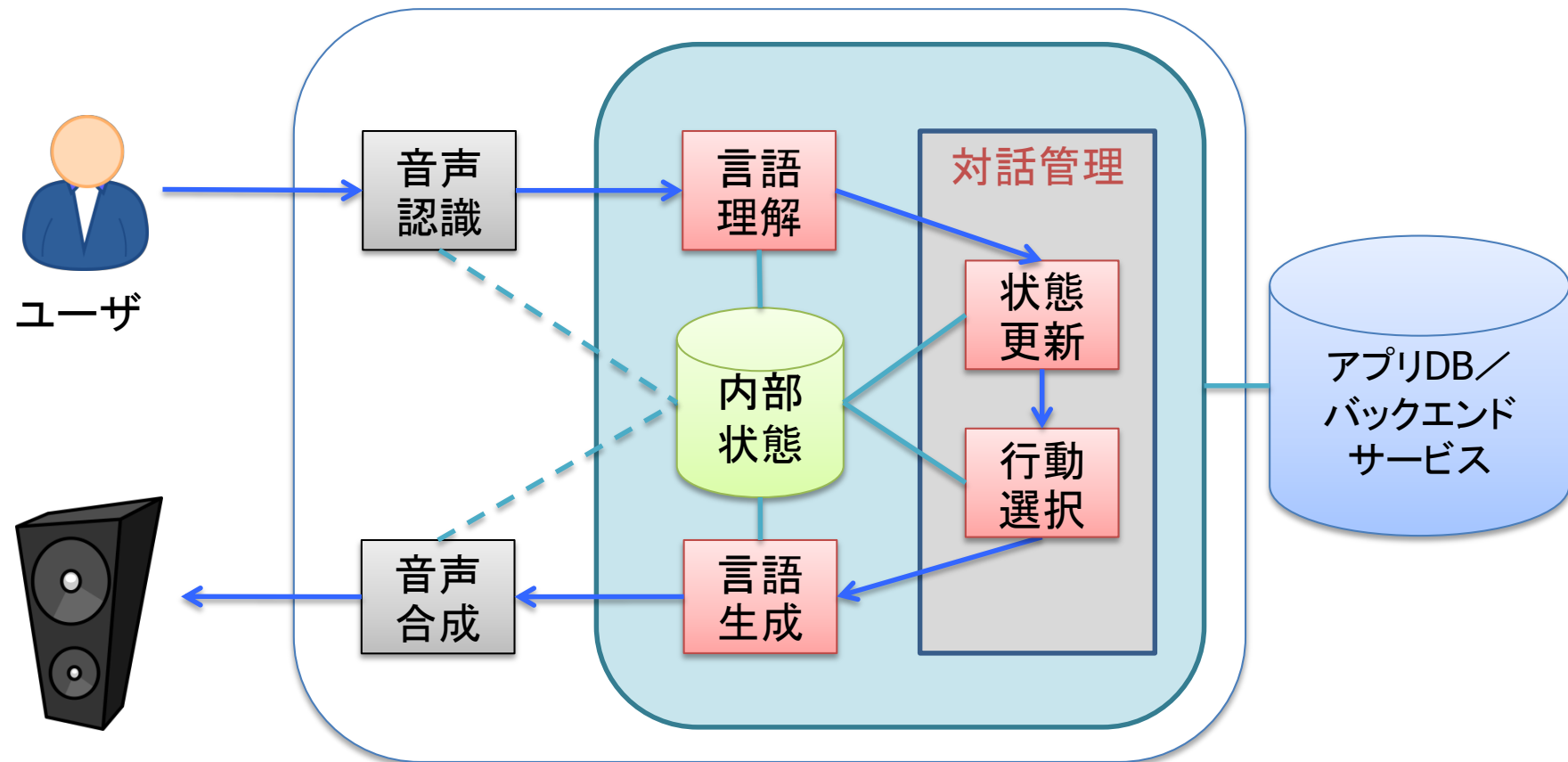
Virtual Discharge Nurse :

退院後の注意点・指示をインタラクティブに説明

・再入院する患者の数が実際に減少

- ・ 指示書を渡すだけでは、ちゃんと読んで理解してもらえない
- ・ ナースの説明では、患者が自分のペースで納得行くまで聞きづらい

対話システムの基本構造



対話システムについて知っておくべきこと

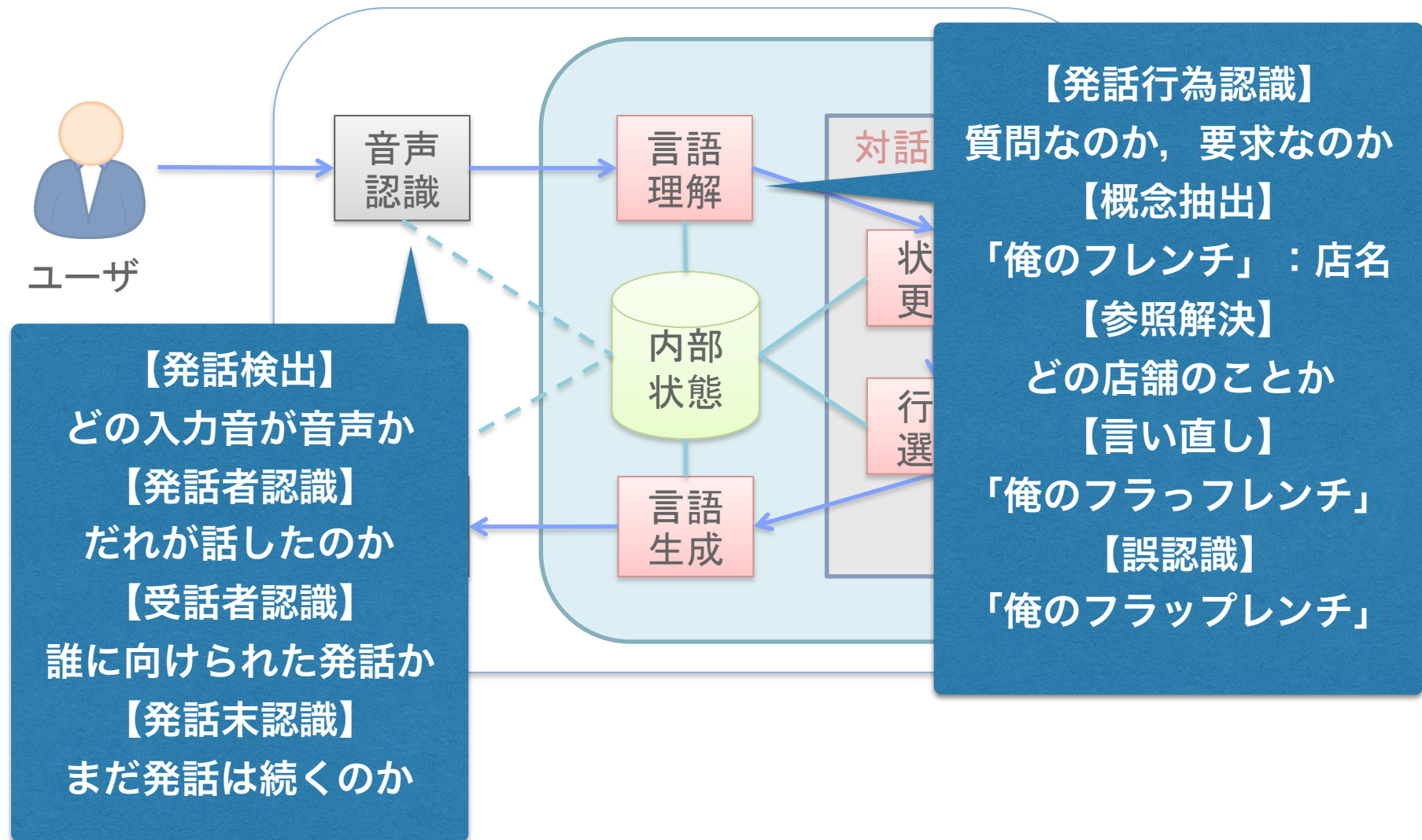
- ・ **音声認識と音声言語理解は別の段階・問題**
 - ・ 音声認識できるからといって理解できるわけではない
 - ・ 音声認識自体，雑音環境ではまだまだ人に及ばない
- ・ **なんでもできる対話システムはない**
どんな対話でも実現できるわけではない
 - ・ 技術的に実現可能な範囲を見定め，システムを設計する
 - ・ 目指す対話によって適切なアプローチが違う（万能薬はない）
- ・ **雑談システムのほうが（ちゃんと）作るのが難しい**
 - ・ タスク指向システム：対話は短ければ短いほどよい
 - ・ 雑談（非タスク指向）システム：対話が長く続くほどよい

対話システムについて知っておくべきこと

- ・ **商用システム・サービス開発では、人手作業に頼るところ大**
 - ・ 入力処理では機械学習は使えるところでは使うが、品質管理上、出力処理では使いづらい
 - ・ 開発に人手が入る要素が大きいので、開発支援ツールの構築やソフトウェア工学・開発方法論の重要性が高い
- ・ **人と同じようには話せない**
 - ・ 設計の段階で人の対話の仕方とは異なる
- ・ **言葉を話せばだれでも使える、わけではない**
 - ・ ユーザが適切な使い方を把握している・できる必要がある
(UI/UX設計が大事)

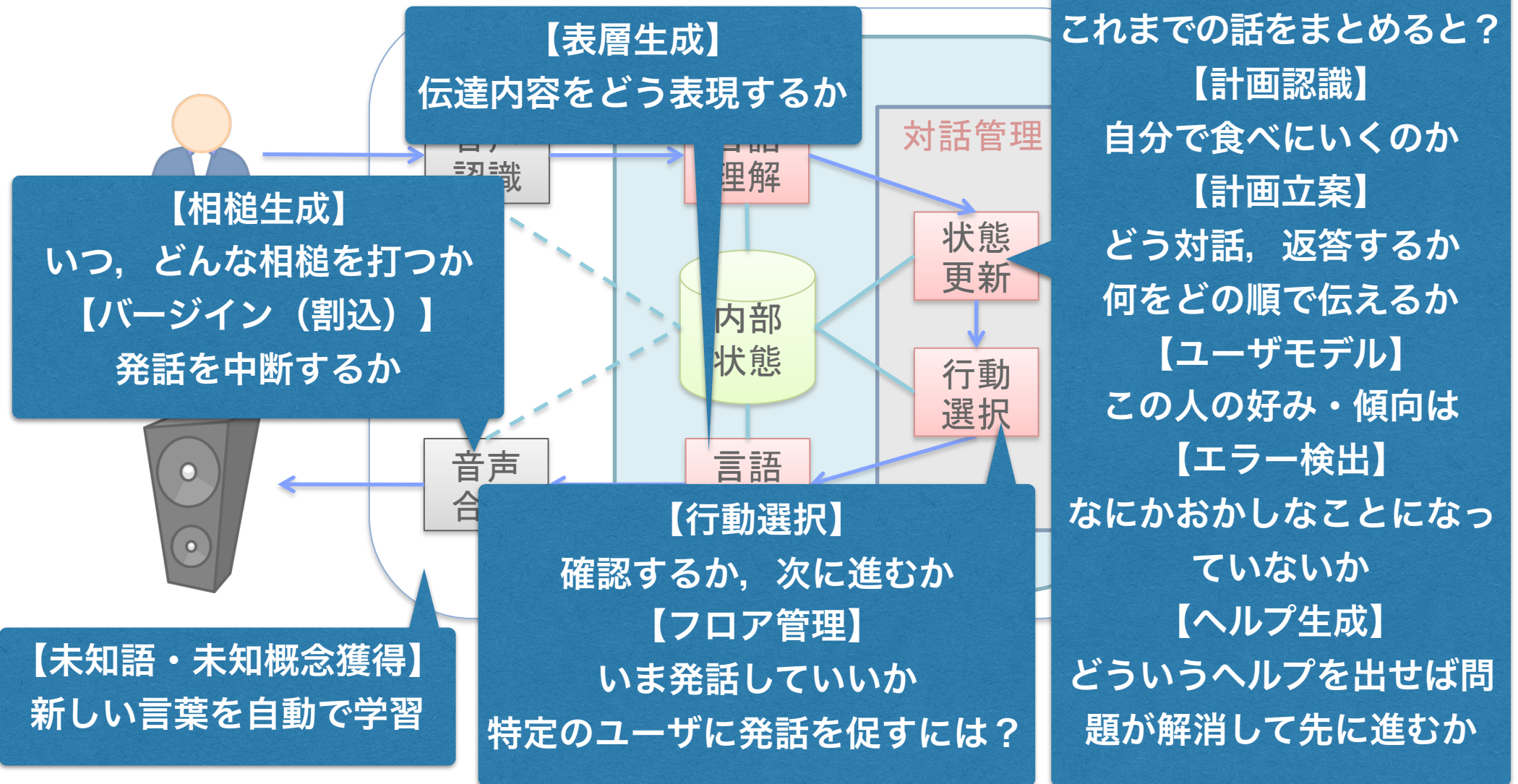
対話システムが扱わなければならない問題

ユーザ：俺のフレンチ予約できる？



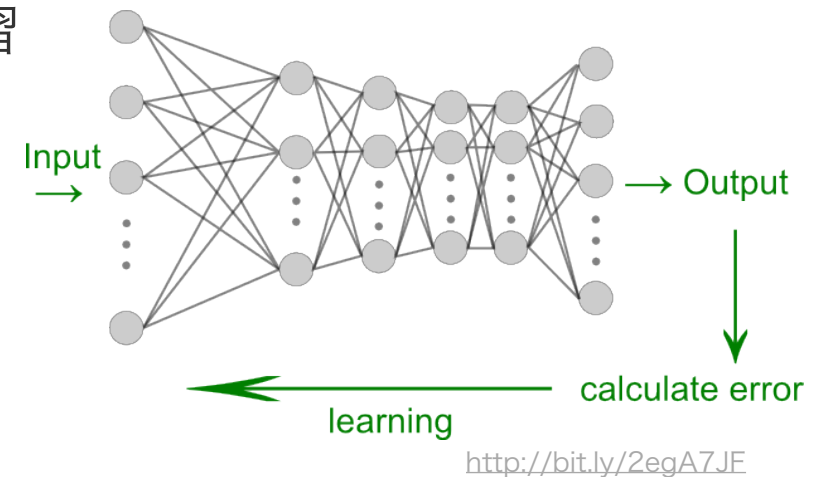
対話システムが扱わなければならない問題

注：これら全てをきっちり実装しなければ実用化できないのではない
いろいろな前提条件を設定をすることで、回避できる



深層学習と対話システム

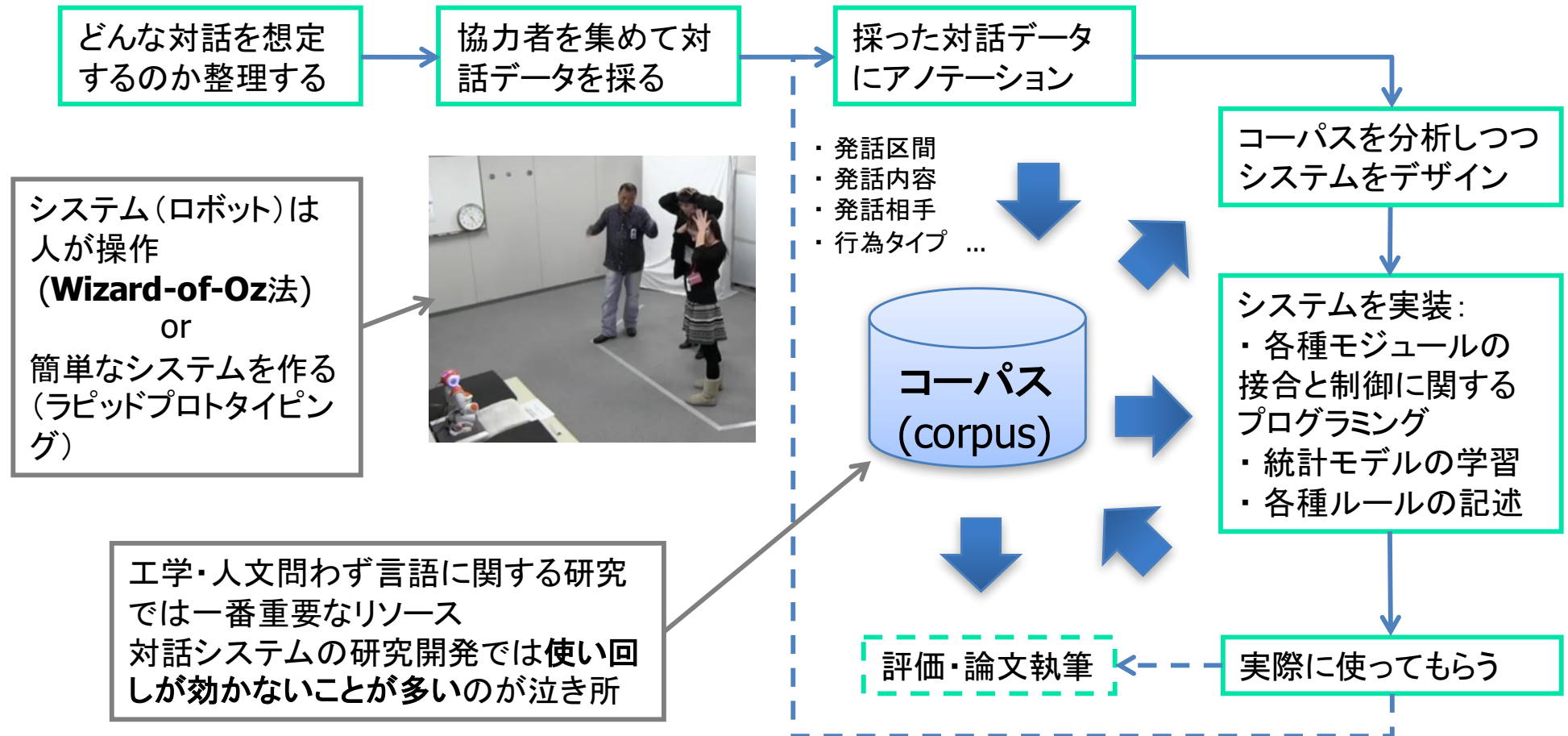
- ・ 深層学習 (Deep Learning)
 - ・ 多階層ニューラルネットや、再帰的ニューラルネットを用いた関数近似手法： $y = f(x)$ をデータから学習
 - ・ 画像・音声認識で大成功
(飛躍的な性能向上：75% $\Rightarrow$ 95%)
 - ・ 様々な言語処理のタスクでも盛んに研究



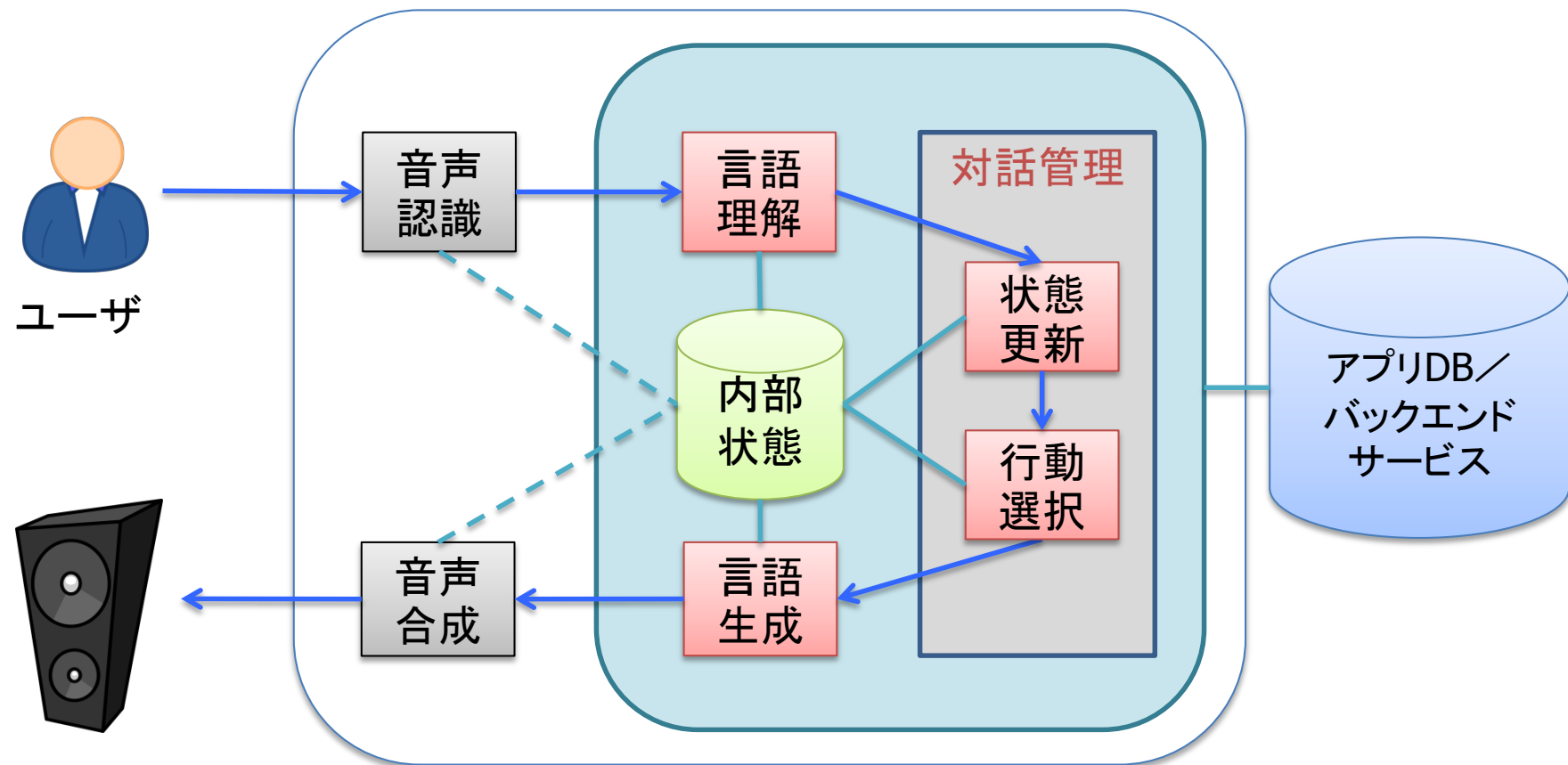
- ・ 対話：言語理解，言語生成で適用事例
 - ・ 特に言語生成での適用が盛ん
- ・ 深層学習で対話システム研究が全部解決するか？
今ある対話システムより格段に優れたものがすぐにできるか？
 - ・ NO
 - ・ 今まで使われていた機械学習手法（の一部）が置き換わっていくだけ

対話システムの作り方

対話システムの開発サイクル



対話システムの基本構造



言語理解

- ・ 文法方式

- ・ パターンマッチング

(正規表現: `*{place}への*{info}を教えて`)

- ・ 係り受け解析・句構造解析

- ・ 統計方式 (機械学習)

- ・ 発話分類 (SVM)

- ・ 系列ラベリング (CRF)

入力

東京/へ/の/フライト/
の/情報/を/教えて

言語理解

出力

対話行為タイプ	情報要求
出発地	null
目的地	東京

内部状態と状態更新

- ・ 辞書構造（key-value構造）でデータを保持し，ルールで更新
- ・ ベイジアンネットワークで確率的に保持・更新する場合もある

対話履歴	過去の発話のリストに新しい発話を加える
談話理解結果	これまでの話しを総合して理解した結果
基盤化状態	理解結果中で，明示的に了解が取れているものと，そうでないもの
注意状態	現在焦点があたっている話題
談話義務	システムがすぐ行わなければいけないこと（例：質問への返答）

U: 明日どれくらい雨？

S: 降水確率ですね

U: そう

S: 場所はどちらですか？

情報：降水確率
（非焦点，確認済）
日時：明日
（非焦点，未確認）
場所：—
（焦点，—）

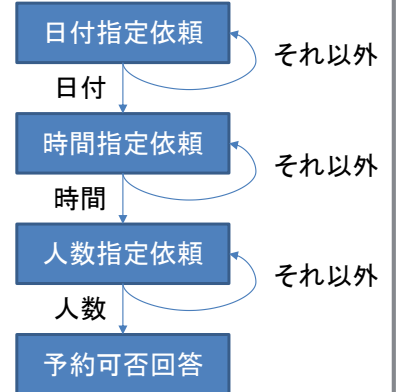
行動選択

- ・ いろいろなアプローチがある
 - ・ 対話システムのタイプにより適切なものを選択する
 - ・ （詳しくは最後に紹介する教科書を参照ください）

状態遷移ネットワーク

- ・ すべての**状態を明示化**する（有限個に限定）
- ・ ある**状態からごく限られた別の状態へのみ**遷移する（そうでないとメンテできない）

⇒ 結果として、硬直的で非効率な対話しかできない



（すぐろくのイメージ）

商用の主流

フレーム+ルール

- ・ 状態を**明示化しない**（原理的には無限個の状態）
- ・ 状態を構成する変数のうち、重要なものだけに注目したルールで行動選択

⇒ 柔軟性が増す（メンテは大変）

研究の主流

マルコフ決定過程

- ・ すべての状態を明示化する
- ・ ある状態から**別の任意の状態への遷移をゆるす**
- ・ **強化学習**により、各状態における最適な行動を学習する

⇒ 効率の良い対話（但しほぼメンテ不能）

言語生成

- ・ **テンプレート生成が一般的**

- ・ 品質が重要
- ・ 多様性の無さはそれほど重要でない
- ・ 手作業でカバーできる範囲の対話しかしない

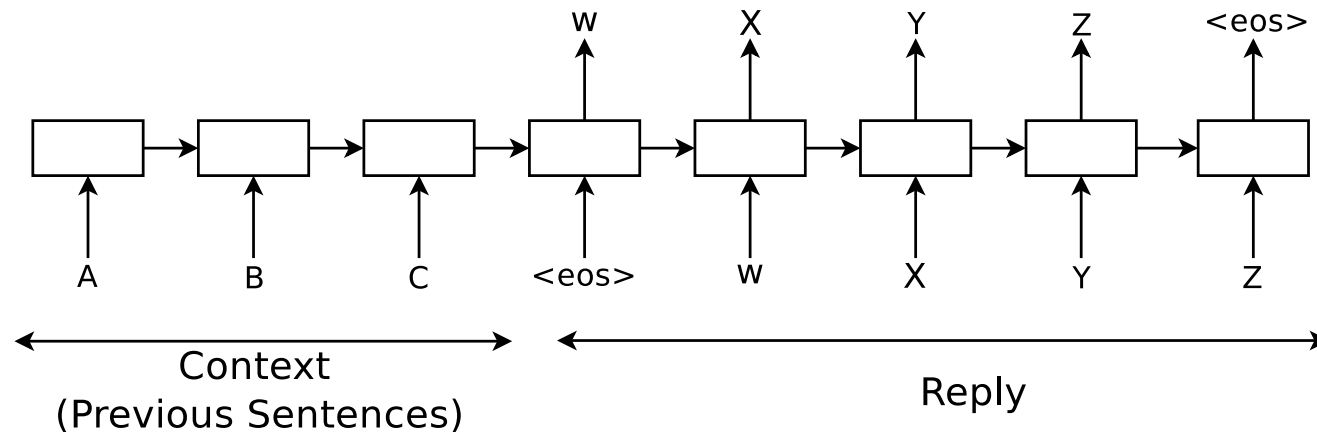
「{date}の{place}の天気は
{forecast}です」

- ・ **コーパス・統計ベース生成**

- ・ ランキング学習
 - ・ 発話候補を自然さを基準にランキング
- ・ 統計的生成
 - ・ 短いテンプレートを統計的に連結して生成
- ・ 深層学習
 - ・ 再帰的ニューラルネットを用いて逐語生成

Neural Conversational Model

Vinyals and Le (2015): <https://arxiv.org/pdf/1506.05869v3.pdf>



- ・ sequence-to-sequence (encoder-decoder) model の応用
 - ・ もともとは機械翻訳の手法
- ・ データさえあれば、それなりの精度で「それっぽい」応答を返す雑談システムが比較的手軽に作れる

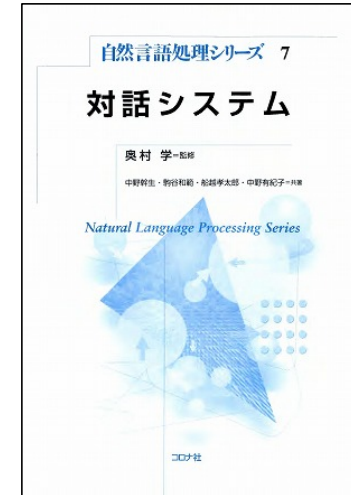
まとめと展望

- ・ 音声認識の性能向上により様々な便利な実用サービス・製品
 - ・ 限られた範囲・場面では、本格的に普及していく可能性
- ・ 対話システム自体にはまだまだ多くの課題
 - ・ 深層学習によって状況が一変する見込みは非常に低い
 - ・ 人のように話し人を凌駕するシステムが現れるのはまだ遙か未来
- ・ 音声認識はコモディティ化する一方、計算資源とデータがあれば圧倒的に優位な対話システムを作れるわけでもない
 - ・ 中小・ベンチャーでも大手にまけないヒットを飛ばせる可能性
- ・ 対話システムは、人口減少・高齢化社会でメリット大
 - ・ 社会実装のためには、どうやったら性能が上がるかより、どうやったら使いやすくなるか、使ってもらえるかを考えることが大事

更に詳しく

- ・ 教科書

- ・ 中野, 駒谷, 船越, 中野: 「対話システム」, コロナ社 (2015)



- ・ 国内イベント・組織

- ・ 人工知能学会全国大会 OS 「知的対話システム」
- ・ 言語・音声理解と対話処理研究会／対話システムシンポジウム
- ・ 言語処理学会, 情報処理学会, 音響学会

- ・ 国際イベント・組織

- ・ SIGDIAL (Special Interest Group on Discourse & Dialogue)

- ・ Manning: “Computational Linguistics and Deep Learning”

- ・ http://www.mitpressjournals.org/doi/pdf/10.1162/COLI_a_00239