

対話システム構築入門

2018年度人工知能学会全国大会（第32回）
チュートリアル
2018年6月5日

(株)ホンダ・リサーチ・インスティテュート・ジャパン
中野 幹生
nakano@jp.honda-ri.com

はじめに

本チュートリアルの目的

- 背景
 - 対話システムの実用化が進んでいる
 - それに伴い, さまざまな研究が行われている
 - 実際のシステム構築に何が有効なのかはわかりにくくなってきている
- 聞いていただきたい方
 - 対話システムを作ろうとしている方
 - ビジネス応用, 社会実装を目的としている方
 - 実際にシステムを作って使ってもらって評価したい方
- 伝えたいこと
 - 現状の対話システムの構築の一般的な方法

本チュートリアルの目的でないこと

- 最新の論文のサーベイ
- 対話システムの理論・モデル・最新の研究
 - 東中竜一郎・船越孝太郎：対話システムの理論と実践，言語処理学会第22回年次大会(NLP2016)チュートリアル資料，2016.
 - 東中竜一郎・南 泰浩：対話システム，2017年度 人工知能学会全国大会（第31回）チュートリアル資料，2016.
- 研究テーマの見つけ方
 - 中野幹生：対話システムの研究課題，SLUD第81回研究会（第8回対話システムシンポジウム）チュートリアル資料，2017.
 - <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkaWFsb2dzeXN0ZW1zanB8Z3g6N2JIN2lyZmQ3YjhiMGJkZg>
- 研究の進め方，論文の書き方
 - 中野幹生：対話システム研究の進め方(2015年8月改訂版)，SLUD第60回研究会（第1回対話システムシンポジウム）チュートリアル資料，2010.
 - <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxkaWFsb2dzeXN0ZW1zanB8Z3g6MThmNGZjNWZiMjQ5OTVhMg>

伝えたいメッセージ

- 論文と実際に使ってもらえるシステムの構築は違う
- 万能な方法やツールはない. システムのタイプ・目的・制約条件によって作り方は変わってくる
- システムを実際に使ってもらうことが大事



最新の論文に書いてある手法をそのまま実装するんじゃだめなの？

難しい会議に通ったからといって汎用的に使える技術というわけではないのよ

提案法が特定の条件下で従来法より良くて、きちんと論文を書けば通るのよ

大量のデータを仮定してたりして、実用には壁のある手法も多いわ



- 本チュートリアルの内容には主観がかなり入っています
- 自分の経験は限られているので、研究コミュニティの皆様から伺った話に基づいているものも多いですが、最終的には自分の責任で決めています。
- 間違いの指摘，質問，ご意見等は中野までお気軽にお寄せください。

今日のお話の流れ

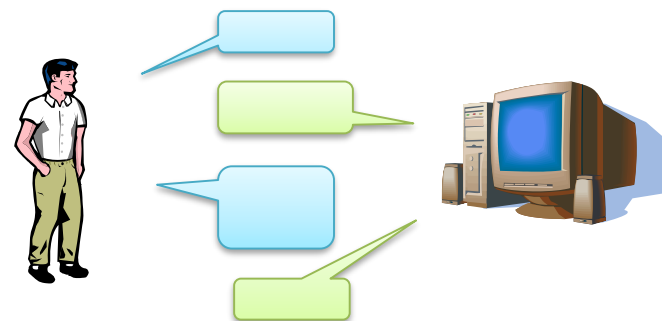
- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

対話システムとは

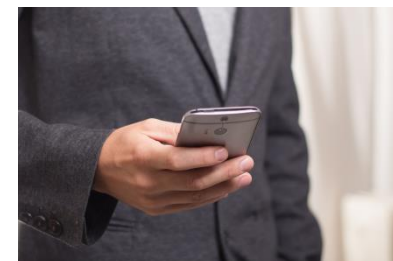
- 人間と言語で情報を授受する機械
- 複数のやりとりを行う（1問1答ではない，履歴を使う）



- システムとユーザの知識に齟齬があってもやりとりの中で解消
- 対話することの楽しみを提供
- ⇒人工知能と人のインタフェースとして有望
社会の様々な問題に適用可能

対話システムのタイプ

- 入出力の**モダリティ**
 - テキスト
 - 音声
 - マルチモーダル（音声，画像，その他のセンサ，CGエージェント，ロボット）
- 達成すべき**目標**の有無・種類
 - タスク指向型（情報検索，説明，インタビュー，説得，交渉，クイズ，ゲーム，...）
 - 非タスク指向型（雑談，傾聴，...）
 - 混合型
- 対話の**ドメイン**
 - 単一ドメイン（フライト，ホテル，野球，レストラン，...），マルチドメイン，オープンドメイン
- 対話**参加者の数**
 - 1対1，マルチパーティ

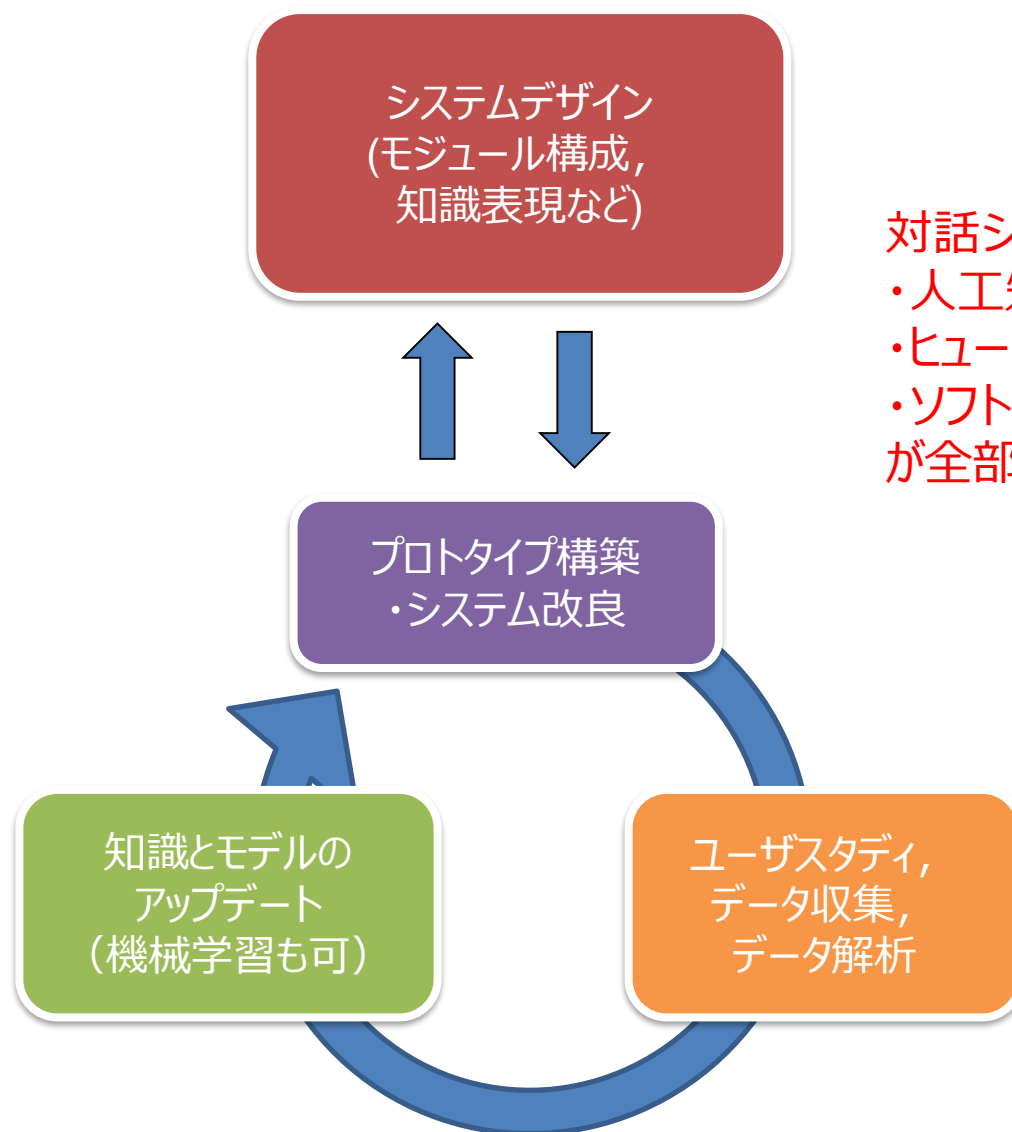


| 製品例 | 入力モダリティ | 出力モダリティ | タスクとドメイン | 参加者の数 |
|-------------------------------|---------|------------|----------------------------------|-------|
| 音声カーナビ | 音声 | 音声・画像 | 目的地設定など | 1対1 |
| 音声アシスタント | 音声 | 音声・画像 | メール, アラーム等のタスク+ オープンドメインの雑談 | 1対1 |
| AIスピーカー | 音声 | 音声・LEDシグナル | 音楽検索, タイマーなどのタスク+ オープンドメインの雑談 | 1対1 |
| 雑談チャットボット (例: マイクロソフトのりんな) | テキスト・画像 | テキスト・画像 | オープンドメインの雑談 | 1対1 |
| タスク指向型チャットボット (FAQなど) | テキスト | テキスト・画像 | 質問応答・クローズド | 1対1 |
| 対話ロボット (Sotaなど) | 音声, 画像 | 音声・ジェスチャ | 音楽検索, タイマーなどのタスク+ オープンドメインの雑談 | 1対1 |

対話システム構築の際の注意

- タイプが異なっても，基本的な構成は共通
- 最適な構築プロセスはさまざまな条件によって変わってくる！
 - タイプ（モダリティ，タスク，ドメイン，参加者数）
 - 開発リソース
 - ハードウェアの制約
 - ターゲットユーザや設置場所
 - 使えるデータの有無
 - FAQのDB，コールセンターログ

対話システム構築・改良のサイクル



対話システム構築には
・人工知能技術
・ヒューマンコンピュータインタラクション技術
・ソフトウェアエンジニアリング
が全部必要！



人間同士の対話データをたくさん集めてから作ればいいんじゃないの？

人間同士のデータを集めるのもコストがかかるのよ

それに、人間同士の対話と人間とシステムの対話はかなり違うのよ

まず簡単なシステムを作って、人とシステムとの対話データを集めたほうが早いことが多いのよ



今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

システムがやるべきこと

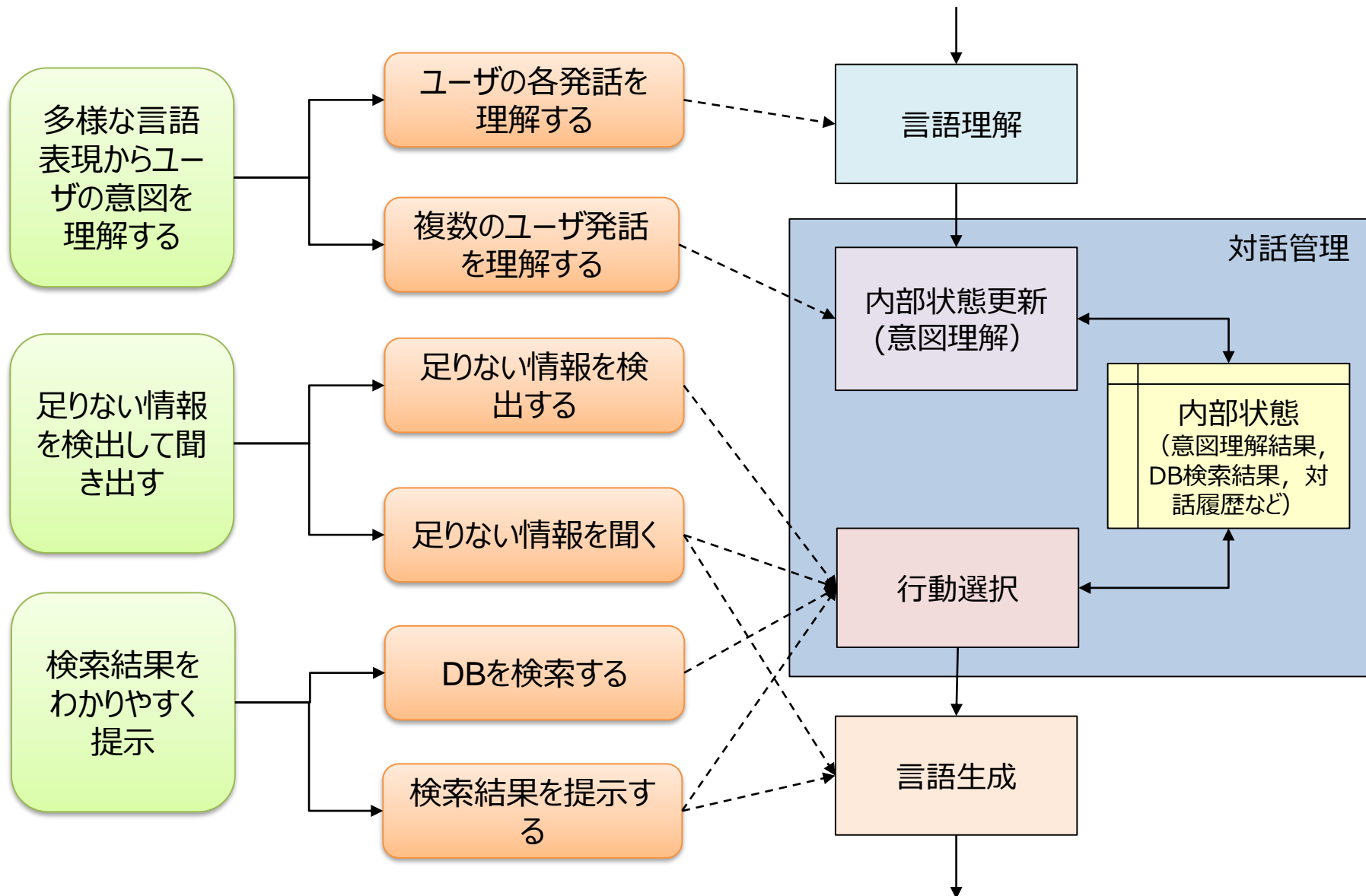
- 作りたいもの：レストラン検索のテキスト入出力対話システム
 - エリア, ジャンル, 喫煙・禁煙の別を指定してDBを検索
 - 検索結果を提示

ユーザ：渋谷のイタリアン
 システム：禁煙・喫煙の指定
 がありますか？
 ユーザ：禁煙で
 システム：50件見つかりました。
 例えば. . .

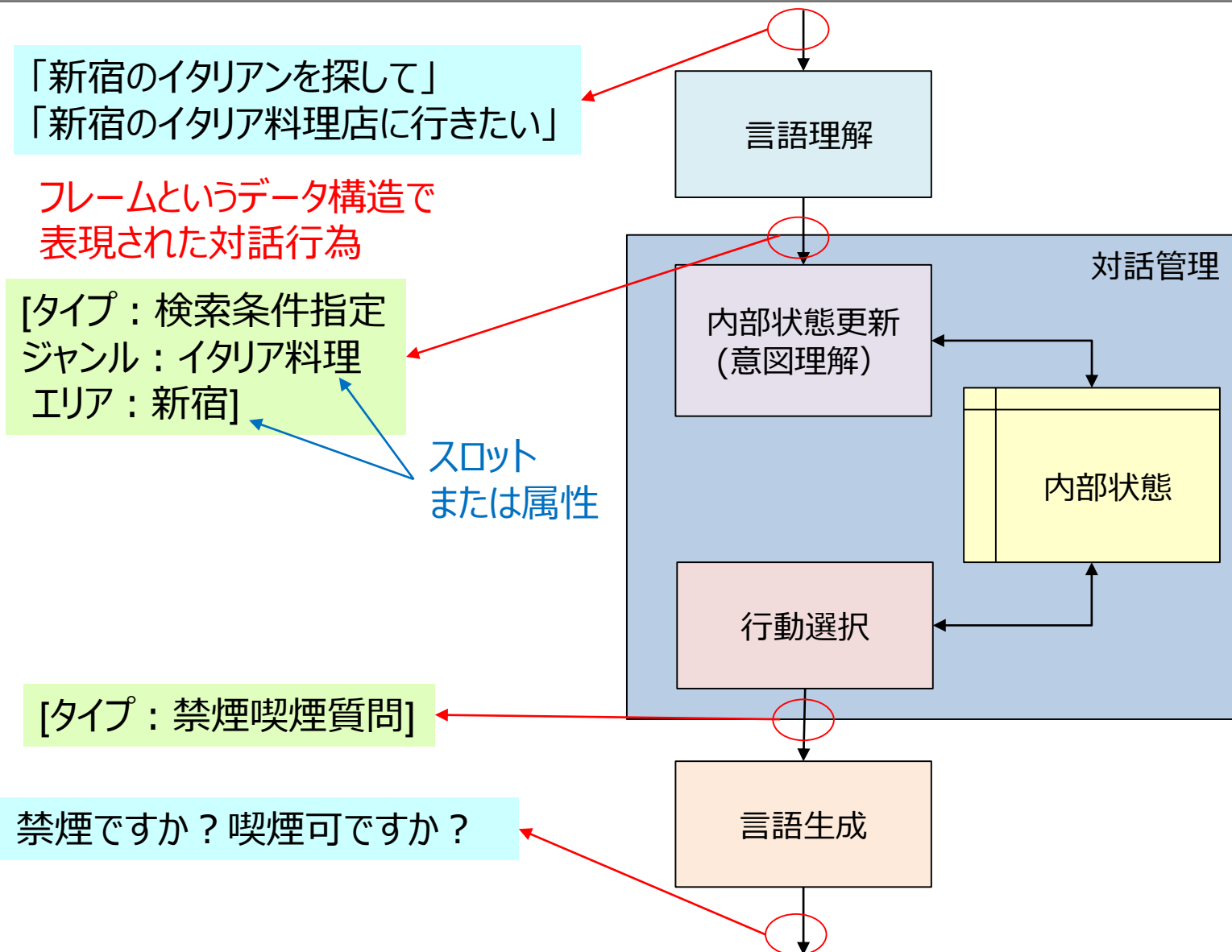
- システムは何をしないとイケないのか？
 - 多様な言語表現からユーザの意図を理解
 - 足りない情報を検出して聞き出す
 - 検索結果をわかりやすく提示

イタリア料理が食べたい
 渋谷でパスタとかいいかも
 やっぱり恵比寿

問題の分割とモジュール化



対話行為：文の意味の抽象化



- DBで用いられる記号（意味）と言語表現は必ずしも一致しない
- 言語表現を意味に変換する必要
 - 辞書を用意する

言語表現

「新宿のイタリアンを探して」
「新宿のイタリア料理店に行きたい」

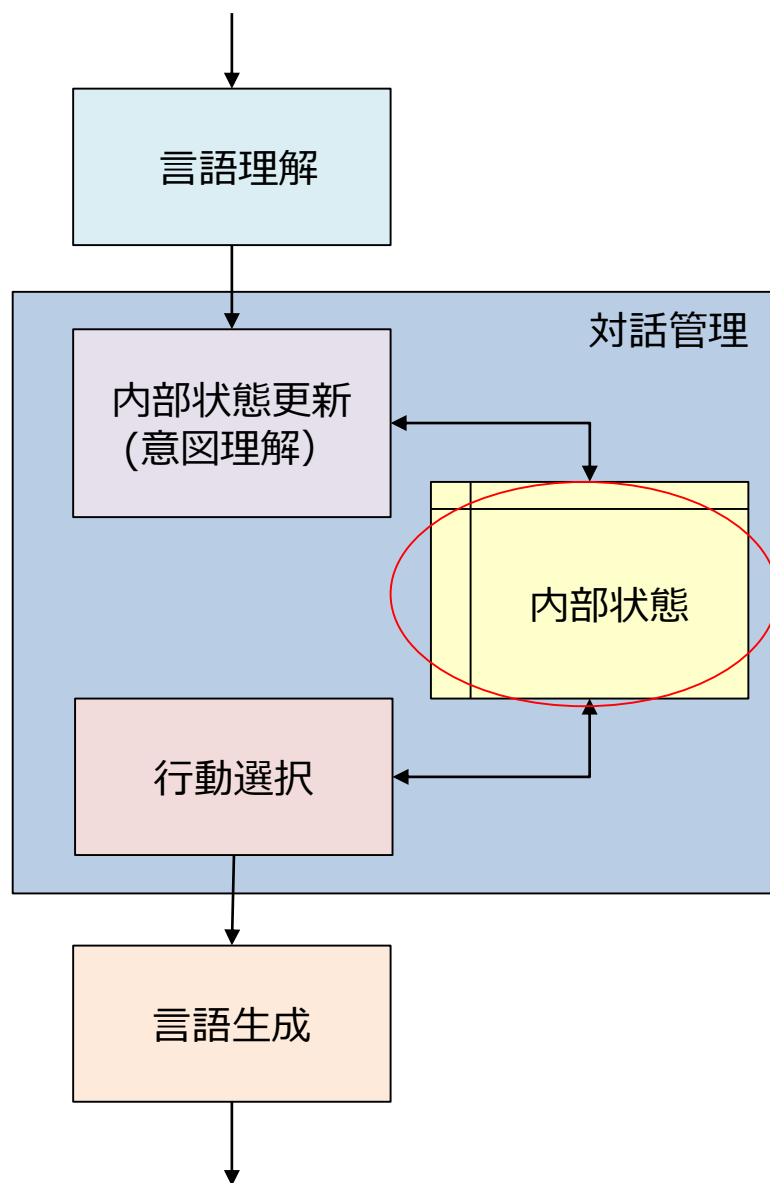
対話行為

[タイプ：検索条件指定
ジャンル：イタリア料理
エリア：新宿]

辞書

| 言語表現 | 意味 |
|---------|--------|
| イタリアン | イタリア料理 |
| イタリア料理 | イタリア料理 |
| イタリア料理店 | イタリア料理 |
| 日本料理 | 日本料理 |

対話管理：内部状態の表現

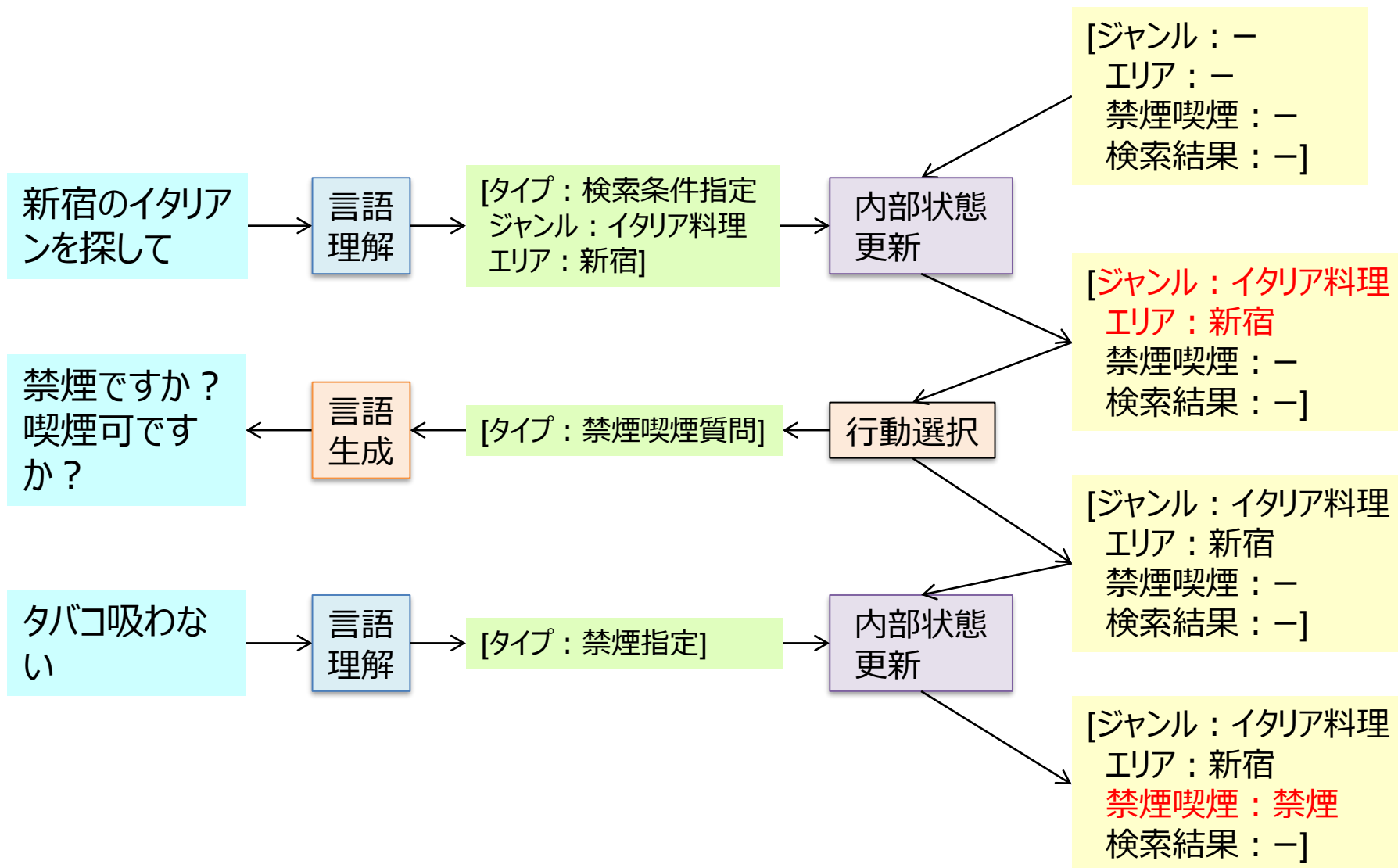


フレームで表現
 (Pythonのdictionaryなどで実装)

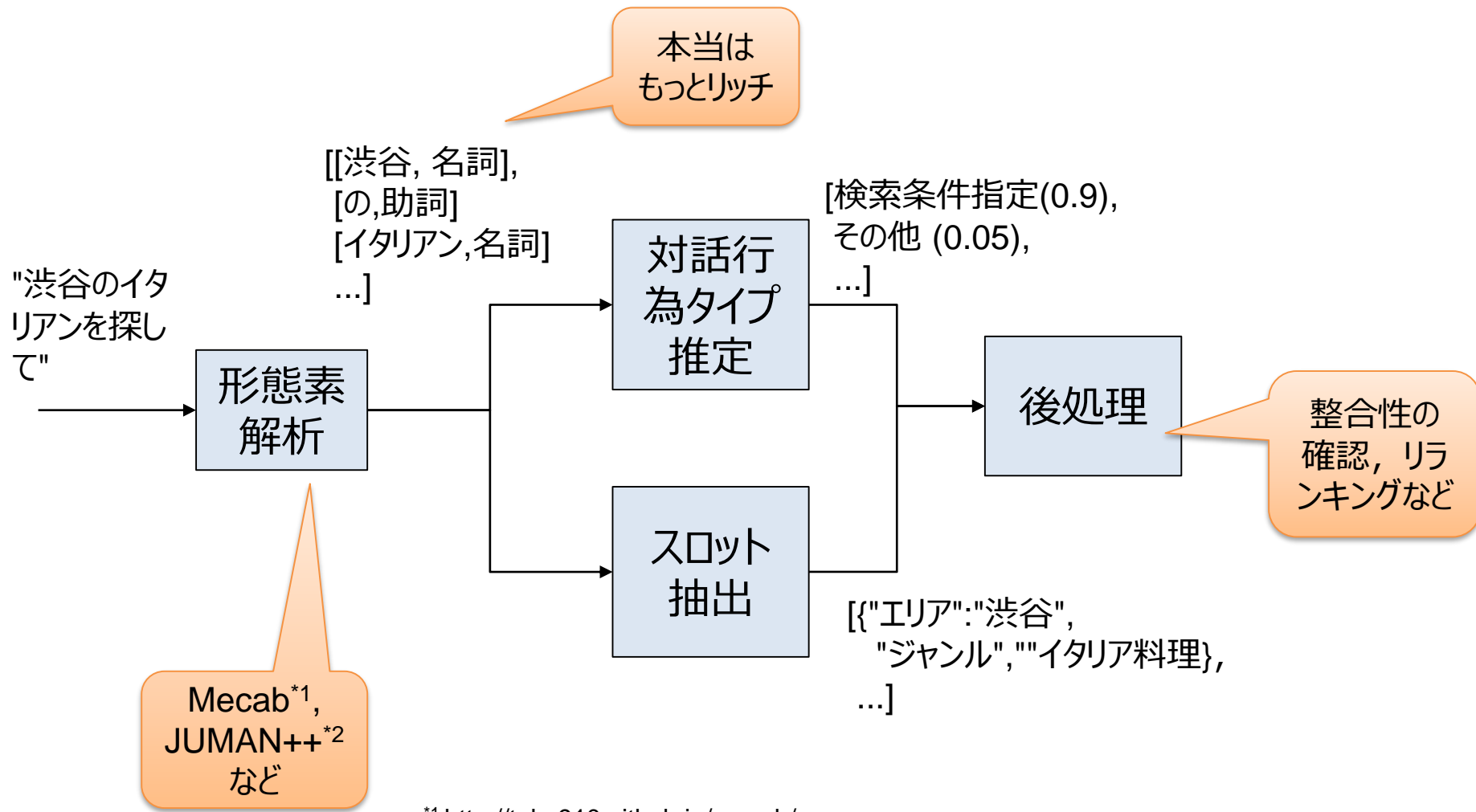
[ジャンル：イタリア料理
 エリア：新宿
 禁煙喫煙：禁煙
 検索結果：<リストオブジェクト>]

フレームベース対話管理

対話管理：内部状態更新と行動選択



言語理解の実装 (1) 概要



*1 <http://taku910.github.io/mecab/>

*2 <http://nlp.ist.i.kyoto-u.ac.jp/index.php?JUMAN++>

- 機械学習で推定モデルを教師有り学習
 - サポートベクトルマシン (SVM) やロジスティック回帰 (最大エントロピー法)
 - LibLinear^{*1}&LibSVM^{*2}, scikit-learn^{*3}などがツールとして使える
 - 特徴量
 - Bag of Words(入力文に現れる単語の集合)
 - 辞書中*i*番目の単語が現れたら*i*次元の値が1であるような特徴量ベクトル [0,.....,0,1,0,.....0,1,0,...0]
 - tfやtf・idfを用いても良い
 - 学習データ

データが集まるまでは手作業で作ればよい

| 文 | 正解タイプ |
|----------|-------|
| こんにちは | あいさつ |
| 新宿のイタリアン | 条件指定 |
| そうだよ | 肯定 |



文を形態素解析して
特徴ベクトルに変換

^{*1} <https://www.csie.ntu.edu.tw/~cjlin/liblinear/>
^{*2} <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>
^{*3} <http://scikit-learn.org/>

言語理解の実装 (3) スロット抽出

- キーフレーズ抽出
 - 辞書にある言語表現を抽出
 - キーフレーズ以外の単語が少なくなるようにする

新宿 御苑 前 の イタリア 料理 店を 教えて

エリア

ジャンル

エリア

ジャンル

- 推定後辞書引きを行い, 意味の表現に変換
 - [新宿,御苑,前] ⇒ 新宿御苑前
 - [イタリア,料理,店] ⇒ イタリア料理
- FSTツールを使えば実装は楽だが, 使わなくても可能
 - OpenFST ^{*1}など
- 問題: 未知語に対処できない

*1 <http://www.openfst.org/>

言語理解の実装 (4) スロット抽出 (続)

- 系列ラベリング問題として解く
 - テキスト処理の固有表現抽出と同じ

B-エリア O B-ジャンル I-ジャンル I-ジャンル O O O
 渋谷 の イタリア 料理 店 に 行き たい

- 確率的条件場(CRF: Conditional Random Fields)を学習し推定
 - CRF++^{*1}, CRFSuite^{*2}などが使える
 - 特徴量
 - 単語, 品詞, 単語の連鎖, 品詞の連鎖, ラベルの連鎖など

^{*1} <https://taku910.github.io/crfpp/>
^{*2} <http://www.chokkan.org/software/crfsuite/>

- 推定後辞書引きを行い, 意味に変換
 - イタリア料理店 ⇒ イタリア料理

- 学習データ
 - 右のようなものを用意すれば
ツールが変換してくれる

| | | |
|------|----|--------|
| 渋谷 | 名詞 | B-エリア |
| の | 助詞 | O |
| イタリア | 名詞 | B-ジャンル |
| 料理 | 名詞 | I-ジャンル |

言語理解の実装 (5) さまざまな工夫

- キーフレーズ抽出結果と系列ラベリングの結果を統合
 - 未知語がある場合とない場合の両方でよい精度を出す
- 発話行為タイプとスロットの整合性がある結果を選ぶ



[タイプ：検索条件指定
ジャンル：イタリア料理
エリア：品川]



[タイプ：禁煙喫煙指定
エリア：新宿]

内部状態更新の実装

- if 対話行為のタイプ=検索条件指定
 then 内部状態のジャンルの値=対話行為のジャンルの値

更新規則はプログラムで書けばよい

「イタリアンがいい」



[タイプ：検索条件指定
 ジャンル：イタリア料理]



「ジャンル：-
 エリア：新宿
 禁煙喫煙：-」



「ジャンル：イタリア料理
 エリア：新宿
 禁煙喫煙：-」

補足 (アップロード版)

- 内部状態にシステム発話の内容を保持しておいて内部状態更新で利用することで、以下のような対話でユーザ発話を正しく理解できるようになる

システム：特急券の予約を承ります。
乗車駅をおっしゃってください
ユーザ：新宿
システム：降車駅をおっしゃってください
ユーザ：小田原

どちらが乗車駅でどちら
が降車駅かは直前のシ
ステム発話に依存する

行動選択の実装

- 内部状態に規則を適用し，DB検索を行ったり，発話行為を出力したりする

- 禁煙喫煙が空 ⇒ "禁煙喫煙の別を尋ねる"
- ジャンル，エリア，禁煙喫煙が空でない ⇒ DB検索を行いもう一度対話行動選択を行う
- 検索結果の数 ≤ 3 ⇒ "検索結果提示"
- 検索結果の数 > 3 ⇒ "最初の候補読み上げ"

規則はプログラムで書けばよい

「ジャンル：-
エリア：新宿
禁煙喫煙：-」



[対話行為タイプ：禁煙喫煙の別を尋ねる]

「ジャンル：イタリア料理
エリア：新宿
禁煙喫煙：禁煙」



[ジャンル：イタリア料理
エリア：新宿
禁煙喫煙：禁煙
検索結果：<リスト>]

対話システムって強化学習を使うのが普通なんじゃないの？行動選択を自動的に学習して最適化してくれるんでしょ？



問題は3つあるわ。

- (1) 学習するためのデータが必要
- (2) 状態と行動の集合を決めるのが難しい
- (3) 報酬設計が難しい

データを集めるシステムができるなら、そこからさらに強化学習で最適化してもあまりよくなるかもしれないのよ

何が最適かがよくわからないと最適化は難しいわ。最適じゃなくても役に立てばいい場合もあるしね

強化学習が本当に使えるようになるまでにはまだまだかかりそうね



言語生成の実装

- テンプレートベース言語生成

テンプレート

[対話行為タイプ：レストラン名の提示]



「<レストラン名>というレストラン
が見つかりました」

実行

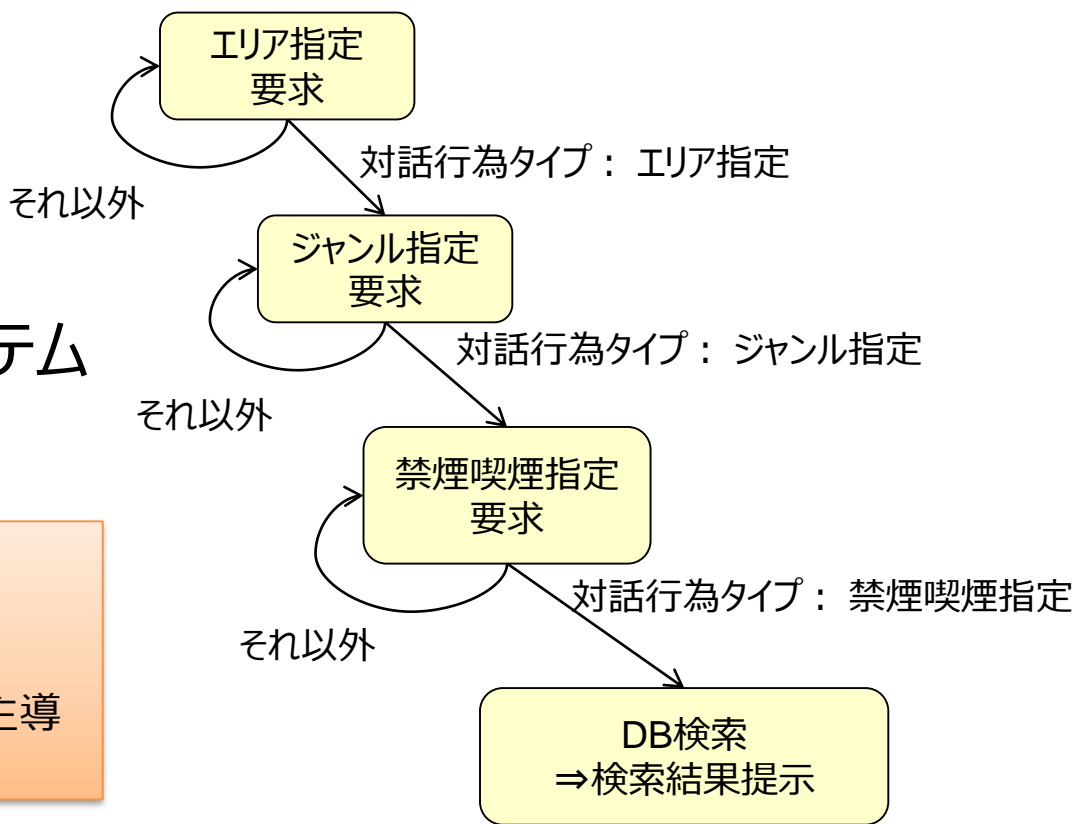
[対話行為タイプ：レストラン名の提示
レストラン名：サイゼリア]



「サイゼリアというレストラン
が見つかりました」

補足：ネットワークベース対話管理

- 内部状態を有向グラフのノード+付加情報（スロット）で表現
- ユーザ発話の対話行為で遷移
- ノードごとにシステムの対話行為を出力
- あとはフレームベース対話管理と同じ
- システム主導の対話システムに適している



主導権

- 次に何を話すかをだれが決めるか
- 対話管理の設計指針
- ユーザ主導，システム主導，混合主導などがある。

補足：アジェンダベース対話管理

- 複雑なタスクの対話管理の見通しをよくする（詳細割愛）
 - 内部表現を小さいフレームに分割し，フレームの並べ替え可能なリスト（アジェンダ）にする [Bohus 09]

今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - **単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–**
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

- DB検索対話システムとの違い
 - タスクが明確ではないので、内部状態を簡単なデータ構造で定義できない
 - 対話行為をどう表現するかが自明ではない
 - 属性集合の定義が難しい
- ユーザ発話とシステム発話の対 + 文脈の簡単な表現
- ここでは二つの方法を紹介
 - ルールベース&用例ベース
 - 統合も可能[目黒 14]

ルールベースシステム

AIML (Artificial Intelligence Markup Language)などでルールを記述

```
<category>
  <pattern>* 眠い * </pattern>
  <template>少し休んだらいかがですか?</template>
</category>
```

ユーザ：眠いな～
システム：少し休んだらいかがですか？

ルールの構築は高コストだが、
間違った応答はしにくい

```
<category>
  <pattern>うん</pattern>
  <that>ラーメンは好きですか？ </that>
  <template>どんなタイプのラーメンが好きですか？ </template>
</category>
```

システム：ラーメンは好きですか？
ユーザ：はい
システム：どんなタイプのラーメンが好きですか？

アップロード版での訂正

当日はAIMLの日本語インタプリタはないと言いましたが、形態素解析を入れればProgramDなどの既存のプログラムが使えます。また(株)NTTドコモの自然対話APIでもAIMLが使えます。

用例ベースシステム(1)

近いものを
検索

「いい天気だな～」

| ユーザ発話 | システム発話 |
|--------------|-----------------------------------------------------------------------------------------|
| こんにちは | <ul style="list-style-type: none"> こんにちは！お元気ですか？ |
| いい天気ですね | <ul style="list-style-type: none"> そうですね。気持ちがいいです。 はい、すがすがしいです。 |
| あまり眠れませんでした | <ul style="list-style-type: none"> それはつらいですね。 お体に気をつけてください |
| おなかがすいた | <ul style="list-style-type: none"> 早くご飯食べたいですね |
| あなたは誰？ | <ul style="list-style-type: none"> 私はボットちゃんです ボットちゃんといいます |
| いいラーメン屋知ってる？ | <ul style="list-style-type: none"> ラーメンなら一蘭が大好きです |

文間の距離計算手法

- tf, tf·idf
- WMD (word mover's distance [Kusner 15] など)

用例は自分で作っても良いし、
Webからとってきて人手でフィ
ルタリングすることも可能

システム発話だけのこともある
その場合は入力とシステム発話
例の距離を計算

用例ベースシステム(2) 発話選択

- 単純な方法の問題
 - 直前のユーザ発話しかみない
⇒ 文脈を利用できないので対話が不自然になる
 - 用例のうちユーザ発話しか見ない
⇒ 応答の良し悪しを考慮しない


- 複数の候補から発話を選択

- 選択基準の例

- 同じ発話を繰り返さない
- それまでの対話で使われている
単語集合との類似度 [Banchs 12]

| ユーザ発話 | システム発話 |
|-------------|---------------------------------------------------------------------------------------------|
| こんにちは | <ul style="list-style-type: none"> • こんにちは！お元気ですか？ |
| いい天気ですね | <ul style="list-style-type: none"> • そうですね。気持ちがいいです。 • はい、すがすがしいです。 |
| あまり眠れませんでした | <ul style="list-style-type: none"> • それはつらいですね。 • お体に気をつけてください |
| 眠いな | <ul style="list-style-type: none"> • 昼寝した方がいいかもしれません |
| おなかがすいた | <ul style="list-style-type: none"> • 早くご飯食べたいですね |

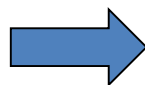
どれを選ぶか？



用例ベースシステム(3) 抽象化

| ユーザ発話 | システム発話 |
|---------|------------|
| *x*が好き | *x*いいですね |
| 私は*y*です | *y*さんこんにちは |

「ラーメンが好き」



「ラーメンいいですね」



オープンメイン雑談システムといえばseq2seqが流行ってるみたいだけど？対話データを用意するだけで学習してくれるんだよね？

応答文を生成するモデルだと、変な応答が生成されるかもしれないし、意味のない応答ばかりが生成されやすくて使えるシステムはできないわ

しかも訓練に大量のデータが必要だし、その応答が良い応答とは限らないでしょ

なので今のところは発話選択の方がいいのよ



今日のお話の流れ

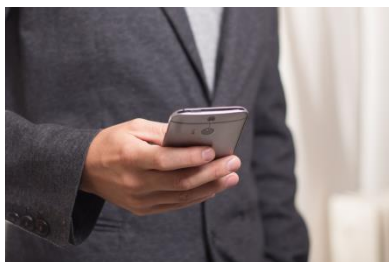
- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - **さまざまな対話システムの構築**
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

- 音声認識APIを使うのが簡単
- 留意点
 - 口語に対応した言語理解が必要
 - 誤認識や漢字の間違いを考慮する必要がある
 - くだけた発話の音声認識はまだ難しい
 - 韻律は認識しない
 - 音声認識結果の単語区切りと言語理解部が想定する単語区切りをあわせる必要

- 利用できるもの
 - 商用の音声合成ライブラリ
 - ビルトインの音声合成（Android TTSなど）
 - 音声合成API
- 留意点
 - 漢字を読み間違ふことがある
 - ⇒ 文脈や常識が必要な場合もあり簡単ではない
 - 「アンジャッシュの渡部」
 - ⇒ 対話システム側で読みを与えるしかない
 - イントネーションを間違えることがある

マルチモーダル入力

- タッチパネルなど
 - 言語理解部で対話行為に統合



(タッチパネルを押しながら)
 このレストランの値段は？



[タイプ：値段を聞く
 レストランID: <GUIから得られたID>]

マルチモーダル出力

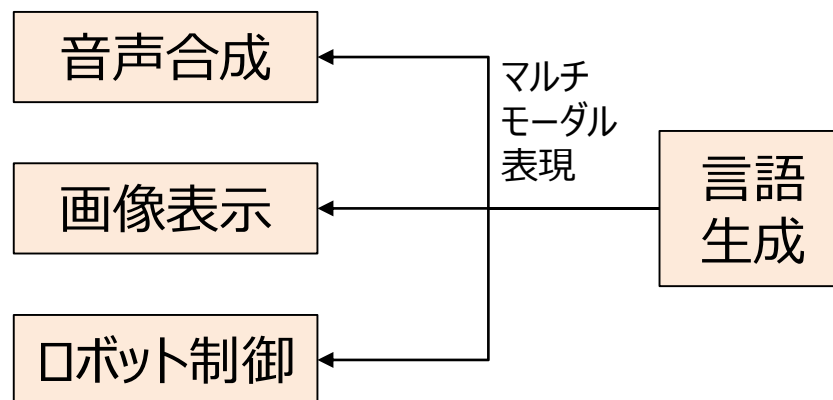
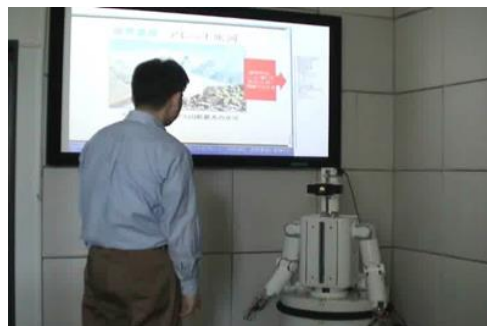
- 画像表示, ロボットジェスチャ
 - 言語生成部でマルチモーダル表現を生成

[西村 06]

[タイプ: レストラン名の提示]



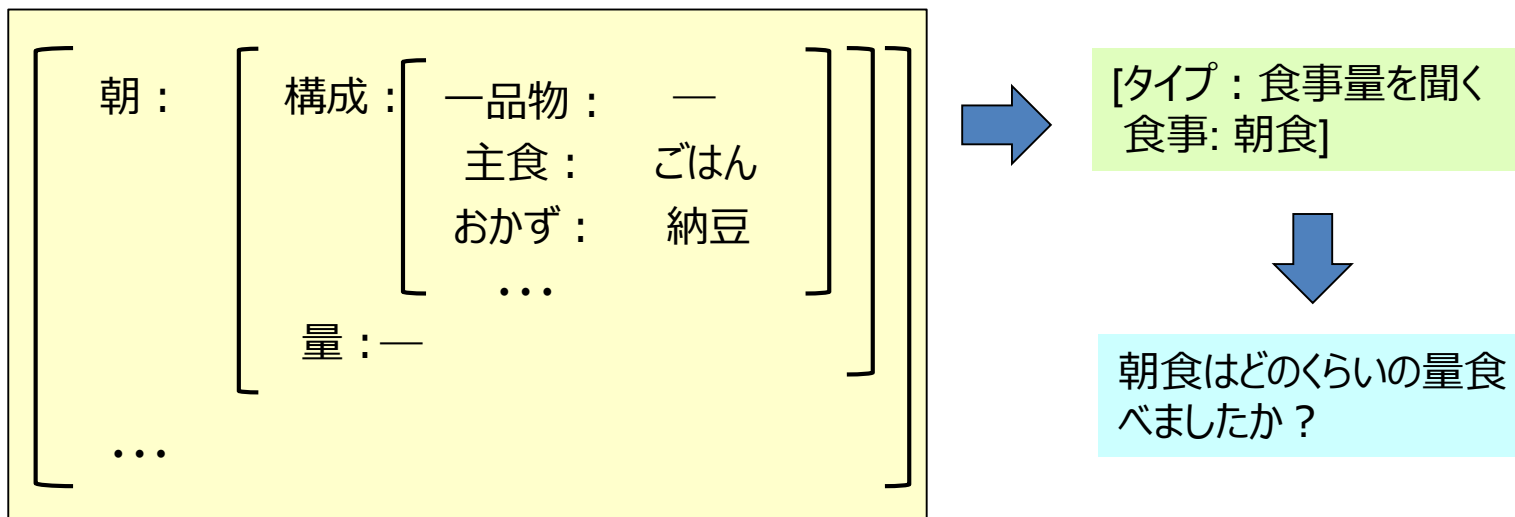
発話: 「<レストラン名>というレストランが見つかりました」
 表示: show_picture(<レストランID>)
 ロボットジェスチャ: 右手を上げる



さまざまなタスクのシステム (1)

- インタビュー, 説明, クイズ, ゲームなど
⇒ フレームベース対話管理やネットワークベース対話管理
で実装可能

食事内容を聞くインタビュー対話システムのフレームの例 [小堀 16]



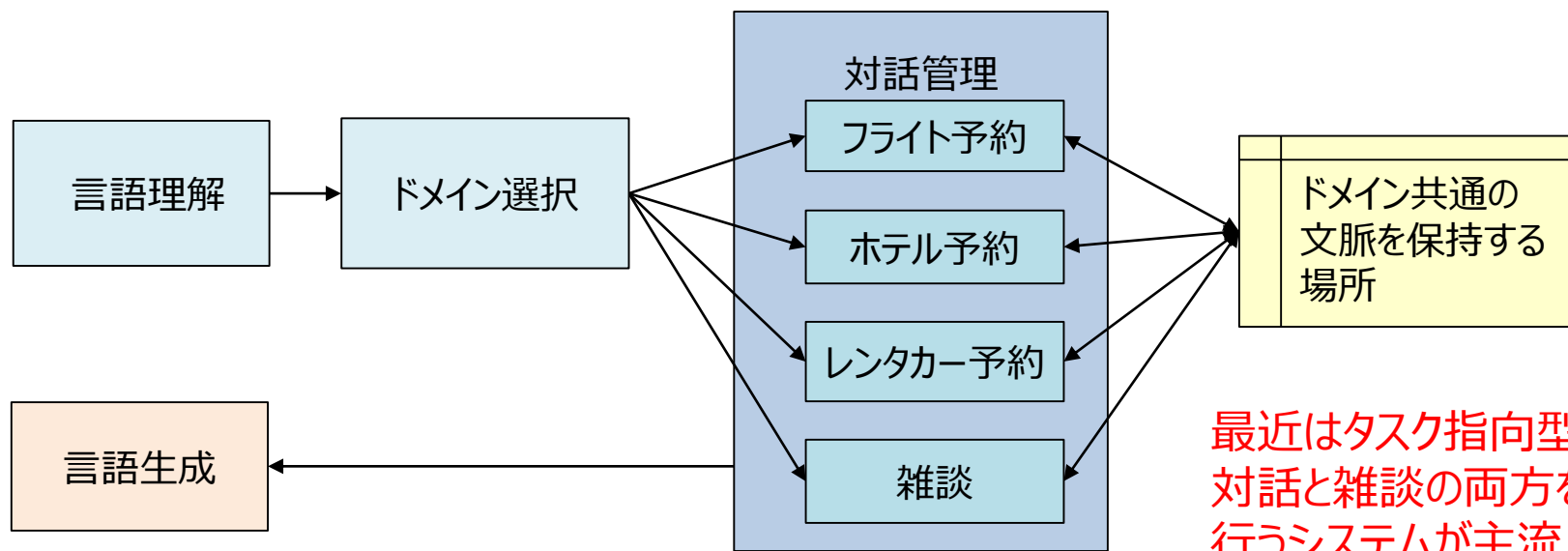
さまざまなタスクのシステム (2)

- 文書検索対話システム
 - 構造化されていない文書データを検索
 - FAQ検索など
 - 用例ベースと同様に構築できる
- 傾聴対話システム
 - 相槌や問い返しをしながらユーザが話すことを聞く [下岡 17][Meguro 13][Lala S17]
 - ユーザ発話のタイプ + 文脈 (対話行為タイプの統計モデルなど)
 - 構成としては雑談対話に近い

マルチドメイン対話システム

- 一つの対話システムがさまざまなタスクをこなす
- 複数の対話管理部を統合
 - 分散型アーキテクチャ
 - システム発話のドメインを推定
 - 一度あるドメインに入ったら、明確に他のドメインであるとわかる発話に来るまでドメインを継続

ユーザ：鹿児島までのフライトを予約したい
 システム：往路の日付と時間帯を教えてください
 ユーザ：6月4日の5時ごろ
 . . .
 システム：予約が完了しました
 ユーザ：城山観光ホテルを予約したい
 システム：6月4日チェックイン6月8日チェックアウトでいいですね？



最近では**タスク指向型**対話と**雑談**の両方を行うシステムが主流



概念的にはだいたいわかったけど、最初は何から手をつけたらいいんだろう？

まずは何か入力すると応答が返ってくるものを動かしてみるといいんじゃない？

公開されているプログラム^{*1}を動かすのが一つの手ね。あとはRepl-AI^{*2}などのAPIを使うのもいいわ。

そのあとで作りたいシステムを作り始めるといいでしょう。



- ^{*1} @Hironasan「機械学習を使って作る対話システム」 Qiita (2017年04月19日更新)
<https://qiita.com/Hironasan/items/6425787ccb75dfae36>
- ^{*2} Repl-AI (インターメディアプランニング(株) & (株)NTTドコモ)
<https://repl-ai.jp/>

- 各社が提供しているAPIを部分的に用いることもできる
- (株)NTTドコモのAPIの例（他社もいろいろある）

| API名 | | 機能 |
|---------|------------------------------------------------|-----------------------|
| 自然対話API | 意図解釈API | ドメイン推定 + 言語理解 |
| | FAQチャットAPI | FAQの検索対話. 曖昧な質問には聞き返す |
| | 雑談対話API | オープンドメインの雑談を行う |
| | 知識検索API | いわゆるテキストベース質問応答を行う |
| 言語解析API | 形態素解析, 固有表現抽出, 語句類似度算出, ひらがな化, 商品評判要約, キーワード抽出 | |
| 音声認識API | 音声をテキストに変換 | |
| 音声合成API | テキストを音声に変換 | |

今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - **対話システムの評価**
 - 少し進んだ構築法
 - 今後の展望
 - Q&A

実用的な対話システムの評価軸

- 動くのか？（何人まで同時にアクセスしても大丈夫か？）
- ユーザやクライアントの役に立つのか？満足させたか？
- たくさんのユーザに継続的に使ってもらえるのか？
- 開発費・運用費を引いても儲かるか？
 - 人件費削減，宣伝効果，広告料収入など

論文の評価法とは異なる

評価方法

- 主観評価
 - アンケート
- 客観評価 (アップロード版で補足)
 - リピート率
 - 対話の長さ
 - 言語理解・意図理解正解率
 - 訂正発話の頻度
 - タスク達成率
 - 運用・保守コスト
 - など
- 評価しつつ公開範囲を拡大する
 - 自分たちで使ってみる
 - 社内で使ってもらう
 - お金を払って使ってもらう (クラウドソーシングも可能)
 - 公開して実証実験

モジュール毎の評価

- 正解が決められるモジュール（言語理解など）
⇒ 正解率などでオフライン評価が可能

- 正解がよくわからないモジュール（行動選択など）はモジュール単独での評価は無理
⇒モジュールを置き換えてみてシステム全体を評価する

今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - **少し進んだ構築法**
 - 今後の展望
 - Q&A

- 技術的アプローチ
 - 対話ログに基づく各モジュールの性能向上
 - 構築コストの削減
- デザインに基づくアプローチ
 - 現在の言語理解技術はかなり限定的⇒**デザイン**で補う
 - 人間同士の対話のモデルがヒントになる
- ここでは比較的すぐに使えるものだけ紹介する

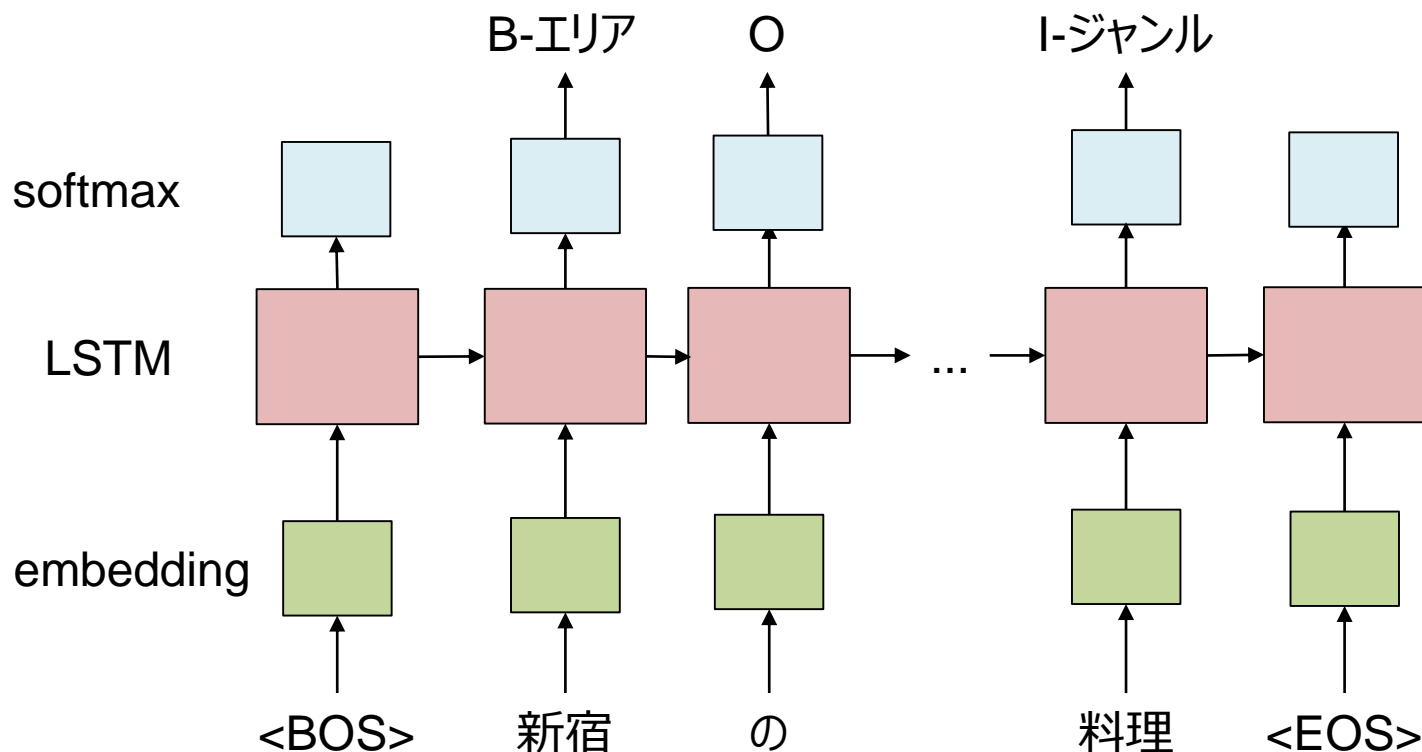
言語理解の精度向上 (1)

- データが集まってきたら, Deep Neural Network (DNN) を利用するのが一つの手
- DNNのメリット: 特徴量の推定
 - CRFで使われる特徴量: 単語, 品詞, 単語の連鎖, 品詞の連鎖, ラベルの連鎖など
 - ⇒ 長期の依存関係が学習できない, 単語の表現が離散的
 - ⇒ Feature engineeringで解決するのは大変
- 対話タイプ推定, スロット抽出にDNNが使われている [颯々野 18]
 - ACL2017のチュートリアルもわかりやすい
 - Yun-Nung Chen, Asli Celikyilmaz, and Dilek Hakkani-Tur, Deep Learning for Dialogue Systems, tutorial at ACL2017
 - https://www.csie.ntu.edu.tw/~yvchen/doc/DeepDialogue_Tutorial_ACL.pdf

言語理解の精度向上(2)

RNNを用いたスロット抽出

- DNNの一種, RNN (Recurrent Neural Network) で系列をモデル化
 - 正解ラベルが出力されるようにネットワークを訓練
 - LSTM (Long Short-Term Memory) を用いる



言語理解の精度向上(2)

LSTMを用いたスロット抽出

- さまざまなバリエーション (Chenらのチュートリアル (前掲) や[颯々野18]参照)
 - 対話行為タイプ推定とスロット抽出を同時に行う
 - Bi-directional LSTMを使う
 - 文全体のencodingを使う
 - LSTMの出力をCRFの入力にする
 - 文脈情報をエンコーディングして用いる

- どのようなタイプのシステムにどのような方法が良いかはまだ不明
 - 学習データ量
 - 学習コーパスにない単語の割合
 - ユーザ発話のバリエーション

- ユーザ発話の理解結果は完全ではない（特に音声入力の場合）

システム：〇〇銀行です． ご用件をどうぞ
ユーザ： 〇〇に1万円振り込んで
システム：かしこまりました． 〇〇さんに百万円振り込みました
ユーザ：えー！？

- 間違いがクリティカルな場合は確認が必要⇒基盤化
 - 基盤化：対話参加者間で共有の知識を持つこと

ユーザ： 〇〇に1万円振り込んで
システム： 〇〇さんへの100万円のお振込みですね

- ユーザ発話の理解結果の確信度を求めそれに応じて確認するかどうかを決めることも可能
- どのくらいの確信度のときに確認するかはタスクにもよる
 - 天気予報vs銀行

エラーハンドリング

- 人間同士でも誤解はある ⇒ 誤解があっても対話で解消すべし
- エラーハンドリング
 - エラー検出
 - ユーザ発話の繰り返し
 - ユーザの話し方の変化
 - エラーリカバリ
 - ヘルプ生成

ユーザ：新宿のイタリアン
システム：渋谷のイタリア料理ですか？
ユーザ：し・ん・じゅ・く

理解しやすい発話をしてもらう

- ヘルプ生成

- ユーザはどう話したらよいかわからない
- 困ったときにヘルプを出す

もう少し自然に話してみてください

明日の天気は、などと聞いてみてくださいね！

- 理解用辞書にない単語は発話しない

- ユーザはシステムが話した単語を話す

ユーザ：新宿のイタリアン

システム：新宿のイタリア料理店は○件見つかりました

ユーザ：やっぱり渋谷のイタリア料理店

- 話題の制御

- システム側から話題を絞り込んでいく

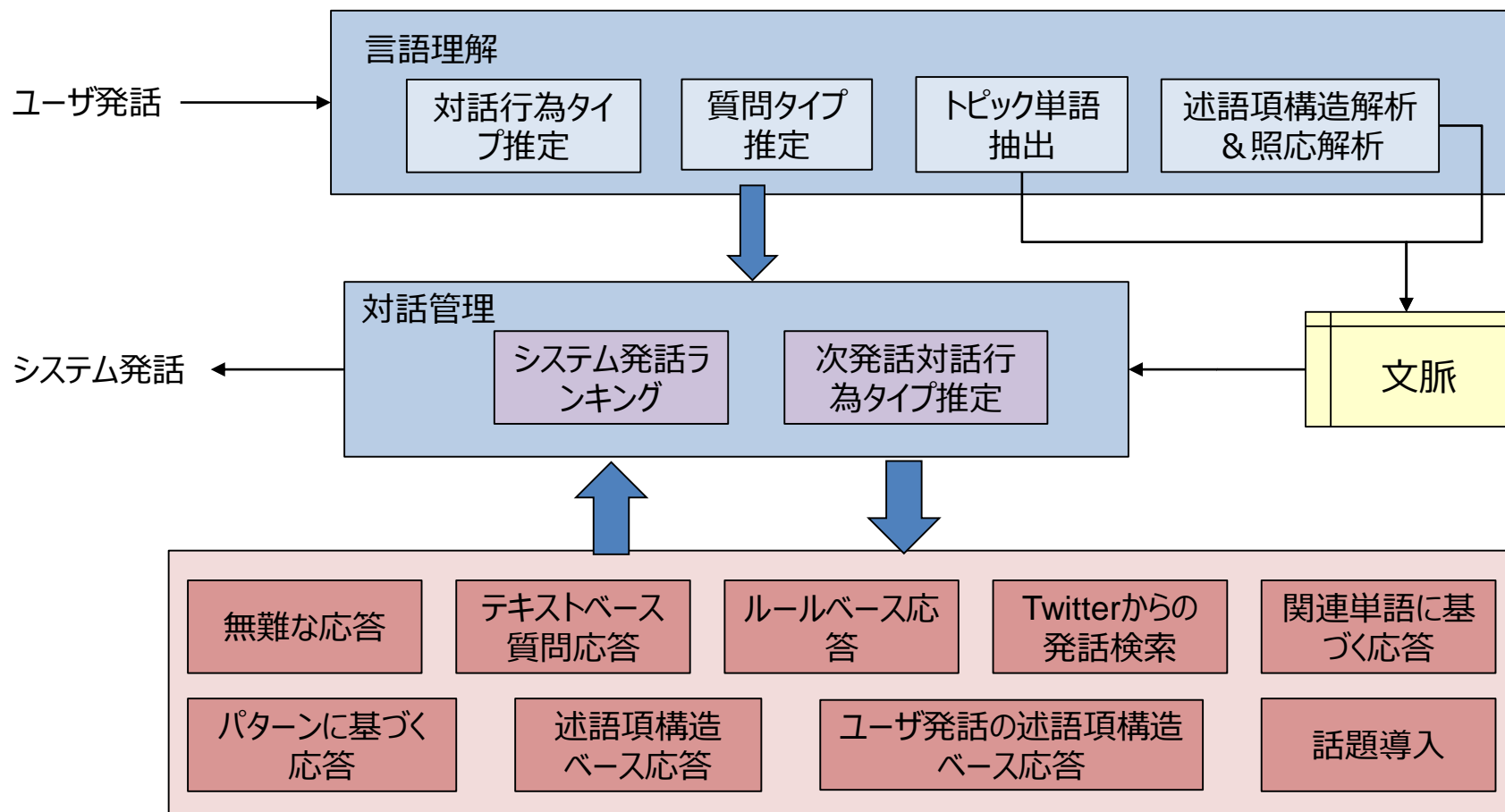
スポーツの話しませんか？

雑談対話システムの高度化（1）

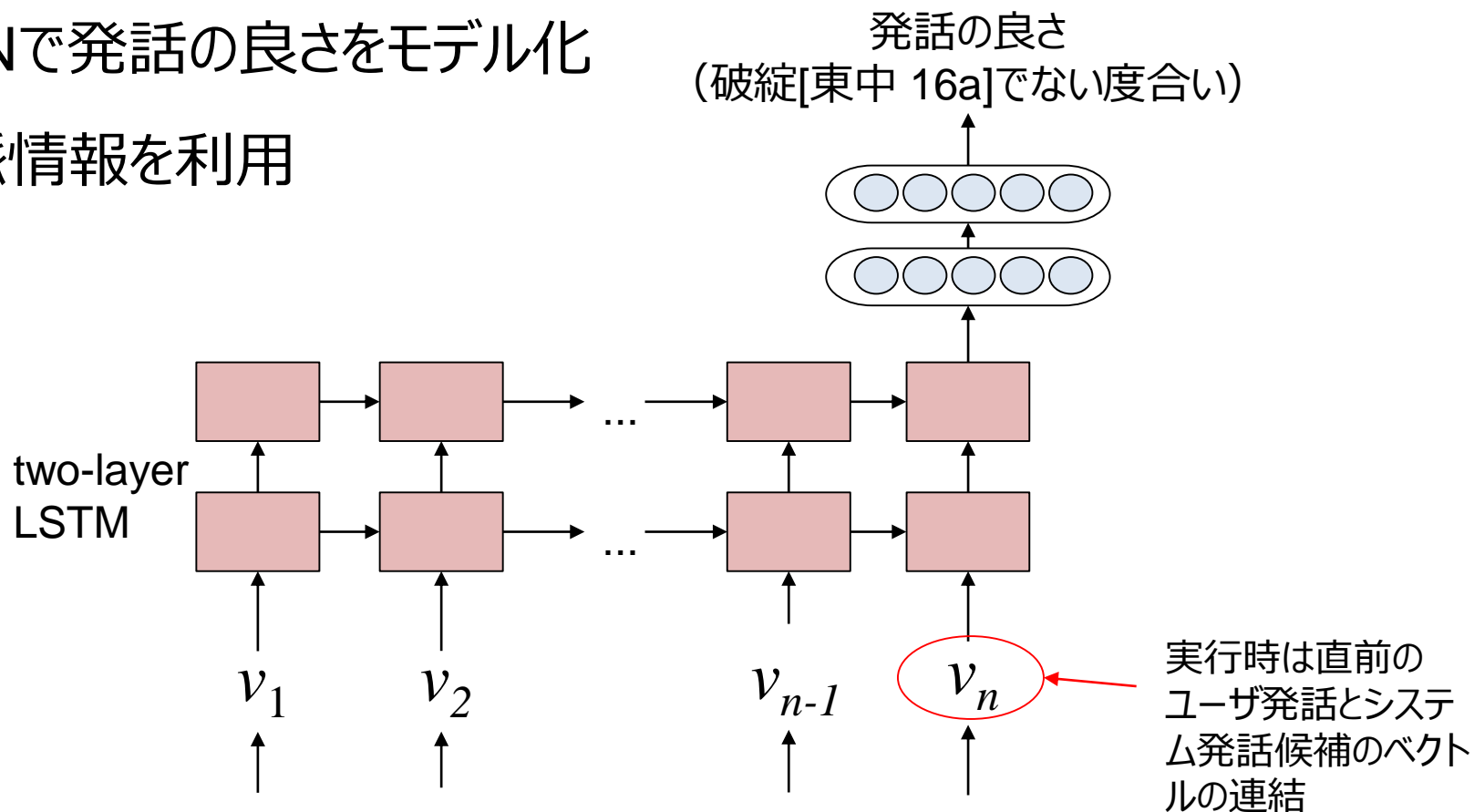
言語理解技術の導入

言語理解を行うことで
より良い応答を取り出す

[Higashinaka 14]をベースに作成



- RNNで発話の良さをモデル化
- 文脈情報を利用



v_i i 番目のユーザ発話のベクトルとシステム発話のベクトルを連結したもののベクトル化はtwo-layer LSTM

パーソナリティ表出・印象の向上

- 一貫性のあるシステムパーソナリティ
- 間違えても許されるキャラクタ
- 外見・話し方をシステムのレベルに合わせる
 - 適応ギャップ [小松 09]
- ユーザに共感する
- ゲーミフィケーション

今日のお話の流れ

- はじめに
- Part 1: とりあえず作ってみる
 - 対話システム構築の概要
 - 単純な対話システムの構築 (1) –DB検索テキスト対話システム–
 - 単純な対話システムの構築 (2) –オープンドメイン雑談対話システム–
 - Q&A
 - さまざまな対話システムの構築
 - Q&A
- Part 2: 対話システムの改良
 - 対話システムの評価
 - 少し進んだ構築法
 - **今後の展望**
 - Q&A

- 韻律の利用
 - 疑問文と平叙文の区別, 「態度」, 「感情」などの検出 [森 14] [東中 16b] [森 18]
- 画像入力の利用
 - ジェスチャや表情からユーザの状態を検出 (感情, 社会的信号) [中野 13] [東中 16b] [岡田 17]
- 逐次処理とスムーズな話者交替
 - 話すそばから理解して適切なタイミングで相槌や応答 [島津 14][駒谷 18]
- 統計的な対話状態推定
 - 複数のユーザ発話から最も尤もらしい意図理解結果を得る [吉野 18]
- 多人数対話
 - 誰が誰に話しているのかを認識して適切なタイミングで適切な人に話す [高瀬 17][杉山 16][藤江 12]
- 対話中の知識獲得
 - 対話に出てきた未知の事物の知識をユーザに聞いて獲得 [大野 18]

- システム構成が複雑になりすぎる
- 統計モデルの学習データをつくるのが大変 (アノテーションコストが大)

まだまだ解決しそうにない問題もたくさんある

- 構文意味的に複雑な発話の理解
- 文脈に依存した発話
- 複雑な話題構造の対話
 - 話題の埋め込み
 - 複数の話題の同時進行
- 言語表現の実世界へのグラウンディング

おわりに (1) 個人的な考え

- 良い対話システムを作るには、人が**対話システムをどう使うか**を知ることが**大事**
 - 研究者の数が少なく、知見が足りない
- 対話システムの汎用的な評価法を決めるは難しいが、**問題ない**
 - システムや条件によってさまざま
 - 対話システムは要素技術ではない
 - 評価軸が固定的だと研究開発の幅が狭まってしまう
- 対話システムにおける**機械学習の利用は簡単ではない**
 - モジュールによる
 - データが入手できるかどうか
 - 正解アノテーションが容易かどうか
 - ⇒ うまく組み合わせることが重要
- **雑談の目的**をはっきりさせる必要がある
 - タスク指向型対話においても雑談は大事
 - 人間は何かしらの目的を持って雑談をしている

おわりに (2) お願い

- どのような対話システムにどのような技術が有効か有効でないかは不明な点も多い

⇒ 是非**対話システムシンポジウム**で発表して共有してください！

対話システムシンポジウム：
人工知能学会 SLUD研究会が毎年秋に行っているシンポジウム。

アップロード版での訂正
今年は11/29,30の予定とお伝えしましたが、11月の別の日変更になる可能性が高いです。

- 今回のチュートリアルに際してご意見を伺った方々に感謝します。
 - 東中竜一郎氏 (NTT)
 - 大庭隆伸氏 (NTTドコモ)
 - 稲葉通将氏 (広島市立大学)
- 日ごろよりディスカッションさせて頂いている皆様に感謝します。
 - HRI-JPの対話システム関連の研究者・エンジニア・インターンの皆様
 - 共同研究者の皆様
 - 電子情報通信学会誌 2018年9月号 小特集「音声言語理解のこれまでとこれから」執筆者・編集チームの皆様

付録

役に立つイベント

- 国際会議SIGDIAL
 - 対話システムを「つくる」発表が比較的多い
 - PresidentやGeneral Chairがデモしたりする
- 対話システムシンポジウム
 - 人工知能学会 SLUD研究会
 - デモもたくさんある

参考になる書籍・記事

- 中野 幹生, 駒谷 和範, 船越 孝太郎, 中野 有紀子 (奥村学監修) : 対話システム, コロナ社, 2015
 - あまりわかりやすく書けなかったが網羅的に載っている
 - 雑談対話システムの話が少ない
- 川添 愛 : 働きたくないイタチと言葉がわかるロボット, 朝日出版社, 2017
 - 現在の言語理解技術と限界がわかりやすく書かれている
- 山田 誠二 編 : ロボットと人の間をデザインする, 東京電機大学出版局, 2007
 - エージェントのデザインに関してさまざまな話が載っている
- 荒木 雅弘 : フリーソフトでつくる音声認識システム(第2版), 森北出版, 2017
 - 音声認識から対話まで, 使えるツールが解説されている
- 東中 竜一郎 : 対話システムの作り方 (仮), 近代科学社, 近刊
 - 非常にわかりやすく書いていただけるとのこと
- 「小特集:音声言語理解のこれまでとこれから」電子情報通信学会誌, 2018年9月号掲載予定

対話システム構築の参考になる文献

- オープンドメイン雑談対話システム
 - [Banchs 12] 文脈を考慮した用例ベースシステム
 - [Higashinaka 14] [大西 14] NTTの雑談対話システム。さまざまな方法で文脈をモデル化している
 - [Fang 17] Alexa Prize 2017で優勝したシステム。Alexa Skillをうまく組み合わせている
 - [稲葉 17] 破綻検出を応用した発話選択
- その他のシステム
 - [Bohus 09] Agenda-based対話管理。事例が豊富
 - [Yan 17] オンラインショッピングの対話システム。機械学習を使うところとそうでないところをうまく組み合わせている
 - [小堀 16] 簡単な構造のインタビュー対話システム。雑談も行う。
 - [Traum 15] ホロコーストの生存者のインタビュー映像を用いた対話システム。簡単な構成だがリアリティのある対話ができる
 - [下岡 17] 傾聴システムの構築法に関して詳しく書かれている

- [Banchs 12] Banchs, R. E. and Li, H.: IRIS: a Chat-oriented Dialogue System based on the Vector Space Model, in Proc. of the ACL 2012 System Demonstrations, pp. 37-42 (2012)
- [Bohus 09] Bohus, D. and Rudnicky, A. I.: The RavenClaw dialog management framework: Architecture and systems, Computer Speech and Language, 23(3), pp. 332-361 (2009)
- [Fang 17] Fang, H., Cheng, H., Clark, E., Holtzman, A., Sap, M., Ostendorf, M., Choi, Y., and Smith, N. A.: Sounding Board - University of Washington's Alexa Prize Submission, in Proc. Alexa Prize 2017 (2017)
- [藤江 12] 藤江 真也, 松山 洋一, 谷山 輝, 小林 哲則 : 人同士のコミュニケーションに参加し活性化する会話ロボット, 電子情報通信学会論文誌, J95-A(1), pp. 37-45 (2012)
- [Higashinaka 14] Higashinaka, R., Imamura, K., Meguro, T., Miyazaki, C., Kobayashi, N., Sugiyama, H., Hirano, T., Makino, T., and Matsuo, Y.: Towards an open-domain conversational system fully based on natural language processing, in Proc. 25th COLING, pp. 928-939 (2014)
- [東中 16a] 東中 竜一郎, 船越 孝太郎, 稲葉 通将, 荒瀬 由紀, 角森 唯子 : 対話破綻検出チャレンジ 2, 人工知能学会言語・音声理解と対話処理研究会 (第7回対話システムシンポジウム), Vol. 78, pp. 64-69 (2016)
- [東中 16b] 東中 竜一郎, 岡田 将吾, 藤江 真也, 森 大毅 : 対話システムと感情, 人工知能学会誌, 31(5), pp. 664-670 (2016)

- [稲葉 17] 稲葉 通将, 高橋 健一: 非タスク指向型対話システムのためのリカレントニューラルネットワークを用いた発話候補ランキング, 電子情報通信学会論文誌, J100-D(6), pp. 661-671 (2017)
- [小堀 16] 小堀 嵩博, 中野 幹生, 中村 友昭: 雑談によりユーザに良い印象を与えるインタビュー対話システム, 人工知能学会 言語・音声理解と対話処理研究会 (第 7 回対話システムシンポジウム), Vol. 78, pp. 19-26 (2016)
- [駒谷 18] 駒谷 和範: 円滑な対話進行のための音声からの情報抽出, 電子情報通信学会誌, 101(9) (2018), (掲載予定)
- [小松 09] 小松 孝徳, 山田 誠二: 適応ギャップがユーザのエージェントに対する印象変化に与える影響, 人工知能学会論文誌, 24(2), pp. 232-2401 (2009)
- [Kusner 15] Kusner, M. J., Sun, Y., Kolkin, N. I., and Weinberger, K. Q.: From word embeddings to document distances, in Proc. 32nd ICML, pp. 957-966 (2015)
- [Lala 17] Lala, D., Pierrick, M., Inoue, K., Ishida, M., Zhao, T., and Kawahara, T.: Attentive listening system with backchanneling, response generation and Flexible turn-taking, in Proc. 18th SIGDIAL Conference, pp.127-136 (2017)
- [Meguro 13] Meguro, T., Minami, Y., Higashinaka, R., and Dohsaka, K.: Learning to control listening-oriented dialogue using partially observable Markov decision processes, ACM Trans. on Speech and Language Processing, 10(4) (2013)

- [目黒 14] 目黒 豊美, 杉山 弘晃, 東中 竜一郎, 南 泰浩 : ルールベース発話生成と統計的発話生成の融合に基づく対話システムの構築, 2014 年度人工知能学会全国大会 (第29回) 論文集 (2014)
- [森 14] 森 大毅, 前川 喜久雄, 粕谷 英樹 : 音声は何を伝えているか - 感情・パラ言語情報・個人性の音声科学 -, コロナ社 (2014)
- [森 18] 森 大毅 : 音声からの感情・態度の理解, 電子情報通信学会誌, 101(9) (2018), (掲載予定)
- [中野 13] 中野 有紀子 : 会話ヒューマノイドにおける非言語コミュニケーションシグナルの理解, 人工知能学会 言語・音声理解と対話処理研究会, Vol. 68, pp. 35-40 (2013)
- [中野 15] 中野 幹生, 駒谷 和範, 船越 孝太郎, 中野 有紀子 (著) 奥村 学 (監修) : 対話システム, コロナ社 (2015)
- [西村 06] 西村 義隆, 簗津 真一郎, 土肥 浩, 石塚 満, 中野 幹生, 船越 孝太郎, 竹内 誉羽, 長谷川 雄二, 辻野 広 司 : インタクション機能を有するプレゼンテーション記述言語の開発, 電子情報通信学会技術研究報告. HCS, ヒューマンコミュニケーション基礎, Vol. 106, pp. 43-48 (2006)
- [大西 14] 大西 可奈子, 吉村 健 : コンピュータと自然な会話を実現する雑談対話技術, NTT DOCOMO テクニカルジャーナル, 21(4), pp. 17-21 (2014)
- [大野 18] 大野 航平, 武田 龍, ニコルズ エリック, 中野 幹生, 駒谷 和範 : 対話を通じた未知語のクラス獲得に向けた暗黙的確認の提案, 人工知能学会論文誌, 33(1), pp. DSH-E 1-10 (2018)

参考文献

- [岡田 17] 岡田 将吾, 石井 亮: 社会的信号処理とAI, 人工知能学会誌, 32(6) pp. 915-920 (2017)
- [颯々野 18] 颯々野 学: 音声発話からの意味理解, 電子情報通信学会誌, 101(9) (2018), (掲載予定)
- [島津 14] 島津 明, 堂坂 浩二, 中野 幹生, 川森 雅仁: 話し言葉対話の計算モデル, 電子情報通信学会 (2014)
- [下岡 17] 下岡 和也, 徳久 良子, 吉村 貴克, 星野 博之, 渡部 生聖: 音声対話ロボットのための傾聴システムの開発, 自然言語処理, 24(1) (2017)
- [杉山 16] 杉山 貴昭, 船越 孝太郎, 中野 幹生, 駒谷 和範: 多人数対話におけるユーザの状態に着目したロボットの応答義務の推定, 人工知能学会論文誌, 31(3), p.C-FB2_1-9, (2016)
- [高瀬 17] 高瀬 裕, 吉野 堯, 中野 有紀子: 多人数会話における調整・介入機能を有する対話ロボット, 情報処理学会論文誌, 58(5), pp. 967-980 (2017)
- [Traum 15] Traum, D. R., Georgila, K., Artstein, R., and Leuski, A.: Evaluating Spoken Dialogue Processing for Time-Offset Interaction, in Proc. 16th SIGDIAL Conference (2015)
- [Yan 17] Yan, Z., Chen, N. D. P., Zhou, M., Zhou, J., and Li, Z.: Building Task-Oriented Dialogue Systems for Online Shopping, in Proc. 31st AAAI (2017)
- [吉野 18] 吉野 幸一郎: 音声発話系列からのユーザの意図の理解, 電子情報通信学会誌, 101(9), (2018), (掲載予定)