

知的情報処理論 (第5回)

2023年5月16日(火)
産業科学研究所
駒谷 和範

レポート(第2回)を出題しました

- CLE上にあります
- 締切: 6月12日(月)
 - 第9回講義(武田先生担当開始日)の前日
 - 考察や説明を, 他人が読んでわかる日本語(または英語)で書くこと.
 - 実行結果のみを送りつけてきた場合は, 極めて低く評価する, または受理しない.
- 駒谷担当分のレポートはこの2回で完了

汎化性能

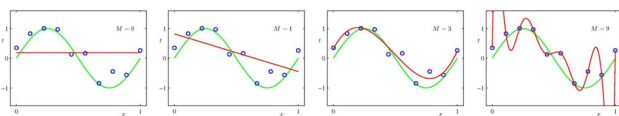
機械学習／パターン認識の目的

- 機械学習の本来の目的
 - × 学習データを正しく分類する
 - 学習データにない未知のデータを正しく分類する

汎化 (generalization) 性能

過学習 (overfitting)

- 学習データは常に有限
 - 統計的機械学習では, 有限のデータから真の分布を推定
- 過学習 (overfitting)
 - 学習データに過剰に適応し汎化性能が低くなること
 - 学習データの量に対して, 自由度の大きい(次元数, パラメータ数が多い)モデルで起こりやすい



上記グラフの出典はPRML:
C.M. ビショップ著, 元田浩, 他訳: "パターン認識と機械学習(上・下)", 丸善出版, 2012.

- モデル選択 (model selection)
 - 適切な学習モデル(ハイパーパラメータを含む)を選ぶ
- 正則化 (regularization)
 - 目的関数に正則化項(制約)をつける. 重みが複雑な値になるのを避ける

得られる重みの値に関する制約

- 得られた重みの値は「最適」なのか
何を以て「最適」とするか？

- 最適になるように損失関数を工夫する

① 重みの値が極端に大きい値を取らない

→ 正則化項 (regularization term) の導入

- 損失関数に重みの値を含める。
L1正則化(Lasso), L2正則化(Ridge)

- L1正則化(Lasso)

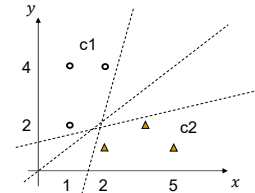
$$\min\{(\text{誤差の総和}) + \lambda \sum_{i=1}^n |w_i|\}$$

得られる重みの値に関する制約

何を以て「最適」とするか？

② 識別面が各クラスの点のちょうど間くらい

→ マージン(margin)最大化



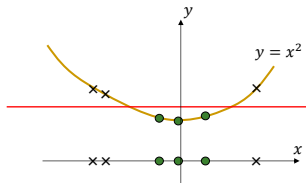
- SVM (support vector machine) の特徴

1. マージン最大化
2. カーネル(kernel)関数の導入

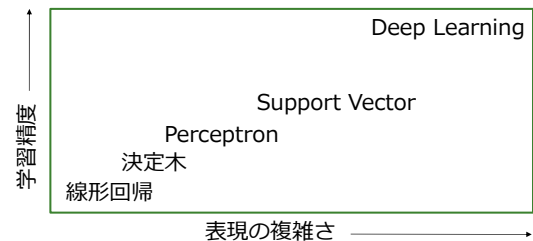
SVMの特徴

2. カーネル(kernel)関数の導入

- 非線形な高次元空間に写像することで分類可能とする



学習に必要なデータ量



- 表現(モデル)が複雑 ← パラメータ数が多い
 - きちんと学習させるには学習データ量が必要
 - 学習データを少し変えた場合にもsensitive
- 「何でも最新のDeep Learningを使えばいい」はウソ

分類問題への3つのアプローチ

入力 x に対してクラスラベル C_i を決める

$$\hat{C} = \operatorname{argmax}_i p(C_i|x) \text{ を求めたい}$$

① 識別関数

- 入力からクラスラベル C_i への写像関数 $f(x)$ を直接求める
- 確率や統計は不使用

② 識別モデル

③ 生成モデル

使われる確率モデルによる大別

データからの統計的な予測(確率分布の推定)に基づく

② 識別モデル (discriminative model)

- 事後確率 $p(C_i|x)$ を直接モデル化

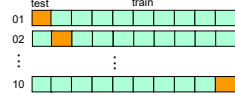
③ 生成モデル (generative model)

- $p(C_i|x) \propto p(x|C_i)p(C_i)$ (ベイズの定理より)
- クラスの条件付き密度 $p(x|C_i)$ と事前クラス確率 $p(C_i)$ をモデル化

「損失関数が最小となるようパラメータ更新」は共通

評価方法

- 汎化性能を測るには評価データと学習データを分けて実験
 - 分割学習法 (open test) → 例えば半分を学習, 半分を評価に使用
 - closed test: 学習データをそのまま評価データにすること
- ただし, それだと学習データが不十分な場合
 - ⇒ 交差検証法 (cross validation)
 - 10分割交差検証
 - 学習データを10分割する
 - 9個を使って学習, 残り1個で評価
 - 2を10個分繰り返し, その結果の平均を用いる
 - 同じデータを分割することから, (おそらく) 均質なデータの中での評価



ハイパーパラメータをデータで最適化する
なら, 検証用セット (validation set) も必要

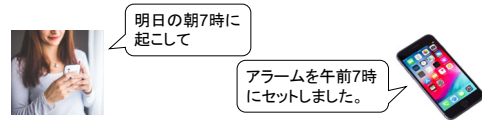
Mini Quiz #1

- Closed test と Open test (もしくは交差検証) では, 一般に前者の方が性能が高い。なぜか。
- 以下のような実験結果が得られた場合どういう状況にあることが考えられるか
 - Closed test も Open test もそこそこ高性能が得られた
 - Closed test の性能は高いが, Open test ではとても低い
 - Closed test の性能が (Open test の性能も) とても低い

本日 (第5回) の内容

- 第1回 人工知能の歴史
- 第2回第3回第4回 機械学習の基礎
- 第5回
 - (音声)対話システムの構成
 - 言語理解 (Language Understanding, LU)
- 第6回 言語理解の利用法, 対話管理
- 第7回 言語以外の要素
- 第8回 人工知能と社会

音声対話システム



- 人間の音声による問いかけに対して, それを理解し応答を返すシステム
 - Apple Siri
 - NTTドコモ シャベッてコンシェル → my daiz
 - Yahoo! Japan 音声アシスト
 - ...

教科書



対話システム
 ■ 中野, 駒谷, 船越, 中野
 ■ 2015年1月
 理論的背景
 網羅的なトピック列挙



Pythonでつくる対話システム
 ・ 東中, 稲葉, 水上
 ・ 2020年3月
 テキスト対話を, サンプルプログラムを使って説明



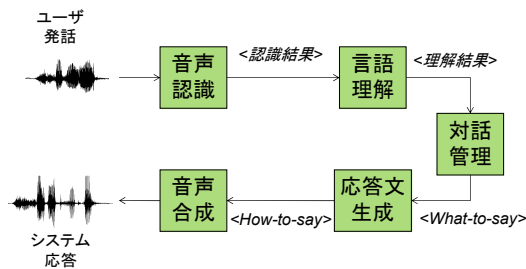
音声対話システム
 基礎から実践まで
 ・ 井上, 河原
 ・ 2022年10月
 音声対話全般



対話システムの作り方
 ・ 東中
 ・ 2023年2月
 テキスト対話システムの
 具体的な構築事例
 最新の情報も含まれる

音声対話システムの構成と分類

音声対話システムの基本構成

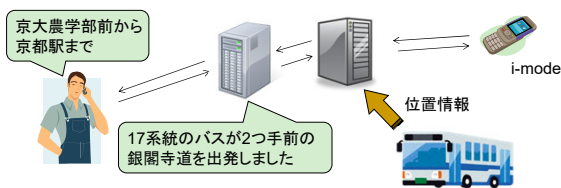


音声対話システムの分類

1. 対話型(タスク指向)
 - システムが状態を持つ
2. 一問一答型

京都市バス運行情報案内システム

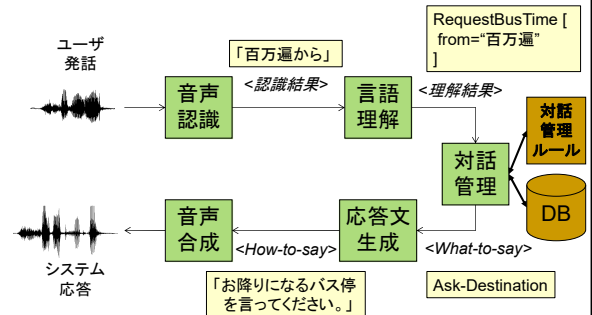
- 電話で利用可能: 一般市民に公開
(2007/03で終了; 34か月間運用)



- バスの接近情報:
京都市交通局がWeb上にリアルタイムで提供



音声対話システムの基本構成(対話型)



基本的な対話例

S: ご利用になるバス停名、系統番号をどうぞ

U: 百万遍から 音声認識

FROM

言語理解

S: お降りになるバス停を教えてください

U: 京都駅 TO

対話管理

S: 百万遍から京都駅まででよろしいですか

U: はい

S: 京都駅行きのバスは、6つ手前の高木町を出発しています

Slot-filling型

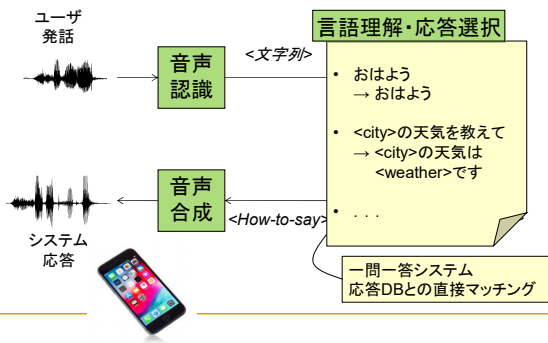
FROM	百万遍
TO	京都駅
ROUTE	

- FROMは必須
- TOまたはROUTEのいずれかが必要

音声対話システムの分類

1. 対話型(タスク指向)
 - システムが状態を持つ
 - 対話管理が必要(同じ入力に対して応答が異なる)
2. 一問一答型

音声対話システムの基本構成(一問一答型)



応答データベースの例

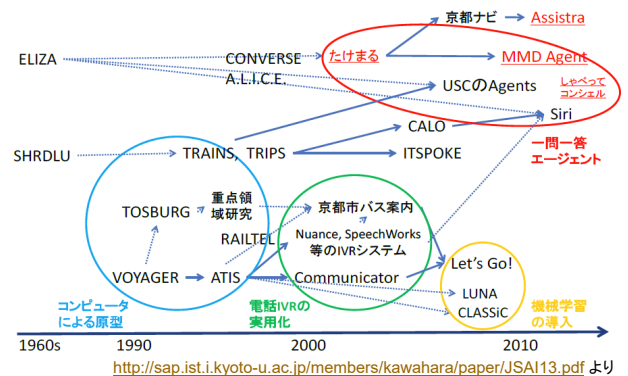
m	質問 Q_m	応答 A_m (動作, 出力)
1	こんにちは	「どうもこんにちは」
2	愛してるよ	「ありがとう。でも仕事に戻りましょうよ。」 「まあ！でもみんなにそう言ってるでしょ。」
3	週末の大阪の天気は	天気検索 (大阪, 週末)
4	天気はどうなるかな	天気検索 (現在地, 今日)
5	タイマーをDURIにセット	タイマー起動 (DUR)
6	富士山の高さは?	質問応答モジュール (富士山の高さは?)
7	STATIONに行きたい	乗換案内 (STATION, 最寄駅)
...

- デフォルトルールを多用することで一問一答化
 - 場所=現在地, 日時=今
- 一問一答なので, 音声認識誤りがあっても確認しない
 - 間違った結果でも, 出力した方が速いから

音声対話システムの分類

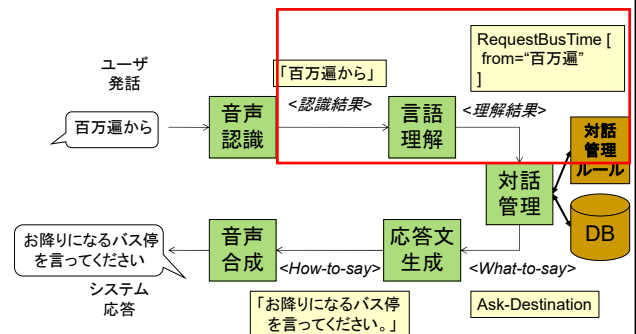
1. 対話型(タスク指向)
 - システムが状態を持つ
 - 対話管理が必要(同じ入力に対して応答が異なる)
2. 一問一答型
 - 入力に対して最適な応答をすればよい
 - スマートフォン上の多くのアプリケーション

代表的な音声対話システムの系譜 [河原13]



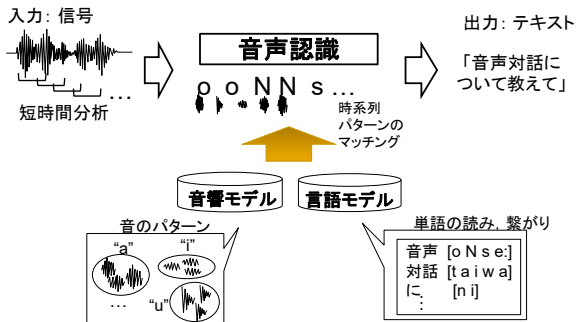
言語理解 (LANGUAGE UNDERSTANDING)

音声対話システムの基本構成(対話型)



音声認識

武田先生担当部分で解説



実用例

■ Microsoft 会話言語理解 (旧 LUIS)

<https://azure.microsoft.com/ja-jp/products/cognitive-services/conversational-language-understanding/>



<http://download.microsoft.com/download/7/D/7/D239418-DAA3-4FBF-BD10-4F099A5E1BC5/u-tokyo.pdf>

言語理解

- 入力: テキスト (音声認識結果の場合単語列)
- 出力: 計算機が理解可能な意味表現
 - タスク・ドメインによって出力形式は異なる



34

言語理解部の中身

理解結果 (中間表現: フレーム表現) を出力

1. ドメイン
 2. 意図
 3. スロット値
- ⇒ 後段の対話管理部に渡される (いわゆる「対話型」)

35

1. ドメイン同定 / 2. 意図同定



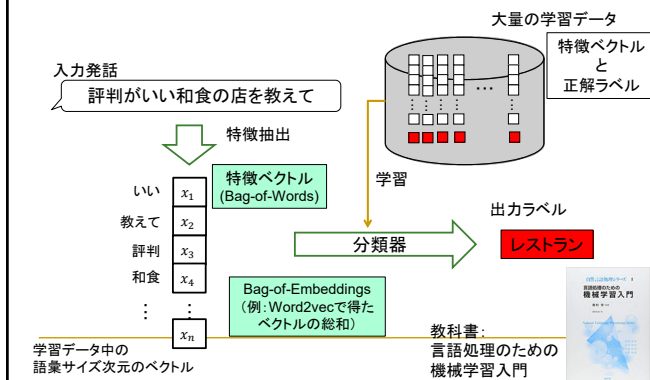
36

機械学習 (多クラス分類問題)

- 入力: 単語列
- 出力: 定義されたクラス
- 分類器
 - Naïve Bayes, SVM, Logistic Regression (LR), ...
 - Neural Network

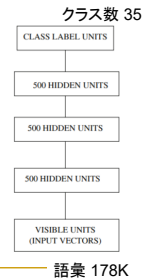
37

多クラス分類問題の例



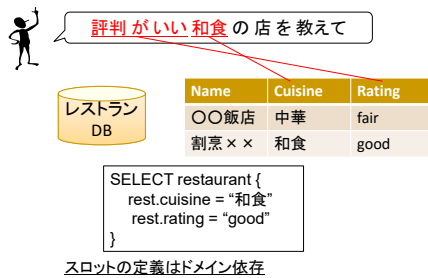
ニューラルネットワークに基づく 言語理解(多クラス分類問題)

- Deep belief nets (DBN) [Sarikaya et al., 2011]
 - Call routingタスク
 - 入力: 語彙サイズV次元ベクトルのうち、各単語に対応する要素に値があるもの
 - DNNに基づく言語理解の草分け
- Deep convex networks (DCN) [Tur et al., 2012; Deng et al., 2012]
- RNN and LSTMs [Ravuri & Stolcke, 2015]
- RNN and CNNs [Lee & Derroncourt, 2016]



39

3. スロット値同定



40

Mini Quiz #2

- (ここまでで説明した)対話システムにおける「言語理解」とは何か
 - 3つ
 - それぞれどういう単位で? (単語? 発話全体?)
- それは、あなたが考える「言語の理解」と同じか
 - 工学的な(ここでの)「理解」/Scientificな「理解」
 - あるなら相違点は何か

ちなみに(第1回で既出の話題)

ALBERT Japan > ニュース > 企業・業界

アリババのAI、読解カテストで人間を超える

Source: Yu, C. (2018, Nov 14). 阿里集团: 2018年11月14日 09時31分

入力: Wikipediaの段落(平均140単語)

In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravity. The main forms of precipitation include drizzle, rain, sleet, snow, graupel and hail. Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. Short, intense periods of rain in scattered locations are called "showers".

What causes precipitation to fall?
gravity

What is another main form of precipitation besides drizzle, rain, snow, sleet and hail?
graupel

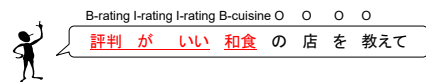
Where do water droplets collide with ice crystals to form precipitation?
within a cloud

Figure 1: Question-answer pairs for a sample passage in the SQuAD dataset. Each of the answers is a segment of text from the passage.

<https://aclweb.org/anthology/D16-1264>

- 人間=クラウドワーカー
 - テスト時は120秒/5問で回答
 - 1.5時間程度連続作業
 - 報酬 9ドル/時
 - EM (Exact Match): 82.304
- SQuAD (Stanford Question Answering Dataset)
 - 現在2.0
 - <https://rajpurkar.github.io/SQuAD-explorer/>
- 2018年10月 BERT (by Google)
 - EMスコア: 87.433 -> SQuAD 1.0 終了

機械学習によるスロット値同定



- 系列ラベリング問題
 - 系列内の各要素にラベルをつける
 - 各ラベルは系列の前後の要素に依存
- IOB2ラベリング
 - ラベルが複数語からなる場合を表現
 - Bは開始点(beginning), Iはその内部(inside), Oは外部(outside)

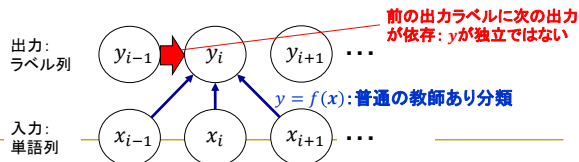
43

スロット値同定

週末の大阪の天気を教えて

[Date] [City]

- キーワードの抽出
 - 単純手法: 辞書中の単語とマッチング
 - 単純手法では誤りとなる例: 「大阪駅のそばにある」
- 単語文脈を考慮した機械学習: 系列ラベリング問題



機械学習による系列ラベリング

- 入力: 単語の系列
- 出力: IOB2ラベルの系列
 - 単語列のどの部分がスロット値かがわかる

系列ラベリング問題

- 分類器
 - Conditional Random Field (CRF), Hidden Markov Model (HMM), ...
 - ニューラルネットワーク (RNN, LSTM)

46

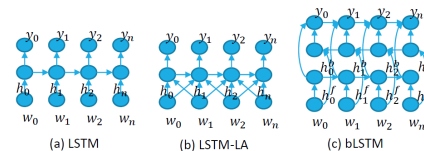
固有表現抽出

- 自然言語処理のタスクとしてポピュラーな系列ラベリング問題
- テキストの中から、固有表現 (Named Entity: NE) を抽出
 - 組織名, 人名, 地名, 日付など

田中 将大 が 神戸 の マウンド に 立つ
B-NAME I-NAME O B-PLACE O O O O

ニューラルネットワークによるスロット値同定(1)

- RNNの利用 [Mesnil et al., 2013] など
 - LSTM
 - n-gramに窓をかけて入力に使用
 - 双方向LSTM

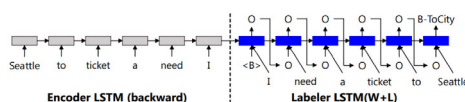


<http://deepdialogue.milulab.tw/> より

48

ニューラルネットワークによるスロット値同定(2)

- 機械翻訳で用いられる系列変換モデルを適用
 - Encoder-Decoderネットワーク [Kurata et al., 2016]



- 注意機構 (attention mechanism) の利用 [Simonnet et al., 2015]

- 系列のj番目の出力時の, i番目の隠れ状態による重み
- (例えば) feed-forward NNにより学習

教科書:
深層学習による自然言語処理 (2017)

RNN/LSTMはTransformerへ

- Transformer [Vaswani+ (Google), 2017]
<https://arxiv.org/pdf/1706.03762.pdf>
 - RNNやCNNを用いず, アテンションを用いたエンコーダデコーダ型モデル
 - 機械翻訳で当時の最高性能を更新

