

Человеко-машинное взаимодействие

Создание текстовых интерфейсов

Часть 2

ИВТ и ПМ
ЗабГУ

2017

Содержание

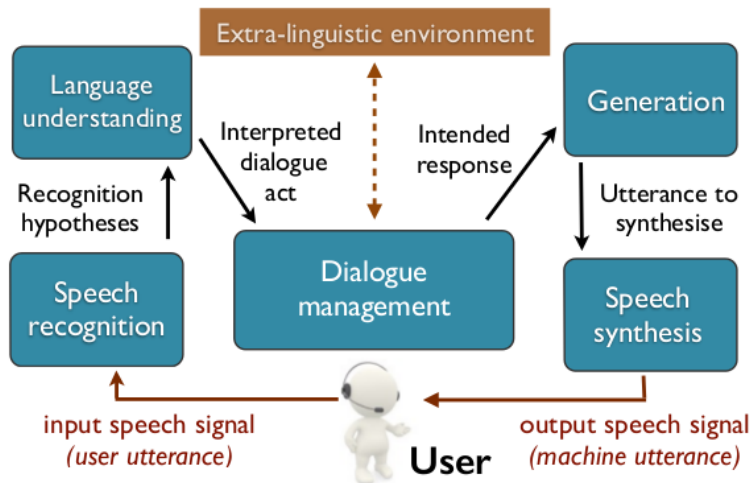
Повторение

Применение word2vec

Разработка Чатбота. Поиск намерений пользователя

Разработка Чатбота. Поиск сущностей

Архитектура чатбота



Архитектура чатбота

- ▶ **Speech recognition:** Распознавание речи. Голос в Текст
- ▶ **Language understanding:** Распознавание намерений пользователя
- ▶ **Dialogue manager:** Управление диалогом, контекстом и принятием решения
- ▶ **Generation:** Подготовка текстового ответа
- ▶ **Speech synthesis:** Генерация голосового ответа. Текст в Голос

Какие из компонентов самые сложные для реализации?

Векторная модель слов

Что такое векторная модель слов?

Векторная модель слов

Что такое векторная модель слов?

- ▶ **dog**: [0.033, -0.009, 0.037, 0.068, ..., 0.053]
- ▶ **cat**: [0.067, 0.069, -0.015, -0.0249, ... , 0.0317]
- ▶ **home**: [-0.0004, 0.0199, 0.03, -0.04, ... , -0.031]
- ▶ **computer**: [-0.0016 , 0.0872, -0.0314, -0.0523, ... , 0.0228]
- ▶ **love**: [-0.169, 0.0382, 0.0084, -0.0182, ... , -0.0209]

Центроид

- ▶ Дано: список слов или предложение
- ▶ Задача: найти слово близкое по смыслу к словам списка

Центроид

- ▶ Дано: список слов или предложение
- ▶ Задача: найти слово близкое по смыслу к словам списка

$$L = [word_1, \dots, word_n]$$

$$Centroid(L) = \frac{\sum_{i=1}^n word2vec(w_i)}{|L|}$$

Применение. Близкие слова

кофе

коффе

чая

чай

капучино

кофн

какао

эспрессо

кофя

цикорий

кофэ

копучино

шоколад

капучинно

Применение. Кластеризация запроса

мобильный телефон

сотовый
телефона
смартфон
телефон
мобильного
мобильник
телефо
телефон
тлф
мобильный
телеон
телефоны
телефрн
сотового

Применение. Отношения между словами

франция пари́ж германия

мюнхен
берлин
дюссельдорф
гамбург
кельн
амстердам
франкфурт
прага
ашаффенбург

сша доллар украина

гривне
доллар
гривны
рубля
доллара
гривна
рублю
рубль
доллару

Применение. Опечатки

преключение

приключение

преключения

приключения

приключеия

прключение

приключеня

прключения

приключени

Применение. Транслитерация

avito

awito

авито

fvito

авита

irr

овито

avito

Применение. Не та раскладка

пщщпду

Применение. Не та раскладка

пщщпду

пщщп

пгпд

гугл

поопду

ппщпду

пщщплу

нфтвуч

пщщпдуюкг

пщщпу

gogle

иштп

google

Применение. Названия сайтов, сайты-аналоги

mail

rambler

meil

inbox

maill

yandex

maii

myrambler

zmail

mefr

jandex

gmail

vkmail

Основная проблема

Какой компонент чатбота самый сложный для разработки?

Основная проблема

Какой компонент чатбота самый сложный для разработки?

Задачи компонента NLU (понимания языка)

- ▶ Intent recognition (понимание намерений человека из реплики)
- ▶ Information extraction (понимание сущностей - деталей реплики)

Решение проблем

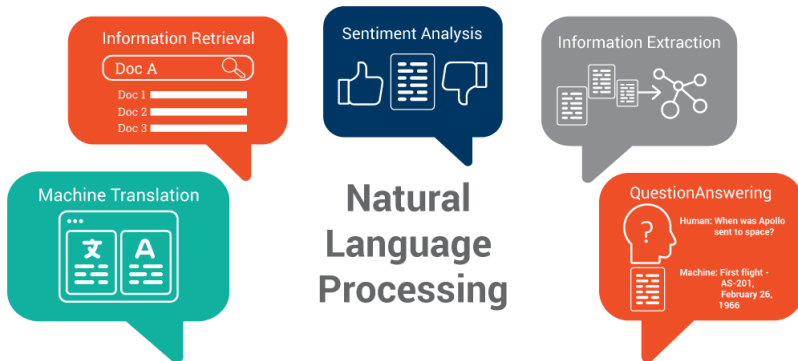
Решить эти проблемы мы сможем с помощью NLP

Решение проблем

Решить эти проблемы мы сможем с помощью NLP
А что такое NLP?

Решение проблем

Решить эти проблемы мы сможем с помощью NLP
А что такое NLP?



Промышленные решения

- ▶ LUIS.ai - By Microsoft
- ▶ Wit.ai - By Facebook
- ▶ Api.ai - By Google
- ▶ Watson - By IBM

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

"Yes", "I agree" - **Acceptance**

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

"Yes", "I agree" - **Acceptance**

"No", "I disagree" - **Reject**

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

"Yes", "I agree" - **Acceptance**

"No", "I disagree" - **Reject**

"Bye" - **Quit**

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

"Yes", "I agree" - **Acceptance**

"No", "I disagree" - **Reject**

"Bye" - **Quit**

- ▶ Специфические намерения (Business Intents)

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию

Что такое намерения (Intents)?

Intent — намерения пользователя выраженные в предложении, могут быть двух типов.

- ▶ Намерения общего назначения (Casual Intents)

"Hello", "How are you", "hi" - **Greetings**

"Yes", "I agree" - **Acceptance**

"No", "I disagree" - **Reject**

"Bye" - **Quit**

- ▶ Специфические намерения (Business Intents)

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию

"Когда сняли 'Марсианина'?"

Поиск сущностей (Entities)

"Когда 'Марсианин' вышел на экран?" -
ДатаВыпускаПоНазванию

Чего нам не хватает для того чтобы ответить на запрос?

Поиск сущностей (Entities)

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию

Чего нам не хватает для того чтобы ответить на запрос?

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию('Марсианин')

Поиск сущностей (Entities)

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию

Чего нам не хватает для того чтобы ответить на запрос?

"Когда 'Марсианин' вышел на экран?" -

ДатаВыпускаПоНазванию('Марсианин')

"Когда 'Прибытие' вышел на экран?" -

ДатаВыпускаПоНазванию('Прибытие')

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

"Кто является главой Apple?"

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

"Кто является главой Apple?" (Apple, **ОРГАНИЗАЦИЯ**)

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

"Кто является главой Apple?" (Apple, ОРГАНИЗАЦИЯ)

"Сколько стоит аренда в Сан-Франциско?"

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

"Кто является главой Apple?" (Apple, **ОРГАНИЗАЦИЯ**)

"Сколько стоит аренда в Сан-Франциско?"

(Сан-Франциско, **ЛОКАЦИЯ**)

Какие бывают сущности

- ▶ Общие сущности (common entities): ДАТА, ЧИСЛА, ОРГАНИЗАЦИИ, ЛОКАЦИИ

"Кто является главой Apple?" (Apple, **ОРГАНИЗАЦИЯ**)

"Сколько стоит аренда в Сан-Франциско?"

(Сан-Франциско, **ЛОКАЦИЯ**)

- ▶ Специфические сущности (special entities): НАЗВАНИЕ ФИЛЬМА

Пример

Сколько стоит пара красных кроссовок адидас 43 размера.

Намерения (intents)?

Пример

Сколько стоит пара красных кроссовок адидас 43 размера.

Намерения (intents)?

Intents: **Запрос****Цены**

Пример

Сколько стоит пара красных кроссовок адидас 43 размера.

Намерения (intents)?

Intents: **Запрос****Цены**

Сущности (entities)?

Пример

Сколько стоит пара красных кроссовок адидас 43 размера.

Намерения (intents)?

Intents: **Запрос****Цены**

Сущности (entities)?

Размер - 43

Фирма - Adidas

Цвет — Red

Поиск сущностей

"Я хочу найти ресторан индийской кухни"

Поиск сущностей

"Я хочу найти ресторан индийской кухни"
Какие здесь сущности?

Поиск сущностей

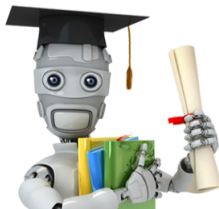
"Я хочу найти ресторан индийской кухни"

Какие здесь сущности?

А как их найти?

Поиск намерений

По аналогии, как найти намерения



k Ближайших Соседей (k Nearest Neighbor)

k Ближайших Соседей

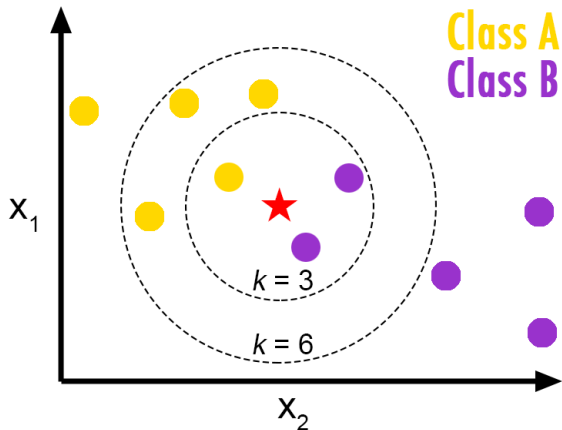
Дано: обучающая выборка

k Ближайших Соседей

Дано: обучающая выборка

- ▶ Вычислить расстояние до каждого из объектов обучающей выборки
- ▶ Отобрать k объектов обучающей выборки, расстояние до которых минимально
- ▶ Класс классифицируемого объекта — это класс, наиболее часто встречающийся среди k ближайших соседей

k Ближайших Соседей



к Ближайших Соседей

Гипотезой Компактности

Если мера сходства объектов введена достаточно удачно, то схожие объекты гораздо чаще лежат в одном классе, чем в разных.

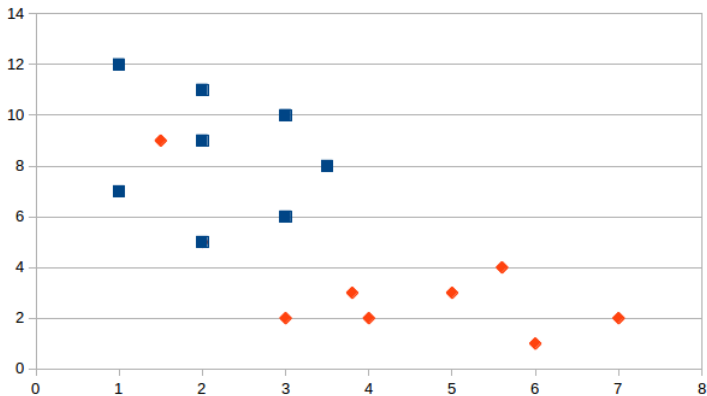
к Ближайших Соседей

Гипотезой Компактности

Если мера сходства объектов введена достаточно удачно, то схожие объекты гораздо чаще лежат в одном классе, чем в разных.

Какие меры сходства мы используем?

к Ближайших Соседей



к Ближайших Соседей

Типы объектов

- ▶ **эталоны** - типичные представители класса
- ▶ **периферийные** - на границе классов
- ▶ **шумовые** - объекты расположенные среди объектов другого класса

Домашняя работа

: capital-common-countries

Athens Greece Baghdad Iraq

Athens Greece Bangkok Thailand

...

: currency

Algeria dinar Angola kwanza

Algeria dinar Argentina peso

...

: city-in-state

Chicago Illinois Houston Texas

Chicago Illinois Philadelphia Pennsylvania

...

: family

boy girl brother sister

boy girl brothers sisters

...

Домашняя работа

king — queen \approx man — woman

Домашняя работа

$$\textit{king} - \textit{queen} \approx \textit{man} - \textit{woman}$$

$$? \approx \textit{king} - \textit{queen} + \textit{woman}$$

Домашняя работа

$$\textit{king} - \textit{queen} \approx \textit{man} - \textit{woman}$$

$$? \approx \textit{king} - \textit{queen} + \textit{woman}$$

$$\textit{man} == ?$$

Домашняя работа

Определение точности

- ▶ Top-1
- ▶ Top-5
- ▶ Top-10

Домашняя работа

- ▶ работа в парах
- ▶ отчеты на практике

Задание:

- ▶ Вычисление **третьего** элемента
- ▶ Полное понимание того, что происходит в коде
- ▶ Дописать функцию `ktop`, `centroid`
- ▶ Показать и обосновать какой вариант расчета расстояния лучший (с примерами)