

Софийски университет
„Св. Климент Охридски”

Дипломна работа

Катедра „Софтуерни технологии”



Прилагане на методи от машинния превод за изграждане на диалогова система за отговаряне на често задавани въпроси

Дипломант:

Мартин Боянов

Специалност:

**Извличане на информация и откриване на
знания**

Факултетен номер:

25187

Научни ръководители:

Д-р Преслав Наков, Проф. Иван Койчев

София 2017г.

Съдържание

1	Увод	2
1.1	Актуалност на проблема и мотивация	2
1.2	Цел и задачи на дипломната работа	2
1.3	Структура на дипломната работа	3
2	Преглед на предметната област	4
2.1	Въведение	4
2.2	Диалогови системи: мотивация и история	4
2.2.1	ELIZA	5
2.2.2	PARRY	5
2.2.3	ALICE	5
2.2.4	games	5
2.2.5	assistants	5
2.3	Машинен превод	5
2.4	Отговаряне на често задавани въпроси	6
3	Използвани технологии и платформи	7
3.1	intro	7
3.2	tensorflow	7
3.3	python	7
3.4	kelp	7
3.5	BPE	7
3.6	that person with the chatbot design	7
3.7	Data	7
3.7.1	UDC	7
3.7.2	QL	7
4	Реализация, експерименти и анализ на резултатите	8
4.1	Реализация	8
5	Заклучение	9
5.1	Обобщение на изпълнението на началните цели	9
6	Референции	10
7	Приложения	11

1 Увод

1.1 Актуалност на проблема и мотивация

Бързо разрастващата се сфера на мигновените съобщения (instant-messaging) поражда голям интерес за създаване на интелигентни системи за водене на диалог, още известни като чатботове.

Изграждането на такова приложение е трудна задача, поради липсата на структура в текста. Трудността се засилва и от факта, че при чат приложенията правописът и граматиката не са приоритет за потребителите. При зададен въпрос от потребител, системата трябва да генерира смислен и релевантен отговор.

Подобна задача се решава при машинния превод. При дадена поредица от думи, трябва да се генерира друга поредица от думи. Целта на тази дипломна работа е да се адаптира модел от машинния превод за изграждане на чатбот.

1.2 Цел и задачи на дипломната работа

Основната цел на тази работа е изследване на възможностите за изграждане на диалогова система базирана на модела seq2seq. Този модел е в основата на съвременните системи за машинен превод, използващи изкуствени невронни мрежи. Ще се изследват и възможностите за интегриране на модела в система за автоматично отговаряне на въпроси.

Задачи, произтичащи от целта:

- Обзор на подходите за решаване на този вид задачи
- Набавяне на набор от данни - Ubuntu Dataset Corpus и Qatar-Living
- Предварителна обработка на данните
- Имплементация и интеграция на seq2seq модела

-
- Анализ на процеса на машинно самообучение
 - Тестване, настройка и оценяване на работата на системата
 - Анализ на резултатите
 - Интеграция в системата за автоматично отговаряне на въпроси
 - Разработване на интерактивно демонстративно приложение

1.3 Структура на дипломната работа

2 Преглед на предметната област

2.1 Въведение

Изграждането на отворена диалогова система е предизвикателна задача, с която са се сблъскали много екипи. През годините са опитвани различни подходи за изграждане на подобна система и в тази глава ще се запознаем с основните методи и тяхната еволюция. Благодарение на прогреса в сферата на машинното самообучение и на нарастващата изчислителна мощ бяха постигнати подобрения в задачата за автоматичен машинен превод. Ще обобщим тяхната архитектура и ще коментираме приложимостта на подобен модел за изграждане на диалогова система. Ще разгледаме и проблема за отговаряне на често задавани въпроси и връзката му с диалоговите системи

<https://research.fb.com/the-long-game-towards-understanding-dialog/>

2.2 Диалогови системи: мотивация и история

Диалоговите системи са ултимативния интерфейс за употреба на компютърно устройство. Комуникацията с машина не трябва да бъде по-различна от комуникацията с друг човек. Постигането на тази цел ще направи устройствата по-достъпни и потребителите по-склонни да ги направят част от ежедневието си живот.

Идеята за диалогова система датира още от зората на компютърната индустрия. В следващите секции ще представим различните стадии на развитие на диалогови системи през годините.

2.2.1 ELIZA

ELIZA [1] е първата известна диалогова система. Разговорите с ELIZA симулират интервю с психотерапевт. Тъй като при такова интервю целта е потребителят да разкаже за собствения си мироглед, програмата няма нужда от знания за обиколния свят. Достатъчно е да бъде пасивна и да рефлектира на репликите на потребителя, насърчавайки го да разкаже повече за себе си.

Системата разчита на правила, които определят нейното поведение. Всяко правило се състои от ключова дума, тежест на ключовата дума и рецепта за трансформиране на входните данни в изходни. Едно примерно правило изглежда така:

```
["remember", 5, [  
  ["* i remember *", [  
    "Do you often think of (2) ?",  
    "Does thinking of (2) bring anything else to mind ?",  
    "What else do you recollect ?",  
    "Why do you remember (2) just now ?",  
    "What in the present situation reminds you of (2) ?",  
    "What is the connection between me and (2) ?",  
    "What else does (2) remind you of ?"  
  ]]  
],
```

При новопостъпило съобщение ELIZA извлича думата с най-голяма тежест и преобразува входните данни спрямо правилото на съответната дума. Примерен разговор с ELIZA може да се намери в 2.1.

2.2.2 PARRY

2.2.3 ALICE

2.2.4 games

2.2.5 assistants

2.3 Машинен превод

- statistical machine translation - neural machine translation - seq2seq
- seq2seq with attention

ФИГУРА 2.1: Примерен разговор с ELIZA

```
Welcome to

          EEEEE LL      IIII ZZZZZZZ AAAA
          EE      LL      II      ZZ  AA  AA
          EEEEE LL      II      ZZZ  AAAAAA
          EE      LL      II      ZZ  AA  AA
          EEEEE LLLLLL IIII ZZZZZZZ AA  AA

Eliza is a mock Rogerian psychotherapist.
The original program was described by Joseph Weizenbaum in 1966.
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:   █
```

[caption]

2.4 Отговаряне на често задавани въпроси

- проблем - решение

3 Използвани технологии и платформи

3.1 intro

3.2 tensorflow

3.3 python

3.4 kelp

3.5 BPE

3.6 that person with the chatbot design

3.7 Data

3.7.1 UDC

3.7.2 QL

4 Реализация, експерименти и анализ на резултатите

4.1 Реализация

5 Заключение

5.1 Обобщение на изпълнението на началните цели

6 Референции

- [1] Joseph Weizenbaum. “ELIZA—a Computer Program for the Study of Natural Language Communication Between Man and Machine”. In: *Commun. ACM* 9.1 (Jan. 1966), pp. 36–45. ISSN: 0001-0782. DOI: 10.1145/365153.365168. URL: <http://doi.acm.org/10.1145/365153.365168>.

7 Приложения