User Intention Classification

MI, Lan

MSc in Big Data Technology

Mentor: Prof. Lin Fangzhen

Catalog

Abstract

User intention classification is a hot topic today, with many kinds of applications like automatic online customer server and various chat-bot appears. And the most popular and robust way to do intention detection is rules-based method. With a detailed knowledge graph, the automatic intention classification could have a good performance. However, with the application cases becoming more and more complicated, it's hard to build such a knowledge graph for each cases. So in this project, we want to find out an automatic way to detect user's intention without pre-knowledge. After cleaning the original data and linking request and response together in a multi-turn communication, we implemented clustering on response first, which is more structured than request, and used the clustering label as the user intention of request to train a classification model based on RNN.

Key words

Word2vector, Sentence Embedding, RNN, User intention

Introduction

In the past few years, natural language processing has been a hot topic both in research and industry. A great number of chat-bot appears in market providing chitchat, booking or other services, like Google Assistant, Alexa and Cortana. Some of them can easily handle single-turn conversation. Take Cortana as an example, you can ask her to sing a song or set am alarm for you few hours later. Those three applications have their own platform and entry. There are also some background service which can be linked to some social network platform like Wechat, Skype and etc., which can also provide various services.

And there are also some chat-bot built for customer service especially in some e-commerce platform. Some simple question like updated logistic information can be solved automatically, which saved a lot of time compared with traditional human server. So this is a market with huge potential in the future and a great number of companies has been worked on this field.

To build a chat-bot, the first step is to understand user intentions when queries come in. For customer service chat-bot, the most common way is to build a knowledge graph for a certain kinds of business and then leverage pattern recognition technology to find out the matched intention for a specific query. And there are many challenges to do intention classification, for example, different user have different descriptions about one question and usually, there are more than one intentions in one request. Besides, the original data without label can not be put in a supervised-learning model.

In our project, we plan to do it in three step. Firstly, we will define the user intention class and make some labeled data manually. Then we will try to create automatic label by clustering, with limited number of labelled data. In the last step, we will build the classification model in neural network.

User Intention Classification

Data Preparation

Our original data comes from the customer service platform of China Mobile. And it contains multi-round of communication between customer and server. And we need to combine the request and response together and filter out some useless part in our row data.

And then we extract the user intention in the following form based on more than 5000 piece of data. There are two levels of classification which are domain and user intention separately.

```
{
    "domain":"通话",
    "intention":[“通话失败”,"话费充值","通讯服务”,”话费查询”,”呼叫转移","信号异常”,”话费异常"],
}
{
    "domain":"套餐",
    "intention":["套餐升级","套餐介绍","套餐更改”,”套餐办理","套餐取消”,“套餐异常”],
}
{
    "domain":"流量",
    "intention":[“流量介绍”, “流量购买”, "流量领取","流量办理","流量使用","流量取消","流量查询”,”流量异常”,”活动领取”],
}
{
    "domain":"宽带",
    "intention":["宽带办理","宽带申请”,”宽带查询”],
}
{
    "domain":"漫游",
    "intention":["漫游办理”,”漫游查询","漫游取消"],
}
{
    "domain":"业务",
    "intention":["业务查询","业务退订”,”呼叫转移”,”短号服务”],
}
{
    "domain":”积分”,
    "intention”:[“积分抽奖”,”积分结算”],
}
```

Figure 1. Label rules

Based on this label rules, we add tags to around 100 piece of data to evaluate our model performance in the following steps.

request	domain	intention
你好，能帮我开通国际漫游吗	漫游	漫游办理
所以我的号每个月都要扣多少钱呢	通话	话费查询
你好，我想问问我这几天话费是怎么扣的？你好，我想问问我这几天话费是怎么扣的？	通话	话费查询
我想查下话费	通话	话费查询
请问现在每个月扣的钱是多少是只有刚刚那个畅听卡扣的钱吗	通话	话费查询
我要新增号码到家庭短号	业务	短号服务
要加三个号，直接发给你吗	业务	短号服务
我想设置呼叫转移我手机没电了能设置呼叫转移吗？	业务	呼叫转移
请问充50送15的前一万名，是当天的零点开始计算的吗	通话	话费充值
交话费，也是在支付宝怎么行？流量就不行了？	通话	话费充值
你们现在有充值100元=360元流量年包+送20元话费的活动吗	通话	话费充值
我想问一下为什么话费没得那么快8号刚充了50元现在就只剩4元了	通话	话费异常
我在7月抽中2元话费券，提示在8月5日至31日到广东移动手机营业厅APP我的卡包领取，怎么会没有了	通话	话费异常
所以费用比较乱搞不清楚到底怎么回事昨天查我有700多块钱的余额今天只要200多块钱余额怎么回事？	通话	话费异常
就是我手机无缘无故收到100块话费	通话	话费异常
我月租是这个月的11号扣，哪个流量也是按这个月扣的吗，要是没用完，。还是在月尾的时候就已经扣掉了	通话	话费异常
为什么我31号还有25元现在直接欠费停机了？	通话	话费异常
我在天天抽奖中抽中1G流量包0元办，怎么领取？怎样领取？	流量	活动领取
我7月份参加“小和要火”，那个活动获得赠送的流量什么时候到帐？说话呀？	流量	活动领取
昨天我微信充直了100元，什么时候送200M流量	流量	活动领取
我刚才用微信充话费怎么样没有流量送	流量	活动领取
你好我想问一下怎么现在老是摇不到E分	积分	积分抽奖
那积分每月都要清0吗？	积分	积分结算
请问WLAN有什么套餐	宽带	宽带办理
收到短信说有免费宽带，前段时间这里没有宽带覆盖，不知现在有没，深圳南山区四达大厦A座有没有	宽带	宽带查询
办理4G、500M流量怎么办理，	流量	流量办理
办理流量年包办理流量年包	流量	流量办理
10元1G聊7天怎么办理	流量	流量办理
那帮我开通一个8月份的500M的省外流量包吧，就8月份这个月呀！	流量	流量办理
360年费流量包，上个月已经申请了充100元送360元年费流量包了，这个还可以申请吗	流量	流量办理

Fighre 2. An overview of labeled data

Data Processing

The second step is word segmentation . There are many methods to do word segmentation, which can be mainly divide into two part: character-based approaches and word-based approach.(Schubert, 2004) In character-based approaches, certain number of characters are extracted from texts. And based on the number of characters, it can be further divided into single character-based approach and multi-character-based approaches which is known as N-gram. Character-based approach has some advantages like simplicity and easy to application, but sometimes it has poor sematic performance.

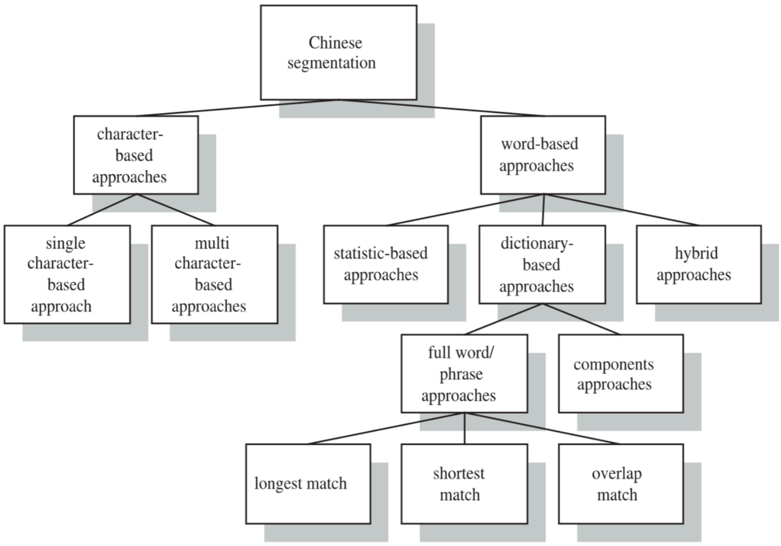

Figure 3. Chinese word segmentation

While word-based approach attempts to extract complete words from sentences and it have three methods, which are statistic-based, dictionary-based and hybrid-based. And dictionary-based method is commonly used in most segmentation system, which match the text with prior built dictionary. And it can be divided into long match and short match. Obviously, in this approach, the accuracy of segmentation depends on the completeness of the dictionary, and it's hard to get a complete dictionary in practice, which is the mainly drawback of dictionary-based method. In hybrid approaches, different approaches are combined to get a good segmentation performance, for example the combination of statistics-based method and word-based method. However, the time and space complexity could be very high when using hybrid approach.

In this project, Jieba is used to complete this work. Jieba is a Chinese word segmentation tools which can be implemented in various platform and support both simple and traditional Chinese segmentation. It has three types of segmentation mode: accurate mode as default mode, full mode and search engine mode. When running the segmentation, it uses word graph scanning based on the dictionary structure in dynamic programming, to find the most probable combination based on the word frequency. And for unknown words, a HMM-based model is used with the Viterbi algorithm.

Word Embedding

To build a language model, word embedding is an important part that could influent the final performance a lot. In this project, I use word2vector to represent the word in corpus. Word2vector is an model architectures for computing continues vector representations in a large dataset.(Mikolov, 2013) Unlike the traditional neural network architecture, Mikolov proposed two new architecture which is called Continuous Bag of Words Model and Continuous Skip-gram Model based on the neural net language model. Unlike the standard bag-of-words model, they use continuous distributed representation of the context. The architecture is as follows:

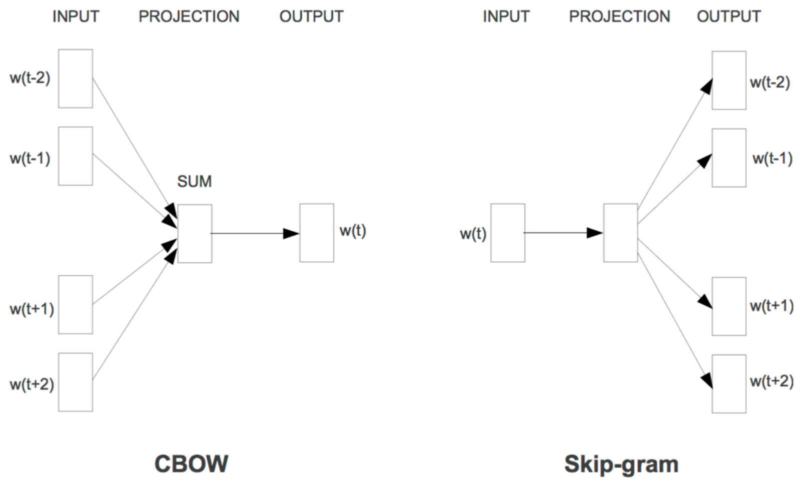

Figure 4. CBOW and Skip-gram

A Word2vec model can be trained with hierarchical softmax and/or negative sampling. To approximate the conditional log-likelihood a model seeks to maximize, the hierarchical softmax method uses a Huffman tree to reduce calculation. The negative sampling method, on the other hand, approaches the maximization problem by minimizing the log-likelihood of sampled negative instances. According to the authors, hierarchical softmax works better for infrequent words while negative sampling works better for frequent words and better with low dimensional vectors. As training epochs increase, hierarchical softmax stops being useful.

In our project, we firstly leverage two pre-trained word2vector model, and one built in character and the other built in word. The word3vector model based on character is a special case for Chinese text, which don't need word segmentation. After comparing the clustering performance, we find out that the output of character-based method group those sentences with similar structure together. And the word-based model will have better performance when the length of sentences is much different. So we choose word-based method in our project.

Sentence Embedding

In this project, we have tried different embedding method. As the input of prediction or classification model, sentence embedding is an important part in this project. And my work is focus on figuring out the best way of sematic representation using vector. The baseline model is average combine of word vector, where we simple add the word vector in this sentence together and calculate the average as the sentence vector. And then we tried another embedding method based on keywords, IF-IDF and encoder-decoder structure in RNN. And we use evaluate the embedding performance based on the clustering ouput.

Weighted-average using TF-IDF

This method is an improvement for baseline model, which multiply the word vector with the weight of each word. And we made an assumption that the word could be important to this sentence if it only appears here. So the weight is calculated by inverse sentence frequency like Inverse Document Frequency.

$$Weight(word) = \frac{p}{p + SF(wor\square)}$$

Where SF is how many times the word appears in other sentence, and p is a none-zero parameter.

After adding up the weighted word vector, the primary sentence vector is ready. And usually, some dimension could not contain so much information which can be reduced by Principle Component Analysis. There is a parameter named “PCA or not”, if the value is set as “False”, it will release the result without PCA.

The clustering performance is as follows. We can find out that when using inverse document frequency, the difference between sentences are reduced. Which means that the model is not suitable here.

Keywords-based model

In this method, we want improve the sentence embedding performance by adding weight to the keywords. In our label rule, there are 7 domains as the first level class. So firstly, we use the domain name as the keywords to group those sentences related to the same domain. And if the performance is good, we can use the keywords from the second level class to do hierarchical clustering.

In the first experiment, we use all first level keywords, and allocate weight of keyword as 2, which means that the importance in sentence represent is doubled compared with other words.

Keyword-based sentence embedding

```
Weight = 2
Sentence_lengths = len(sentence)
For word in sentences
    If word in keywords:
        Sentence_vector += weight * Word2vector(word)
    Else:
        Continue
Sentence_vector = Sentence_vector / sentence_length
```

And the performance is as follows:

And the clustering performance is good in the long tails. Like in cluster “0”, there are sentence related to package introduction. Cluster '3' is about package changing, while cluster '4' is about cellular data.

```
1 df_cluster[df_cluster.cluster ==0]
```

	Unnamed: 0	sessionid	request	response	cluster
1	1	13410000422T16082318212113APP	查询我现在的套餐	，您改了38元4G飞享套餐，9-13号生效哦，还在线吗，还有其他可以帮到您吗?	0
4	5	13410001228T160803125048	是的，我要求取消的昨天你们客服打电话来，我没接到	显示订单回退了，就是承诺使用38元套餐2个月的那个订单	0
13	20	13410001228T160807211011	能否帮我查下现在用的是不是六块月租的	，您现在是6元的月租的轻松卡的套餐的	0
27	42	13410006681T160808172432	我想把套餐升级 现在不是18一个月的吗我想换成38那个	2016年7月1日-8月31日，已实名无合约、合约三个月内到期且无套餐或者升档套餐全球通\预...	0
28	46	13410006681T160808172432	帮我查查是不是那个一年流量包就是一个月最低消费58对吧	目前2.3元正常情况是要用58的套餐的，但一般只有系统改套餐时能通过也可以改，但仍有最低消费	0

```
1 df_cluster[df_cluster.cluster ==3]
```

	Unnamed: 0	sessionid	request	response	cluster
36	56	13410006941T16081915080041APP	现在怎么把这个套餐取消而且我已经办理流量包了	，您的58元套餐是之前一直在使用的哟保持这个套餐继续用就可以了哦	3
62	149	13410013735T160818151135	今天开通的呀，说不要就直接取消	，只能更改套餐哦[微笑]	3
69	158	13410013735T160818151135	还有那个5元30m是全国的流量，还是就广东的?	那个套餐已经取消了哦	3
100	231	13410022435T16082712000050APP	帮我换下套餐 把现在的套餐取消	，要换什么套餐?	3
137	306	13410028479T160809204157	你好，我想质询下本机号是否开通了香港日漫游因为到现在还未有反应	，本机刚才办理了香港日套餐了	3
164	383	13410035760T16080509403099APP	我要取消58元套餐改用38元套餐	，请问取消的话，要改什么套餐呢?，对38元飞享套餐了解吗?	3

```
1 df_cluster[df_cluster.cluster ==4]
```

	Unnamed: 0	sessionid	request	response	cluster
65	153	13410013735T160818151135	好的，我还有多少流量可以查吗	的剩余流量 流量类型剩余总和省内4G体验流量1024MB省内2/3/4G灵活账期流量100M...	4
146	342	13410029511T16080311371717APP	为什么一扣话费，你们就直接扣了1024的流量? ? ?	省内2/3/4G流量1321MB国内2/3/4G流量1724MB仅供参考	4
151	351	13410030944T16080118152890APP	查流量查话费余额	本机剩余国内2/3/4G流量105MB省内2/3/4G流量200MB么么哒，请问还有什么可以...	4
155	362	13410034146T16083013112147APP	那能不能帮忙查询下为什么啊? 不然作为用户怎么放心	，确实是扣得流量费哦	4
167	390	13410036699T16080311435302APP	请问流量的使用有地区限制吗	经查你的流量只要在国内都能使用的	4
192	473	13410051086T16081518275017APP	是不是360块一年的流量	嗯 每个月有一个G流量	4
194	476	13410051658T16081416100209APP	集群网共享流量是什么意思?	就是您在网上买的流量	4

Figure 5. Clustering performance

Encoder-decoder model

Encoder-decoder structure is a common structure in sequence processing task like machine translation. And it will encode the input sequence into a fixed dimension vector and then a decoder will decode the vector into target sequence as showed in following.

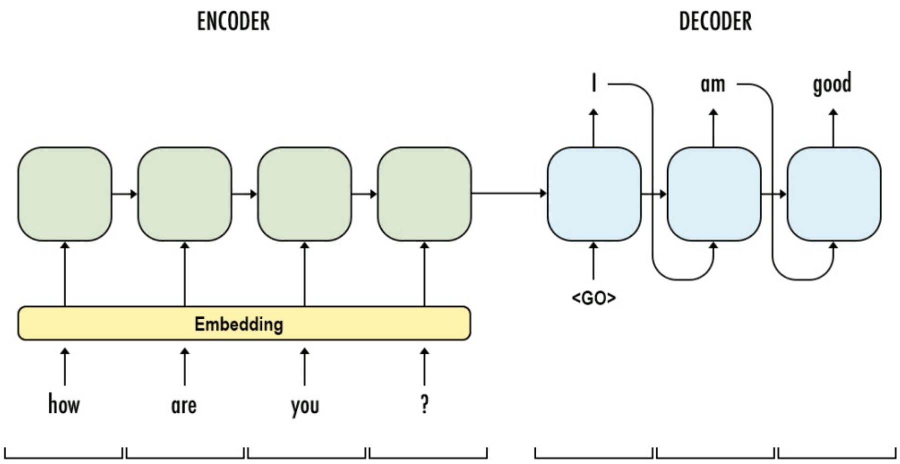

Figure 6. Encoder-Decoder structure

And there are several options for encoder and decoder, like CNN, RNN, LSTM and etc. Encoder and decoder don't need to be the same kind for network here.

In our project, we want to build a connection between word representation – word2vector and sentence itself. So we use the set of words vector as the input of the network and one-hot encoding of sentence as the output, using encoder-decoder model to build the relation which is represented by the vector in middle of this model. Then we implement clustering based on the sentence vector and use labeled data to evaluate the clustering performance.

And the clustering distribution using the four kinds of sentence embedding method is as follows. We can find out thart

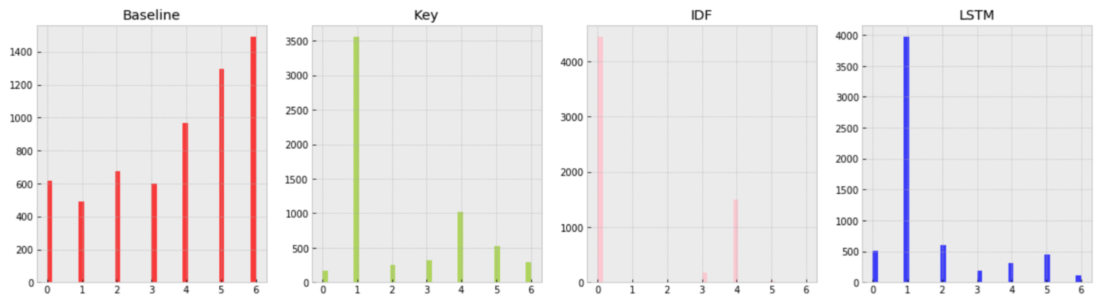

Figure 7. Clustering distribution

User Intention Classification

After clustering based on server response, we have labels for user request. And in this step, what we need is to build a model to identify the intention of new come request.

	Unnamed: 0	sessionid	request	response	cluster
0	0	13410000258T16081911402229APP	实名制了怎么还不能打电话?	请问是本机吗？我帮您开机，请稍等哈	0
1	1	13410000422T16082318212113APP	查询我现在的套餐	，您改了38元4G飞享套餐，9-13号生效哦，还在线吗，还有其他可以帮到您吗?	0
2	2	13410001228T160803125048	我前天办的飞享套餐，我要取消38元的	，请问是在什么时候办理的呢	0
3	4	13410001228T160803125048	电话办理的	您是不是之前要求取消了呢?	0
4	5	13410001228T160803125048	是的，我要求取消的昨天你们客服打电话来，我没接到	显示订单回退了，就是承诺使用38元套餐2个月的那个订单	5

Figure 8. Data overview

We use bidirectional RNN as the classification network. The network structure is as follows. We divide our data into training and testing set, and train this model use the different output of different sentence embedding method.

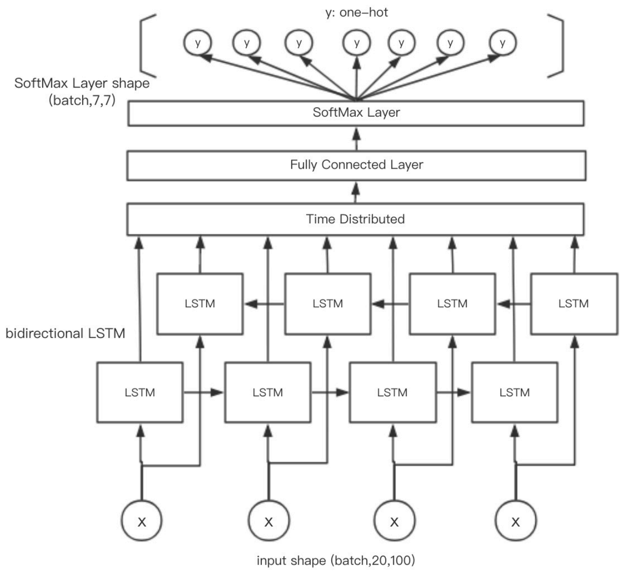

Figure 9. Classification network structure

And the classification precision of different model is as follows.

Table 1. Classification accuracy

Model	Training Accuracy	Testing Accuracy
Baseline model (Average)	0.441	0.372
Weighted Average based on IDF	0.625	0.586
Weighted Average based on Keywords	0.674	0.647
Encoder-decoder model	0.705	0.709

Conclusion

In this project, we use customer server's response to do intention detection by clustering and add it as a label to customer request. Then based on the labeled dataset, we trained a sequence network to make classification for new come request. And my work focus on different method of sentence embedding for intention clustering. A good representation of sentence can improve the clustering performance a lot. And in the last step, we compare the different classification performance which is trained on different clustering output. And based on the result, we can find out that the encoder-decoder model has good performance when doing sentence embedding.

Reference

Chelba, Ciprian, Milind Mahajan, and Alex Acero. "Speech utterance classification." Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 1. IEEE, 2003.

Chen, Yun-Nung, et al. "End-to-End Memory Networks with Knowledge Carryover for Multi-Turn Spoken Language Understanding." INTERSPEECH. 2016.

Chen, Yun-Nung, et al. "Matrix factorization with knowledge graph propagation for unsupervised spoken language understanding." Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). Vol. 1. 2015.

Haffner, Patrick, Gokhan Tur, and Jerry H. Wright. "Optimizing SVMs for complex call classification." Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). 2003 IEEE International Conference on. Vol. 1. IEEE, 2003.

Mesnil, Grégoire, et al. "Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding." Interspeech. 2013.

Acknowledge

Thanks for Prof. Lin to provide the chance of independent project linked with industry, during which I learned a lot about NLP and sentence embedding.

Thanks for Yilun Huang and Gunan Lu’s support in our project.

Thanks a lot!

Attachment: Meeting minutes

Date: Feb 7

Participants: Prof. Fangzhen Lin, all registers of the project.

Content:

- Introduction and intuition of the project
- Data source introduction and preview
- Q&A session.

Date: Feb 28

Participants: Prof. Fangzhen Lin, Huang Yilun, Lu Guannan, Mi Lan

Content:

- Project details discussions, including each steps of user intention classification, model candidates, automatic tagging using clustering
- Feedback from Prof.Lin: take rules into consideration, learn from IBM knowledge grapf building process, ensemble learning.

Date: Apr 9

Participants: Prof. Fangzhen Lin, Xiao I Robot, Lu Guannan, Mi Lan

Content:

- Middle report of our project.
- Prof.Lin and Xiao I Robot raised question about data cleaning and automatic tagging part of our project.
- Xiao I Robot introduced their methods to solve this problem in current stage

Date: May 11

Participants: Prof. Fangzhen Lin, Huang Yilun, Lu Guannan, Mi Lan

Content:

- Report the progress of the project. It mainly includes four parts: data preprocessing, word/sentence embedding, clustering, classification. We show our results of part 1-3. Prof. Lin carry out some ideas and questions of our solution.
- Discuss about how to enhance the performance of clustering (labeling).
- Prof. Lin emphasize the classification is the main part of the project. We should pay more attention on it.