1. 百度贴吧、微博社区'滴滴出行'网约车评论文本信息展示

```
df['comment'].head()
```

1	经常用，一款实用的 APP，希望能再多点优惠
2	并且打车的发票在手机端就可以直接开出，而且我觉得打车的补贴还是比较多的，滴滴打车的优势是方便...
3	滴滴出行是个不错的出行平台，最大限度的提高了出行者的权益，但在司机的准入上要求不严，以致连续...
4	滴滴就是出了事情或者责任就会逃避
5	很想给滴滴好评的，因为它确实方便我出行了，而且滴滴司机给我的印象也不错。无奈最近它的大众形象...

2.消费者关于网约车评论文本预处理

（1）文本去重

剔除重复的文本数据，加强数据的可用性，利用 python 的 drop_duplicates（）函数对 dataframe 结构进行去重处理

文本数据去重之前数据为 2517 条

```
data = pd.read_csv('E:\Desktop\semantic analysis\didicomment.csv',encoding='gbk')
data.shape[0]
```

```
2517
```

文本数据去重之后数据为 2482 条

```
data_last = data.drop_duplicates()
data_last.shape[0]
```

```
2482
```


（2）停用词剔除处理

加载库中的停用词表

```
def get_stopword_list():
    stop_word_path = './stopword.txt'#
    stopword_list = [sw.replace('\n','')for sw in open(stop_word_path,encoding='utf8').readlines()]
    return stopword_list
```

（3）干扰词过滤处理

过滤词中词以及长度<2 的

```
def word_filter(seg_list,pos=True):
    stopword_list = get_stopword_list()
    filter_list = []


    for seg in seg_list:
        if not pos:
            word = seg
            flag = 'n'
        else:
            word = seg.word
            flag = seg.flag
        if not flag.startswith('n'):
            continue

        if not word in stopword_list and len(word)>1:
            filter_list.append(word)
    return filter_list
```

3.基于语义网络的消费者网约车评论分析

加载需要处理的文本数据，利用 Python 的 jieba 库进行分词处理

文本数据分词部分结果:

[['一款', '实用', 'APP', '希望', '多点', '优惠'],['打车','发票','手机',开出','打车','补贴','滴滴','打车','优势','方便快捷','一竿子','推翻','滴滴','社会','行业','贡献','网友','历史','出门在外','流量','变得','完善','支付','方式','一会','专车','来接','滴滴','车子','外出','遭受','侵害','例子','确实','出行','价格','明朗','十点','县城','打]

```
def seg_to_list(sentence,pos=False):
    if not pos:
        seg_list = jieba.cut(sentence)
    else:
        seg_list = psg.cut(sentence)
    return seg_list
```


4.利用 Ucient6 软件对文本数据进行词频分析

将文本数据处理成以分号分隔的格式，再用 Ucient6 软件生成词频数，取其中 top20 生成 20x20 的共现矩阵，top20 词频数如下:

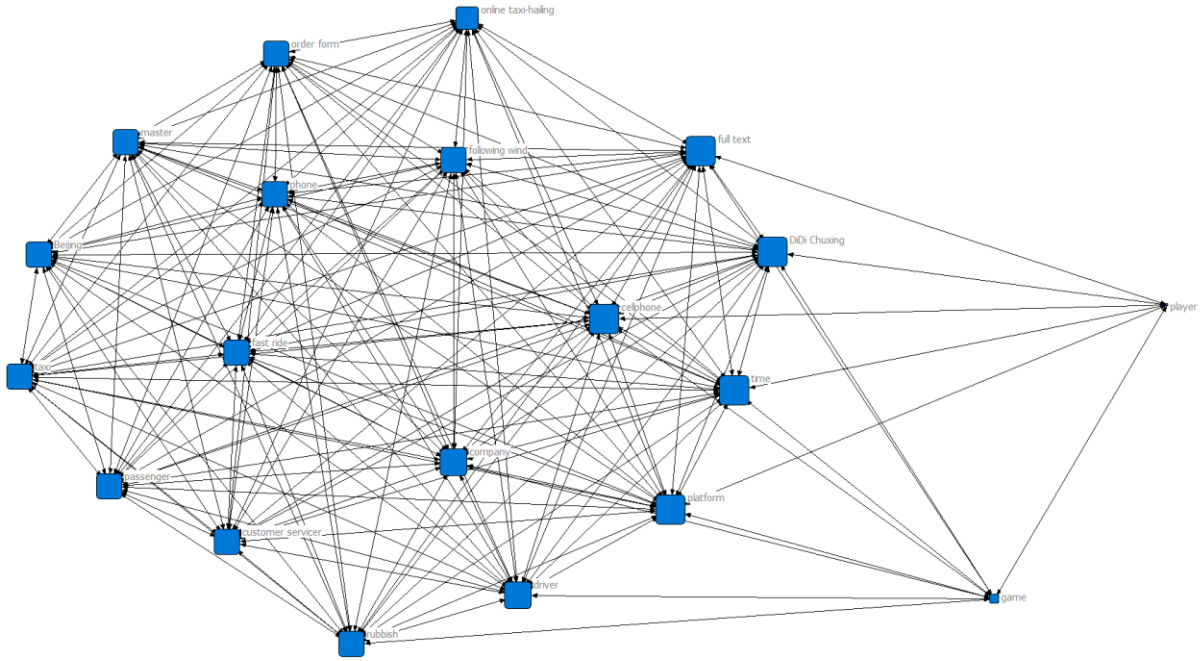
1	DiDI Chuxing	5415
2	driver	1670
3	passenger	584
4	customer servicer	530
5	game	509
6	company	507
7	full text	408
8	platform	408
9	master	227
10	fast ride	220
11	phone	216
12	player	210
13	following wind	209
14	cellphone	207
15	order form	204
16	time	190
17	taxi	189
18	online taxi-hailing	172
19	rubbish	169
20	Beijing	159

Top20 关键词生成的 20x20 共现矩阵:

		DiDi Chuxir	driver	passenger	customer servicer	game	company	platform	full text	master	fast ride
2	DiDi Chuxir	0	757	295	288	261	270	232	408	138	171
3	driver	757	0	229	159	1	134	140	208	68	89
4	passenger	295	229	0	56	0	72	74	90	24	39
5	customer s	288	159	56	0	0	54	59	105	21	51
6	game	261	1	0	0	0	1	7	59	0	0
7	company	270	134	72	54	1	0	38	73	10	30
8	platform	232	140	74	59	7	38	0	88	17	34
9	full text	408	208	90	105	59	73	88	0	38	53
10	master	138	68	24	21	0	10	17	38	0	11
11	fast ride	171	89	39	51	0	30	34	53	11	0
12	phone	146	103	42	82	0	30	32	71	17	25
13	player	157	0	0	0	133	0	2	43	0	0
14	following w	101	35	23	16	0	11	22	29	4	17
15	cellphone	90	57	24	28	5	9	13	38	14	12
16	order form	129	90	44	46	0	21	38	50	18	18
17	time	133	63	33	31	14	20	20	47	8	15
18	taxi	119	58	20	9	0	24	18	24	14	24
19	online taxi-	90	39	20	4	0	25	26	43	4	8
20	rubbish	94	52	17	31	1	30	24	13	8	4
21	Beijing	98	32	10	9	0	13	14	34	8	8

fast ride	phone	player	following wind	cellphone	order form	time	taxi	online taxi-hailing	rubbish	Beijing
171	146	157	101	90	129	133	119	90	94	98
89	103	0	35	57	90	63	58	39	52	32
39	42	0	23	24	44	33	20	20	17	10
51	82	0	16	28	46	31	9	4	31	9
0	0	133	0	5	0	14	0	0	1	0
30	30	0	11	9	21	20	24	25	30	13
34	32	2	22	13	38	20	18	26	24	14
53	71	43	29	38	50	7	24	43	13	34
11	17	0	4	1	1	8	14	4	8	8
0	25	0	17	1	1	15	24	8	4	8
25	0	0	13	3	32	26	9	3	11	9
0	0	0	0	4	0	10	0	0	0	0
17	13	0	0	4	9	18	12	12	7	12
12	25	4	4	0	10	15	7	4	5	6
18	32	0	9	10	0	22	10	7	14	3
15	26	10	18	15	22	0	11	7	3	11
24	9	0	12	7	10	11	0	11	5	6
8	3	0	12	4	7	7	11	0	0	13
4	11	0	7	5	14	3	5	0	0	3
8	9	0	12	6	3	11	6	13	3	0

将 Ucient6 生成的 20x20 共现矩阵导入 Netdraw 中进行可视化节点中心度分析，形成网约车在线评论语义网络图



6.基于凝聚子群的滴滴网约车客户评论关键词 CONCOR 分析

将从数据中提取的共现矩阵导入 Ucient6 软件进行 CONCOR 分析，形成聚类图表如下：

Topic 1

Class1:DiDi Chuxing

Class2:platform、passenger、customer service、order form、phone、rubbish、cellphone

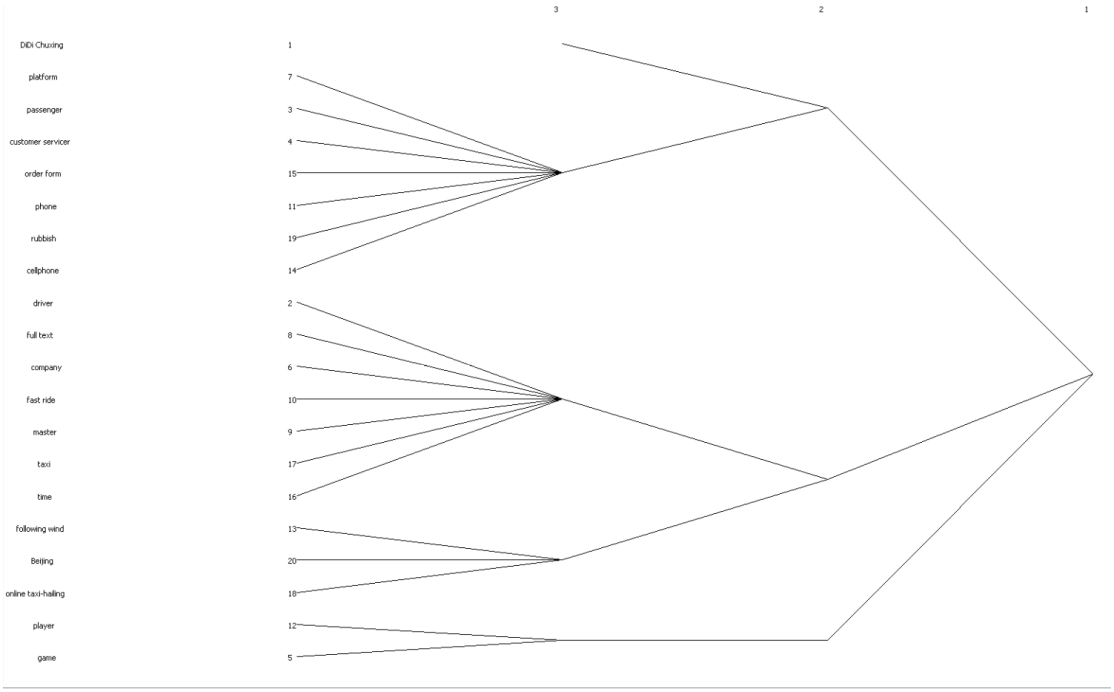
Topic2

Class1:driver、full text、company、fast ride、master、taxi、time

Class2:following wind、Beijing、online taxi-hailing

Topic3

player、game



形成的相关系数矩阵如下:

	DiDi	drive	passe	custo	game	compa	platf	full	maste	fast	phone	playe	follo	cellp	order	time	taxi	onlin	rubbi	Beiji
DiDi Chuxing	1.00	0.74	0.92	0.83	0.05	0.90	0.90	0.96	0.88	0.88	0.83	0.15	0.72	0.88	0.87	0.90	0.84	0.74	0.82	0.74
driver	0.74	1.00	0.99	0.97	0.78	0.99	0.99	0.98	0.98	0.99	0.93	0.63	0.97	0.96	0.98	0.98	0.97	0.93	0.95	0.95
passenger	0.92	0.99	1.00	0.96	0.61	0.96	0.98	0.95	0.96	0.96	0.93	0.47	0.91	0.95	0.97	0.94	0.95	0.91	0.95	0.90
customer servicer	0.83	0.97	0.96	1.00	0.68	0.97	0.97	0.97	0.97	0.98	0.98	0.54	0.94	0.98	0.98	0.97	0.96	0.92	0.94	0.94
game	0.05	0.78	0.61	0.68	1.00	0.72	0.67	0.76	0.74	0.69	0.56	1.00	0.76	0.67	0.60	0.76	0.72	0.71	0.64	0.79
company	0.90	0.99	0.96	0.97	0.72	1.00	0.99	0.98	0.99	0.98	0.92	0.57	0.97	0.96	0.96	0.97	0.97	0.94	0.96	0.95
platform	0.90	0.99	0.98	0.97	0.67	0.99	1.00	0.97	0.98	0.98	0.95	0.54	0.95	0.97	0.98	0.97	0.95	0.94	0.94	0.93
full text	0.96	0.98	0.95	0.97	0.76	0.98	0.97	1.00	0.98	0.98	0.93	0.65	0.95	0.97	0.95	0.99	0.96	0.93	0.96	0.95
master	0.88	0.98	0.96	0.97	0.74	0.99	0.98	0.98	1.00	0.98	0.92	0.59	0.96	0.97	0.95	0.98	0.97	0.92	0.93	0.96
fast ride	0.88	0.99	0.96	0.98	0.69	0.98	0.98	0.98	0.98	1.00	0.95	0.54	0.96	0.97	0.97	0.98	0.96	0.93	0.96	0.95
phone	0.83	0.93	0.93	0.98	0.56	0.92	0.95	0.93	0.92	0.95	1.00	0.43	0.87	0.96	0.97	0.93	0.86	0.84	0.92	0.85
player	0.15	0.63	0.47	0.54	1.00	0.57	0.54	0.65	0.59	0.54	0.43	1.00	0.60	0.54	0.45	0.64	0.57	0.57	0.50	0.64
following wind	0.72	0.97	0.91	0.94	0.76	0.97	0.95	0.95	0.96	0.96	0.87	0.60	1.00	0.92	0.92	0.96	0.96	0.94	0.90	0.98
cellphone	0.88	0.96	0.95	0.98	0.67	0.96	0.97	0.97	0.97	0.97	0.96	0.54	0.92	1.00	0.98	0.97	0.92	0.88	0.91	0.91
order form	0.87	0.98	0.97	0.98	0.60	0.96	0.98	0.95	0.95	0.97	0.97	0.45	0.92	0.98	1.00	0.95	0.92	0.88	0.94	0.89
time	0.90	0.98	0.94	0.97	0.76	0.97	0.97	0.99	0.98	0.98	0.93	0.64	0.96	0.97	0.95	1.00	0.95	0.93	0.93	0.96
taxi	0.84	0.97	0.95	0.96	0.72	0.97	0.95	0.96	0.97	0.96	0.86	0.57	0.96	0.92	0.92	0.95	1.00	0.93	0.93	0.96
online taxi-hailing	0.74	0.93	0.91	0.92	0.71	0.94	0.94	0.93	0.92	0.93	0.84	0.57	0.94	0.88	0.88	0.93	0.93	1.00	0.88	0.96
rubbish	0.82	0.95	0.95	0.94	0.64	0.96	0.94	0.96	0.93	0.96	0.92	0.50	0.90	0.91	0.94	0.93	0.93	0.88	1.00	0.88
Beijing	0.74	0.95	0.90	0.94	0.79	0.95	0.93	0.95	0.96	0.95	0.85	0.64	0.98	0.91	0.89	0.96	0.96	0.96	0.88	1.00