In [ ]:

```
```

预处理

In [1]:

```
import jieba
import jieba.posseg
import numpy as np
import pandas as pd
import math
import operator
```

获得停用词列表

In [2]:

```
# 获取停用词
def Stop_words():
    stopwords = []
    data = []
    f = open('./data/stopword.txt', encoding='utf8')
    for line in f.readlines():
        data.append(line)
    for i in data:
        output = str(i).replace('\n', '')
        stopwords.append(output)
    return stopwords

stopwords=Stop_words()
print('停用词表的大小为：',len(stopwords))
```

```
停用词表的大小为： 767
```

对文档集过滤词性和停用词

In [3]:

```
# 加载文档集，对文档集过滤词性和停用词
def Filter_words(data_path,stopwords):
    documents = []
    for line in open(data_path, 'r', encoding='utf8'):
        segment = jieba.posseg.cut(line.strip())
        filter_words = []

        for word, flag in segment:
            if flag.startswith('n') is False:
                continue
            if  word not in stopwords and len(word) > 1:
                filter_words.append(word)
        documents.append(filter_words)
    return documents

data_path = './data/corpus4keyword.txt'
documents=Filter_words(data_path,stopwords)

print('文档的数量为',len(documents))
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\fengl\AppData\Local\Temp\jieba.cache
```

```
Loading model cost 0.458 seconds.
Prefix dict has been built successfully.
文档的数量为 164
```

TF-IDF抽取关键词

In [4]:

```
# TF-IDF 算法
def tf(filter_sent):
    # 统计TF值
    tf_dict = {}
    for word in filter_sent:
        if word not in tf_dict:
            tf_dict[word] = 1
        else:
            tf_dict[word] += 1
    for word in tf_dict:
        tf_dict[word] = tf_dict[word] / len(filter_sent)
    return tf_dict

def idf(documents):
    # 统计IDF值
    idf_dict = {}
    for doc in documents:
        for word in set(doc):
            if word not in idf_dict:
                idf_dict[word] = 1
            else:
                idf_dict[word] += 1
    return idf_dict

def tf_idf(documents):
    # 计算TF-IDF值

    idf_dict=idf(documents)
    print('idf词典中词的个数为：',len(idf_dict))

    for filter_sent in documents:
        tf_idf_dict = {}
        tf_dict=tf(filter_sent)
        for word in tf_dict:
            tf_idf_dict[word] = tf_dict[word] * math.log(len(documents) / (idf_dict[w

        # 提取前10个关键词
        topk = 10
        sort_kwords=sorted(tf_idf_dict.items(), key=operator.itemgetter(1),reverse=Tr

        kword_list=[kword for kword,value in sort_kwords[:topk]]

        print('TF-IDF模型结果：','/ '.join(kword_list))

tf_idf(documents)
```

```
idf词典中词的个数为： 4064
TF-IDF模型结果： 地铁/ 坐票/ 座位/ 车厢/ 行李/ 老街/ 轨道交通/ 赵鹏林/ 大件/ 档次
TF-IDF模型结果： 同心县/ 核心区/ 春暖迟/ 秋凉/ 少雨/ 风大沙/ 沙尘暴/ 干热风/ 霜冻/ 冰
雹
TF-IDF模型结果： 永康/ 病痛/ 出院/ 病情/ 医院/ 疝气/ 腹股沟/ 疝嵌顿及/ 肠梗阻/ 监护室
TF-IDF模型结果： 康师傅/ 茶叶/ 商行/ 废料/ 商家/ 吉安/ 饲料/ 枕头/ 刘俊/ 名茶
TF-IDF模型结果： 咨询电话/ 邮箱/ 项目/ 报名表/ 资助/ 方案/ 办法/ 方式
TF-IDF模型结果： 全民/ 血压/ 拉开帷幕/ 年度/ 主题/ 方式
TF-IDF模型结果： 节目/ 中央电视台/ 服务项目/ 电视/ 哪些项目/ 项目/ 频道/ 资助/ 大家/
慈善
TF-IDF模型结果： 图书/ 学生/ 童话/ 标致/ 东风/ 孩子/ 小学/ 乡村/ 志愿者/ 武汉市
```

TF-IDF模型结果： 纪念/ 国际/ 汶川/ 索尔/ 弗利/ 联合会/ 中国红十字会/ 战役/ 红新月会/ 人道
TF-IDF模型结果： 世间/ 大赛/ 内心/ 艾滋病/ 高校/ 主题/ 中国
TF-IDF模型结果： 阿坝州/ 茂县/ 羌族/ 合唱团/ 小学/ 四川省阿坝藏族羌族自治州/ 凤仪/ 凤仪镇/ 团员/ 商议
TF-IDF模型结果： 汽车/ 文明/ 作品名称/ 作者/ 事故/ 情景/ 短片/ 驾车/ 中山大学/ 女生
TF-IDF模型结果： 小学/ 学校/ 兰州/ 运动鞋/ 白银/ 爱心/ 组委会/ 甘肃/ 山区/ 临夏
TF-IDF模型结果： 野草/ 全额/ 所有人/ 一分钱/ 成都/ 策划/ 象征性/ 水平/ 市民/ 项目
TF-IDF模型结果： 晋江市/ 城市/ 大会/ 爱心/ 中华/ 基金会/ 许嘉璐/ 巡视员/ 重庆市/ 人大常委会
TF-IDF模型结果： 奶粉/ 光明/ 菌落/ 合格/ 南山/ 工商部门/ 总数/ 沙湾/ 配方/ 产品
TF-IDF模型结果： 中国儿童少年基金会/ 日照/ 安康/ 人为/ 搜狐/ 灾区/ 儿童节/ 娃娃/ 山东/ 家园
TF-IDF模型结果： 流浪/ 人员/ 灾民/ 条例/ 住宿/ 本市/ 基本/ 居民/ 异地/ 职业
TF-IDF模型结果： 马来西亚/ 国籍/ 无国籍/ 无法/ 华人/ 人球/ 无权/ 人权/ 英国/ 基本
TF-IDF模型结果： 指标/ 直饮水/ 水机/ 专人/ 毒理学/ 放射性/ 感官/ 性状/ 小区/ 水质
TF-IDF模型结果： 父亲/ 老人/ 儿女/ 香烛/ 病危/ 大街/ 道德/ 有点/ 食物/ 公众
TF-IDF模型结果： 水流/ 财经频道/ 栏目/ 困局/ 极端/ 利益冲突/ 案例
TF-IDF模型结果： 头枕/ 车窗/ 汽车/ 保险/ 缝隙/ 窗玻璃/ 弹簧/ 撞针/ 工具/ 金属杆
TF-IDF模型结果： 女婴/ 心脏/ 手术/ 医院/ 大院/ 法式/ 动脉/ 导管/ 大动脉/ 安静
TF-IDF模型结果： 老年人/ 敬老/ 爱老/ 老龄/ 老人/ 委员会/ 中信银行/ 吴玉韶/ 养老院/ 敬老院
TF-IDF模型结果： 资产/ 黄怒波/ 计划/ 官办/ 信任/ 商铺/ 过户/ 双方/ 慈善机构/ 办理
TF-IDF模型结果： 百姓/ 企业/ 经济/ 绿色/ 技术/ 明星/ 国家/ 个人/ 政协会议/ 分组讨论
TF-IDF模型结果： 关系/ 医疗保险/ 年限/ 养老保险/ 纲要/ 人社部/ 规划/ 制度/ 医疗保障/ 办法
TF-IDF模型结果： 运动鞋/ 人民币/ 网友/ 孩子/ 刘春/ 爱心/ 批量/ 献爱心/ 总编辑/ 崔永元
TF-IDF模型结果： 植树/ 金山岭/ 上山/ 生长/ 绿化/ 义务/ 网友/ 植下/ 许愿树/ 顺利开展
TF-IDF模型结果： 餐具/ 消费者/ 餐饮企业/ 南京/ 人们/ 部分/ 商业道德/ 陷阱/ 杯具/ 安全感
TF-IDF模型结果： 农民工/ 兄弟/ 崔永元/ 饭局/ 慈善/ 残疾人/ 救人/ 礼物/ 现场/ 搜狐
TF-IDF模型结果： 强奸/ 妇女/ 律师/ 奸淫幼女/ 当众/ 公共场所/ 暴力/ 网友/ 环球时报/ 英文版
TF-IDF模型结果： 孩子/ 潘小胡/ 大山/ 土豆/ 老师/ 校长/ 铁索桥/ 学校/ 甘洛县/ 牛吾村
TF-IDF模型结果： 血液/ 血站/ 献血者/ 中心/ 各省市/ 刘江/ 团体/ 参观/ 志愿者/ 价格
TF-IDF模型结果： 女性/ 妇女/ 公正/ 性别/ 联合国/ 报告/ 旗舰/ 宪法/ 睁大眼睛/ 点滴
TF-IDF模型结果： 捐款人/ 问用/ 钱用/ 现象/ 地方
TF-IDF模型结果： 公共场所/ 烟草/ 法规/ 场所/ 吸烟者/ 法律法规/ 疾病/ 规定/ 报告/ 图形
TF-IDF模型结果： 索尼/ 交流/ 学生/ 访日之旅/ 回国/ 旅程/ 晚宴/ 中华全国青年联合会/ 全国青联/ 友好使者
TF-IDF模型结果： 瑞丽/ 阳光/ 杂志社/ 弱势/ 儿童/ 基金/ 影响力/ 价值观/ 专项基金/ 河北
TF-IDF模型结果： 日照/ 安康/ 人为/ 搜狐/ 灾区/ 儿童节/ 娃娃/ 山东/ 家园/ 礼物
TF-IDF模型结果： 济慈/ 家具/ 思源/ 焦点/ 焦点网/ 基金/ 物资/ 曲美/ 粟凡/ 福利事业
TF-IDF模型结果： 运动鞋/ 孩子/ 孙冕/ 健康成长/ 发票/ 穿鞋/ 中西部/ 贫困山区/ 贫困地区/ 长征
TF-IDF模型结果： 地铁/ 杭州/ 守则/ 市民/ 乘客/ 集团/ 杭州市/ 折叠车/ 盲人/ 导盲犬
TF-IDF模型结果： 书包/ 基会/ 新生/ 京华/ 故事/ 时报/ 搜狐/ 北京/ 中国/ 公益
TF-IDF模型结果： 养老/ 退休年龄/ 网民/ 缺口/ 民众/ 公平/ 问题/ 养老保险/ 制度/ 人力资源
TF-IDF模型结果： 养老金/ 缺口/ 网民/ 双轨制/ 个人账户/ 人口老龄化/ 账户/ 问题/ 养老/ 养老保险
TF-IDF模型结果： 基金会/ 总收入/ 资金/ 秘密/ 卫生局/ 商议/ 副理事长/ 公益事业/ 公司/ 年度
TF-IDF模型结果： 联合国开发计划署/ 联合国/ 愿景/ 先生/ 潘基文/ 协调员/ 罗黛琳/ 周迅/ 画册/ 邵忠
TF-IDF模型结果： 石油/ 蓝星/ 小清河/ 油污/ 污水/ 气味/ 柴油/ 梁王河/ 漏油/ 济南市
TF-IDF模型结果： 工作日/ 志愿/ 个人信息/ 通信地址/ 管理工作/ 电子邮箱/ 特长/ 技能/ 鞋子/ 姓名
TF-IDF模型结果： 孤儿/ 费用/ 地址/ 中国扶贫基金会/ 心理/ 项目/ 执行机构/ 海淀区/ 双榆树/ 小灵通
TF-IDF模型结果： 盖茨/ 现场/ 创办人/ 好友/ 股神/ 巴菲特/ 二人/ 财富/ 论坛/ 视频
TF-IDF模型结果： 公益活动/ 总支出/ 孩子/ 善款/ 总结
TF-IDF模型结果： 艾滋病/ 红丝/ 中华/ 全国工商联/ 艾滋/ 仪式/ 家园/ 情系/ 主席/ 工人
TF-IDF模型结果： 李冰冰/ 健康成长/ 运动鞋/ 孩子/ 亲笔签名/ 穿鞋/ 中西部/ 贫困山区/ 贫困地区/ 长征
TF-IDF模型结果： 孤儿/ 费用/ 地址/ 中国扶贫基金会/ 心理/ 项目/ 执行机构/ 海淀区/ 双榆树/ 小灵通
TF-IDF模型结果： 志愿/ 历史进程/ 团中央/ 社区服务/ 成人/ 社会公众/ 事业/ 环境保护/ 小时/ 志愿者

TF-IDF模型结果： 创新奖/ 世界/ 韩国/ 代表团/ 公益/ 组委会/ 成员/ 中国科学院/ 文献/ 情报中心
TF-IDF模型结果： 油脂/ 油水/ 分离器/ 地沟油/ 餐馆/ 北京市/ 餐厨/ 企业/ 市政/ 垃圾
TF-IDF模型结果： 房屋/ 新源县/ 和静县/ 地震/ 有序/ 新疆维吾尔自治区/ 伊犁哈萨克自治州/ 昭苏县/ 特克斯县/ 察布查尔锡伯自治县
TF-IDF模型结果： 幼儿园/ 小伊伊/ 事故/ 项余遇/ 动车/ 幸存者/ 左腿/ 温州/ 叔叔/ 女孩
TF-IDF模型结果： 孩子/ 儿童节/ 爸妈/ 礼物/ 孤儿/ 棉花/ 游乐场/ 幻想/ 公主/ 小王子
TF-IDF模型结果： 贵阳市/ 规模/ 群众/ 标准/ 命案/ 嫌疑人/ 贵阳/ 基金/ 市委书记/ 风尚
TF-IDF模型结果： 方向/ 申报/ 领域/ 缺水/ 水系/ 项目/ 主办方/ 江河/ 个人/ 问题
TF-IDF模型结果： 车门/ 汽车/ 司机/ 郑州/ 水位/ 侧窗/ 车窗/ 车厢/ 无法/ 过程
TF-IDF模型结果： 地球/ 全球/ 小时/ 时区/ 网友/ 城市/ 搜狐/ 时间/ 北京/ 公益
TF-IDF模型结果： 西藏/ 阳光工程/ 基金/ 委员会/ 农牧民/ 太阳能/ 领导/ 资金/ 家庭/ 宗旨
TF-IDF模型结果： 公益活动/ 志愿/ 助医/ 民族/ 出生日期/ 电子邮箱/ 特长/ 地址/ 姓名/ 性别
TF-IDF模型结果： 垃圾/ 产生/ 规划/ 垃圾处理/ 设施/ 农村/ 地下水/ 速度慢/ 负荷/ 镇村
TF-IDF模型结果： 退休年龄/ 劳动力/ 人口/ 老年人/ 建议/ 研究所/ 所长/ 劳动者/ 年龄/ 老龄化
TF-IDF模型结果： 垃圾/ 垃圾袋/ 分类/ 计量/ 收费/ 李廷贵/ 广州/ 模式/ 台北/ 厨余
TF-IDF模型结果： 池塘/ 救人/ 小弟弟/ 小男孩/ 女儿/ 萧山/ 前镇/ 韩老师/ 女孩/ 爸爸
TF-IDF模型结果： 老龄化/ 中国/ 问题/ 两鬓/ 爱尔兰/ 诗人/ 叶芝/ 爱情/ 态势/ 催债
TF-IDF模型结果： 运动鞋/ 小学/ 鞋款/ 象山/ 网友/ 总额/ 善款/ 学校/ 孩子/ 爱心
TF-IDF模型结果： 童童/ 肛门/ 先天性/ 心脏病/ 手术/ 酒店/ 后门/ 外表/ 常人/ 被遗弃
TF-IDF模型结果： 魏先生/ 男子/ 女子/ 谢谢/ 黄衣/ 儿子/ 靳鸽/ 何杰/ 衣角/ 妻儿
TF-IDF模型结果： 水井坊/ 技艺/ 水井街/ 酒坊/ 遗址/ 先生/ 酒厂/ 白酒/ 陈剑/ 和古
TF-IDF模型结果： 中华环境保护基金会/ 李伟/ 报社/ 副社长/ 李瑞农/ 共青团/ 中国人民大学/ 乔昆/ 做客/ 书记
TF-IDF模型结果： 范松池/ 松山/ 部队/ 炮兵/ 战役/ 远征军/ 神炮手/ 时刻/ 主堡/ 碉堡
TF-IDF模型结果： 利益/ 志愿者/ 志愿/ 部门/ 长效机制/ 研究生/ 变味/ 激励机制/ 经历/ 入学
TF-IDF模型结果： 艾滋病/ 红丝/ 中华/ 全国工商联/ 艾滋/ 仪式/ 家园/ 情系/ 主席/ 工人
TF-IDF模型结果： 大学生/ 绿色/ 偶像/ 中心/ 评价/ 心目/ 榜样/ 远洋/ 地产/ 美国
TF-IDF模型结果： 血液/ 常态/ 郭燕红/ 献血者/ 卫生部/ 人次/ 成本/ 水平/ 人口/ 手术
TF-IDF模型结果： 苍南县/ 福利院/ 孤儿/ 保育员/ 铁链/ 事情/ 苍南/ 县政府/ 护工/ 小群
TF-IDF模型结果： 创新奖/ 世界/ 韩国/ 代表团/ 公益/ 组委会/ 成员/ 中国科学院/ 文献/ 情报中心
TF-IDF模型结果： 指数/ 因素/ 心理压力/ 幸福感/ 家庭收入/ 毕业生/ 夫妻/ 年龄/ 家庭/ 本市
TF-IDF模型结果： 地铁/ 网友/ 上海/ 女性/ 抗议/ 女权/ 常识/ 女人/ 争议/ 官方
TF-IDF模型结果： 饮用水/ 水源/ 水源地/ 委员/ 法律/ 部门/ 建议/ 机制/ 香港/ 工程
TF-IDF模型结果： 白领/ 公益/ 群体/ 主题/ 春夏秋冬/ 花会/ 时尚/ 聚会/ 高达/ 社会
TF-IDF模型结果： 广发/ 贫困家庭/ 祝福/ 云南/ 慈善/ 志愿者/ 代表/ 楚雄彝族自治州/ 学习用品/ 书架
TF-IDF模型结果： 人口/ 菲律宾/ 全球/ 丹妮/ 卡马乔/ 聚光灯/ 马尼拉/ 两极/ 彭希哲/ 零点
TF-IDF模型结果： 索尼/ 交流/ 中国/ 主题/ 植根/ 社会/ 索尼公司/ 高中生/ 意义/ 贡献
TF-IDF模型结果： 手术/ 爱纯/ 室间隔/ 肺动脉/ 右心室/ 精心/ 笑容/ 心脏/ 出院/ 小女孩
TF-IDF模型结果： 艾滋病/ 规划署/ 世界/ 联合国/ 企业/ 主题/ 合作伙伴/ 礼包/ 手套/ 围巾
TF-IDF模型结果： 象牙/ 国人/ 私带/ 制品/ 入境/ 投票/ 获奖者/ 眼霜/ 卫士/ 金额
TF-IDF模型结果： 伊拉克/ 波及/ 巴格达/ 红新月/ 章程/ 调解者/ 公正/ 武装冲突/ 红十字国际委员会/ 红新月会
TF-IDF模型结果： 小学/ 乡村/ 广东省/ 标致/ 东风/ 世纪/ 经济/ 条件/ 孩子/ 学校
TF-IDF模型结果： 爱鸟周/ 生态/ 野生动物/ 观鸟/ 全国/ 两国间/ 候鸟/ 陆生/ 中国野生动物保护协会/ 鸟网
TF-IDF模型结果： 尹成基/ 人社部/ 养老保险/ 农民工/ 问题/ 城镇/ 退休年龄/ 养老金/ 总体/ 政策
TF-IDF模型结果： 大熊猫/ 种群/ 自然保护区/ 四川省/ 四川/ 缅甸/ 越南/ 一带/ 熊猫/ 栖息地
TF-IDF模型结果： 周岩/ 法院/ 民事/ 陶汝坤/ 附带/ 李智贤/ 合肥市/ 法庭/ 代理律师/ 精神
TF-IDF模型结果： 运动鞋/ 孩子/ 健康成长/ 项目/ 基金会/ 穿鞋/ 潘石屹/ 中西部/ 贫困山区/ 董事长
TF-IDF模型结果： 慰问金/ 村民/ 红十字会/ 房山/ 米面/ 烈士/ 群众/ 红会/ 门头沟/ 灾民
TF-IDF模型结果： 作品/ 音乐/ 群众/ 音乐网/ 参赛/ 参赛者/ 爱好者/ 官网/ 机会/ 组委会
TF-IDF模型结果： 条例/ 列支/ 信息/ 规定/ 备案/ 胡增耆/ 比例/ 亮点/ 方案/ 上海
TF-IDF模型结果： 教师/ 乡村/ 教学/ 崔永元/ 冷酸灵/ 飞翔/ 中国国际广播电台/ 结业/ 倒数/ 爱迪
TF-IDF模型结果： 自行车/ 租车/ 一卡通/ 市民/ 北京市/ 试运营/ 功能/ 小时/ 交通/ 东城区
TF-IDF模型结果： 蔡屯/ 粉笔/ 精彩/ 小学/ 老师/ 孩子/ 互联网/ 课程/ 支教/ 河南
TF-IDF模型结果： 都灵/ 金三银/ 二金四银/ 冬奥/ 军团/ 奖牌/ 成绩/ 韩国/ 历史/ 中国
TF-IDF模型结果： 男子/ 女子
TF-IDF模型结果： 包子/ 老公/ 早饭/ 主题

```
TF-IDF模型结果：  姑姑/ 儿子/ 媳妇/ 身体/ 极差/ 姑父/ 回老家/ 样子/ 脑袋/ 有点
TF-IDF模型结果：  营养/ 孩子/ 饮料/ 家长/ 方便面/ 早餐/ 维生素/ 错误/ 牛奶/ 习惯
TF-IDF模型结果：  市场开发部/ 袁斌/ 搜狐公司/ 首席/ 财务/ 余楚媛/ 部长/ 北京
TF-IDF模型结果：  蒋效愚/ 北京奥运/ 官方网站/ 主席/ 北京
TF-IDF模型结果：  食品/ 农药/ 马志英/ 菜叶/ 危害/ 市民/ 虫眼/ 教授/ 塑料盒/ 羊肉串
TF-IDF模型结果：  学子/ 王老吉/ 加多宝/ 主旨/ 集团/ 大学生/ 公益活动/ 社会
TF-IDF模型结果：  黄河/ 母亲河/ 江河/ 福气/ 环境/ 家居/ 气候变化/ 健康状况/ 决策/ 水资
源
TF-IDF模型结果：  叙利亚/ 摩洛哥/ 大使/ 红十字国际委员会/ 人道主义法/ 政府/ 武器/ 哈桑/
局势/ 外交部
TF-IDF模型结果：  盖茨/ 现场/ 创办人/ 好友/ 股神/ 巴菲特/ 二人/ 财富/ 论坛/ 视频
TF-IDF模型结果：  指标/ 直饮水/ 水机/ 专人/ 毒理学/ 放射性/ 感官/ 性状/ 小区/ 水质
TF-IDF模型结果：  鞋子/ 孩子/ 马学良/ 妹妹/ 运动鞋/ 老师/ 团县委/ 爷爷/ 学校/ 妈妈
TF-IDF模型结果：  娃娃/ 环境/ 剧社/ 儿童/ 热点/ 孩子/ 视界/ 舞台/ 视角/ 呼声
TF-IDF模型结果：  农民工/ 工人/ 工程项目/ 兄弟/ 英雄/ 志强/ 工装/ 暴雨/ 洪水/ 衣服
TF-IDF模型结果：  全球/ 时区/ 地球/ 搜狐/ 官方/ 平台/ 小时/ 网友/ 公益/ 城市
TF-IDF模型结果：  环卫工/ 记者/ 居民楼/ 门窗/ 环卫所/ 水电/ 房租/ 赵理/ 居民/ 关键
TF-IDF模型结果：  市场/ 监管局/ 食用油/ 地沟油/ 餐厨/ 食品/ 办法/ 企业/ 监管部门/ 垃圾
处理
TF-IDF模型结果：  尘肺病/ 农民工/ 兄弟/ 作坊/ 矿山/ 村庄/ 病重/ 妻离子散/ 倡议书/ 粉尘
TF-IDF模型结果：  大病/ 医院/ 市人/ 社局/ 人力/ 定额/ 标准/ 政策/ 基础/ 本市
TF-IDF模型结果：  乘客/ 地铁/ 站台/ 暴雨/ 公交/ 预案/ 列车/ 车站/ 晚点/ 值班员
TF-IDF模型结果：  山路/ 小康/ 代步/ 孩子/ 土墙/ 茅屋/ 手脚/ 冻疮/ 身子/ 箩筐
TF-IDF模型结果：  住房/ 人员/ 待遇/ 烈士/ 条件/ 工伤保险/ 民政部/ 补助金/ 意见/ 情形
TF-IDF模型结果：  女性/ 女人/ 人群/ 慈善家/ 母性/ 妇女节/ 热心/ 姐妹/ 天生/ 人才
TF-IDF模型结果：  三聚氰胺/ 食品/ 标准/ 牛奶/ 法典/ 含量/ 液态/ 限量/ 婴儿/ 奶粉
TF-IDF模型结果：  飞镖/ 女士/ 居民/ 射中/ 小区/ 小猫/ 流浪/ 报案/ 弓弩/ 杀伤力
TF-IDF模型结果：  公厕/ 手纸/ 卫生纸/ 青岛/ 浪费/ 游客/ 景点/ 过度/ 岛城/ 市南区
TF-IDF模型结果：  绿化/ 条例/ 绿地/ 行政处罚/ 居住区/ 树木/ 公共绿地/ 广州市/ 古树名/
案件
TF-IDF模型结果：  春苗/ 营养/ 安利/ 计划/ 关工委/ 基金会/ 农村/ 伙食/ 湖北/ 中西部
TF-IDF模型结果：  危机/ 老龄化/ 纳税人/ 反观/ 浪潮/ 制度性/ 结构/ 进程/ 隐患/ 局面
TF-IDF模型结果：  网民/ 强降雨/ 销量/ 常识/ 工具/ 灾害/ 暴雨/ 大雨/ 雷雨/ 钥匙扣
TF-IDF模型结果：  联合国开发计划署/ 联合国/ 愿景/ 先生/ 潘基文/ 协调员/ 罗黛琳/ 周迅/
画册/ 邵忠
TF-IDF模型结果：  桔灯/ 图书馆/ 志愿者/ 乡村/ 小学/ 江苏/ 计划/ 面向社会/ 贵州/ 省份
TF-IDF模型结果：  上海/ 空气质量/ 数据/ 淀山湖/ 浓度/ 监控点/ 刘代玲/ 环保部门/ 上海市/
小时
TF-IDF模型结果：  中国儿童少年基金会/ 日照/ 安康/ 人为/ 搜狐/ 灾区/ 儿童节/ 娃娃/ 山东/
家园
TF-IDF模型结果：  节能灯/ 节电/ 准则/ 效果/ 灯泡/ 呼唤/ 代价/ 浙江省/ 电冰箱/ 冰箱
TF-IDF模型结果：  弱势/ 儿童/ 计划/ 黄英男/ 人人/ 题材/ 电影/ 种类/ 素材/ 张元
TF-IDF模型结果：  高尔夫/ 财智/ 邢傲伟/ 体操/ 搜狐/ 峰会/ 冠军/ 运动员/ 成龙/ 金牌
TF-IDF模型结果：  小丽/ 哥哥/ 施暴/ 男孩/ 家长/ 强奸/ 女孩/ 父母/ 孩子/ 钟落潭
TF-IDF模型结果：  粉笔/ 报系/ 老师/ 一堂课/ 雪佛兰/ 小手/ 世纪/ 乡村/ 同学/ 支教
TF-IDF模型结果：  邓锦杰/ 晏建伟/ 救者/ 邓秋琼/ 娄底/ 藏獒/ 弟弟/ 市民/ 救人/ 老板
TF-IDF模型结果：  全民/ 血压/ 拉开帷幕/ 年度/ 主题/ 方式
TF-IDF模型结果：  英特尔/ 社会福利/ 慈善事业/ 民政部/ 有限公司/ 信息化/ 工作坊/ 怀柔/
创新奖/ 公益
TF-IDF模型结果：  水灾/ 救援队/ 基金/ 云南/ 技能/ 民间/ 灾害/ 现场/ 广西南宁/ 指挥系统
TF-IDF模型结果：  银行/ 提款机/ 英国/ 苏格兰/ 皇家/ 选项/ 客户/ 慈善/ 恒生/ 交易
TF-IDF模型结果：  附加费/ 燃油/ 运次/ 出租车/ 油价/ 城市/ 出租汽车/ 厦门/ 价格/ 南昌
TF-IDF模型结果：  传统/ 慈善/ 心被/ 初级阶段/ 行政化/ 境界/ 符号/ 五彩/ 纸片/ 平民
TF-IDF模型结果：  住址/ 小组/ 出生日期/ 户口/ 联系人/ 健康状况/ 助医/ 电话/ 病情/ 姓名
TF-IDF模型结果：  指挥部/ 身份/ 北京/ 新闻/ 冯强/ 张驰/ 潘安/ 遗体/ 日讯/ 通报
TF-IDF模型结果：  肌肤/ 美白/ 黑色素/ 羽西/ 女性/ 净肤/ 圣品/ 抗氧化/ 产品/ 功效
TF-IDF模型结果：  报告/ 产品/ 字数/ 编辑/ 会员/ 个人资料/ 黑名单/ 用户/ 体验/ 体会
TF-IDF模型结果：  肌肤/ 凝胶/ 主张/ 彩妆/ 深层/ 润泽/ 皮肤/ 皮脂/ 肌醇/ 养分
TF-IDF模型结果：  感觉/ 试用装/ 正品/ 泡沫/ 盒子/ 瓶子/ 面膜/ 套装/ 洗面奶/ 珍珠
TF-IDF模型结果：  报告/ 个人资料/ 频道/ 几率/ 产品/ 评论/ 字数/ 精华/ 体验/ 编辑
```

TextRank抽取关键词

In [1]:

```
import jieba.analyse
text = '广州地铁集团工会主席钟学军在开幕式上表示，在交通强国战略的指引下，我国城市轨道
```

n 名词　　nr 人名　　nr1 汉语姓氏　　nr2 汉语名字　　nrj 日语人名　　nrf 音译人名　　ns 地名　　nsf 音译地名　　nt 机构团体名　　nz 其它专名　　nl 名词性惯用语　　ng 名词性语素

In [2]:
```
keywords= jieba.analyse.extract_tags(
    sentence=text,
    topK=10,
    allowPOS=('n','ns','nr','nt','nz'),
    withWeight=True)

print(keywords)

#for keyword,weight in keywords:
#    print(keyword,weight)
```

```
Building prefix dict from the default dictionary ...
Loading model from cache C:\Users\fengl\AppData\Local\Temp\jieba.cache
Loading model cost 0.376 seconds.
Prefix dict has been built successfully.
[('轨道交通', 0.4728850527157576), ('城市', 0.30216050785575754), ('密集型', 0.2880675302212121), ('行业', 0.2812424269787879), ('里程', 0.24700862624454545), ('强国', 0.24389318703636365), ('钟学军', 0.18113284095303028), ('国赛', 0.18113284095303028), ('交通', 0.18015916857757577), ('城市轨道', 0.1773250465848485)]
```

In [3]:
```
### 使用pyecharts画词云
from pyecharts.charts import WordCloud

#所需格式： [('python', 23),('word',10),('cloud',5)]

mywordcloud = WordCloud()
mywordcloud.add('',keywords,shape='circle')#词云形状可以选择，也可以自定义背景图片，
### 渲染图片
mywordcloud.render()
```

Out[3]: 'F:\\jupyter_project\\NLP_Course\\render.html'