

# 选题二：基金会舆情影响力衡量报告

团队：ThreeTimeZone

Cao Lu, Yunhan Yang, Stephanie Zhang

曹璐，杨云汉，张梦雯

## 1. 项目综述

通过对衡量基金会影响力的四项指标：微博影响力，谷歌搜索影响力，新闻媒体影响力和基金会关联度影响力赋予不同的权重，再结合基金会的地域与领域分类，我们可以算出基金会的综合影响力指标。我们通过互动地图页面和指标权重调节页面进行互动性的结果展示。

[交互性分析结果展示 https://lucaoshinyapps.io/fundhackathon\\_app/](https://lucaoshinyapps.io/fundhackathon_app/)

## 2. 项目目的

我们的目的为测量各个基金会的舆情，我们的方式是，从多个媒体媒介的最基本的新闻覆盖和影响出发，计算出微博影响力、新闻媒体影响力、google 搜索出现信息数、分别的分数，并计算总分，以此刻画出捐赠方最感兴趣的信息。

## 3. 思路综述

1. 因为绝大多数基金会新闻都为正面新闻，我们没有尝试使用 sentiment analysis ( 因 99.9% 的关于基金会的新闻均属于正面新闻，不纳入我们的考量舆论声量范畴 ) ；
2. 由于数据是 unlabeled data, 从机器学习的角度打分是不有效的 ；
3. 因此我们的分数采用 domain knowledge 的裁定，在微博新媒体，新闻媒体，和 google 搜索数几个角度立体的描述各基金会的舆情面貌。我们认为新闻的覆盖人群的广度，和对话题关心和讨论的参与度等的次数的采集，是刻画基金会类的舆情最有利的出发点。

## 4. 数据来源

1. 新闻数据来源为官方提供的数据 ；
2. 新浪微博的数据来源来自于爬取的基金会的实时官方微博统计，爬取代码如附件 ；
3. 谷歌搜索条数来自于实时的谷歌统计，爬取代码如附件。

## 5. 算法过程

为了衡量基金会的舆情声量，我们从三方面：微博影响力，新闻媒体影响力，和谷歌搜索树进行衡量，采集数据如下图 Key metrix。我们分别取得三方面的各指标，计算 Log 值作为单项分数（并且设定最低分），分配相应权重得出三方面的分数衡量，最后通过用户通过交互界面赋值权重计算出总分数进行衡量。详细信息请见 Key Matrix 表格，算法过程和介绍代码如附件（ipython notebook）。

Key Matrix:		计算分数方法	分项：
1. 微博影响力 15%	是否有主页 10%	登录各基金会微博官方页面，抓取出评论数、粉丝数、关注数、转发数、点赞数、原创微博数的数值计算 $\log(\ln)$ 值作为分项的分数，将低于 10 条的数字转换分数时按照 $\ln 10 = 2.3$ 分计算，并按照左边所示的比例计算加权，成为总分为 10 分的分数。	微博影响力分
微博具有实时性，极强的互动性和亲和力，衡量互动者的影响度，是在其他媒体不可替代的	主页关注数 20%		
	主页原创内容数 20%		
	被点赞总数 10%		
	被评论总数 10%		
	被其他微博主提及次数 10%		
2. 新闻媒体影响力 60%	新闻包中被提及次数汇总 70%	新闻媒体影响力包括：1. 新闻关联分数，我们认为基金会如果在新闻中被其他基金会提及次数越多，说明影响力越大。新闻关联分数由每个基金会其他基金会提及的次数加总，算出 Log 得到分数(总分 10 分),将低于 10 条的数字转换分数时按照 $\ln 10 = 2.3$ 分计算(最低分)；2. 新闻提及次数，此基金会在新闻中总共出现的次数，算出 $\ln$ 得到分数(总分 10 分) 将低于 3 条的数字转换分数时按照 $\ln 3 = 1.09$ 分计算（最低分）	新闻媒体影响力分
在采集所得的新闻库中被提及的次数是衡量一个基金会舆情最基本和重要的指标, 我们考察 1.新闻提及，2.和其他基金会相互提及的次数	新闻包中与其他基金会的 connection 数（如何转化成分数？） 30%		
3. Google 搜索数 25%	被 Google 精准搜索到的新闻数量统计 100%	谷歌搜索分数衡量谷歌搜索此基金会所得出的结果条数进行统计，以搜索出现次数的角度概括舆情。计算方法为将各基金会的谷歌搜索数取 Log 作为得分，将低于 10 条的数字转换分数时按照 $\ln 10 = 2.3$ 分计算	搜索引擎信息覆盖分数

## 6. 算法的优势

### 1. 多维度准确考量舆情：

因微博属于交互类社交，新闻媒体属于权威新闻范畴，谷歌搜索数是宏观测量网络总信息量的简单工具，本算法考量因舆情从三个维度互不干涉，我们从 3 个角度分别设分衡量，以赋予权重给总分，给用户多维度的考量。

### 2. 算法的轻量级和直切重点：

算法采用直观的处理，精准从几个维度共同估量舆情，并未考虑 sentiment analysis（因 99.9% 的关于基金会的新闻均属于正面新闻，不纳入我们的考量舆论声量范畴），或 machine learning 等相对能发挥实质功效的方法。

### 3. 算法的交互性：

- a. 算法的展示部分可让用户选择自己对三个维度的考量和侧重，
- b. 交互界面让用户和地理、基金会性质、等全方位配合了解综合情况。

## 7. 应用展示

通过对衡量基金会影响力的四项指标：微博影响力，谷歌搜索影响力，新闻媒体影响力和基金会关联度影响力赋予不同的权重，再结合基金会的地域与领域分类，我们可以算出基金会的综合影响力指标。我们通过互动地图页面和指标权重调节页面进行互动性的结果展示。

[交互性分析结果展示 https://lucaoshinyapps.io/fundhackathon\\_app/](https://lucaoshinyapps.io/fundhackathon_app/)

**功能 1：**通过对指标权重进行调节，使用者可以定制基金会影响力排名表，获取在自己认为重要的指标上具有较高影响力的基金会信息。（附录 2 图 1，图 2，图 3）

**功能 2：**通过对基金会地域和领域的选择，使用者可以获取在自己关注的地域和领域具有较高影响力的基金会信息。（附录 2 图 2）

1. 按照地理分类	东部沿海，东北，中部和西部，对综合舆情分数排名
东部沿海	ranking 1-10
东北	ranking 1-10
中部	ranking 1-10
西部	ranking 1-10

2. 按照领域分类	国际事务，文化教育，资源环境，社会福利，对综合舆情分数排名
国际事务	ranking 1-10
文化教育	ranking 1-10
资源环境	ranking 1-10
社会福利	ranking 1-10

**功能 3：**我们还提供了交互式的地图展示，可视化的展现了具有较高影响力的基金会的地域分布。同时根据浮标中的具体信息，如基金会网址等，使用者可以选择进一步了解自己关注的基金会。（附录 2 图 4）

**功能 4：**我们列表的方式展示基金会的用户画像，通过基金会成立天数，原始资本，志愿者数，员工数，项目总收入，项目总支出，受捐赠金额，资助金额，项目数，专项基金数等指标从多个维度展示了基金会的特性。（附录 2 图 5）

## 附录：

### 附录 1：主要文件结构和描述如下图所示（ README ）

1. DataPreprocess/
  - 1.1 Data Exploration.ipynb – 研究最终影响力指标特征的分布，异常值等问题
  - 1.2 DataPreprocess.py – 预处理最终的影响力指标特征
  - 1.3 Full\_Data.xlsx – 最终影响力指标特征的原始数据
2. FundStat/
  - 2.1 Fund\_Stat.csv – 所有可能的影响力指标特征的原始数据
  - 2.2 FundRelation.py – 统计每个基金会与其他所有基金会共同出现在新闻里的次数
  - 2.3 Funds\_Cooccurrence\_Matrix – 基金会共同出现在新闻中的次数的矩阵
  - 2.4 Funds\_Mentioned – 出现在每则新闻中的基金会列表
  - 2.5 FundStat.py – 特征工程
  - 2.6 Provice.csv – 中国所有省，直辖市，自治区列表
  - 2.7 Sector.csv – 基金会所有领域分类列表
3. Shiny/My\_Shiny/my\_app/
  - 3.1 fundhackathon\_app /
    - 3.1.1 data/ – 最终分析使用的数据
    - 3.1.2 helpers.R | server.R | ui.R – 创建交互性分析报告
  - 3.2 deploy.R – 部署交互性分析报告
4. WeiboSum/
  - 4.1 Weibo/ | Weibo1/ | Weibo2/ – 从微博移动版爬取的基金会微博内容
  - 4.2 WeiboStat/ – 粉丝数，追踪数，微博数，转发数等衡量微博影响力的指标
  - 4.3 AppendWeiboData.py – 合并微博影响力指标与其他指标
  - 4.4 CleanWeiboData.py – 微博影响力指标数据清洗
  - 4.5 ExtractWeiboUserID.py – 从文件中获取微博 ID
  - 4.6 WeiboSpyder.py – 创建微博爬虫类
  - 4.7 WeiboSpyderImplement.py – 实行微博爬虫

附录 2：图示



图 1



图 2

TopFunds

Show10entries

Search:

	基金会名称	影响力总分	微博影响力得分	谷歌影响力得分	新闻媒体影响力得分	基金会关联度影响力得分	Action
1	中国扶贫基金会	9.594	8.985	7.48	10	10	
2	中国儿童少年基金会	9.3552	8.508	7.89	8.69	9.92	
3	中国红十字基金会	9.3008	8.762	7.78	8.76	9.66	
4	中国青少年发展基金会	9.2408	8.132	8.02	9.41	9.98	
5	中国妇女发展基金会	9.1556	8.324	7.69	9.49	9.71	
6	爱佑慈善基金会	9.0988	9.232	6.54	7.17	9.01	
7	南都公益基金会	8.9756	7.649	6.42	7.96	9.86	
8	中华社会救助基金会	8.944	8.035	8.5	7.65	9.55	
9	中华思源工程扶贫基金会	8.8772	7.703	5.65	8.12	9.66	
10	安利公益基金会	8.7364	8.641	7.14	6.33	8.8	

Showing 1 to 10 of 10 entries

Previous1Next

Showing 1 to 10 of 10 entries

Previous 1 Next

图 3

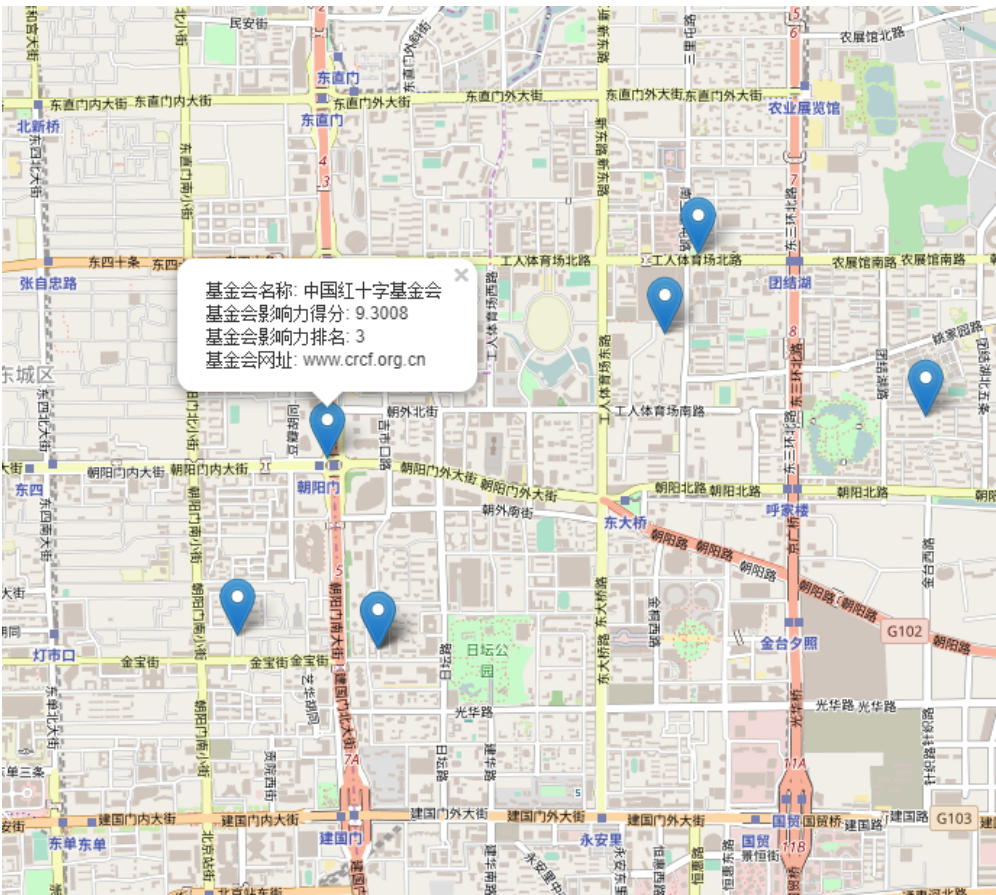


图 4

基金会名称

爱佑慈善基金会 北京大学教育基金会  
传媒大学教育基金会  
北京交通大学教育基金会 华民慈善基金会  
凯风公益基金会

Show 10 entries

Search:

	名称	成立天数	地址	领域	原始资本	志愿者数	员工数	项目收入	项目支出	受捐赠金额	资助金额	项目数	专项基金数	Action
1	爱佑慈善基金会	3373	北京市朝阳区	儿童，医疗救助	20000000	221	39	17841000	13151000	314726588.47		17	0	
2	北京大学教育基金会	8063	北京市海淀区	教育，科学研究	20000000	19	30	7754000	69709528	1205500000	49547930	18	0	
3	传媒大学教育基金会	3693	北京市朝阳区	国际事务，教育，科学研究	20000000	30	7	20810461	17364138	53261520	5500000	74	0	
4	北京交通大学教育基金会	2924	北京市海淀区	教育，科学研究	20000000	33	10	65266617	124170192	41940000	0	29	0	
5	华民慈善基金会	3429	北京市东城区	安全救灾，创业就业，扶贫助困，老年人	50000000	25	15	3400000	50208621	2994549.22	360000	8	0	
6	凯风公益基金会	3805	北京市朝阳区	扶贫助困，公益事业发展	50000000	0	5	0	29388050	52500000	0	24	0	

Showing 1 to 6 of 6 entries

Previous1Next

图 5