

# 基金会中心网推荐系统

从公司与基金会的需求出发

队伍：Steeve 和他的朋友们

队员：黄功详 林舒雅 姚昕婕

18/08/2017

## 1. 摘要

基金会是我国公益慈善领域的独特社会组织，具备与众不同的运营模式；传统的运行型基金需要资助型基金为其提供运作资本，解决社会问题。资助型基金一方面要负责筹集基金帮助项目运行，另一方面需要根据自身偏好，找到即能带来社会价值，又可以创造经济价值的项目执行方。如何在省时省力地找到具有具有潜在资助兴趣的执行型基金，如果确保匹配结果能够符合科学客观的遴选成果，成为此方案的出发点。最终提议方案建立了在战略、执行力以及影响力这三个维度的社会价值综合评价体系，并且将资助型基金的投资偏好在匹配中列入考量。测试数据来源于主办方基金会中心网提供的原始数据以及通过网络爬虫获得的新闻数据。此推荐系统客观科学地评价了受助方，具备综合性，灵活性的特点。

## 2. 背景

在各种类型的社会组织中，基金会占据相对特殊的地位。资金的筹募、运作模式，以及公益慈善的核心理念，都标志着其独特的生存、发展和管理模式。

在我国四千多家基金会中，大部分是传统模式的运行型基金，它们在筹集善款之后，自己运作项目，并在项目完成之后向理事会、捐赠方、公众进行报告。然而，运行型基金的发展壮大带来了运作成本上升，在政府政策对成本有一定的限制的背景下，项目执行效果往往受到影响。同时，我国公益行业产业链仍不完善，很多运行型基金难以筹集到发展所需的资金。在这种情况下，资助型基金的存在显得尤为必要。并且，资助型基金的多少又是体现公益行业整体发展成熟度的晴雨表，也是我国慈善基金会发展的方向之一。

作为慈善基金的提供者，资助型基金关注的核心问题在于如何使公益项目既能创造出社会价值，又能提供商业价值。创造共享价值是和社会合作共赢的关键所在。资助型基金的积极发展，有助于发展壮大项目执行方和充分发挥广大慈善基金会的作

用，推动慈善救助工作的进程以及覆盖范围，更符合被救助者的核心需求，以此来撬动公益投资的杠杆，实现“小球推动大球转”的经济效应。

### 3. 需求

资助型基金会通过引入商业投资的经验和公益投资的体系，为公益行业提供资金支持，并承担公益组织进行社会创新而带来的风险，以创新和高效来完善公益生态链。如何将筹集到的公益基金最大程度地造福社会，创造最多的社会价值，成为了众多资助型基金会的难点之一。

除了强有力的资金实力做后盾之外，资助的公益项目科学遴选制度也是保障资助型基金会健康发展的强有力因素之一。结合贯彻创新、高效的公益理念，完善项目的评审流程与标准，精准识别与验证项目执行方对项目的设计理念。

如何通过精准的需求评估与匹配，为公益项目提供针对性的支持和个性化的资源对接，是基金会的瓶颈所在；如何在大数据时代，融入创新思维，利用先进的人工智能技术来进一步优化评估、匹配体系，也是基金会的机会所在。遴选制度基于基金会自身的关注领域，以及行业导向，挑选项目的标准和习惯，来更好地发掘潜在合作伙伴，实现效益的最大化。

### 4. 最终方案及数据来源

我们的最终方案涵括了在战略目标、执行力以及影响力这三个方面对于基金会的评估。测试数据来源于主办方基金会中心网提供的原始数据以及通过网络爬虫获得的新闻数据。

在战略这一方面，我们将基金会的信息透明度和工资福利占总资产的比例作为两个主要指标，并且将资助型基金会的投资偏好在匹配中列入考量。信息透明度的得分来源于基金中心网提供的原始数据，透明度越高，评分越高。并且，根据民政部的规定，工资福利占总资产的比例不得超过百分之十。通过将投资信息与项目信息两个表单联合，我们可以获得资助型基金会曾经投资过的基金会名单，以及相应的关注领域和性质，从而了解到每一个资助型基金会独特的投资偏好并且将项目按照匹配度的高低进行排序。

```
In [36]: bi_invest_df = invest_df.loc[:,['受助单位名称','项目编码']].join(project_df.loc[:,['基金会名称','项目编码']].set_index('项目编码'))
bi_invest_df
```

Out[36]:

	受助单位名称	项目编码	基金会名称
0	爱佑慈善基金会	63990	腾讯公益慈善基金会
1	爱佑慈善基金会	64114	招商局慈善基金会
2	爱佑慈善基金会	65610	北京百度公益基金会
3	爱佑慈善基金会	70032	中华思源工程扶贫基金会
4	爱佑慈善基金会	80714	万科公益基金会
5	爱佑慈善基金会	80643	万科公益基金会
6	爱佑慈善基金会	83721	中国儿童少年基金会

图表 4.1 投资信息以及项目信息联合表格

而在执行力的评价中，公益支出和总资产两者的趋势以及前者占后者的比例都是重要的考量标准。在计算公益支出和总资产的趋势时，我们从原始数据中提取出 2014 和 2015 年的数据，两者的差值即是趋势得分。如果某一基金会 2014 或 2015 年的数据有缺失，则将其趋势赋值为所有基金会趋势的均值，从而避免了极端型预估的风险。同时，负责人数也是基金会在项目中投入的人力的一个体现，在此项指标中，人数越多，则评分越高。

对于影响力的判断则主要基于负责人中曾任及现任国家工作人员的数量和项目在新闻中的曝光程度。我们运用网络爬虫的技术，在搜索引擎中提取并计算项目的曝光量，曝光量越高，则意味着项目的影响范围更广，因而在影响力方面的得分更高。

资助型基金会可以根据自己的需求设定各项指标的权重，以期获得最精准的匹配。

## 5. 研究方法

在相似度的评估中，我们将投资信息与项目信息联合，运用 Multiple One-Hot Encoding 表示曾投资项目的关注领域，而后计算其余弦相似度。将相似度包含于评估标准中能够清晰地了解不同基金会的不同投资偏好并做出更为个性化的精准判断。

```
In [16]: SDGList = ['无贫穷', '零饥饿', '良好健康与福祉', '优质教育', '性别平等', '清洁饮水和卫生设施', '经济适用的清洁能源',
                  '体面工作和经济增长', '产业、创新和基础设施', '减少不平等', '可持续城市和社区', '负责任消费和生产', '气候行动',
                  '水下生物', '陆地生物', '和平、正义与强大机构', '促进目标实现的伙伴关系', '其他']
```

图表 5.1 Multiple One-Hot Encoding 例子

```
In [18]: SDGdf['SDG'] = SDGdf['SDG'].apply(SDGMultiHotEncoding)
SDGdf
```

```
Out[18]:
```

	受助单位名称	SDG
0	杭州市生态文化协会	[0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
1	帝斯曼维生素（上海）有限公司	[1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
2	左云县林业局	[0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
3	北京市华远集团公司	[0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
4	北京富润德斯文化发展有限公司	[0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...
5	北京修实公益基金会	[1.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, ...

图表 5.2 Multiple One-Hot Encoding 例子

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|_2 \|\mathbf{B}\|_2} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

图表 5.3 余弦相似度计算公式（引自维基百科词条：余弦相似度）

```
Out[19]:
```

受助单位名称	杭州市生态文化协会	帝斯曼维生素（上海）有限公司	左云县林业局	北京市华远集团公司	北京富润德斯文化发展有限公司	北京修实公益基金会	北京世纪银通投资有限公司	中国科学院上海微系统与信息技术研究所	云南商学院	米易县麻陇乡财政所	...	汶川县教育局	中国少先队事业发展中心	北京中科青云实验学校	昆山育才学校	北京欧亚美艺术展览服务有限公司
杭州市生态文化协会	1.000000	0.000000	0.000000	1.000000	1.000000	0.707107	0.000000	1.000000	0.000000	0.000000	...	0.000000	1.000000	0.000000	0.707107	1.000000
帝斯曼维生素（上海）有限公司	0.000000	1.000000	0.000000	0.000000	0.000000	0.408248	0.000000	0.000000	0.577350	0.000000	...	0.000000	0.000000	0.000000	0.408248	0.000000
左云县林业局	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000	0.000000
北京市华远集团公司	1.000000	0.000000	0.000000	1.000000	1.000000	0.707107	0.000000	1.000000	0.000000	0.000000	...	0.000000	1.000000	0.000000	0.707107	1.000000

图表 5.4 余弦相似表

信息透明公开是现代文明社会的重要标志，也是公益行业发展的重要目标。透明度的提升有助于加强民众对我国公益基金会的信任与支持，因此，透明度也是评估基金会时的重要指标。

基金会的总资产只能表明其规模的大小，却不能展现其发展的前景，因而在对基金会进行考量时，我们将总资产的变化趋势而非数值作为指标，从而观测基金会的运行状况和进步空间。

同样的，在考虑基金会的公益支出时，为了消除规模对于支出数值的影响，我们将公益支出占总资产的比例作为指标，从而检验基金会的执行能力。

舆论与新闻的曝光量则主导了影响力的衡量，媒体宣传力度以及民众的讨论热度是项目所带来的社会影响的最佳体现。

得到所有指标的评分数值后，我们将每一项指标中的最高分作为该指标的满分，将所有分值按比例转换为 0 至 1 之间的数值，从而得到标准化的评估结果。

$$\begin{aligned} & norm(): a \text{ function that scales the values of an array between } 0 \text{ and } 1. \\ & field \text{ similarity} = cosine \text{ similarity}(one \text{ hot encoded}(field)) \\ & invest \text{ similarity} = M_{invested} * cosine \text{ similarity}(one \text{ hot encoded}(SDG)) \\ & strategy = norm(field \text{ similarity}) + norm(invest \text{ similarity}) + norm(transparency) \\ & influence = norm(total \text{ asset}) + norm(total \text{ service expenditure}) + norm(scraper \text{ result}) \\ & efficiency = norm(total \text{ asset trend}) + norm(total \text{ service trend}) + norm(service \text{ portion}) \\ & \quad + norm(man \text{ force}) + norm(number \text{ of leader}) \\ & Final \text{ Score} = strategy * Weight_{strategy} + influence * Weight_{influence} \\ & \quad + efficiency * Weight_{efficiency} \end{aligned}$$

图表 5.5 最终评分数值计算公式

6. 使用者介面

进入我们系统的页面后，使用者可以从下拉列表中选择一个需求方基金会，然后输入此需求方对于战略、影响力以及执行力这三方面（从左至右）的考量权重，单击按钮，即可获得运行方基金会的推荐分数排名。具体的运行及使用方法请参考压缩包内的 Readme.md 文档。

杭州市余杭区老年人体育协会

60

20

20

GO!

基金会名称	分数
浙江大学教育基金会	65.48177191699072
中国扶贫基金会	62.1983938471821
中国青少年发展基金会	60.59737769818803
东莞市东莞理工学院教育发展基金会	60.30212272634631
中华少年儿童慈善救助基金会	57.83995918089881
上海唐君远教育基金会	56.77243230629681
北京科技大学教育发展基金会	56.62191722376018
榆林学院教育发展基金会	55.80659428143146
中国教育发展基金会	55.687832837988054
阿里巴巴公益基金会	55.68055238527812
广东省华美教育慈善基金会	55.39725018853721
南都公益基金会	55.0756310688004

图表 6.1 使用者界面截图

## 7. 研究用例

针对中国红十字基金会，前十五名评分结果如下图所示。根据分数从高到低，表示项目执行方和中国红十字基金会投资偏好的吻合度大小。从而帮助资助型基金会做出更好的资助选择。

```
finalScoreDf.sort_values(['score'], ascending=False)
```

	score
清华大学教育基金会	51.7559
广东省东风日产阳光关爱基金会	50.0824
中国医药卫生事业发展基金会	49.956
安徽金牛公益基金会	49.5384
河仁慈善基金会	48.8386
北京大学教育基金会	46.9552
广东省同心圆慈善基金会	46.2784
中国扶贫基金会	45.3461
北京中金公益基金会	42.5117
深圳市博时慈善基金会	42.1454
中国青少年发展基金会	41.1577
中华思源工程扶贫基金会	39.6071
陕西省红十字基金会	39.5715
广东省茂华慈善基金会	39.5201
老牛基金会	39.304

图表 6.1 评分结果

## 8. 讨论

我们所提出的解决方案匹配推荐系统科学客观地评价了资助型基金会，以及项目执行方，把用户模型中的兴趣需求信息和推荐对象的特征信息相匹配，同时使用对应的推荐算法进行计算筛选，为资助型基金会找到符合遴选标准的执行方，为项目执行型基金会找到合适的资金提供方，然后推荐给资助型基金会，以供参考。

### **a. 综合性**

尽管不同资助型基金会根据自身性质特点，具有量身定制的项目筛选标准。我们的最终方案综合不同的评分细则，找出相似点，归纳整理不同点；最终从基金会战略、基金会影响力、基金会执行力三个纬度构建社会价值综合评分体系。这套体系能够从三个角度全面考察执行基金会：以更公平、可持续地有效解决社会问题为目标导向；鉴于以前的工作成效，考察组织和团队实施创新解决方案，达成社会目标的影响能力；具备在合作中整合资源，进行资源最优化配置和有效利用的执行能力。这样的阶梯式上升思考模式：以解决社会问题为核心需求，基于团队以往表现、影响力范围为坚实基础，用数据客观考察对项目的执行能力和领导力，预估执行效果的前景，确保社会价值和商业价值的双实现。

### **b. 灵活性**

该提议考虑到不同资助型基金会依然具有自己的考察侧重，例如有些基金会更关注项目的规模化潜力，有些基金会更看重项目是否可以有效递送给最终服务对象。不同的业务需求使得评判打分标准需要个性化定制，从而迎合不同基金会的遴选口味。该提议在维持数据和训练模型的不变的基础框架下，为三个维度增加权重系统，资助型基金会从而具备了灵活调整评判标准，更能找到适合自身基金会价值定位的推荐结果。此推荐系统真正做到从客户角度出发，以为提高推荐质量，提升投资效率贯穿始终，体现了人性化的设计理念。

### **c. 相对局限性**

现有的一些资助型基金会在对项目执行度考察中，增设创新性的衡量角度，旨在发现项目执行方对解决问题的模式、方法、工具和机制的创新。此推荐系统的数据来源是主办方基金会中心网提供的原始数据和网络爬虫的新闻数据，其中缺少对项目执行方创新性方案的量化标准。项目简介的描述中覆盖相关角度的特点也为数不多，鉴于数据特征点的缺少以及数据信息量的庞大，我们无法做到人工打分评判执行性基金会的创新思维。当然如果数据资源中可以增加对项目执行方的执行细节的描述，提议的推荐系统将可以利用语义理解来训练相关模型，实现对项目创新性的评定。

此推荐系统客观科学地从影响力、战略目标、执行度三个维度评价项目执行方，开发出综合考察的社会价值评估体系；考虑到资助型基金会的投资偏好差异，允许其调整三个维度的侧重，有一定的灵活性；方案从构思到完成，利用双重数据来源，使得评价更能反映出社会看法,更具有社会价值。