# 指标的体系贡献率方法研究设计报告

## 灵敏度（贡献率）分析的意义

灵敏度分析在很多领域中又叫重要性分析，是复杂系统不确定性分析问题中的一个重要方面，旨在研究复杂系统输出性能不确定性向各个输入变量分配的问题，进而识别出影响复杂系统输出性能的主、次输入变量，为系统的可靠性设计和评估模型的简化提供丰富的信息。

灵敏度分析技术可以从输入变量的整个分布范围来定量的衡量输入变量的不确定性对工程设计中所感兴趣的输出量的贡献程度。对于一个输出响应函数形式为的结构系统，灵敏度分析可将基本输入变量的不确定性与输出响应量的不确定性联系起来。它可以指出各基本输入变量的不确定性对响应量影响程度的大小，继而可以依次确定它们的实验或研究的优先级别，即确定模型输入变量的重要性次序，甚至还可以帮助更好地定义未知参数，以最终减少响应的不确定性范围，直到获得要求的响应性能的不确定性范围，也就为有针对性地改善结构模型提供了一种可行的新途径。

在天基信息网络中，其系统性能的影响不仅与底层的网络设计参数相关，也与所选取的评估系统综合性能的指标变量有关，因此对于天基信息网综合效能的灵敏度的分析可分为网络设计参数对综合评估结果的贡献率的分析，以及评估指标体系变量对综合效能评估结果的贡献率的分析。

天基信息网网络设计参数的灵敏度分析主要解决的问题是如何有效甄别各输入网络设计参数变量的不确定性对系统性能响应量的影响，从而提供网络系统性能设计所需要的重要信息，对提高网络运行质量和可靠性、简化分析模型以及减小管理决策失误率都具有重要意义。

对天基信息网进行评估指标体系变量的灵敏度分析可以有效的分别指标体系中各变量对系统综合效能的影响程度，从而对天基信息网综合指标体系设计与构建，以及优化综合评估方法提供一定的指导。同时也对针对性地提高天基信息网综合效能，和天基信息网性能优化分析有重要的指导意义。

### 灵敏度分析总体思路

天基信息网的贡献率的分析分为指标评估体系的灵敏度分析和网络设计参数变量的灵敏度分析。因此在进行天基信息网灵敏度分析时分别进行指标体系的灵敏度分析和网络设计参数变量的灵敏度的分析。

指标体系的灵敏度分析可以通过天基信息网进行时变图数据指标的提取，同时对各组数据指标采用了综合效能评估之后就形成了输入变量和输出响应的数据集，为了评估分析天基信息网络中不同的指标对综合评估效能的影响程度，以及不同评估方法下网络指标的贡献情况，因此可将提取出的天基信息网性能指标和综合评估的效能值进行关系建模，进而分析出各个指标的灵敏度的变化情况，从而对天基信息网络的指标体系的设计和综合评估方法的优化提供一定的指导作用。

天基信息网网络设计参数变量的灵敏度分析可通过对不同的网络设计参数运行多次的网络场景仿真，每次仿真通过对网络的性能指标进行综合评估，从而得到该组网络设计参数下网络系统的综合效能的评估值，通过多次的运行仿真并对仿真数据进行综合评估，也可将网络设计参数变量当做输入变量而综合性能的评估结果当做输出响应，从而可以构建分析网络设计参数和综合评估结果的关系模型，进而分析得出各个网络设计参数对系统综合效能的贡献率的情况，根据各个网络设计参数对天基信息网的综合效能的影响情况，则可进一步有针对性地进行网络的设计和优化。

因此天基信息网灵敏度分析的整体思路如图1所示：



**图 1 灵敏度分析思路**

### 基于随机森林的灵敏度分析方法

对输入变量对输出响应的影响进行灵敏度分析，需通过对输入和输出的关系进行建模，所以可以使用机器学习的方法构建和的关系模型，同时又从该模型还得具备能够从中得出输入变量对输出响应的贡献率重要性的情况，因此对比了各种机器学习的方法选取了基于随机森林的灵敏度的分析方法。

随机森林通过建立一组随机决策树（回归树或分类树，取决于具体问题）组成随机森林，然后根据每棵树的预测结果进行投票得到最终预测结果，是一类特别适用于高维小样本问题的分析方法，并且对非线性和交叉作用显著的问题也适用。该算法通过统计学习可以自动识别出对输出影响较小的输入变量，因而具有灵敏度分析功能。随机森林是近年在生物学领域的研究热点，目前，该方法在数据挖掘和生物信息学领域得到广泛应用，将其引入到天基信息网络综合评估中，具有一定的实用性和创新性。

随机森林由一系列的决策树组成，而这些树是通过递归分裂法则（如分类和回归树法则）建立起来的。作为一种非参数回归和分类的机器学习方法，随机森林对低维和高维数据分析均适用，特别是针对输入变量的维度大于数据规模的问题以及“黑盒”问题。当用于分类问题时，随机森林可看做由树结构分类器的集合构成的分类器，给定一个输入向量值，每个决策树都会预测出一个投票。单个决策树对于输入的小扰动较为敏感，但是随机森林能够有效克服这一缺点，因而具有更加稳健的分析结果。

随机森林算法还可用于衡量输入变量的相对重要性，也即可以用于灵敏度分析。使用最广泛的是扰动重要性指标(PVIM)，因此在天基信息网性能评估分析中使用扰动重要性指标刻画指标的灵敏度。

#### 随机森林模型的建立

复杂系统用确定型模型来表示，其中是输出变量,是维相互独立的随机输入变量。用表示用于建立随机森林的训练样本集,其中表示训练样本的规模。从总训练样本中重复抽样组成一系列样本集合，第个样本集合用于建立第个决策树, 表示的样本规模,表示需要建立的总决策树的个数。建立随机森林的过程简要如下:

1. 通过实验或抽样的方法得到一组样本规模为的训练样本。
2. 随机的（替换或非替换的方法）从样本集中选取个训练样本组成新的规模为的训练样本,推荐取.
3. 用CART 法则建立第个决策树。CART 法则事实上是一种二元递归分裂法则,它将输入变量的空间分割成越来越多的同类别的小矩形区域，所建立的决策树根节点包含了所有 个训练样本集，在每一个分裂节点（包括根节点）有如下分裂准则：
4. 从输入变量中随机选取个变量作为候选分裂变量。对于回归问题取,对于分类问题。分裂变量和分裂阈值的选取准则是使得节点不纯度的减小量达到最大，其中表示父节点的 Gini 不纯度，和分别表示左右子节点的Gini不纯度,和分别表示父节点落入左右子节点的概率。
5. 将父节点中的数据分配到子节点中。变量的值小于阈值的样本分配到左子节点中，剩余的样本分配至右节点中。按此准则不断向下分裂，直到满足停止条件（如 Gini 不纯度指标低于某个给定值）。
6. 在叶子节点，模型输出值通常用该叶子节点内数据的平均值来确定。也就是说，每个叶子节点内的输出值是常数，因此决策树也可以看成是分段常值函数。

用以上步骤建立起随机森林后，就可以用其进行预测和重要性分析。给定一个新的输入向量值，输出量的预测值就是对个决策树的预测值的平均。

假定有个样本，个特征，训练的时间复杂度为.如果是用个树来预测，时间复杂度为.同时随机森林作为一种集成学习的组合模型，很容易做成多核并行处理的算法，所以其时间的复杂度可以通过并行处理进一步降低。

#### 基于随机森林扰动重要性指标的灵敏度

由于第个决策树是根据训练样本建立的，将未被用于建该树的数据叫做袋外数（OOB）。袋外数据可用于衡量第个树的预测精度。

对于第个树，其袋外数据为, 其中表示第个树的袋外数据，是输入变量的精确值。用表示第个树的对的估计值,则第个树的均方误差表示为:

(1)

将袋外数据中的第列的数值进行随机扰动，表示随机扰动后的第个树的预测结果,则扰动后的均方误差为:

(2)

变量扰动后的均方误差相对于扰动前的均方误差为:

(3)

扰动重要性指标定义为所有树的的平均:

(4)

通过随机扰动的值，其边缘分布没有发生变化，但是破坏了变量和输出响应量的相关性，以及变量与其所有其他输入变量之间的相关性。如果变量越重要，那么扰动后的均方误差增量就会越大，因此指标的值也越大。因此，包含了变量的独立作用以及和所有其它变量的交互作用对输出性能的影响，从而可以用来进行变量筛选和变量重要性排序。因此，由上分析可知指标可以直接作为输入变量对输出响应影响灵敏度的刻画。

### 基于随机森林的天基信息网指标灵敏度分析

#### 总体框架

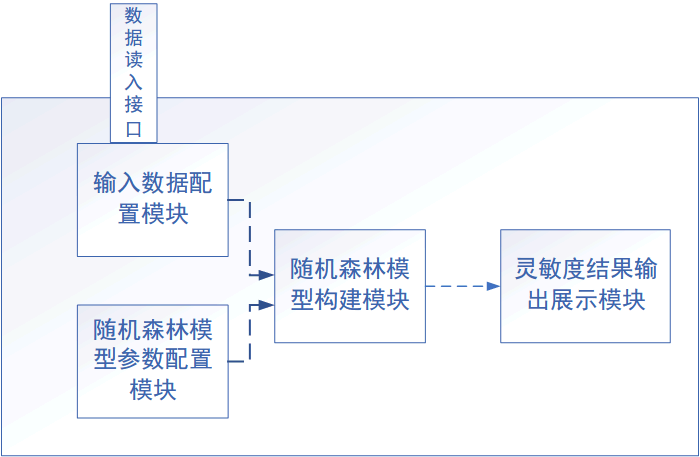
基于以上随机森林的构建过程和变量重要性得分的计算过程，因此将随机森林算法用于天基信息网指标灵敏度分析的过程如下：

输入个原始指标数据和对应的指标数据的效能评估值，为了能有更好的效果，样本量越多越好；

以效能评估值为目标值，指标数据为输入变量，训练随机森林回归模型；

通过训练得到的随机森林模型，输出各指标数据的重要性得分。

基于上述过程，则该算法的软件设计的框架如图 2所示：



**图 2 基于随机森林的灵敏度分析框架**

其中输入数据配置模块，可配置目标输出值，和输入变量值，即能够选择不同的自变量和因变量；随机森林模型配置参数配置模块可配置随机森林的模型参数，如树的个数，树的深度，并行处理的进程数等参数；输出结果的展示模块，可以输出展示一次或者多次的不同输入参数的一个或者多个的模型的灵敏度的分析结果。

#### 具体过程

针对分析网络设计参数对综合效能评估结果贡献率分析和网络指标数据对综合效能评估结果的贡献率影响分析均可采用同样的建模方法，针对不同的目标灵敏度分析的输入和输出不一样。

##### 网络性能指标的灵敏度分析

网络性能指标对综合评估结果影响分析：



**图 3性能评估指标灵敏度分析**

分析天基信息网性能指标数据对综合评估效能的影响程度的流程图如图3所示，其具体实施过程为：对于天基信息网仿真场景，将网络仿真场景的性能指标（不同时刻的指标，或者不同场景设置下的指标）和指标对应的综合评估结果，作为模型的输入变量，通过对指标数据和评估结果的关系建模，从而输出各个指标对综合评估结果的重要性的影响程度，同时也可以作为对网络性能指标数据的综合评估结果的预测模型，亦即对于新来的网络性能指标，可以不通过复杂的综合评估方法的流程，而通过直接输入到训练的模型里面，即可得出该组指标数据在该评估方法下的综合评估结果。

##### 网络设计要素的灵敏度分析

网络设计要素对综合评估的影响分析：



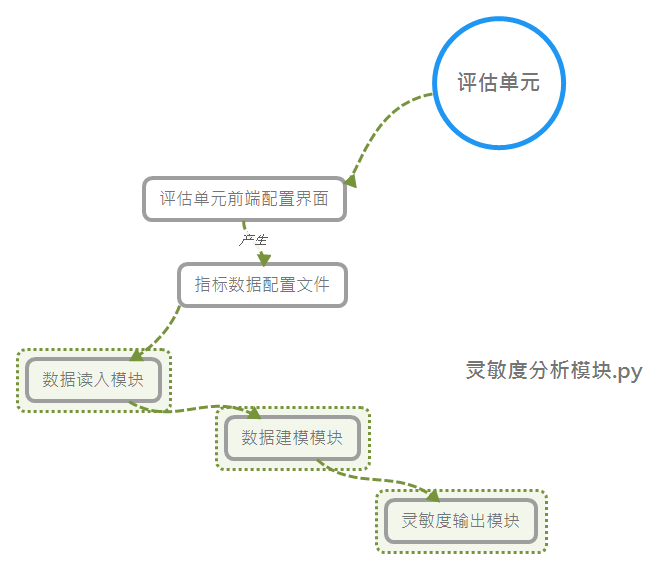
**图 4 网络设计参数灵敏度分析**

分析天基信息网网络设计参数对综合效能评估结果的的重要性程度分析如上图4所示，整体分析流程与性能指标数据对综合评估效能的影响程度的分析类似，但其具体实施过程有一定的差别，具体实施过程为：对于多组网络设计参数，分别进行仿真场景设置，运行不同的仿真场景，记录不同设计参数场景下网络的整体性能指标，并对整体性能指标进行综合效能评估，将评估结果和网络设计参数输入训练模型以训练出网络设计参数和综合评估结果的关系模型，从而得出各个网络参数对综合评估结果的重要性程度。该模型也可预测在新的网络设计参数条件下，该组网络参数的综合效能的评估结果。

由上面的分析过程可以看出，提出的基于随机森林的天基信息网指标灵敏度分析，具有较强的通用性，可适用于同一评估方法下的指标的灵敏度分析，或者网络设计参数的灵敏度分析，也适用于不同的评估方法下同样的指标数据或者网络设计参数灵敏度分析，可以进行横向的评估方法的对比，也可以进行纵向的指标或者网络设计参数的分析对比；算法的较强的通用性也为该分析方法提供了灵活的配置，且该配置可根据目标的需求进行指定。

通过该方法进行指标或者网络设计参数对综合评估结果的贡献率分析之后，算法模型输出了各个指标或者各个网络参数对综合评估效能结果的重要性影响程度，则可针对性的对网络性能指标或者网络设计参数进行优化设计，同时也可以对综合评估方法进行对比分析。

##### 灵敏度分析模块设计



**图 5 网络设计参数灵敏度分析模块**

1. 数据指标文件配置格式设计

灵敏度分析模块首先需要对数据指标和相应的指标的评分的数据进行读取，之后基于读取的数据进行随机森林回归模型建立，因此数据的读入方式决定了后续的设计的代码和模型的设计。所以数据指标文件的配置文件需要指定具体的分析的指标的名称，初步采用json的文件格式设计。

指标配置文件格式为：

{

"tableName1": {

"type": 0,

"Index": [

"IndexName1",

"IndexName2",

"IndexName3"

]

},

"tableName2": {

"type": 1,

"Index": [

"IndexName4",

"IndexName5"

]

}

}

指标配置文件使用json格式，其中tableName1，tableName2，…,表示存储指标数据的表名；type表示数据表里面存储的指标数据的类型（0：标量型，1：非标量型）；Index表示每个表里面在进行评估的时候使用到的指标列的名字，即表示参与评估的指标。

由于使用时变图评估对时间进行了分段，所以指标评估结果文件需要存储各个分段的评估结果，其文件格式为：

SegmentId, StartTime, Endtime, EnvaluetedScore

使用csv逗号分隔符的形式存储各个分段的评估值，其中：

SegmentId：分段编号；StartTime：分段的开始时间；EndTime：分段的结束时间；EnvaluetedScore：该分段下的总体评估值。

1. 数据读取接口数据结构设计

数据读入接口需要首先解析配置文件，获取到相应的指标名称，同时基于读到的指标的数据名称构建相应的sql语句查询出相应的数据，或者更加配置文件指定的excel文件进行数据的读取，存储数据的时候使用list的数据结构。

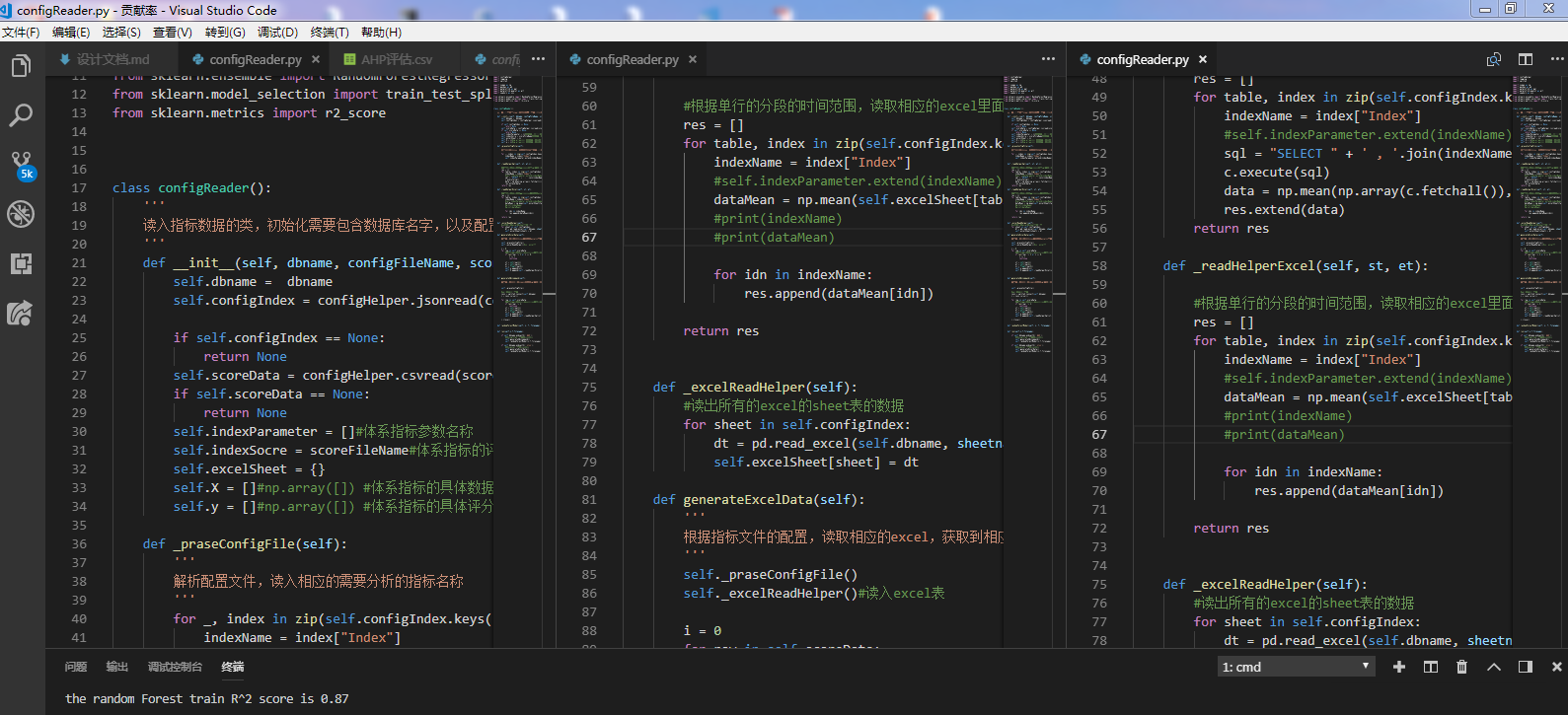
1. 数据分析模型设计

数据分析建模的设计主要基于读取到的数据进行随机森林回归模型的训练和构建，首先需要传入随机森林回归模型的参数进行模型的初始化，传入参数为：n：表示构建的随机森林的树的个数（树越多模型越准确），f：表示模型训练是使用的处理器的个数（-1表示使用所有的处理器。

1. 灵敏度分析结果输出模块设计

通过训练好的随机森林模型，输出其对应各个指标下的重要性得分即为各个指标的灵敏度定量分析结果，同时各个指标的灵敏度的结果以柱状图的的形式输出，保存到特定的文件路径，以供后续的界面显示使用，因此输出模型需要传入的参数是保存灵敏度结果的图片的filename，filename需要包含路径名（绝对路径或者相对路径）和保存生成的文件名，如一个示例为：filename=./out/image/灵敏度.png，表示为当前目录下的out目录下的image目录下的灵敏度.png的文件名。

灵敏度分析模块的部分代码截图如下：



**图 6 开发代码截图**

#### 天基信息网灵敏度分析总结

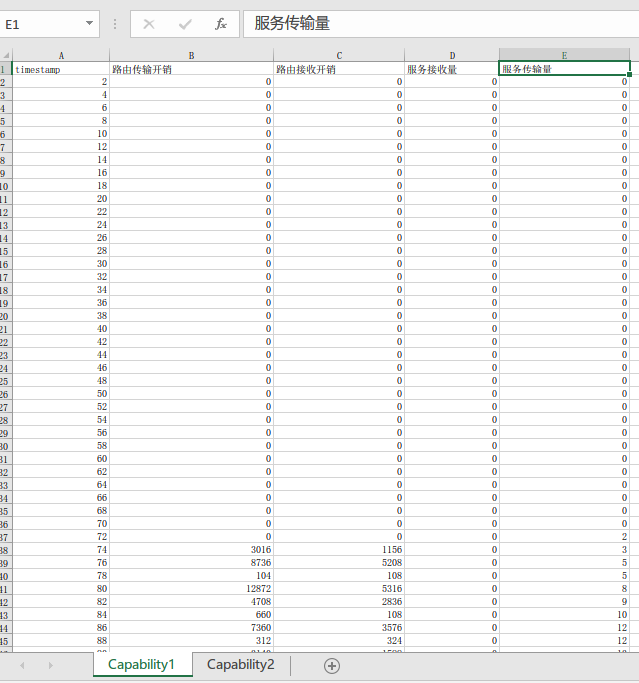
综上所述，将天基信息网灵敏度的分析总结为下表1：

表1 灵敏度分析总结

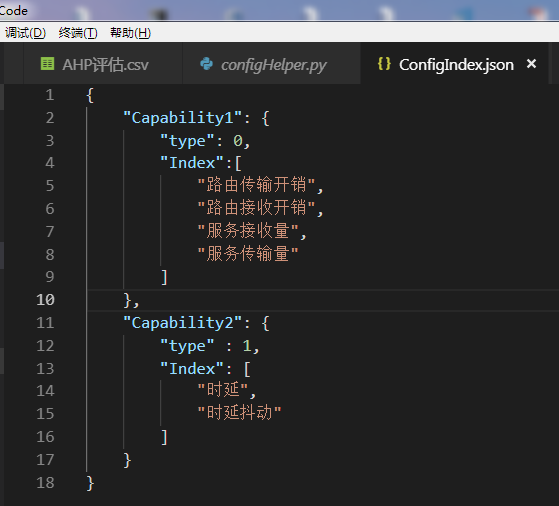
|  |  |  |
| --- | --- | --- |
| **配置说明** | **评估指标灵敏度分析** | **设计要素灵敏度分析** |
| 输入 | 多组指标数据和指标数据对应的综合评估结果。 | 多组网络设计参数和对应的综合效能的评估结果。 |
| 输出 | 各个指标变量对综合评估结果的影响的重要性程度的得分。 | 各个网络设计参数对综合评估结果的影响的重要性程度的得分。 |
| 算法 | 输入变量和输出的综合效能评估结果的随机森林回归模型，需要配置随机森林的训练的树的个数，树的深度，训练运行的进程数。 | |
| 说明 | 算法可以变换不同的输入（如改变评估方法），则可以输出不同情况下的指标灵敏度各网络设计参数灵敏度的情况，具有通用性和灵活的配置性，可针对性的对网络性能指标或者网络设计参数进行优化设计，同时也可以对综合评估方法进行对比分析。 | |

#### 计算样例

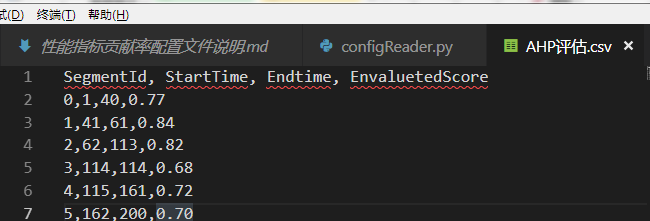
如图7， 8， 9所示，分别表示为指标数据的截图，指标数据配置文件的截图，以及指标数据分段评估后的结果



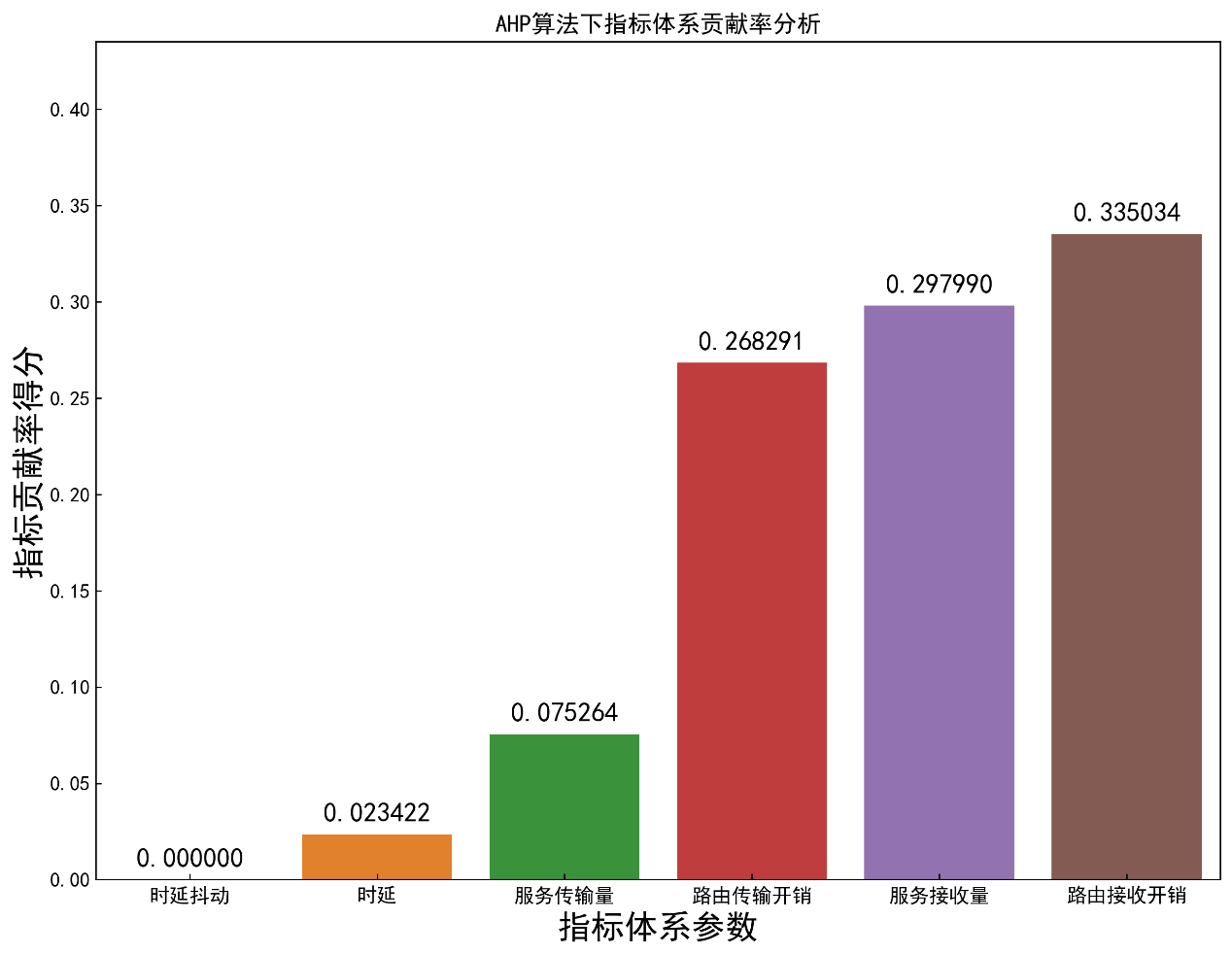
**图 7 部分指标数据截图**



**图 8 指标配置文件截图**



**图 9 分段评估结果截图**



**图 10 贡献率分析结果**

如图10，由模型灵敏度分析的结果可以看出，在使用AHP算法进行综合评价时，服务接收量，路由接收开销和路由传输开销的指标占有比较高的重要性，同时也能看出各个指标参数的重要性的排名。该灵敏度分析的模型中只需要指定自变量的参数和因变量的参数和相应的数据，就可以对自变量对因变量的影响进行灵敏度分析。

对于指标体系综合评估的灵敏度的定量的分析结果，可以明显看出各个指标在综合评估方法下的重要性的得分情况，因此根据灵敏度分析的结果，可以将对系统综合评估优化的方向，具体到要实际优化提高的性能指标上，即根据灵敏度分析的结果优化重要性比较高的指标，可以更容易的使系统的综合评价得分升高，亦即综合效能更优。