# Title ：基于仿真数据驱动的空间信息网络建模方法

## 摘要 ：

随着复杂系统复杂度的不断增加，传统的复杂系统建模与仿真的方法已经难以满足系统仿真研究的需求，因此对于空间信息网这样的复杂网络整体，本文提出了一种基于仿真数据驱动的空间信息网络设计建模方法，分析了空间信息网络的性能包含网络的底层构建要素和网络的业务的运行类型两个因素的整体性关系模型，通过多次仿真获取业务输出指标和网络设计要素参数的大量的数据，通过仿真数据的建模分析，构建网络设计关键要素和网络系统效能的关系模型。并给出了构建承载遥感数据业务传输的空间信息网络设计的案例分析，通过对业务输出指标的综合效能评估对业务的效能给出一个量化的分析结果，同时将效能评估的结果和网络设计参数进行机器学习的模型的构建，从而直接通过网络设计参数映射到业务性能的综合效能的表现，可以更快的解释各种网络设计参数对承载的业务效能的影响，进而对空间信息网络设计给出一个更为明确的指导方向。

## 一：介绍和引言

### 1 空间信息网的复杂性

空间信息网络是以空间平台（如同步卫星或中、低轨道卫星、平流层气球和有人或无人驾驶飞机等）为载体，实时获取、传输和处理空间信息的网络系统。空间信息网络通过组网互联，实时采集、传输和处理海量数据，实现卫星遥感、卫星导航和卫星通信的一体化集成应用与协同服务[1]。所以，空间信息网络是一个规模巨大、时空跨度大、异质异构的复杂网络，在构建的过程中，其顶层架构、网络模型、通信机制、网络协议设计、网络管理、安全机制、网络性能分析方面都面临巨大的挑战[2]。

### 2 传统复杂系统建模仿真的瓶颈

复杂系统建模与仿真已经成为研究各类复杂系统的最佳手段之一[3]，目前常用的复杂系统建模与仿真的方法[4] 有: 基于智能技术的复杂系统建模与仿真, 如遗传算法, 神经网络等方法; 基于数学手段的复杂系统仿真方法, 如参数优化方法、模糊仿真方法等; 基于离散事件动态系统的复杂系统建模与仿真, 如Petri 网、任务/资源图建模等等.传统的复杂系统建模仿真的研究思路如图 [1](#fig:SystemSim)所示, 通常采用类比方法,即基于相似性理论, 将复杂系统化解成多个简单的系统, 先进行子系统的构建, 再形成一个大系统. 但是, 这种通过局部子系统仿真还原全局系统、逐渐逼近原系统的传统建模方式, 是站在还原论角度，把复杂的系统简单化, 用简单的运动规律来代替高效运动的规律, 将复杂系统不断分解到最小单位, 通过解构系统还原单个节点和链接的理论, 随着系统复杂度的增强, 传统的仿真建模方法已不能从本质上正确认识复杂系统。因此，对于空间信息网络这样一个复杂系统的研究，传统的复杂系统建模与仿真的方式，并不能完全反应其真实的情况。

图 1: 传统的复杂系统建模仿真研究思路

图 1: 传统的复杂系统建模仿真研究思路

### 3 引出本文的方法

随着信息技术的不断发展，大数据，人工智能等新技术的出现，也提出了将大数据方法与仿真建模方法相融合的新的建模仿真的思路，基于大数据对复杂系统进行整体性的研究,两者结合将使仿真建模方法更能胜任于复杂系统研究[5]。所以对于空间信息网络的复杂系统研究，本文提出了基于仿真数据驱动的空间信息网络建模方法，将仿真与数据结合起来对空间信息网进行整体性建模分析。

由于影响复杂通信网络性能的因素可以主要分为两个方面：一是网络本身的特性，如网络拓扑结构、节点设备性能、以及通信链路和通信协议等；二是网络业务类型和业务量的因素，而且在实际的网络运行过程，网络本身的因素和网络业务的因素又会相互作用相互影响，如图 [2](#fig:Intro)所示。因此本文在空间信息网建模分析的过程中，从网络本身特性的因素和业务特性的因素的相互影响的关系出发，在网络的关键设计要素参数不断变化的情况下，通过网络仿真系统搭建的空间信息网络仿真场景，不断地输出其上层业务的性能指标，并以网络设计要素参数和仿真系统产生的业务性能输出指标的数据为驱动，分析空间信息网关键网络设计要素对承载业务性能的整体性的关系模型，从而对空间信息网的设计提供一定的指导方向。

图 2: 网络和业务相互影响关系

图 2: 网络和业务相互影响关系

## 二：基于仿真数据驱动的空间信息网络建模方法

空间信息网络结构复杂，在推进空间信息网建设中需要对诸多关键技术进行分析验证，空间信息网络面对海量的数据传输及资源需要较高运转速度的网络支持，因此对于空间信息网构建过程中网络关键技术的研究与分析是很关键的环节。所以本文提出了如图 [3](#fig:ModelPro)所示的空间信息网络建模方法。

图 3: 仿真数据驱动的空间信息网建模方法

图 3: 仿真数据驱动的空间信息网建模方法

建模分析的流程如下： 1. 对空间信息网络构建过程中的网络关键技术进行分析（如带宽分配、组网等），提取出关键技术的设计要素（如业务接入带宽、通信路由方式等）； 2. 将提取出的关键设计要素，进行参数设计，搭建空间信息网网络仿真场景，通过空间信息网络仿真平台对场景进行仿真，同时对相应的性能指标（如丢包、时延、抖动、吞吐量等）进行统计； 3. 对关键设计要素的多组参数，进行多次的仿真输出其相应的性能指标， 因此设计要素的参数和相应的仿真的数据输出就构成了即构成了空间信息网设计分析的密集型的数据集； 4. 对仿真输出的性能指标数据进行数据的特征分析，按照设计的目标对仿真输出的性能指标进行综合效能评估； 5. 将仿真输出指标的综合的效能评估和关键设计要素的参数进行机器学习的建模分析，所以，可以从整体上评估分析网络的关键设计要素参数对空间信息网的影响，从而，继续指导空间信息网网络关键技术的研究。

### 1：仿真指标综合效能评估

仿真输出的业务的性能指标往往是一个多维的，并不能通过某一指标单一的综合评估效能的好坏，因此需要对高维的性能指标进行综合评估，用一个综合评估值的变化来反应输出统计指标的变化，同时为了更直接的反应网络效能变化的方向，综合评估的效能值应该是单调递增或者单调递减变化的。同时为了体现不同的设计目标，对系统输出指标进行综合评估的时候，可以根据不同的设计要求和目的，采取不同的主观权重赋值方式，同时统计输出的业务的指标之间本身就有一定的相关的变化关系，比如时延的增大，相应的丢包率也会受一定的影响。所以也可采取客观赋权的方式，从输出的指标之间的数据表现出来的关系出发，而给出输出指标的一个综合的评估值。可使用的典型方法有主观的如：层次分析，客观有：主成分分析，因子分析等。

#### 1 层次分析法的介绍（介绍层次分析法综合评价的过程）

层次分析法(Analytic Hierarchy Process)将主观分析与客观分析相结合，适用于评价因素难以量化且结构复杂的评价问题。在评价过程中，使用AHP 算法可以对定性指标进行定量分析。AHP 算法的基本思路是：首先找出影响某问题所涉及的主要因素，按其重要性将各因素组成一个层次结构模型，通过对问题中各方案的两两比较，从定性的角度考虑，来确定每个方案的相对重要性，并将该定性评价转化为定量评价,其评价过程如图 [4](#fig:AHP)所示。

图 4: 层次分析法综合评价过程

图 4: 层次分析法综合评价过程

层次分析法AHP 结合了定性和定量的分析元素，既考虑到参与人员的偏重，又参照实际数据。AHP 是一种比较有用的决策分析方法，有许多优点，但是AHP算法本身有很多缺点，例如，AHP应用只能从已有的方案中进行选择，无法进行自主提出新方案；同时AHP方法需要专家系统来支持，若专家系统中的评判指标之间的关系不正确的话，则最终的结果将会指向一个错误的方向。

#### 2 主成分分析综合评价的介绍

主成分综合评价是一种通过降维技术把多个指标化为少数几个综合指标的统计分析方法。其基本思想是：设法将原来众多具有一定相关性的指标，重新组合成一组新的相关无关的综合指标。通过主成分分析，将原来相关的各原始变量变换成为相互独立的主成分，进而对这些主成分进行综合评价，可以消除由于指标间相关而在评价时反映的重复信息；其次，主成分综合评价的权重是从评价指标包含被评价对象分辨信息多少来确定的，具有客观性。同时，通过主成分分析，所取主成分个数小于原始指标个数，评价对象个数的减少，不但方便综合评价，也简化计算，其分析过程如图 [5](#fig:PCA)所示。

图 5: 主成分分析综合评价过程

图 5: 主成分分析综合评价过程

但是主成分分析综合评价也有一定的弊端，由于其权重信息是完全从数据中提取的，所以其在综合评价时对数据的样本量有一定的要求，进行主成分综合评价也没有考虑指标本身的相对重要，其权重信息更多的来自于样本数据的量，而非质，同时，实际中主成分的权重可能会出现负权数，与实际情况相违背，虽然可以在对主成分进行解释时进行了探讨，但更多的停留在整体性的非定量的说明上，而且有时候也会出现对于主成分难以解释的情况。

#### 3 因子分析法的介绍

因子分析综合评价是从研究变量内部的相关的依赖关系出发，把一些具有错综复杂关系的变量归结为少数几个综合因子的一种变量统计分析方法。它的基本思想是将观测变量进行分类，将相关性较高，联系比较紧密的分在同一类里面，而不同类之间的相关性较低，每一组分类变量代表一个基本结构，即为公共因子。通过对变量进行公共因子的提取后，即可用综合的公共因子来计算个变量的因子得分，同时各个因子得分又根据各因子的方差贡献率为权重，计算其综合的因子得分，即为因子分析综合评价的综合得分。其分析过程如图 [6](#fig:FA)所示。

图 6: 因子分析综合评价过程

图 6: 因子分析综合评价过程

因子分析的意义在于简化数据结构，通过科学的定量分析构造一个统计上优良的指标体系，然后对评价对象进行综合评价。但是因子分析法只适用于相关性比较强的变量，所以使用其分析方法的准确性与其观测数据的特定的模型结构有关。

### 2：效能评估结果和网络设计参数的建模方法

对于效能评估结果和设计参数，可以使用机器学习的方法学习出两者的映射关系。针对不同设计目标的，效能评估的表现形式不一样，所以可以根据效能评估结果的表现形式，可以很容易将其转化为机器学习的回归或者分类问题的求解。在机器学习中最突出、最常用的算法有线性模型、随机森林、神经网络三类。

#### 1 线性模型：

线性回归是处理回归任务最常用的算法之一。其算法的形式十分简单，期望使用一个超平面拟合数据集（只有两个变量的时候就是一条直线）。如果数据集中的变量存在线性关系，那么其拟合效果就会非常好。在实际使用中，简单的线性回归通常被使用正则化的回归方法（LASSO、Ridge 和 Elastic-Net）所代替。正则化是一种对过多回归系数采取惩罚以减少过拟合风险的技术。当然，还得确定惩罚强度以让模型在欠拟合和过拟合之间达到平衡。

优点：线性回归的理解与解释都十分直观，并且还能通过正则化来降低过拟合的风险。另外，线性模型很容易使用随机梯度下降和新数据更新模型权重。

缺点：线性回归在变量是非线性关系的时候表现很差。并且也不够灵活以捕捉更复杂的模式，添加正确的交互项或使用多项式很困难并需要大量时间。

#### 2 随机森林

随机森林首先随机选取不同的特征(feature)和训练样本(training sample)bagging，生成大量的决策树，然后综合这些决策树的结果来进行最终的分类回归，是一种基于决策树基模型的集成学习方法，其核心思想是通过特征采样来降低训练方差，提高集成泛化能力。适用于数据维度相对低（几十维），同时对准确性有较高要求的情况。

随机森林具有在数据集上表现良好，在前先很多数据集上要优于现有的很多算法；可以并行，训练速度相对较快；防止过拟合；能够处理高维特征，而且不用做特征选择，可以给出特征重要性的评分，训练过程中，可以检测到特征之间的相互影响。同时在树越多的情况下，随机森林的表现才会越稳定，而且在不平衡的数据集上，分类结果或者变量重要性会倾向于样本多的类别，所以训练样本中各类别的数据必须相同。

#### 3 神经网络

神经网络是一种监督学习算法，$ f(.) : R^m R^o$ 通过对数据集进行训练来学习函数，其中是输入的维数，是输出的维数。给定一组特征 和一个目标，它可以为分类或回归学习一个非线性函数逼近器。神经网络在研究的过程中提出了很多模型, 它们之间的差别主要表现在研究途径、网络结构、运行方式、学习算法以及算法的应用上。其通用的结构模型在输入层和输出层之间可以有一个或多个非线性层，称为隐藏层。图 [7](#fig:MPLRes)显示了一个带标量输出的隐藏层神经网络模型。

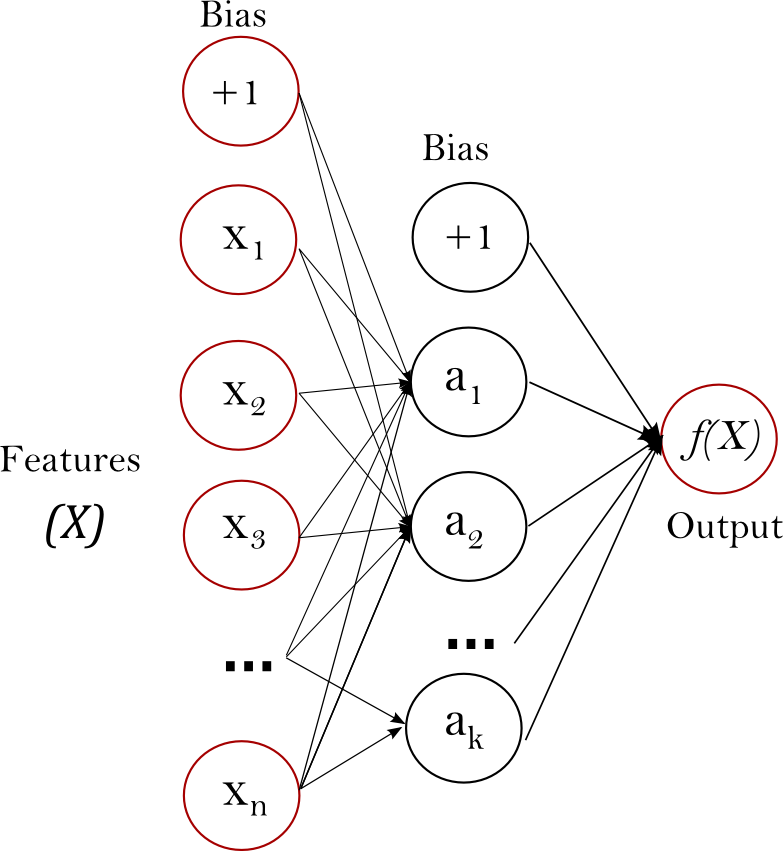


图 7: 一个隐藏层的神经网络模型

神经网络具有适用于分类或者回归问题的精确建模，对数据的关系不做任何假设的优点，但同时也有计算量大，训练缓慢，容易出现过拟合或者欠拟合，同时对于不同超参变化比较大。

#### 4 学习模型的对比和选择（分析对比介绍的方法的选取的考虑）

## 三：案例分析

由于空间信息网提供卫星遥感、卫星导航和卫星通信的一体化集成应用与协同服务，所以基于以上分析提出的建模方法，给出一个低轨卫星通信星座承载遥感数据业务传输性能评估分析建模的案例，以分析遥感卫星和通信卫星联合服务的性能和关键技术要素的关系。

### 1 使用的仿真工具

搭建了基于STK(System Tool Kit)和Exata的空间信息网仿真平台，其中STK进行卫星轨道，星座的仿真，Exata进行网络通信的仿真，并输出统计的性能指标。

### 2 仿真场景的目标

由于需要对低轨卫星通信星座承载遥感数据业务传输性能评估分析建模，所以搭建的仿真场景如图 [8](#fig:Scenario)所示。

图 8: 仿真场景

图 8: 仿真场景

其中遥感数据传输的业务，遥感数据由遥感卫星产生，数据的传输通过遥感卫星接入低轨卫星通信星座，经过低轨卫星通信系统的星间路由的方式，传输到遥感数据目的地面站的接入卫星，并通过该卫星转发到地面站。同时，为了仿真整个系统的运行，所以遥感卫星和通信卫星均以星座的方式存在，而且同一时刻可能有多条遥感数据在发送。同时为了体现系统的随机性，所有每次每个仿真场景进行多次仿真，并将每次仿真的统计数据都记录下来。

#### 1 流量模型的介绍

遥感数据业务的流量模型：遥感数据业务具有传输数据量大，数据包尺寸大，对于快速回传的遥感数据信息，还具有传输速率快的特点。基于以上几点的特征构建了遥感数据业务的流量模型，单个数据包大小为64KB，对于遥感数据包的传输采用巨帧的形式进行传输，每次数据的发送量按照均匀分布设置在200MB–1.2GB的范围，对于遥感数据的传输的起始时间服从指数分布的模型；同时对于遥感数据源的选择，按照均匀分布从遥感卫星星座中进行选择，其目的的地面接收站固定在国内不同的地方。

#### 2 仿真网络的介绍

#### 3 网络设计参数和输出指标参数的选取

将遥感卫星的遥感数据通过接入低轨卫星通信星座进行传输需要涉及到带宽分配、空间链路性能、空间路由方式、遥感数据压缩传输等关键技术，因此考虑的网络设计参数有：遥感业务的接入带宽、星间链路的带宽、空间链路的丢包率、星间路由方式、数据发包速率。同时为了评估分析遥感数据传输的性能，选取的统计输出指标有： 数据吞吐量:

接收数据量，接收数据持续时间。 端到端通信时延:

所以接收数据包的时延，单个数据包的时延为数据包从接收端收到的时间减去数据包从发送端发出的时间，接收到的数据包的个数。 数据丢包率：

数据发送量， 数据接收量。 数据响应时间：

第一个数据包收到的时间， 第一个数据包发出的时间。 数据持续时间:

最后一个数据包收到的时间,第一个数据包发出的时间。

## 四：仿真的数据分析

### 1：数据预处理的方法（简单介绍数据预处理的方法，无量纲化，丢失处理的方法）

### 2：效能评估的结果

### 2：机器学习模型的结果

## 五：结论总结

本文提出的基于数据驱动的空间信息网络建模方法，将空间信息网运行的业务综合效能评估和网络设计参数进行映射关系建模，通过仿真数据的训练可以学习出其关系的模型，通过构建该模型，可以将网络设计的关键设计要素，直接映射到综合效能表现的结果，从而可以在网络设计构建过程中更为直接，便捷的指导网络设计的进行。同时给出了一个构建遥感数据业务传输的空间信息网的设计要素分析的案例，案例分析了，业务接入带宽、业务数据发包速率、链路丢包率、网络路由方式、以及星间链路带宽对遥感数据业务传输性能的影响，同时选取了业务响应时间、吞吐量、时延、丢包率、以及持续时间作为遥感数据业务传输的性能指标参数，同时对该性能指标参数，使用了因子分析的综合评价的方法，不仅可以提取出对遥感数据业务传输的性能影响的隐性的因子结构，分析其指标之间的结构关系，也可以输出的性能指标进行综合的效能评估，并将综合评估效能的结果和网络的设计参数进行机器学习的建模分析，分析指出，在考虑的设计要素中，业务的接入带宽是对遥感数据业务传输性能影响的主要因素，同时为了更高的综合效能的体现，业务的发包速率也要和业务的接入带宽匹配。

## 参考文献

[1] 李德仁, 沈欣, 龚健雅, 等. 论我国空间信息网络的构建[J]. 武汉大学学报(信息科学版), 2015, 40(6): 711–715.

[2] 常青, 李显旭, 何善宝. 我国空间信息网发展探讨[J]. 遥测遥控, 2015, 36(1): 1–10.

[3] 沈爱民中国科协学会学术部. 复杂系统建模仿真中的困惑和思考[M]. 中国科学技术出版社, 2012.

[4] 徐庚保, 曾莲芝. 基于仿真的复杂系统研究[J]. 计算机仿真, 2013, 30(2): 1–4.

[5] 胡晓峰. 大数据时代对建模仿真的挑战与思考[J]. 军事运筹与系统工程, 2013, 27(4): 5–12.