基于仿真数据驱动的空间信息网络建模方法

**摘要：**针对传统的复杂系统建模仿真方法不在适应复杂度极高的空间信息网络系统仿真建模研究的问题，*提出了一种基于仿真数据驱动的空间信息网络建模方法。通过仿真数据的驱动，对空间信息网进行整体性的建模分析，并在此基础上，设计了空间信息网络性能指标的深度学习自编码网络和随机森林回归的混合预测分析模型。针对仿真数据中空间信息网络性能指标维度高，样本分布宽的特点，利用深度学习中深度自编码网络的非线性降维和数据重构的能力，进行空间信息网络性能指标编码解码网络的构建，结合随机森林回归模型构建空间信息网网络设计参数到性能指标的关系模型。最后给出了一个案例分析，表明该混合模型不仅可以直接预测空间信息网络性能指标，也能够对网络设计参数进行灵敏度分析。*

**关键词：**仿真数据驱动；空间信息网建模；深度自编码网络；随机森林回归

Simulation Data-Driven Model Approach for Space Information Network

**Abstract：**Aiming at the problem of non-adaptive and complexity for the research on the simulation and modeling of the Space Information Network, with high complexity, using the traditional complex system modeling and simulation method, *a Space Information Network modeling approach, based on simulation data, is proposed. Through the driving of simulation data, the Space Information Network is modeled and analyzed, and on this basis, a deep predictive Auto-encoder network, for Space information network performance indicators, and a hybrid forecasting analysis model, with random forest regression, are designed. Due to the characteristics of high dimensionality and wide distribution of performance indicators of Space information network in simulation data, the construction deep learning Auto-encoder network, which has the capabilities of nonlinear feature extraction and dimensionality reduction, of Space information network performance index is realized. Combined with the random forest regression model, the relationship model between the network design parameters of Space Information Network and the performance indicators, is built. And a case is given, it shows that the hybrid model can not only directly predict the Space information network performance indicators, but also can analyze the sensitivity of the network design parameters.*

**Keywords：**Driven of Simulation data, Space Information Network Modeling, Deep Auto-Encoder Network, Random Forest Regression

引言

空间信息网(SIN, Space Information Network)作为国家重要基础设施，是以空间平台（卫星、平流层气球、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[1]。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]。

空间信息网建设的目标是作为未来的综合信息服务系统，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]，因此空间信息网是一个复杂度很高的复杂网络系统。其复杂性体现在几个方面：空间信息网系统内各个单元系统的复杂性，构成空间信息网的单个的服务系统本身就是一个复杂系统，如卫星通信系统、遥感卫星系统；系统内各个系统单元之间关系的复杂性，各个单元系统之间需要进行融合，异构的设计，在提供服务时，系统单元系统进行联合工作；系统任务的复杂性，空间信息网提供一体化协同的服务，其业务和种类具有多样性，体现在业务的种类的变化，质量要求的等级，以及动态性和时变性上。空间信息网的复杂性主要表现在其网络系统结构本身的复杂性和空间信息网业务应用的复杂性上。系统结构的复杂性和业务应用的复杂性使得对空间信息网进行直接类比的系统建模研究的任务变得更加复杂；同时也面临在进行系统优化时，目标与约束的表现形式也变得复杂，问题规模选取模糊，仿真局限等方面的问题[3]。空间信息网络作为信息网络学科的发展前沿[4]，对其进行系统建模研究和综合的评估分析空间信息网络各个网络设计组件之间的关系是有效地指导空间信息网络的构建的关键[5]。

# 相关工作

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[6]，关于空间信息网络的建模仿真研究目前主要集中在空间信息网的体系结构建模上，如文献[7]类比了以往对天基综合信息网、空间信息系统等研究对象开展的描述和建模技术，引入数据即中心DaaC(data as a center)思想，综合基于活动、面向服务、面向对象等建模的实现，进行基于DaaC的空间信息网络体系结构建模；文献[8]对空间信息系统仿真系统的组成及体系结构进行了研究，通过构建仿真系统，从仿真的角度给出空间信息系统的体系结构和系统仿真建模开发方法。

在针对具体的空间信息系统单元内的局部问题，文献[9]中对空间信息网络的低轨卫星通信网络中的路由算法进行了研究分析，讨论了不同的基于IP的路由协议在面对空间网络环境的适用情况；在具体的空间任务系统建模研究上，文献[10]针对的空间信息网的支援任务问题，基于空间信息支援体系特点，提出将多Agent、多分辨率和复杂网络相结合的一种建模方法，构建了空间信息支援体系模型框架，建立了适用于空间信息支援体系网络的多Agent 模型、多分辨率模型和复杂网络模型。为了更直观地对空间信息网进行评估展示，文献[11]使用超网理论构建了空间信息网评估分析模型，利用超网理论对空间信息网进行抽象表示，构建空间信息网可视化表示评估模型，通过直觉上的可视化的结果来表示模型的评估结果。

从已有的对空间信息网络的相关的系统建模相关研究中，可以发现目前的空间信息网的建模研究存在一定的局限性：没有对空间信息网建模研究的定量的计算和表示；对面向完整的空间信息网系统任务的分析和优化缺乏完整的系统模型的建模研究；因此，需要探索新的方法和试验不同的工具来解决空间信息网系统建模和优化的研究问题。

随着大数据，人工智能，深度学习等新技术的出现，对复杂系统的研究，可以通过观测系统的数据收集，并运用大数据、机器学习等技术从数据中发现系统的结构和规律，从而为复杂系统的研究提供了新的途径。同时针对复杂系统仿真建模，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[12]。

因此本文结合机器学习、深度学习的方法，提出了基于仿真数据驱动的空间信息网络建模方法。通过局部系统模型的构建与组合，搭建整体的空间信息网的仿真系统，根据空间信息网设计的任务目标，设计仿真实验参数，通过搭建的仿真平台对空间信息网络模拟，并对不同的设计参数和目标对空间信息网进行多次仿真，并不断运行空间信息网的仿真系统，通过多次仿真获取更多的空间信息网不确定性数据，输出相应的空间信息网统计指标，并以仿真统计输出的数据和网络设计参数为驱动，对空间信息网络的系统建模进行整体性的建模分析。

# 基于仿真数据驱动的空间信息网络建模方法

传统的复杂系统仿真建模研究是以还原论为指导，以模型驱动为主要模式，即以模型假设为核心，通过实验验证该假设的合理性。但是随着系统的复杂度越来越高，以模型为驱动的模式已经不在适应其需求。因此对于空间信息网络系统建模研究，本文提出基于仿真数据驱动的空间信息网络建模方法，数据驱动的运行原理类似光场相机，即先对所有数据进行收集、存储，然后面向全样本数据进行处理、挖掘和分析。通过数据驱动的方式不仅可以绕开空间信息网络系统时变性、非线性、不确定性等原因造成的因果性的不易发现的问题，而且对刻画空间信息网络系统性能的指标数据进行了充分的利用，通过深度挖掘分析的方法对空间信息网进行整体性的系统建模分析。

## 主要组成

基于仿真数据驱动的空间信息网络建模的方法主要包括三项内容：空间信息网络的目标任务需求分析；仿真数据采集、处理与建模；目标任务建模预测结果的验证与分析。

目标任务需求分析。对于不同的空间信息网任务需求，研究的范围和层次结构不同，涉及的研究对象也不一样，因此需要对具体的空间信息网设计任务进行需求分析，需要确定其关键的设计要素（网络本身相关的和业务因素相关的），以及具体的目标任务的性能指标参数。

仿真数据采集、处理与建模。在本阶段要根据目标任务需求分析的结果，进行仿真实验参数的设计，并进行仿真场景的搭建，通过运行仿真场景，进行性能指标数据的采集；然后对采集到的数据进行数据处理，特征提取，模型构建和训练。

目标任务建模预测结果的验证与分析。在这一阶段要完成对建模预测结果的合理性、精准度、可信度的验证性分析，基于该模型预测结果，向前可使用一定的综合评价方法继续进行指标数据的效能评估，向后也可进行相关设计参数的灵敏度分析等。

## 建模分析流程

空间信息网络系统建模的目标是在任务目标的需求下，确定任务设计参数对空间信息网络系统的影响，其具体表现在指标参数的变化情况以及整体性能的影响。空间信息网的性能指标是一个多维的指标系统，指标与指标之间的关系有强有弱，从仿真数据中发现空间信息网络性能指标之间的隐含关系，以获得空间信息网整体性的特征和关系模式是基于仿真数据驱动的建模方法的难点。同时，空间信息网的设计参数也是一个多维度的变量，各个设计参数之间也有一定的强弱关系，因此解决这样的高维数据与高维数据之间关系模式的分析问题，是空间信息网络系统建模的重点。

深度学习[13]可以自主辨别数据底层隐含的特征，并且通过层次化结构，逐层获取输入数据间的各级抽象表达，近年来在语音识别、图像分类、自然语言处理等领域的应用取得了很好的效果。其概念源自人工神经网络，通过模仿人脑结构，自动对复杂输入数据进行逐层抽象，实现对数据较为本质的刻画，理论上能够对任何函数进行非线性拟合，同时深层的模型具有更强的泛化和推广能力。因此经过广泛地分析对比后，本文利用了深度学习的方法进行空间信息网的数据驱动的建模分析。

深度自编码网络DAEN(Deep Auto-Encoder Network)[14]是深度学习的一种，它通过构建具有多隐含层的机器学习模型对训练样本进行逐层特征变换，将样本在原空间的特征表示变换到一个新特征空间，达到降维压缩的目的，同时也可以通过解码网络将原本的样本数据进行重构。对于空间信息网的性能指标数据，如果直接对各个指标数据进行相应的建模分析，则会增加建模分析的复杂性，同时针对单个指标进行分析，会丧失空间信息网的整体性能信息。因此本文使用深度学习自编码的技术对空间信息网的性能指标进行编码和解码，通过编码可以提取出空间信息网性能指标的综合特征的整体信息，其编码的结果可以作为空间信息网的整体性能的体现，而经过解码也能重构相应的具体指标数据，从而能对相应的指标数据提供定量预测的功能。

随机森林(RF)[15]是一种非参数回归和分类的机器学习方法，是基于决策树基模型的集成学习方法，在处理高维小样本问题时具有一定的优势。同时由于随机森林扰动重要性指标既考虑了某输入变量在其全域上的变化对输出的影响也考虑了该变量与其它变量在全域上交互作用产生的影响，所以其扰动重要性指标是一种特别适用于高维数据或计算模型的全局灵敏度分析方法。因此，对于空间信息网设计参数也是高维输入变量的问题，本文利用的随机森林回归的方法构建空间信息网设计参数和空间信息网性能指标综合特征信息的关系模型，从而既能够预测在一定设计参数下的空间信息网的综合的特征信息，并将该信息经过自编码解码网络预测具体的性能指标数值，同时利用随机森林回归的扰动性重要评分也能对空间信息网络的设计参数进行灵敏度的分析。

因此本文提出的基于仿真数据驱动的空间信息网的建模分析方法的流程如图 1所示：

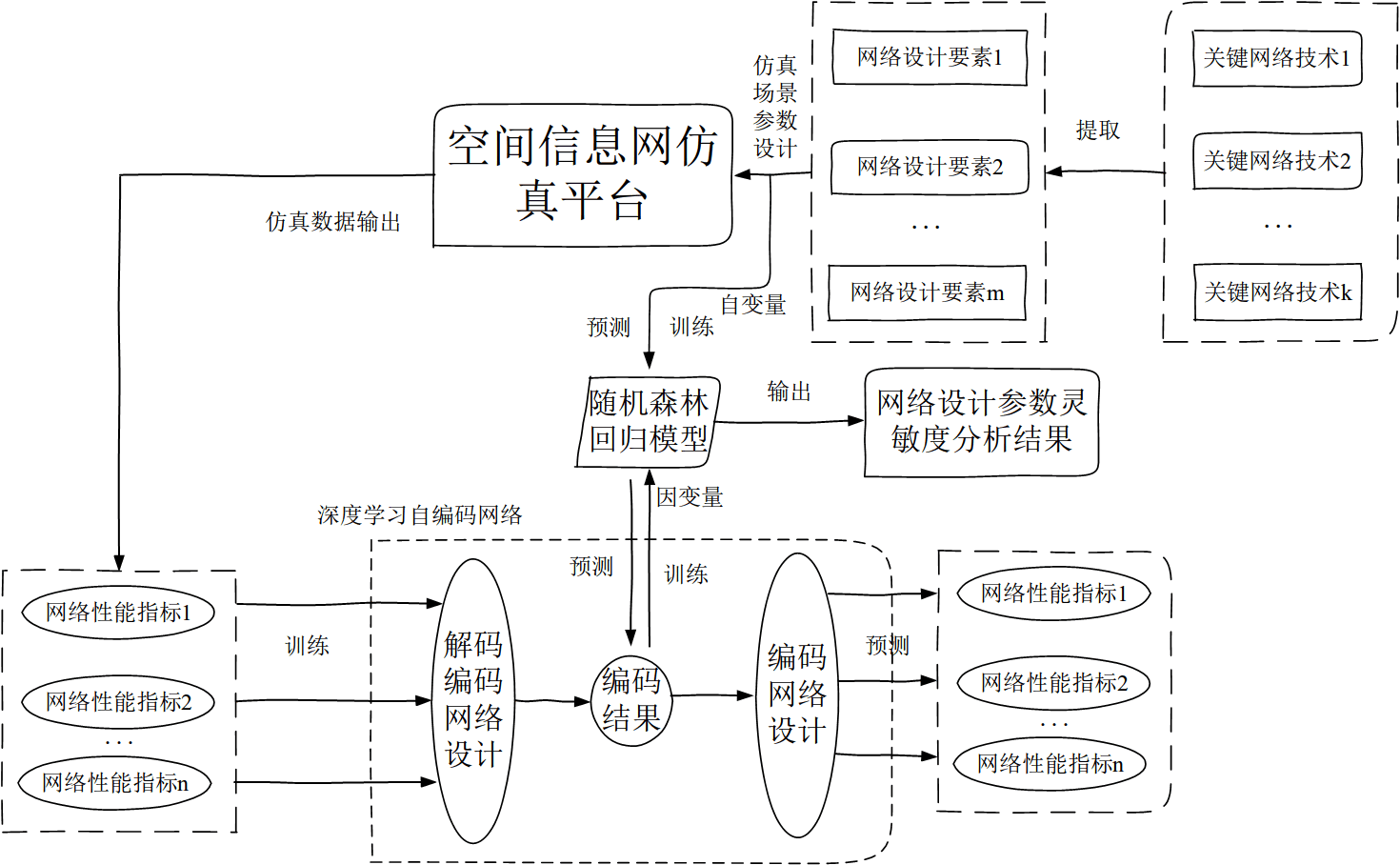


图 1 仿真数据驱动的空间信息网建模分析流程

Fig. The Procedure of Simulation Data-Driven SIN Model

建模分析的流程如下：

1. 根据任务目标需求，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，进行深度学习自编码网络的训练，生成空间信息网的性能指标数据的编码网络和解码网络；

5. 将空间信息网络性能指标数据的编码网络的输出结果和关键设计要素的参数进行随机森林回归的机器学习模型的构建。

由此，建模分析的步骤完成，对于空间信息网络的指标进行预测，通过将新的网络设计参数输入构建好的随机森林回归模型，预测出其相应的指标的编码结果，然后通过深度自编码的解码网络对预测的编码结果进行解码，重构出对应的具体的空间信息网的性能指标数据；对于网络设计参数对空间信息网整体性能的灵敏度的定量评分通过随机森林回归模型的变量重要性得分得到。因此，运用随机森林回归模型结合深度自编码网络的方法，可对在设计参数范围内的网络设计参数进行定量的空间信息网络的性能指标的预测，可避免复杂的网络仿真场景的搭建和运行，同时随机森林能对输入变量进行重要性的评分，所以可以对网络设计参数对空间信息网整体性能的影响进行灵敏度分析，从而对空间信息网设计进行优化提供方向。

## 建模组合模型

### 深度自编码网络简介

深度自编码网络也称栈式自编码网络，是由多层自编码器 AE(Auto-Encoder)堆叠而成的神经网络，训练采用逐层训练的方法，有效地解决传统神经网络训练方法不适用于多层网络训练的问题，整个深度自编码网络的训练分为预训练和微调两个阶段。

预训练为初始化网络参数的过程，采用逐层无监督特征优化的算法，初始化各层神经网络的层与层之间的连接权重以及各层神经元的偏置。以一个一层的编码器为例，其网络结构如图 2所示。

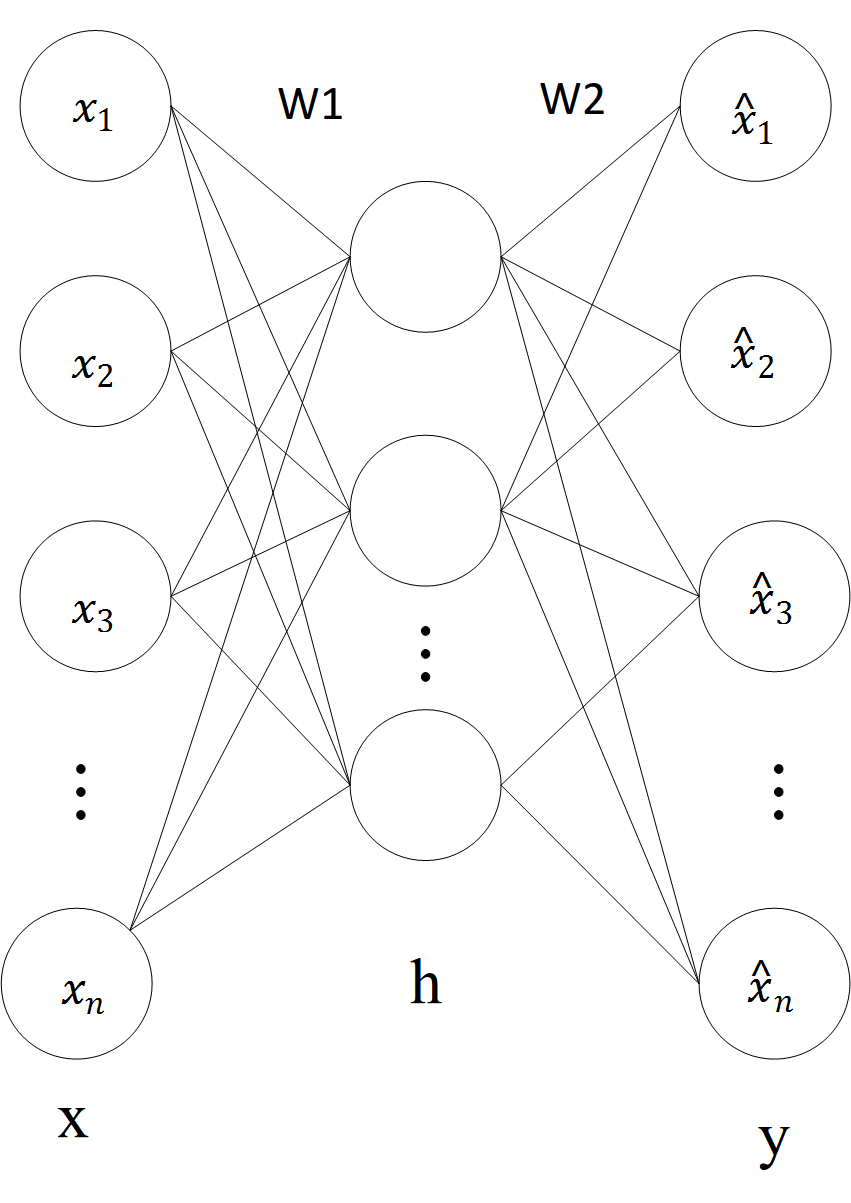


图 2 自编码（单层）网络结构

Fig. A Single Auto-Encoder Network Structure

单层的自编码器可视为一个3层的神经网络结构，即包含输入层，隐藏层和输出层，其中输入输出层的规模相同。从输入层到隐藏层是编码的过程，从隐藏层到输出层是解码的过程。设 和分别表示编码函数和解码函数，则编码过程为:，解码过程为:。

设训练样本集为，预训练过程为通过样本集的训练确定自编码网络权重和偏置参数使得解码后的 与输入 尽可能的接近，输入输出的接近程度同重构误差函数：



基于重构误差函数，针对训练样本集，利用梯度下降的方法对损失函数进行极小化，得到该层的自编码的网络权重和偏置参数。通过逐层对自编码网络进行训练，直到整个深度自编码网络训练完成。

在进行逐层训练时，单独训练某一层的自编码网络参数，会固定其他层的网络参数保持不变。因此为了得到更好的训练结果，在完成预训练后，需要利用的带标签的数据，通过后向传播算法同时调整整个深度自编码网络的所有层的参数，以达到全局最优。

### 随机森林回归模型简介

随机森林是一系列决策树组合，利用重抽样方法从原始样本中抽取多个样本，对每个样本进行决策树建模，然后将这些决策树组合在一起，通过投票得出最终分类或预测的结果。

假设复杂系统用确定型模型来表示，其中是输出因变量,是维相互独立的随机输入变量。用表示规模为的训练样本集。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。

建立随机森林的算法为:

通过实验或抽样的方法得到一组样本规模为 的训练样本 。

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。

用CART[16]法则建立第个决策树。

在叶子节点，模型输出值用该叶子节点内数据的平均值来确定。

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。

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称为袋外数(OOB)，袋外数据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，是输入变量的真实值。用表示第个树的对的估计值,则第个树的均方误差表示为:



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:



变量扰动后的均方误差相对于扰动前的均方误差为:



扰动重要性指标定义为所有树的的平均:



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输入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[17]。

### 组合模型训练分析过程

空间信息网性能指标的各个指标数据各自含有不同的单位量纲和物理意义，如果将各个指标数据直接联合进行分析，则训练分析的模型会受到数据的不一致性的影响，因此需要对指标数据进行预处理。因此对空间信息的性能仿真数据进行采集处理之后，整个模型的训练过程如图 3所示。将全部的样本数据按照一定的比例划分为训练集和测试集，其中训练集首先进行自编码网络的训练，然后，将训练集的自编码网络的编码输出结果，和指标数据对应的设计参数进行随机森林回归模型的训练，其中因变量是自编码网络的输出结果，自变量是网络的设计参数。最后的模型的预测过程是，将网络设计参数输入随机森林回归模型，输出其指标数据的编码结果，然后输入到自编码的解码网络对指标数据进行预测。测试集用于对组合模型的预测性能进行测试。同时对于测试性能满足要求的模型，训练集的随机森林回归模型的扰动性重要指标，作为网络设计参数对空间信息网整体性能影响的灵敏度的分析结果。

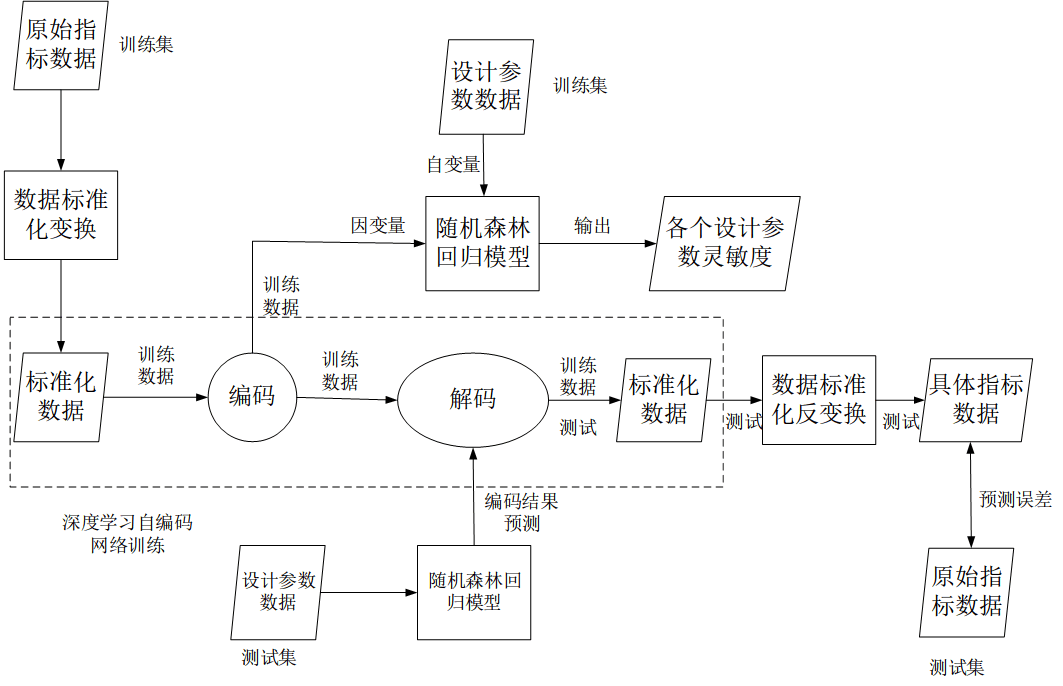


图 3 模型训练流程

Fig. The Procedure of Model Train

# 案例分析

由于目前我国仍无法实现全球建立卫星测控站，遥感卫星的数据回传，需要在遥感卫星经过国土地面上空时才能向地面站进行数据传输，所以传统的遥感数据的传输有较长的间断性，当应对紧急情况时不能及时响应，因此传统的遥感数据传输的方式已经不能应对遥感数据观测的及时响应快速回传的新的需求。所以在空间信息网建设的一个目标是将卫星遥感和卫星通信的一体化集成应用，对于该任务目标需要对低轨卫星通信星座承载遥感数据业务传输的场景进行性能评估和验证分析。所以本文基于上述分析提出的建模方法，以该任务为案例，对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进行建模分析。

## 仿真场景

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 4所示。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进行记录。

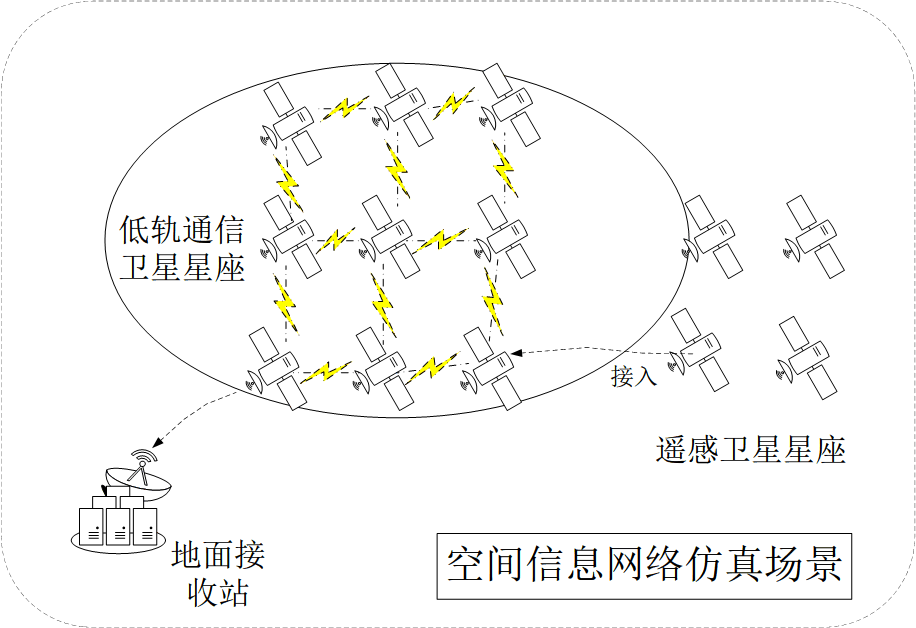


图 4 仿真场景

Fig. Simulation Scenario

## 仿真场景设置

搭建了基于STK(卫星仿真工具箱)和Exata(网络通信离散事件仿真器)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。

### 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；对于每次遥感业务流的数目以指数分布均值为5进行选择。

### 仿真网络的介绍

仿真网络里面的通信星座和遥感卫星信息均选用walker星座，通信卫星之间有星间链路可进行星间路由得转发，遥感卫星传送遥感数据时需要接入通信卫星才能进行数据发送，所以为了减少传送数据过程中频繁的链路切换，在接入通信卫星的选择时，以最长可见通信时间为准则进行选择。

### 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。

对低轨卫星通信星座承载遥感数据业务传输的目标，其性能除了基本的数据传输性能的要求外，还关注了其及时响应和快速回传的能力，所以本文选取了基本的数据传输性能传输指标：数据吞吐量、端到端通信时延、数据丢包率，加上数据响应时间和数据传输持续时间作为遥感数据业务传输性能的指标。各个指标的定义如下：

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。

# 仿真数据分析

## 数据处理和模型训练

本文对仿真场景进行了1200组网络设计参数的实验仿真，每次仿真实现分别修改不同的网络设计参数(遥感业务接入带宽，星间通信带宽，星间通信路由方式，空间链路丢包率)，同时，同样的网络设计参数使用不同的随机种子运行5次仿真，统计输出的指标选取了各条遥感数据流的数据吞吐量、端到端通信时延、数据响应时间、数据持续时间、数据丢包率，作为与网络设计参数相对应的网络性能输出指标的样本点。

实验网络设计参数如表 1所示：

表 1 网络设计参数范围表

Tab Range Table of Network Design Parameter

|  |  |
| --- | --- |
| 网络设计参数 | 参数范围 |
| 遥感业务接入带宽 | 40 ~ 1000 (Mbps) |
| 星间通信带宽 | 1000 ~ 2000 (Mbps) |
| 星间通信路由方式 | 1: 主动式，-1: 被动式 |
| 空间链路丢包率 | 0 ~ 0.01 |
| 数据发包速率 | 5 ~ 512 (Mbps) |

仿真统计的性能指标数据部分如表 2所示。

表 2 部分性能指标数据

Tab Part Data of Performance Indexes

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 序号 | 吞吐量(Mbps) | 时延(ms) | 响应时间(ms) | 持续时间(s) | 丢包率 |
| 1 | 101.90 | 135.19 | 136.44 | 95.06391 | 0.004 |
| 2 | 499.62 | 294.57 | 1039.49 | 10.4191 | 0.092 |
| 3 | 51.149 | 157.70 | 158.75 | 170.59 | 0.001 |
| 4 | 170.69 | 107.87 | 107.819 | 18.008 | 0.000 |
| 5 | 499.66 | 105.23 | 48.78 | 4.69 | 0.000 |
|  | ... | ... | ... | ... | ... |
| 20126 | 10.180 | 150.21 | 145.86 | 504.606 | 0.005 |

按照图 3所示的模型的训练过程，为了保留每个指标的分布范围的信息，使用了最大最小化的线性变换进行数据的标准化，既将数据的取值范围映射到[0,1]的值域，也对数据进行了无量纲化的处理；对于深度学习自编码网络，使用了Keras深度学习库搭建了{5,3,1,3,5}的深度自编码网络；同时，使用了scikit-learn机器学习库对自编码编码网络输出和设计参数的随机森林回归模型的构建。

## 建模结果分析

测试误差分析分别使用了平均绝对误差，解释方差得分和决定系数。

平均绝对误差：



解释方差得分：



决定系数：



由于主成分分析也具有特征压缩线性降维的能力，因此为了对比分析本文提出的方法的有效性，本文同时还利用主成分进行了指标参数进行特征压缩，同时将该特征压缩结果与网络设计参数进行随机森林回归模型建立，预测时先预测出主成分的综合特征，然后进行主成分分析的反变换，计算出对应的具体指标数据，即建立了主成分分析和随机森林的组合模型。

误差计算结果如表 3和表 4所示

表 3 深度自编码和随机森林组合模型误差分析

Tab Tolerance Analysis of Model Combined by Auto-Encoder and Random Forest

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 误差 | 吞吐量 | 时延 | 响应时间 | 持续时间 | 丢包率 |
|  | 4.32(Mbps) | 21.908(ms) | 75.99(ms) | 20.394(s) | 0.0007 |
|  | 2.47 | 3.792 | 8.786 | 11.84 | 3.09 |
|  | 0.8915 | 0.8526 | 0.8006 | 0.8685 | 0.8077 |
|  | 0.8749 | 0.8522 | 0.7404 | 0.8223 | 0.8077 |

表 4 主成分分析和随机森林的组合模型误差分析

Tab 4 Tolerance Analysis of Model Combined by PCA and Random Forest

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 误差 | 吞吐量 | 时延 | 响应时间 | 持续时间 | 丢包率 |
|  | 112.32(Mbps) | 489.908(ms) | 750.99(ms) | 192.8(s) | 0.23 |
|  | 62.1 | 84.72 | 88.006 | 106.63 | 100.09 |
|  | 0.52 | 0.2974 | 0.0013 | 0.0485 | 0.248 |
|  | 0.52 | 0.2962 | 0.009 | 0.0483 | 0.247 |

通过对比主成分分析和随机森林的组合模型预测误差的结果，本文构建的随机森林加深度学习自编码网络的空间信息网的性能指标的预测模型能够对空间信息网的性能指标的分布情况的解释分数均在80%以上，同时决定系数均在74%以上，表明了该模型对空间信息网的性能指标的可预测的比率在74%以上。同时结合平均绝对误差和测试均值的比值，给出该模型的预测的偏移的在11.84%以内。而如果利用主成分分析进行降维提取，由于主成分分析是线性的降维分析，所以进行预测时空间信息网指标参数中的非线性的信息损失较多，而且并不能对所有的指标数据有较好的预测结果，其表现为决定系数和解释分数有明显得高低的差距，同时其偏移量也较高。因此本文提出的自编码网络的模型对空间信息网的非线性的信息具有保留性，同时误差的计算的偏移结果表明，本文的模型对于直接预测一定网络设计参数下的空间信息网的性能指标参数具有一定的参考价值，同时也能给出性能指标预测的定量结果。

利用随机森林的回归模型中变量重要性评分对网络设计参数进行灵敏度分析的结果如图 5和图 6所示：

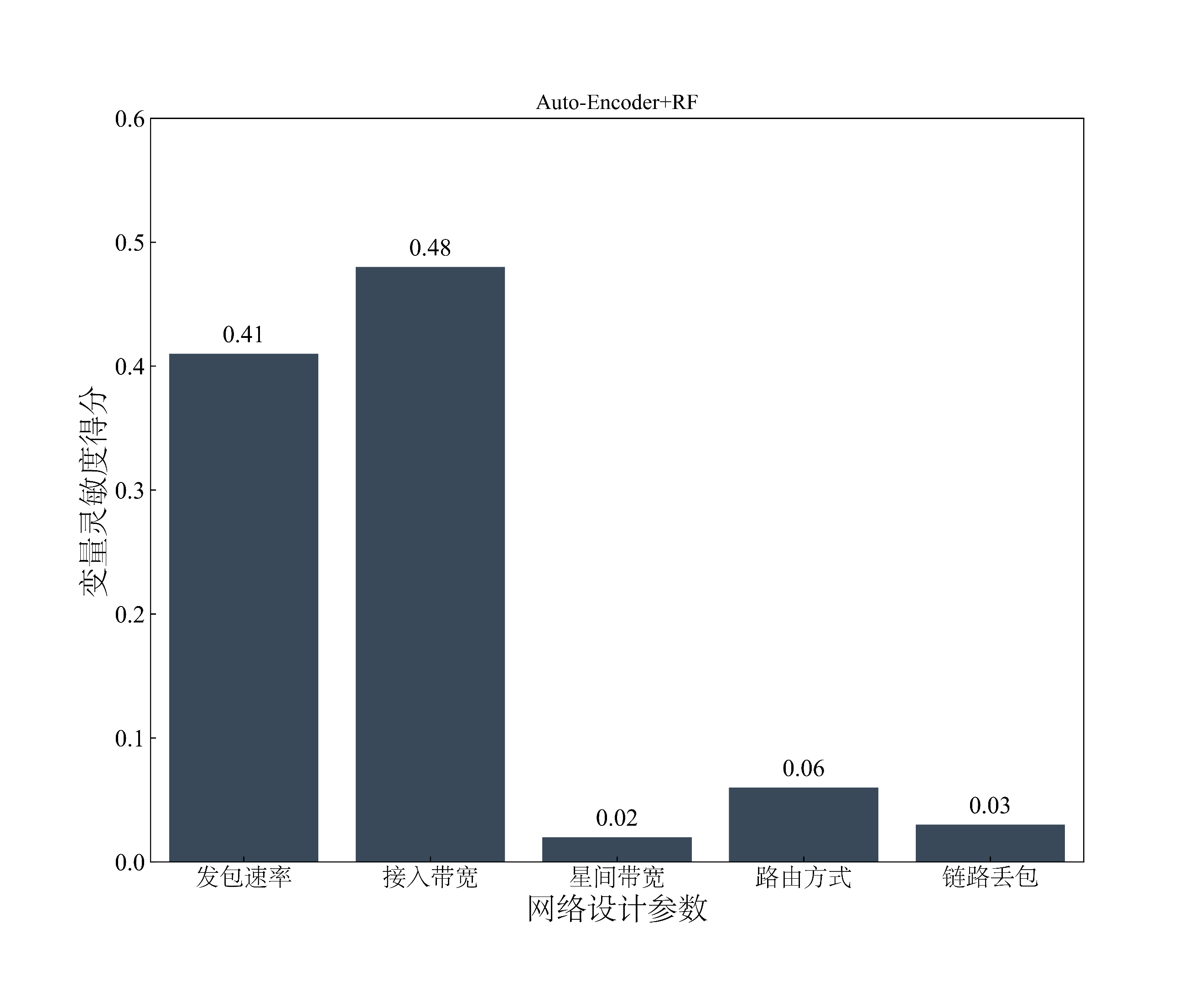


图 5 深度自编码组合模型灵敏度分析

Fig. Sensibility Analysis of Model Combined Auto-Encoder

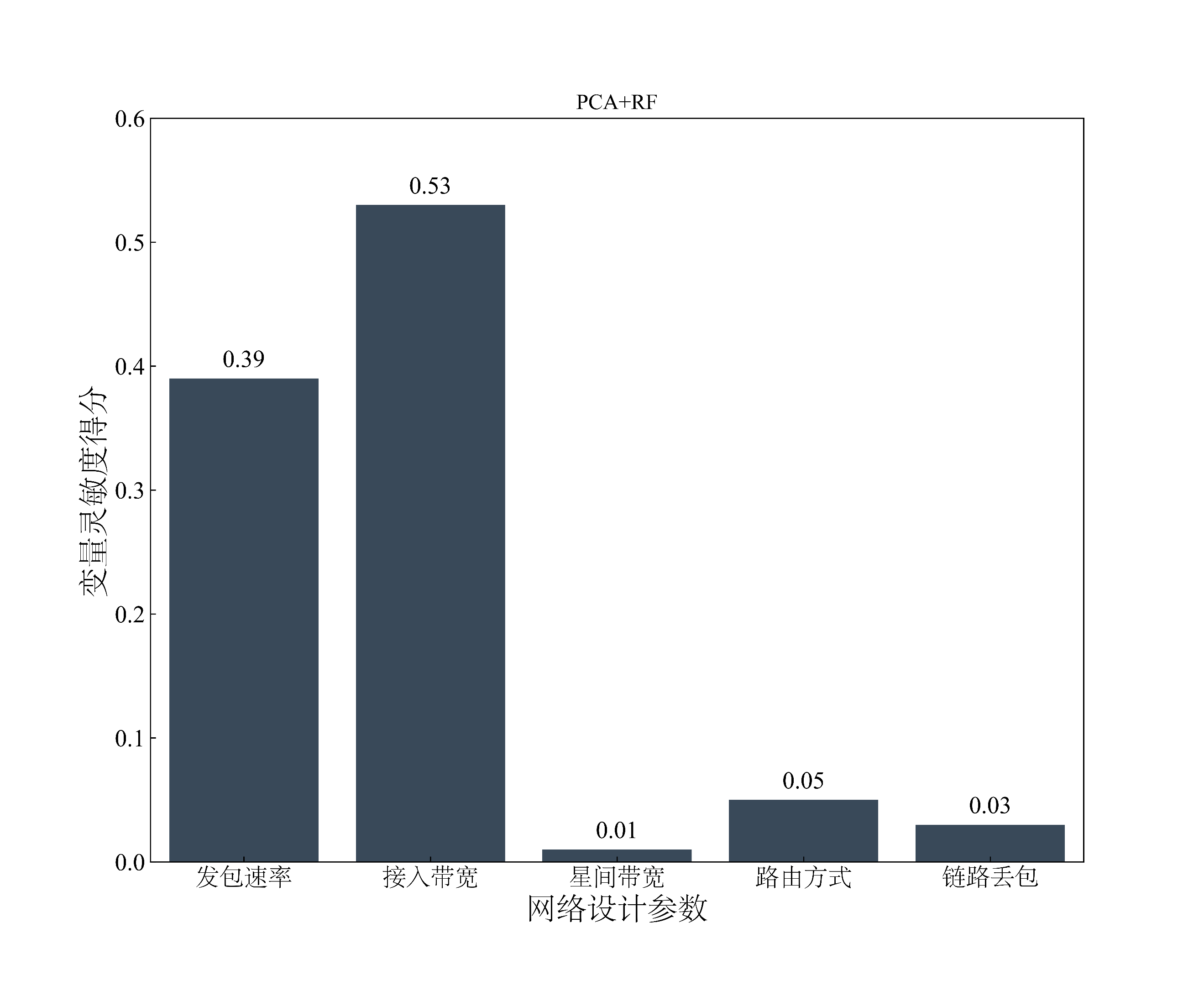


图 6 主成分组合模型中灵敏度分析

Fig. Sensibility Analysis of Model Combined PCA

由灵敏度分析的结果表明，对于低轨卫星通信星座承载遥感数据进行业务传输，不论是进行遥感数据传输业务指标参数的线性的整体性能特征提取（利用主成分分析），还是进行非线性的整体性能的提取（利用深度自编码网络），遥感数据业务传输的整体性能的影响程度排名均为：遥感业务接入带宽 > 遥感数据发包速率 > 通信路由方式 > 空间链路丢包率 > 星间通信带宽 。同时，其量化的重要性的数值表明，遥感业务的接入带宽和数据的发包速率不仅对遥感数据业务传输的性能的线性信息的影响相对较大，而且非线性的性能信息影响也较大，因此，在对低轨卫星通信星座承载遥感数据进行业务传输设计优化时，需要重点考虑其业务速率和业务接入带宽的因素。同时也表明了本文的模型的灵敏度分析的结果对于空间信息网的网络设计参数的设计和优化方向具有定性和定量分析的作用。

# 结束语

本文针对空间信息网建模研究直接进行系统建模分析地复杂性问题，提出的基于数据驱动的空间信息网络建模方法，针对空间信息网的设计目标，设计提取网络设计参数，并根据设计参数进行相应的空间信息网络场景的搭建和仿真，通过空间信息网仿真平台记录进行仿真并统计输出空间信息网性能指标参数，以网络设计参数和仿真统计性能指标参数为驱动，利用了深度学习自编码网络和随机森林回归的技术，对空间信息网进行整体性的系统建模分析。对空间信息网的性能指标进行深度学习自编码网络的编码压缩和解码重构，同时将编码压缩结果和网络设计参数进行随机森林回归模型的构建，不仅能对空间信息网性能指标参数进行直接定量预测，也能提供对网络设计参数进行灵敏度分析的功能，从而指导空间信息网的设计和优化。

对于空间信息网低轨卫星通信星座承载遥感数据进行业务传输性能评估分析的案例，使用本文提出的方法进行了建模分析，预测误差的计算结果表明，本文提出的方法对空间信息网的性能指标的定量预测对空间信息网的性能指标的参数的分布具有较强的解释性，同时其偏差也较低。灵敏度分析的结果表明，在空间信息网建设承载遥感数据业务的低轨卫星通信网络时，遥感业务的接入带宽和遥感数据的发包速率，是该业务整体性能的主要决定因素，同时也提供了对于所考虑的设计参数的灵敏度的定性和定量的结果。

通过案例分析，给出了本文提出的基于数据驱动的空间信息网络建模方法针对具体的任务进行分析的流程，在具体使用的过程中需要针对设计目标进行参数设计和选择，在进行性能指标参数的深度学习自编码网络的训练时需要进行深度学习参数调节以到达相应的预测精度。本文是对基于数据驱动的空间信息网的建模研究的初步的探索尝试，对于复杂的空间信息网建设的任务需求，后续还需采集分析更多的场景数据，使用模型展开进一步的分析验证，同时还将利用各种探索性分析的方法和工具对空间信息网的具体任务场景进行机理回溯的研究。

## 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.

LI D R, SHEN X, GONG J Y, et al. On construction of China’s space information network [J]. Geomatics and Information Science of Wuhan University, 2015, 40(6): 711-715.

1. 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.

CHANG Q,LI X X,HE S B. Confer on the evolution of earth-space integrated information network of china[J]. Journal of Telemetry Tracking and Command, 2015(1):1-10.

1. 杨雅婷. 卫星系统的数学模型以及面向任务的卫星系统优化[D]. 中国科学院大学, 2016.

YANG Y T. Mathematical model of satellite system and task-oriented satellite system optimization[D]. Chinese Academy of Sciences University, 2016.

1. JIANG C, WANG X, WANG J. Security in space information networks[J]. IEEE Communications Magazine, 2015, 53(8): 82–88.
2. LIU S, WANG H M, WANG S T, et al. Space-Based Information Integrated Network Technology and Performance Analysis Based on Cognitive Radio[M]. 2017.
3. 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.

Chinese Science Association. Puzzlement and Thinking in Complex System Modeling and Simulation [M]. Beijing, China: China Science and Technology Press,

1. 于少波, 吴玲达, 张喜涛. DaaC:空间信息网络体系结构建模方法[J]. 通信学报, 2017, 38(a01): 165–170.

YU S B, WU L D, ZHANG X T. DaaC: an architecture modeling of space information network[J]. Journal on Communications, 2017, 38(a01): 165-170.

1. 任昊利, 罗飞, 张凯. 空间信息系统仿真框架及建模研究[J]. 系统仿真学报, 2009(s2):15-17.

REN H L, LUO F, ZHANG K. Research on Space Information System Simulation Framework and Modeling Method[J]. Journal of System Simulation, 2009(s2): 15-17.

1. WEI J, WANG D. Research on Routing Algorithms in Space-Based Integrated Information Network[C]. International Conference on Information Science and Engineering. 2009: 1823–1827.
2. 张庆军, 张明智, 张庆娟, 等. 基于复杂网络理论空间信息支援体系建模研究[J]. 系统仿真学报, 2017, 29(9): 1907–1913.

ZHANG QJ, ZHANG MZ, ZHANG QJ, et al. Research of Modeling of Space Information Support SoS Based on Complex Network Theory [J]. Journal of System Simulation, 2017, 29(9): 1907-1913.

1. LIU C, XIONG W. Research on the Space-Based Integrated Information Network Evolution Model Visualization Methods Based on the Super Network Theory[C]. IEEE Second International Conference on Data Science in Cyberspace. 2017: 355–358.
2. 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.

HU X F. Challenges and thoughts of modeling & simulation is big data era[J]. Military Operations Research and Systems Engineering, 2013, 27(4): 5-12.

1. GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. MIT Press, 2016.
2. Xu P, Ye M, Li X, et al. Dynamic Background Learning through Deep Auto-encoder Networks[C]// ACM International Conference on Multimedia. ACM, 2014:107-116.
3. LIU Y, WANG Y, ZHANG J. New Machine Learning Algorithm: Random Forest[M]. Springer Berlin Heidelberg, 2012: 246–252.
4. Breiman L I, Friedman J H, Olshen R A, et al. Classification and Regression Trees (CART)[J]. Encyclopedia of Ecology, 1984, 40(3):582-588.
5. NICODEMUS K K, MALLEY J D, STROBL C. The behaviour of random forest permutation-based variable importance measures under predictor correlation[J]. Bmc Bioinformatics, 2010, 11(1): 110.