
基于仿真数据驱动的空间信息网络建模方法

摘要

空间信息网的复杂性包括多个方面，从系统结构单元到直接建模都具有很高的复杂度，传统的复杂系统建模与仿真的方法已经难以满足复杂的空间信息网络系统仿真建模研究的需求，因此，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，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特征降维和数据重构的能力，进行空间信息网络性能指标编码网络解码网络的构建，结合随机森林回归模型构建空间信息网网络设计参数到网络性能指标的关系模型，通过该混合模型不仅可以直接预测空间信息网性能指标，也能够对网络设计参数进行灵敏度分析。

1 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（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，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]。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，典型的需要包含各种空天地的服务系统，如遥感、通信、导航一体的服务，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^[3]，对其进行系统建模研究和综合的评估分析空间信息网各个网络设计组件之间关系是有效地指导空间信息网络的构建的关键^[4]。

2 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文献有：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）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；文^[9]的作者使用超网理论^[10]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；^[11]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；文献^[12]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。

从已有的对空间信息网络的相关的系统建模中可以发现目前的针对空间信息网的建模研究主要集中在几个方面：对空间信息网络的定性模型的研究如^[8]和^[9]；对空间信息网系统内部局部的问题的建模研究如^[8]和^[11]；针对具体的任务对空间信息网进行建模研究如^[12]。然而，目前的空间信息网的建模研究存在一定的局限性：缺乏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^[13]。

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。

在整体性的建模分析阶段，使用了深度学习自编码对空间信息网的性能指标数据进行编码和解码，同时也使用了随机森林回归技术对网络设计参数和空间信息网性能指标编码结果进行关系建模，不仅能够对空间信息网系统建模进行定量的计算和预测，同时还提供了网络设计参数对空间信息网整体性能影响的灵敏度分析的功能，对空间信息网设计优化方向提供了一定的指导。

因此本文的主要创新总结如下：

- 1) 提出基于仿真数据驱动的空间信息网的建模方法，将复杂的空间信息网的系统建模问题，通过数据观测与分析的角度进行空间信息网的整体性的建模分析；
- 2) 使用深度学习自编码网络对空间信息网性能指标进行了编码和解码，提供了空间信息网建模的定量描述和预测的功能；
- 3) 使用随机森林回归模型对空间信息网络设计参数和空间信息网络性能指标编码结果进行关系建模，不仅具有对空间信息网性能指标的定量的预测功能，也具有对空间信息网整体性能的灵敏度的分析的功能，从而为空间信息网的设计优化提供优化方向。

3 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和研究的的问题，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。

3.1 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。

3.2 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)^[15]是深度学习方法的一种，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：

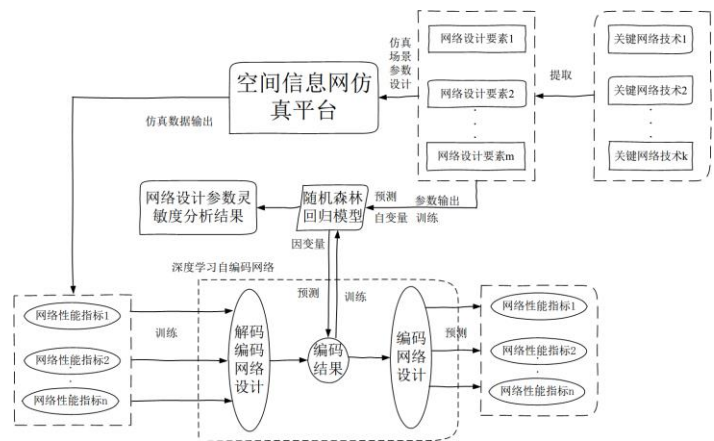


图 1 仿真数据驱动的空间信息网建模分析流程

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。

3.3 建模组合模型

3.3.1 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，训练时采用逐层训练的方法，可以有效地解决传统神经网络训练方法不适用于多层神经网络训练的问题，整个深度自编码网络的训练分为预训练和微调两个阶段 3.3.2。

预训练的过程即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。

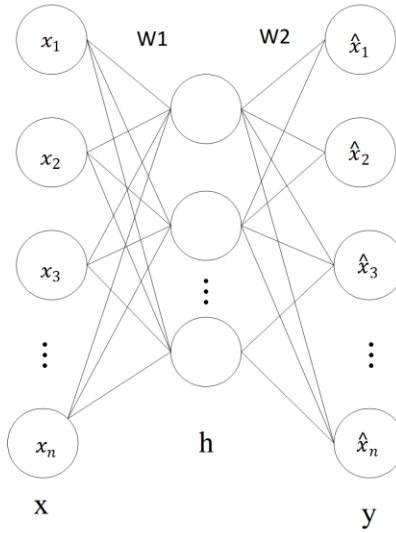


图 2 一层的自编码网络结构

一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 f 和 g 分别表示编码函数和解码函数，则编码过程为: $h = f(x) = S_f(W1x + p)$ ，解码过程为: $y = g(h) = S_g(W2h + q)$ ，其中 S_f 和 S_g 分别为编码和解码的激活函数， $W1$ 为输入层和隐藏层之间的权重矩阵， $W2$ 为隐藏层和输出层之间的权重矩阵，通常有 $W2 = W1^T$ ； p 和 q 分别为隐藏层和输出层的偏置。记自编码网络的参数为 $\theta = W1, p, q$ 。

设训练样本集为 $X = x_1, x_2, \dots, x_n$ ，预训练过程为通过样本集 X 的训练确定自编码网络的参数 θ 使得解码后的 y 与输入 x 尽可能的接近，输入输出的接近程度同重构误差函数：

$$L(x, y) = -\sum_{i=1}^n [x_i \ln(y_i) + (1 - x_i) \ln(1 - y_i)] \quad (1)$$

基于重构误差函数 $L(x, y)$ ，针对训练样本集 X ，利用梯度下降的方法对损失函数进行极小化，即可得到该层的自编码的网络参数 θ 。通过以上方法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。

3.3.2 随机森林回归模型简介

随机森林(RF)^[16]是一系列决策树组合，利用 bootstrap 重抽样方法从原始样本中抽取多个样本，对每个 bootstrap 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 $Y = g(X)$ 来表示，其中 Y 是输出因变量, $X = (X_1, X_2, \dots, X_n)$ 是 n 维相互独立的随机输入变量。用 $T = (x_i, y_i) (i = 1, \dots, N)$ 表示规模为 N 的训练样本集。从总训练样本 T 中重复抽样组成一系列样本集合 $T_k = (x_j^{(k)}, y_j^{(k)}) (j = 1, \dots, N_k, k = 1, \dots, n_{tree})$ ，第 k 个样本集合用于建立第 k 决策树, 表示 T_k 的样本规模, n_{tree} 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 N 的训练样本 T 。

随机的（替换或非替换的方法）从样本集 T 中选取 N_k 个训练样本组成新的规模为 N_k 的训练样本 $T_k (k = 1, \dots, n_{tree})$ 。

用 CART[参考文献加]法则建立第 k 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 n_{tree} 个决策树的预测值的平均。

由于第 k 个决策树是根据训练样本 T_k 建立的，将未被用于建该树的数据叫做袋外数 (OOB) 袋外数据 $B_k = T - T_k$ 用于衡量第 k 个树的预测精度。

对于第 k 个树，其袋外数据为 $B_k = (x_l^{(k)}, y_l^{(k)}) (l = 1, \dots, N_{\bar{k}})$ ，其中 $N_{\bar{k}} = N - N_k$ 表示第 k 个树的袋外数据， $y_l^{(k)}$ 是输入变量 $x_l^{(k)}$ 的精确值。用 $\hat{y}_l^{(k)}$ 表示第 k 个树的对 $x_l^{(k)}$ 的估计值, 则第 k 个树的均方误差表示为:

$$MSE_k = \frac{1}{N_{\bar{k}}} \sum_{l=1}^{N_{\bar{k}}} (y_l^{(k)} - \hat{y}_l^{(k)})^2 \quad (2)$$

将袋外数据 B_k 中的第 X_i 列的数值进行随机扰动， $\hat{y}_l^{(k)}(X_i) (l = 1, \dots, N_{\bar{k}})$ 表示随机扰动后的第 k 个树的预测结果, 则扰动后的均方误差为:

$$MSE_k(X_i) = \frac{1}{N_{\bar{k}}} \sum_{l=1}^{N_{\bar{k}}} (y_l^{(k)} - \hat{y}_l^{(k)}(X_i))^2 \quad (3)$$

变量 X_i 扰动后的均方误差相对于扰动前的均方误差为:

$$VI_i^{(k)} = MSE_k(X_i) - MSE_k \quad (4)$$

扰动重要性指标定义为所有树的 $VI_i^{(k)}$ 的平均:

$$PVIM_i = \frac{1}{n_{tree}} \sum_{k=1}^{n_{tree}} VI_i^{(k)} \quad (5)$$

通过随机扰动 X_i 的值，其边缘分布没有发生变化，但是破坏了变量 X_i 和输出响应量 Y 的相关性，以及变量 X_i 与其所有其他输入变量之间的相关性。如果变量 X_i 越重要，那么扰动 X_i 后的均方误差增量就会越大，因此指标 $PVIM_i$ 的值也越大。因此， $PVIM_i$ 包含了变量 X_i 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^[17]。

3.3.3 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。

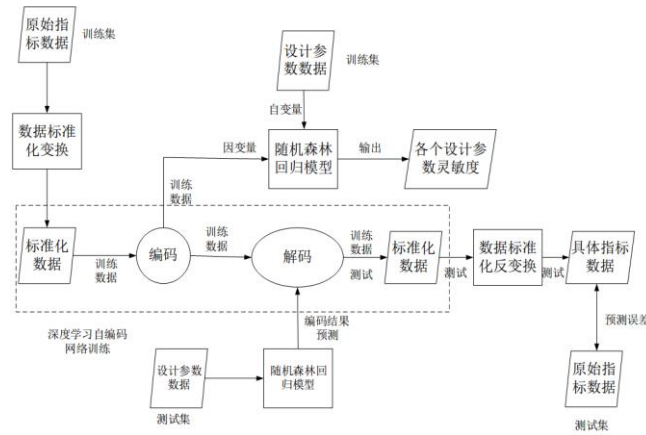


图 3 模型训练与分析流程

4 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。

4.1 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。

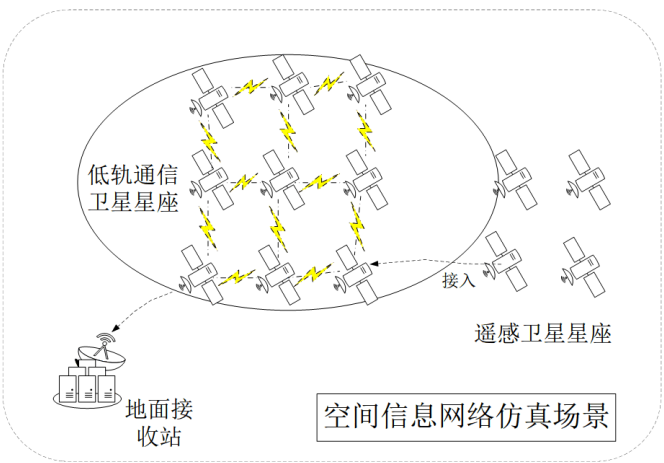


图 4 仿真场景

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记录下来。

4.2 仿真场景设置

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。

4.2.1 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，其目的地地面接收站固定在国内不同的地方。

4.2.2 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。

4.2.3 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。

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：

$$Thro = \frac{Data_{Recv}(bits)}{Time_{Duration}(s)} \quad (6)$$

$Data_{Recv}(bits)$:接收数据量, $Time_{Duration}(s)$:接收数据持续时间。

端到端通信时延：

$$Delay = \frac{All_{Delay}}{Num_{Packets}} \quad (7)$$

All_{Delay} :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， $Num_{Packets}$:接收到的数据包的个数。

数据丢包率：

$$LossRatio = \frac{Data_{Sent} - Data_{Recv}}{Data_{Sent}} \quad (8)$$

$Data_{Sent}$:数据发送量, $Data_{Recv}$:数据接收量。

数据响应时间：

$$Time_{Response} = Time_{Recv} - Time_{Sent} \quad (9)$$

$Time_{Recv}$:第一个数据包收到的时间, $Time_{Sent}$:第一个数据包发出的时间。

数据持续时间：

$$Time_{Duration} = Time_{last} - Time_{first} \quad (10)$$

$Time_{last}$: 最后一个数据包收到的时间, $Time_{first}$: 第一个数据包发出的时间。

4.3 仿真的数据分析

4.3.1 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

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

吞吐量(Mbps)	时延(ms)	响应时间(ms)	持续时间(s)	丢包率
101.90	135.19	136.44	95.063913	0.004
499.62	294.57	1039.49	10.4191	0.092
51.149	157.70	158.75	170.59	0.001
170.69	107.87	107.819	18.008	0.000
499.66	105.23	48.78	4.69	0.000
...
10.180	150.21	145.86	504.606	0.005

按照图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{5,1,5}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。

4.3.2 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。

$$\text{平均绝对误差: } MAE(y, \hat{y}) = \frac{1}{n_{test}} \sum_{i=0}^{n_{test}-1} |y_i - \hat{y}_i|$$

$$\text{解释方差得分: } EVS(y, \hat{y}) = 1 - \frac{Var(y - \hat{y})}{Var(y)}$$

$$\text{决定系数: } R^2(y, \hat{y}) = 1 - \frac{\sum_{i=0}^{n_{\text{test}}-1} (y_i - \hat{y}_i)^2}{\sum_{i=0}^{n_{\text{test}}-1} (y_i - \bar{y})^2}$$

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的误差分析结果

误差	吞吐量	时延	响应时间	持续时间	丢包率
<i>MAE</i>	4.32(Mbps)	21.908(ms)	75.99(ms)	20.394(s)	0.0007
$\frac{MAE}{MEAN} \%$	2.47	3.792	8.786	11.84	3.09
<i>EVS</i>	0.8915	0.8526	0.8006	0.8685	0.8077
<i>R</i> ²	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分析模型

误差	吞吐量	时延	响应时间	持续时间	丢包率
<i>MAE</i>	112.32(Mbps)	489.908(ms)	750.99(ms)	192.8(s)	0.23
$\frac{MAE}{MEAN} \%$	62.1	84.72	88.006	106.63	100.09
<i>EVS</i>	0.52	0.2974	0.0013	0.0485	0.248
<i>R</i> ²	0.52	0.2962	0.009	0.0483	0.247

通过上表的预测误差的结果可知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 R^2 均自 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。由误差的计算结果表明，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。

分析结果对比（说服力解释）

网络设计参数灵敏度分析结果如图 6 所示：

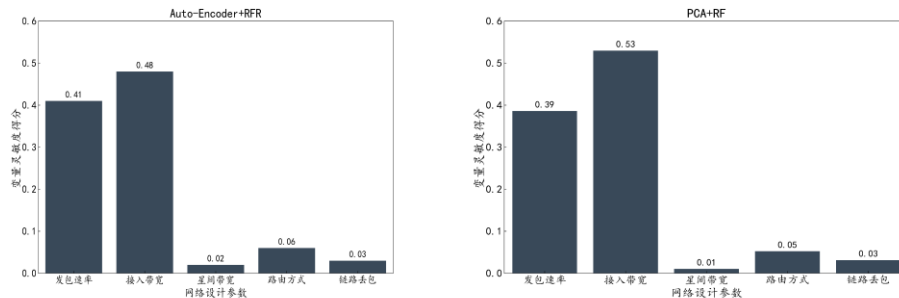


图5 主成分组合模型中灵敏度的分析结果

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,其整体性能的影响程度排名为: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,网络设计参数的主要影响因素是遥感业务的接入带宽和数据的发包速率,所以,在其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。因此本文模型的灵敏度分析的结果对于空间信息网的设计和优化的方向具有定性和定量分析的作用。

5 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调参以到达相应的预测精度。

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] JIANG C, WANG X, WANG J, 等. Security in space information networks[J]. IEEE Communications Magazine, 2015, 53(8): 82–88.
- [4] LIU S, WANG H M, WANG S T, 等. Space-Based Information Integrated Network Technology and Performance Analysis Based on Cognitive Radio[M]. 2017.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 任昊利, 罗飞, 张凯. 空间信息系统仿真框架及建模研究[C]//中国系统仿真学会成立 20 周年大会暨学术年会. 2009: 15–17.
- [9] 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.
- [10] 石福丽, 朱一凡. 基于超网络理论的军事通信网络复杂性度量方法[J]. 通信学报, 2011, 32(12): 51–59.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] 赵勇. 卫星总体多学科设计优化理论与应用研究[D]. 国防科学技术大学, 2006.
- [15] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning[M]. MIT Press, 2016.
- [16] LIU Y, WANG Y, ZHANG J. New Machine Learning Algorithm: Random Forest[M]. Springer Berlin Heidelberg, 2012: 246–252.
- [17] NICODEMUS K K, MALLEY J D, STROBL C, 等. The behaviour of random forest permutation-based variable importance measures under predictor correlation[J]. BMC Bioinformatics, 2010, 11(1): 110.