航空学报
Acta Aeronautica et Astronautica Sinica
ISSN 1000-6893,CN 11-1929/V

《航空学报》网络首发论文

题目：面向稀薄流非线性本构预测的机器学习方法研究
作者：李廷伟，张莽，赵文文，陈伟芳，蒋励剑
收稿日期：2020-06-09
网络首发日期：2020-09-15
引用格式：李廷伟，张莽，赵文文，陈伟芳，蒋励剑．面向稀薄流非线性本构预测的机器学习方法研究．航空学报.
https://kns.cnki.net/kcms/detail/11.1929.V.20200915.1348.020.html

航 空 学 报
Acta Aeronautica et Astronautica Sinica ISSN 1000-6893 CN 11-1929/V
http://hkxb.buaa.edu.cn hkxb@buaa.edu.cn

引用格式：李廷伟，张莽，赵文文，陈伟芳，蒋励剑. 面向稀薄流非线性本构预测的机器学习方法研究[J]. 航空学报，2021，42(4):524386.
Study of machine learning method in the correction of rarefied nonlinear constitutive relations[J]. Acta Aeronautica et Astronautica Sinica, 2021, 42(4): 524386(in Chinese). doi: 10.7527/S1000-6893.2015.2020.24386

面向稀薄流非线性本构预测的机器学习方法研究

李廷伟[1,3]，张莽[2]，赵文文[1,*]，陈伟芳[1]，蒋励剑[1]

1. 浙江大学 航空航天学院，杭州 310027
2. 中国运载火箭技术研究院 研究与发展中心，北京 100076
3. 中国电子科技集团公司第五十四研究所 航天系统与应用专业部，石家庄 050081

摘 要：稀薄非平衡流域内连续介质假设已经失效，主要围绕Boltzmann方程及模型方程对稀薄非平衡流开展理论与计算研究，统一气体动理论格式（UGKS, Unified Gas-kinetic Scheme）是其中一种代表性方法。在稀薄非平衡流数值模拟中，NS方程连续介质假设已经失效，不能有效描述流场非平衡特征。UGKS方法虽然计算精度高，但速度空间离散导致计算效率低下，多维高速条件下数值计算难以开展。基于数据驱动的思想，在NS方程与UGKS方法的研究基础上发展出了一种稀薄非平衡流非线性本构关系求解方法（DNCR，Data-driven Nonlinear Constitutive Relations）。该方法以NS与UGKS求解器获得的流场数值模拟计算结果作为训练数据集，基于流场特征参数采用极端随机树算法生成机器学习模型，对待预测流场中线性粘性应力项与热流项进行非线性修正，并耦合非线性本构关系求解宏观守恒方程得到目标状态稀薄非平衡流动数值解。本文针对DNCR方法中所采用的机器学习方法-极端随机树模型，通过二维顶盖驱动方腔流算例对高维非线性建模涉及的特征参数选取、参数调优开展了相关验证与研究工作，选取若干典型状态对极端随机树模型的泛化性能开展研究，并评估了相关模型与方法的计算精度与计算效率。

关键词：稀薄非平衡流；本构关系；数据驱动；极端随机树；模型参数

中图分类号：V211.3 **文献标识码**：A **文章编号**：1000-6893（2020）XX-XXXXX-XX

钱学森[1]根据努森数定义流体的稀薄程度，将流动区域划分为连续流域（$Kn<0.01$）、滑移流域（$0.01<Kn<0.1$）、过渡流域（$0.1<Kn<10$）以及自由分子流域（$Kn>10$），稀薄非平衡流域包含除连续流域以外的三大流域。

连续流域符合流体力学连续介质假设条件，连续介质流体力学是其理论基础，其中最具有代表性的控制方程为 NS 方程，形成了以广义牛顿定律和傅里叶热传导定律为基础的基于 NSF（Navier-Stokes-Fourier）关系的连续流数值模拟计算方法，此方法的发展为其他各种方法的建立提供了宝贵经验。稀薄非平衡流域中气体稀薄属性逐渐凸显，例如物面出现了十分显著的速度滑移与温度跳跃现象。此时在物面附近的努森层内，连续介质假设已经失效，准确地说是基于三大守恒方程（质量守恒、动量守恒和能量守恒）推导出来的粘性应力项与热流项已经不能再简单使用低阶宏观物理量（速度、温度）的梯度线性表征，即线性本构关系已经不再适用于精准描述稀薄非平衡流动问题。

对稀薄非平衡流问题的研究主要围绕稀薄气体流动控制方程-Boltzmann 方程开展，它是分子气体动力学的基本方程，可以不受努森数的限制对连续流到自由分子流整个流域进行统一描述。Boltzmann 方程是一个复杂的七维积分/微分方程，大部分研究均是对其直接或者间接求解，发展形成了多种数值计算方法与理论，统一气体动

理论格式（UGKS, Unified Gas-kinetic Scheme）[2]是其中一种代表性方法。UGKS 方法使用 Boltzman 方程的 Bhatnagar-Gross-Krook（BGK）[2]类模型方程作为控制方程，用有限个离散速度替代整个速度空间，在一定条件下均能给出较为准确的流动特征，具备求解各流域多尺度问题的能力[3]。UGKS 方法采用 BGK 类模型方程基于当地积分解对通量进行构造，将分子运动与碰撞过程自然地耦合在一起，物理意义更加明确，突破了 DSMC 方法[4]解耦处理所带来的计算时间步长小于分子的平均碰撞时间、计算单元网格尺度小于分子平均自由程的限制。但由于在 UGKS 方法在求解过程中依赖速度分布函数对速度空间进行离散，这对计算与存储的要求非常高，多维与高速条件下计算效率较低。

随着计算机理论与技术的迅猛发展，大数据时代启发了人类思考问题新的思维，基于数据驱动和机器学习的研究方式也应运而生。研究人员开始通过机器学习的方法解决流体力学领域中难以解决的问题，目前在气动优化设计[5-9]、湍流问题[10-15]、非定常气动力建模[16-20]等领域上已经开展了很多卓有成效的研究工作。

以 NS 方程和 UGKS 方法为理论基础，提出了一种适用于稀薄非平衡流数值模拟的基于数据驱动非线性本构关系（DNCR, Data-driven Non-linear Constitutive Relations）数值计算方法。基于流场特征参数采用机器学习方法对 NS 方程线性的粘性应力项与热流项进行非线性离散重构，理论适用范围可以覆盖较大来流努森数条件，通过耦合非线性本构关系求解宏观守恒方程得到待预测状态稀薄非平衡流动数值解。与传统机器学习建模方法摒弃基本物理规律直接对数据进行训练与预测的思想相比，DNCR 方法未抛弃物理规律而是对流体本构关系进行修正后耦合守恒方程迭代求解，同时保留了数据建模与物理建模的优点，方法物理意义更加清晰明确。

在 DNCR 方法中，采用的具体机器学习算法为极端随机树[21,22]（Extremely Randomized Trees）算法。在机器学习算法中，针对所研究的具体问题选取低冗余、高代表性的特征参数对于机器学习算法的泛化性能和运行效率具有重要影响。本文拟通过二维顶盖驱动方腔流算例对高维非线性建模涉及的特征参数选取、参数调优开展相关验证与研究工作，选取若干典型状态对极端随机树模型的泛化性能开展研究，并评估相关模型与方法的计算精度与计算效率。

1 基于数据驱动非线性本构关系计算方法

数据驱动非线性本构方程数值计算方法（DNCR）是基于 NS 方程和 UGKS 方法的数值模拟计算结果作为流场样本训练数据，采用机器学习方法构建热流 / 应力项与流场特征参数的高维复杂非线性回归关系模型，对 NS 方程本构关系进行非线性修正。最终通过耦合数据驱动的非线性本构关系求解宏观量守恒方程得到稀薄非平衡流数值解。其具体的计算流程如下图所示：

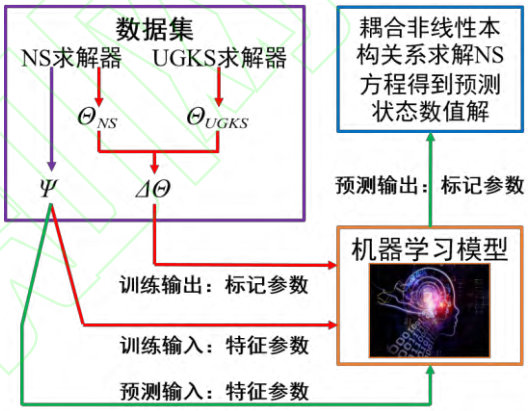


图 1 DNCR示意图

Fig. 1 Schematic of DNCR

DNCR方法计算流程如图1所示。ψ 代表在NS流场数值模拟结果中提取的特征参数，作为机器学习模型的输入量，$\Delta\Theta$ 为标记参数，作为机器学习模型的输出量，即机器学习模型建立了 ψ 与 $\Delta\Theta$ 之间的复杂高维回归关系，与非线性本构物理建模函数的区别在于其本构关系没有具体数学表达式，而是全流域存在离散当地映射 $\psi \rightarrow \Delta\Theta$，所形成的离散本构映射关系在全流场计算域适用，而模型泛化性能取决于训练集特征代表性与特征参数选取，理论上若特征参数选择未涉及到任何与空间网格与当地坐标值直接相关的参数且训练集包含相似的非平衡流动特征，机器学习模型就具备一定的迁移预测能力。

DNCR方法的一个重要特点是训练与预测过程相对独立，图1中红线与绿线分别表示了训练与预示流程。模型训练过程通过对包含不同典型

流动特征的基础流场数据集开展训练，获得ψ与$\Delta\Theta$之间的复杂回归关系。而预测过程则首先采用NS求解器对待预测状态开展初步计算，获得待预测当地特征值ψ，然后采用已训练生成的回归关系$\psi \to \Delta\Theta$得到待预测流场当地标记值$\Delta\Theta$，最终通过时间推进方式求解质量、动量与能量守恒方程，耦合非线性本构关系完成计算，守恒方程如式（1）所示

$$\frac{\partial \boldsymbol{Q}}{\partial t}+\frac{\partial \boldsymbol{E}}{\partial x}+\frac{\partial \boldsymbol{F}}{\partial y}+\frac{\partial \boldsymbol{E}_{\mathrm{v}}}{\partial x}+\frac{\partial \boldsymbol{F}_{\mathrm{v}}}{\partial y}=0 \qquad (1)$$

式中Q为守恒量，E，F为无粘通量，E_v，F_v为粘性通量。控制方程中粘性应力张量与热流项表达式为：

$$\begin{cases}\tau_{xx}=\tau_{xx,Tag}-\mu\left(\frac{4}{3}\frac{\partial u}{\partial x}\Big|_{Tag}-\frac{2}{3}\frac{\partial v}{\partial y}\Big|_{Tag}\right)+\mu\left(\frac{4}{3}\frac{\partial u}{\partial x}-\frac{2}{3}\frac{\partial v}{\partial y}\right)\\ \tau_{xy}=\tau_{yx}=\tau_{xy,Tag}-\mu\left(\frac{\partial u}{\partial y}\Big|_{Tag}+\frac{\partial v}{\partial x}\Big|_{Tag}\right)+\mu\left(\frac{\partial u}{\partial y}+\frac{\partial v}{\partial x}\right)\\ \tau_{yy}=\tau_{yy,Tag}-\mu\left(\frac{4}{3}\frac{\partial v}{\partial y}\Big|_{Tag}-\frac{2}{3}\frac{\partial u}{\partial x}\Big|_{Tag}\right)+\mu\left(\frac{4}{3}\frac{\partial v}{\partial y}-\frac{2}{3}\frac{\partial u}{\partial x}\right)\\ q_x=q_{x,Tag}-\kappa\left(\frac{\partial T}{\partial x}\Big|_{Tag}\right)+\kappa\left(\frac{\partial T}{\partial x}\right)\\ q_y=q_{y,Tag}-\kappa\left(\frac{\partial T}{\partial y}\Big|_{Tag}\right)+\kappa\left(\frac{\partial T}{\partial y}\right)\end{cases} \qquad (2)$$

在式（2）中，下标 Tag 表示待预测流场当地标记值。式（1）与式（2）随着时间推进，相关热流、应力张量与流场梯度量向标记值趋近，最终完成计算收敛，获得待预测流场定常解。

当物体或扰动源的速度大于扰动信息的传播时，扰动无法向上游传播所形成的强压缩间断称为激波。一维激波结构是最简单、最基本的非平衡流动现象之一，是研究本构关系与非平衡流动的典型算例，可以用来对模型进行验证。

NS、UGKS、DNCR 相同网格下，单原子气体 Ar 物性参数取 $\mu_0=2.272\times10^{-5}$ Pa.s，$T_0=300\mathrm{K}$，$\gamma=1.667$，$Pr=0.667$，$R=208.16\mathrm{J/(kg\cdot K)}$，逆幂律幂次取 $\varepsilon=0.72$。计算状态如表 1 所示。

表 1 Ar一维激波结构计算状态

Table1 State of one-dimensional Argon shock structure calculation

类别	计算条件
来流马赫数(Ma)	8.0
激波上游静压(Pa)	73477
激波上游静温(K)	288.15

对于一维激波结构，如前所述将UGKS数值模拟结果中的$q_x,\tau_x,\frac{\partial u}{\partial x},\frac{\partial T}{\partial x}$数据提取出来作为DNCR计算程序的读入数据。计算收敛后对比DNCR、NS、UGKS密度、速度、压强、温度等基本物理量如图2-图5所示。可以看出DNCR的计算结果与NS相比更加接近UGKS的计算结果，表明了本文所提出的本构关系非线性修正方案的正确性与可行性。

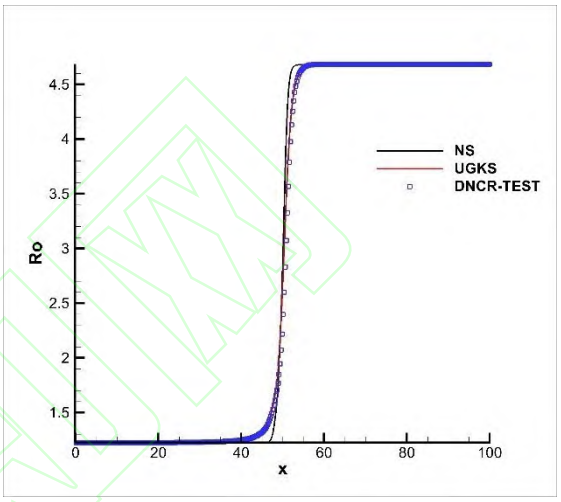


图 2 激波结构密度ρ对比

Fig. 2 Density distributions in shock wave

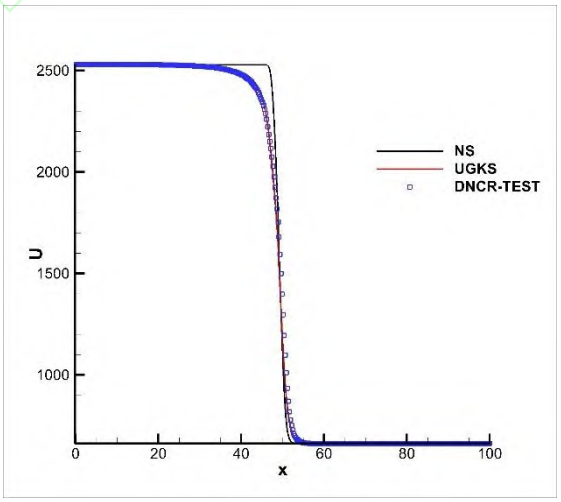


图 3 激波结构速度v对比

Fig. 3 Velocity distributions in shock wave

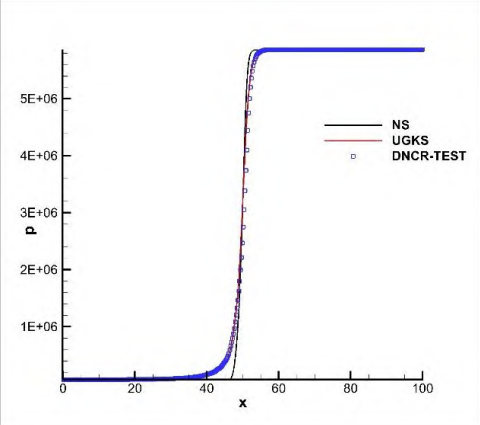


图 4 激波结构压强 P 对比

Fig. 4 Pressure distributions in shock wave

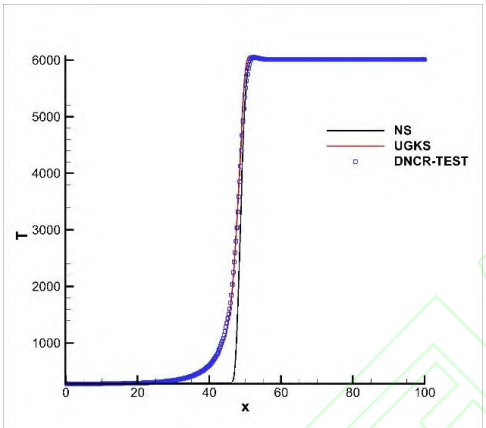


图 5 激波结构温度 T 对比

Fig. 5 Temperature distributions in shock wave

2 训练数据集生成

本文选取图6所示顶盖驱动方腔流动作为测试算例，计算网格如图7所示，计算网格为 61×61 的均匀网格，速度离散为 28×28 。选取5组稀薄非平衡流动状态生成训练数据集，计算状态如表2所示。

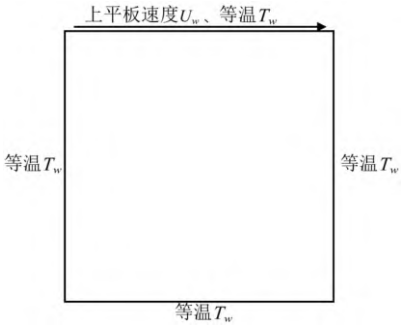


图 6 顶盖驱动方腔流示意图

Fig. 6 Schematic of lid-driven cavity

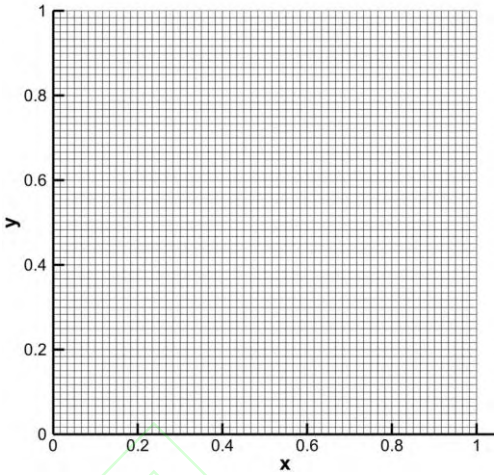


图 7 二维顶盖驱动流计算网格

Fig. 7 Grid of Two-dimensional lid-driven cavity

表 2 二维方腔计算状态

Table 2 State of two-dimensional cavity calculation

方腔尺寸(m)	1
上平板移动速度(m/s)	50
努森数(Kn)	0.5/0.7/1.0/1.3/1/5
来流静温(K)	273
壁面温度(K)	273

3 机器学习模型研究

在二维顶盖驱动方腔流问题中，标记参数 $\Delta\Theta$ 共包括11个量 $\Delta q_i,\Delta\tau_{ij},\Delta\dfrac{\partial u_i}{\partial x},\Delta\dfrac{\partial T}{\partial x},\Delta\dfrac{\partial u_i}{\partial y},\Delta\dfrac{\partial T}{\partial y}$ 。本文回归问题建模使用并行集成学习算法的典型代表--极端随机树，该算法的基学习器本文采用分类与回归树[25]（CART，classification and regression tree）。

3.1 CART与极端随机树算法

X 与 Y 分别为特征输入和标记输出变量，并且 Y 为连续变量，给定训练数据集。$D=\{(x_1,y_1),(x_2,y_2),\cdots,(x_N,y_N)\}$

回归树是用于研究回归问题的决策树模型，决策树是一种树形结构的基本分类与回归算法。在机器学习中决策树是一个预测模型，代表的是对象属性（即特征参数）与对象值（即标记值）之间的一种映射关系。一个回归树对应着输入空间(即特征空间)的一个划分以及在划分的单元上

的输出值。假设在训练数据集所在的输入空间划分为 M 个单元 $R_1,R_2,\cdots,R_M$，这里的单元可以理解为从最顶端往下依次划分构建的 M 个分支，每个分支只对应一个输出值。递归地将每个区域划分为两个子区域并决定每个子区域的输出值 c，构建二叉决策树：

（1）选择最优切分变量 j 与切分点 s：特征空间包含多个特征参数即特征变量，决策树的构建就是每层进行二叉树的划分，划分时采用某特征变量能够使误差最小，此变量即为最优切分变量；切分点指的是切分变量进行左右划分的值，小于此值划分为左子树，大于此值划分为右子树，切分时某切分点能够使误差最小即为最优切分点。通过求解

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right] \tag{3}$$

遍历变量 j，对固定切分变量 j 扫描切分点 s，选择使上式最小(j,s)。

（2）用选定的(j,s)划分区域并决定相应的输出值：

$$R_1(j,s)=\{x|x^{(j)}\le s\},R_2(j,s)=\{x|x^{(j)}>s\}$$
$$\hat{c}_m=\frac{1}{N_m}\sum_{x_j\in R_m(j,s)}y_i,x\in R_m,m=1,2 \tag{4}$$

（3）继续对两个子区域调用步骤（1），（2），直至满足条件。

（4）将输入空间划分成 M 个区域 $R_1,R_2,\cdots,R_M$，生成决策树：

$$f(x)=\sum_{n=1}^{M}\hat{c}_nI(x\in R_M) \tag{5}$$

极端随机树在生成 CART 过程中选择划分属性不再是选择最优属性而是完全随机选择。采用训练数据集进行并行训练得到 T 个 CART，对本文回归问题模型最终采用平均法：

$$H(x)=\frac{\sum_{i=1}^{T}f(x)_i}{T} \tag{6}$$

3.2 特征参数选择

在 DNCR 方法中，流场特征参数构成特征空间$\{\Psi\}$，热流、应力修正所需各项构成标记空间$\{\Delta\Theta\}$。选取低冗余、高代表性的特征参数对于机器学习算法的泛化性能和运行效率具有重要影响。首先从物理机理上考虑，我们选择表征流场稀薄非平衡特征的参数，例如基于当地流场梯度的局部努森数 $Kn_{GLL}(\rho,p,T)$ 作为主要的流场特征参数，然后尝试加入其他参数并最终选取适当参数集作为流场特征参数建立特征空间。

本文采用方差过滤准则进行特征选择，方差越小，表示该特征的数据差异越小，可以认为这个特征对区分样本贡献不大，导致预测的效果就越差，因此可以剔除此特征或降低特征权重。另外考虑到真实物理问题中不同物理量之间数据的量级可能存在较大差异，量级较大导致方差较大，为了避免这一不利因素，在本论文中使用标准差与极差的商值消除量级影响进行特征选择。

$$\sigma^*=\frac{\sqrt{\sum_{i=1}^{n}(x_i-\mu)^2/n}}{\max(X)-\min(X)} \tag{7}$$

其中 $X=\{x_1,x_2,x_3,,...,x_n\}$ 代表某一特征参数，n 代表样本点的数量，μ 为 $X=\{x_1,x_2,x_3,,...,x_n\}$ 的算数平均值。标准差与极差 $\max(X)-\min(X)$ 做商的目的在于消除数据量纲的影响。

采用二维顶盖驱动方腔流的数值模拟结果进行特征选择工作。流场中初始待选特征参数包括 $\rho,u,v,P,T,\frac{\partial\rho}{\partial x},\frac{\partial P}{\partial x},\frac{\partial\rho}{\partial y},\frac{\partial P}{\partial y},Kn_\rho,Kn_P,Kn_T$。$X$ 方向上各特征参数的标准差情况如图 8 所示，特征依次标记为 F1-F12。由于 Y 方向分布与之类似，本文统一使用 X 方向进行对比。

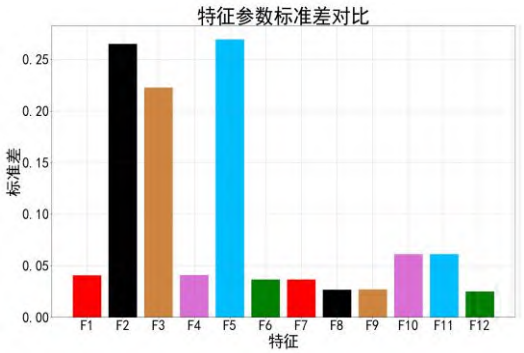


图 8 特征参数标准差对比

Fig. 8 Comparison of standard deviation of characteristic parameters

首先从稀薄非平衡特征上考虑，确定了基于当地流场梯度的局部努森数 $Kn_{GLL}(\rho,p,T)$ 作为主要流场特征参数。由图 8 可知 F2、F3、F5 即 u、v、T 标准差较大，第一组特征参数选取

u,v,T,Kn_ρ,Kn_P,Kn_T，另外为了进行不同参数选取对预测结果的影响，选取$\rho,u,v,P,T,\frac{\partial\rho}{\partial x},\frac{\partial P}{\partial x},\frac{\partial\rho}{\partial y},\frac{\partial P}{\partial y},Kn_\rho,Kn_P,Kn_T$作为第二组特征参数进行对比。本文中采用余弦相似度[26]的另一形式为评价指标用于评估所选特征参数对模型泛化性能的影响，数值越小代表预测准确性越高即泛化性能更好。

$$Cos\_similarity = 1-\frac{\overrightarrow{Y_{predict}}\cdot\overrightarrow{Y_{actual}}}{\left|\overrightarrow{Y_{predict}}\right|\cdot\left|\overrightarrow{Y_{actual}}\right|} \tag{8}$$

其中$Y_{predict}$代表预测的标记值，Y_{actual}代表真实的标记值，均可看作形如$Y=\{y_1,y_2,y_3,...,y_n\}$的向量，$y_i$代表某网格点标记值，$n$代表计算网格点的数量。

采用$Kn0.7$与$Kn1.3$两组状态对机器学习模型进行训练，对$Kn1.0$状态进行预测并与$Kn1.0$的 UGKS 数据进行对比，计算各预测量余弦相似度，第一组特征值（6 特征值）与第二组特征值（12 特征值）结果对比情况如图 9 所示。由图 9 可以看出，虽然训练集中特征参数u,v,T,Kn_ρ,Kn_P,Kn_T的数据差异更大，但极端随机树方法在 DNCR 中采用 12 特征值时泛化性能更优，在本文后续工作中特征参数将选用第二组特征参数$\rho,u,v,P,T,\frac{\partial\rho}{\partial x},\frac{\partial P}{\partial x},\frac{\partial\rho}{\partial y},\frac{\partial P}{\partial y},Kn_\rho,Kn_P,Kn_T$。

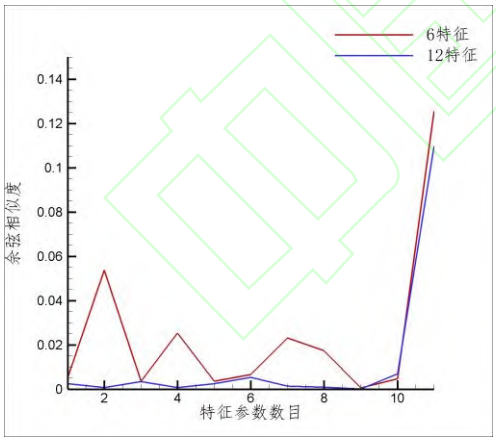


图 9 余弦相似度对比

Fig. 9 Comparison of cosine similarity

3.3 极端随机树参数调优

在研究具体问题时，不同机器学习算法中都有许多的参数需要人为设定，即使是同一种算法，面向不同回归问题的参数配置也有所区别，最终得到的模型性能表现往往也存在十分明显的差别。

对于极端随机树算法来说，主要参数包括框架参数：基学习器的数量 n_estimators、度量分裂的标准 criterion、是否采用袋外样本来评估模型的好坏 oob_score；决策树参数：决策树的每个节点随机选择的最大特征数 max_features、叶子结点上应有的最少样例数 min_samples_leaf、决策树最大深度 max_depth 等参数。在本文中重点对 n_estimators 与 max_features 两个参数进行研究，其余参数均采取默认设置，研究方法为单一变量法。

本文采用可决系数[27]评价参数调节后的模型优劣情况。可决系数表示一个随机变量与多个随机变量关系的数字特征，即可决系数值越大越接近于 1 时说明特征参数对标记参数的解释程度越高。

$$R^2=1-\frac{SS_{res}}{SS_{tot}}=1-\frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \tag{9}$$

其中$y_i\in Y$，$\bar{y}$为Y均值，$\hat{y}_i$为预测值。

采用$Kn0.7$与$Kn1.3$两组状态对机器学习模型进行训练，对$Kn1.0$状态进行预测。选取 n_estimators 在 [1,1000] 范围内，R^2 随 n_estimators 变化趋势如图 10 所示。可以看出最终趋近于一个比较稳定的值，较好拟合且随着基学习器的数量增加未出现严重过拟合现象。如图 11 所示，模型的训练时间与基学习器数目基本满足线性关系，考虑到模型训练的时间成本，在本文中选取 n_estimators=300。

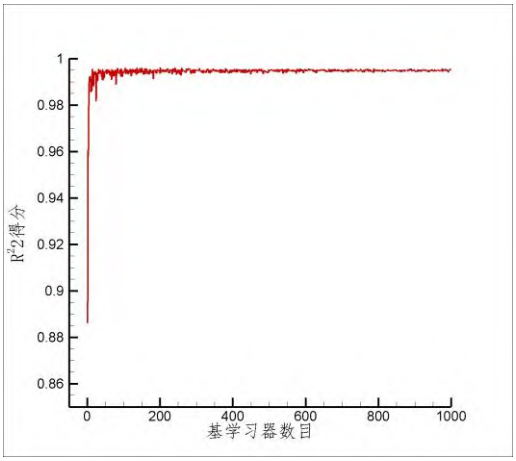


图 10 $R^2 score$随基学习器数目变化曲线

Fig. 10 $R^2 score$ variation curve with n_estimators

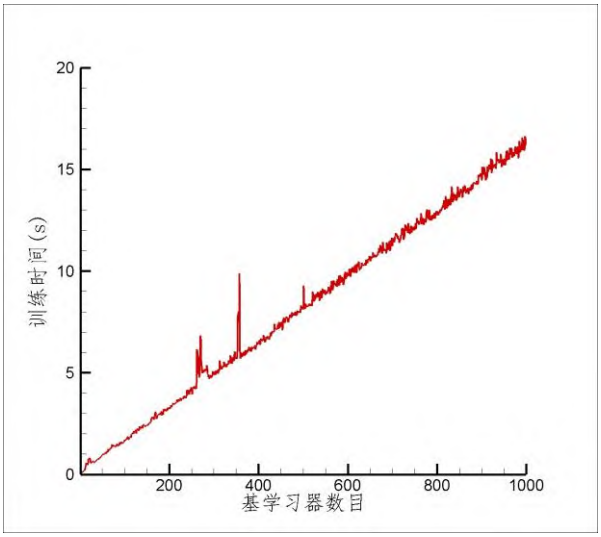


图 11 训练时间随基学习器数目变化曲线

Fig. 11 Train time variation curve with n_estimators

根据前文中确定使用的 12 个特征值，因此 max_features 在[1,12]范围中取值。为避免偶然误差，本文训练 1000 次后取得平均值，得到R^2值随 max_features 变化趋势如图 12 所示。可以看出在 DNCR 方法中极端随机树的预测准确性随着 max_features 变化大致呈现递增的趋势，max_features 的大小对模型训练预测时间的影响可以忽略不计，在本文中选择 max_features=12。其余模型参数及其他待预测物理量的模型参数调优过程与之类似，不做赘述。

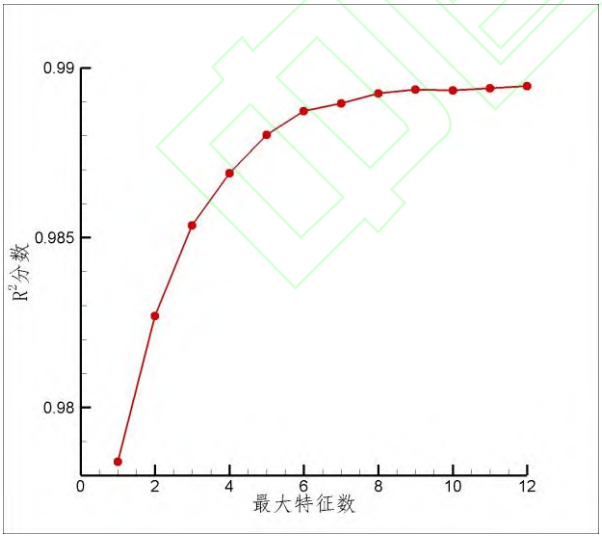
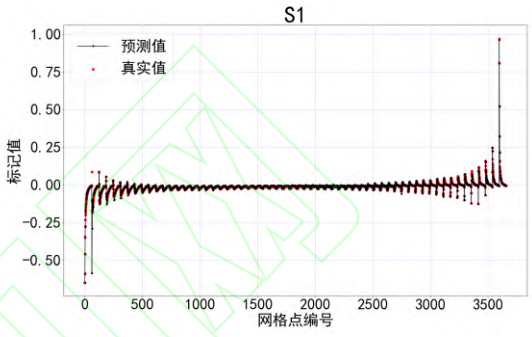


图 12 $R^2 score$随最大特征数变化曲线

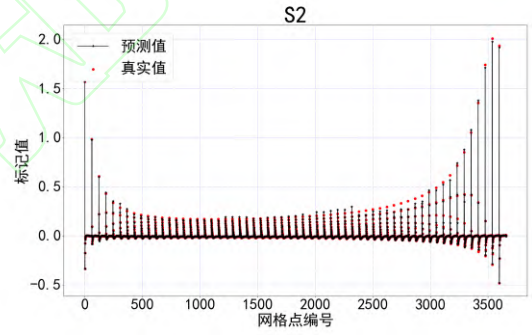
Fig. 12 $R^2 score$ variation curve with max_features

4 DNCR计算精度与极端随机树泛化性能分析

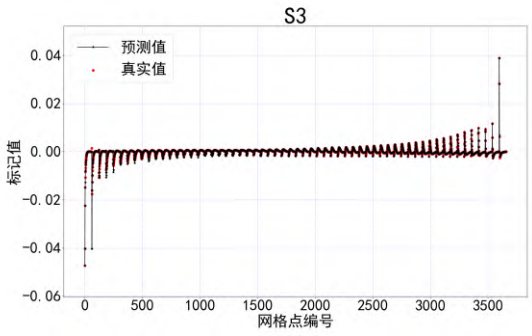
采用$Kn0.7$与$Kn1.3$两组状态对机器学习模型进行训练，选取一组中间状态$Kn1.0$进行预测，$\Delta q_i,\Delta\tau_{ij},\Delta\frac{\partial u_i}{\partial x},\Delta\frac{\partial T}{\partial x},\Delta\frac{\partial u_i}{\partial y},\Delta\frac{\partial T}{\partial y}$标记值预测结果如图13所示，图中横坐标为网格点编号，纵坐标为标记值量值。



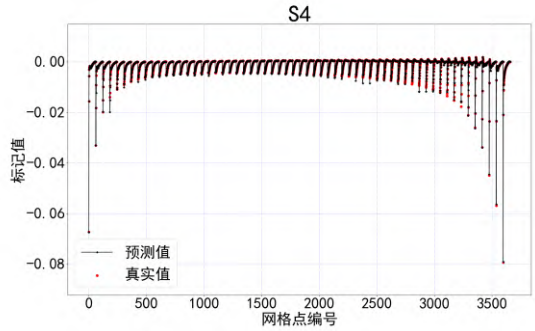
(a)Δq_x 预测结果



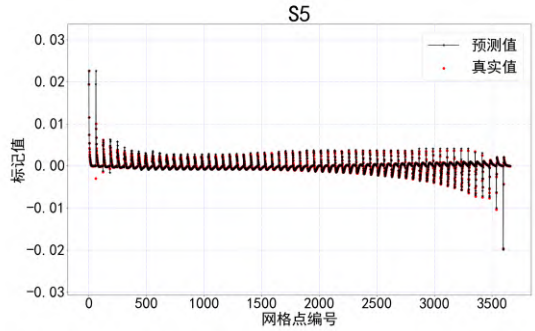
(b)Δq_y 预测结果



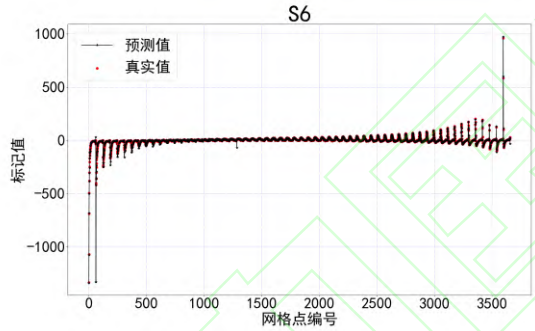
(c)$\Delta\tau_{xx}$ 预测结果



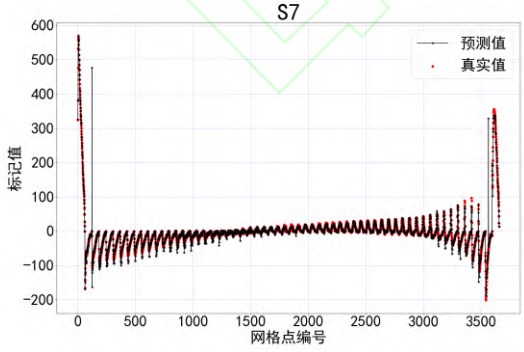
(d) $\Delta\tau_{xy}$ 预测结果



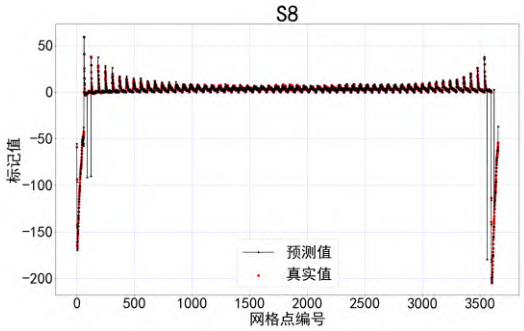
(e) $\Delta\tau_{yy}$ 预测结果



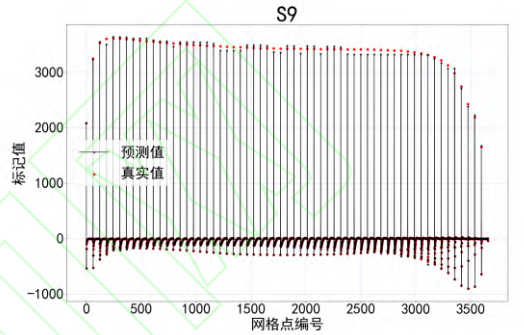
(f) $\Delta\dfrac{\partial u}{\partial x}$ 预测结果



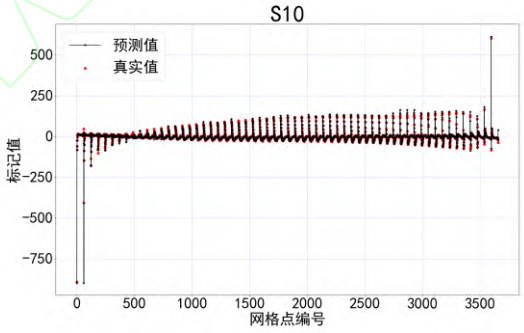
(g) $\Delta\dfrac{\partial v}{\partial x}$ 预测结果



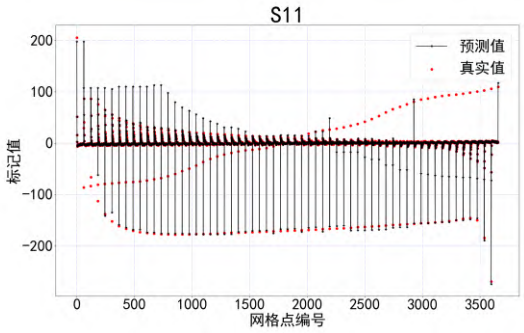
(h) $\Delta\dfrac{\partial T}{\partial x}$ 预测结果



(i) $\Delta\dfrac{\partial u}{\partial y}$ 预测结果



(j) $\Delta\dfrac{\partial v}{\partial y}$ 预测结果



(k) $\Delta\dfrac{\partial T}{\partial y}$ 预测结果

图 13 预测情况

Fig. 13 Prediction situation

为更好地反映预测值与真实值的对比情况，我们采用均方误差（Mean Square Error，MSE）和均方根误差（Root Mean Square Error，RMSE）可以衡量预测值同真值之间的偏离程度，R^2为式（9）可决系数。

$$MSE=\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2 \tag{10}$$

$$RMSE=\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{y}_i-y_i\right)^2} \tag{11}$$

其中$\hat{y}_i$为预测的标记值，y_i为真实的标记值，n为网格点数。S1-S11 的 MSE、RMSE、R^2情况如下表 3 所示。

表 3 标记值误差特性表

Table 3 Error characteristics of predicted values

\	S1	S2	S3	S4	S5
MSE	8.12E-6	1.39E-5	2.76E-8	1.76E-8	1.06E-8
RMSE	2.85E-3	3.72E-3	1.66E-4	1.33E-4	1.03E-4
R^2	0.9943	0.9983	0.9927	0.9983	0.9947

\	S6	S7	S8	S9	S10	S11
MSE	25.48	10.85	0.96	32.91	16.74	175.54
RMSE	5.05	3.29	0.98	5.74	4.09	13.40
R^2	0.9891	0.9971	0.9980	0.9998	0.98577	0.6755

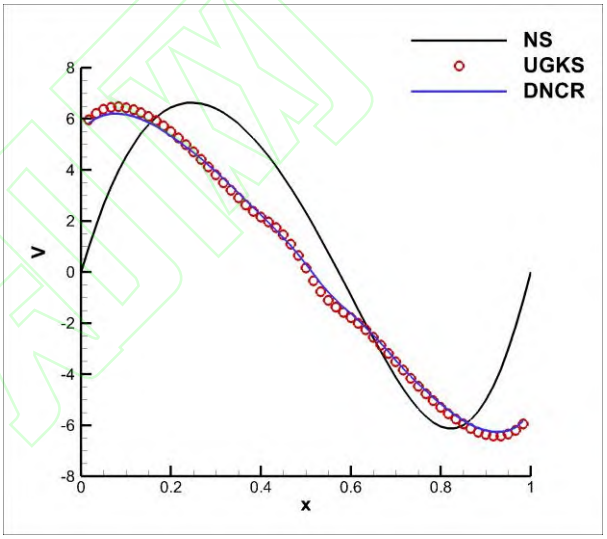
从表中可以看出 MSE 与 RMSE 与标记值本身数据量级有关，并且与其量级相比，MSE 与 RMSE 的值处于较低水平即预测效果较好，通过观察R^2值也以很好佐证，除 S11 预测最差外，其余均能达到 0.98 以上水平。

将每个网格点预测得到的标记值读入 DNCR 计算程序迭代求解守恒方程，最终得到收敛待预测$Kn1.0$流场。为了直观对比计算精度，如图 14 所示，选择$Y=0.5\mathrm{m}$截线上 DNCR 与 NS、UGKS 方法的物理量分布进行对比。采用下式计算评估精度提升情况：

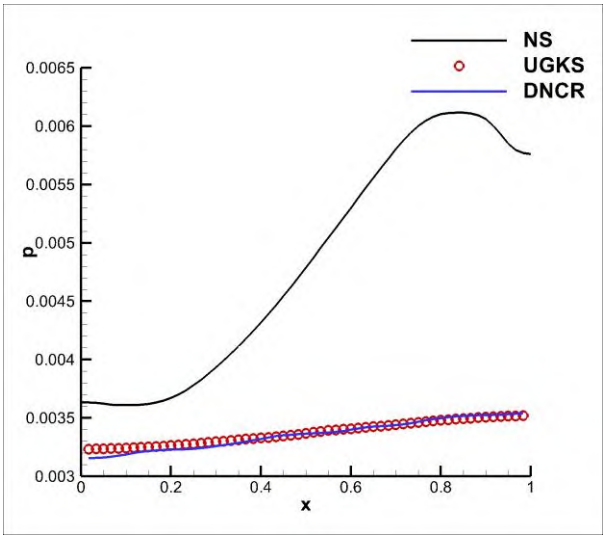
$$\xi=1-\frac{\sqrt{\sum_{i=1}^{N}\left(\mathrm{y}_{DNCR}^{i}-y_{UGKS}^{i}\right)^2}}{\sqrt{\sum_{i=1}^{N}\left(\mathrm{y}_{NS}^{i}-y_{UGKS}^{i}\right)^2}} \tag{12}$$

其中，若ξ趋于 1，则表明 DNCR 方法预示值与 UGKS 基本一致，若ξ趋于 0，则表示 DNCR 预示值与 NS 方程计算结果趋于一致。

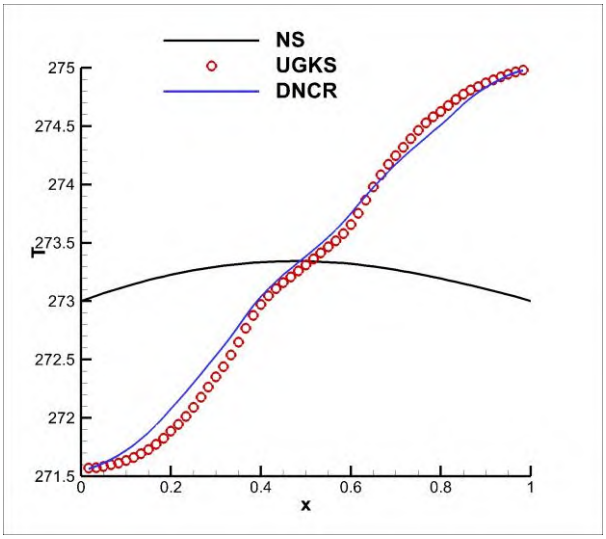
通过计算得到$Y=0.5\mathrm{m}$截线上v、P、T经计算 DNCR 相比 NS 精度提升分别为 92.23%、97.77%、90.80%。



(a) $Y=0.5\mathrm{m}$速度分布对比



(b) $Y=0.5\mathrm{m}$压力分布对比



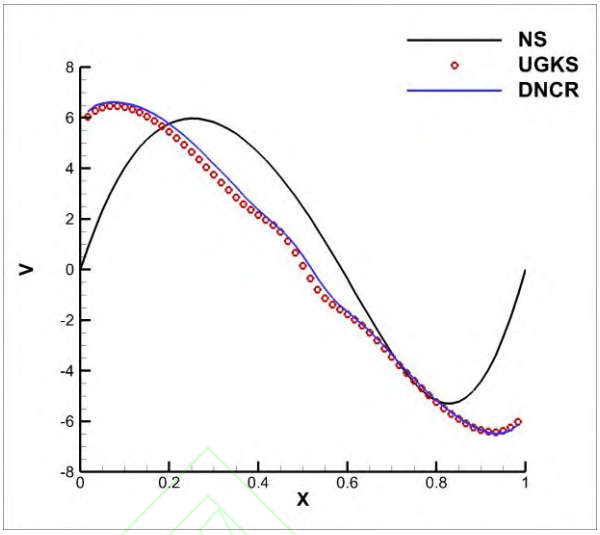
(c) $Y=0.5\mathrm{m}$ 温度分布对比

图 14 $Kn1.0$ 计算结果对比

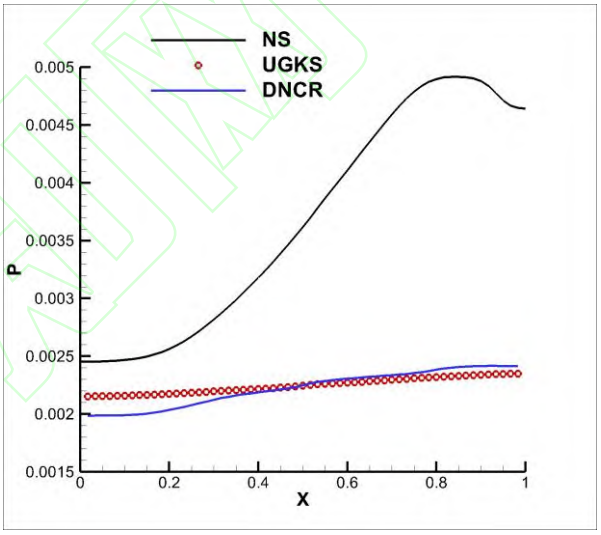
Fig. 14 Comparison of $Kn1.0$ calculation results

为进一步评估极端随机树的泛化能力，选取 $Kn0.5$ 与 $Kn1.5$ 两个超出训练集范围的努森数条件开展 DNCR 计算，重点考察 DNCR 方法的训练集外延数据泛化性能。其中 $Y=0.5\mathrm{m}$ 截线 DNCR 与 NS、UGKS 方法的 Y 方向速度、温度与压力物理量分布对比分别如图 15、图 16 所示。DNCR 较 NS 方程预测精度提升情况依次为 73.59%、95.71%、71.20%、89.12%、94.36%、87.69%，精度计算公式如式（8），可以看出精度提升与预测 $Kn1.0$ 状态时相比虽略有下降，但仍保持较高水平，即 DNCR 方法中所采用极端随机树模型具有较好的泛化性能。

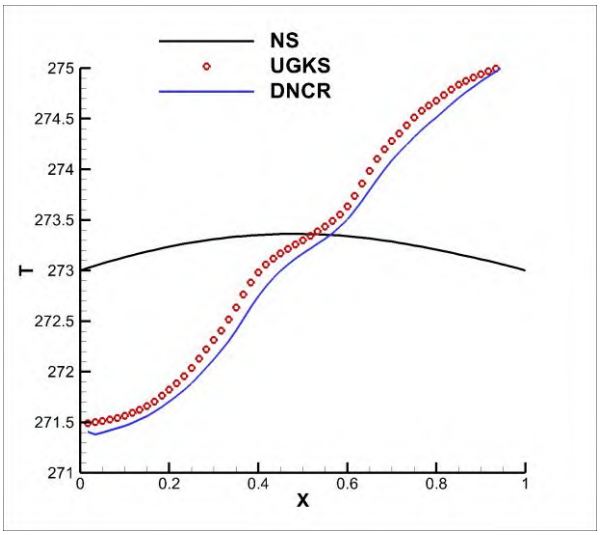
计算效率方面，由于 DNCR 方法与 NS 求解方法的区别仅在于对线性应力与热流项进行了非线性修正，因此其迭代求解速度与 NS 方程基本相当。但 DNCR 方法需要首先采用 NS 方程对待预测状态进行预计算并提取流场特征值作为机器学习模型的输入数据，最终才能在 NS 计算结果上耦合数据驱动的非线性本构关系求解宏观量守恒方程得到稀薄非平衡流数值解，因此 DNCR 计算时间较 NS 方程会有所增加，若不考虑机器学习模型训练时间，DNCR 与 NS 方程计算效率能够保持同一量级。



(a) $Y=0.5\mathrm{m}$ 速度分布对比



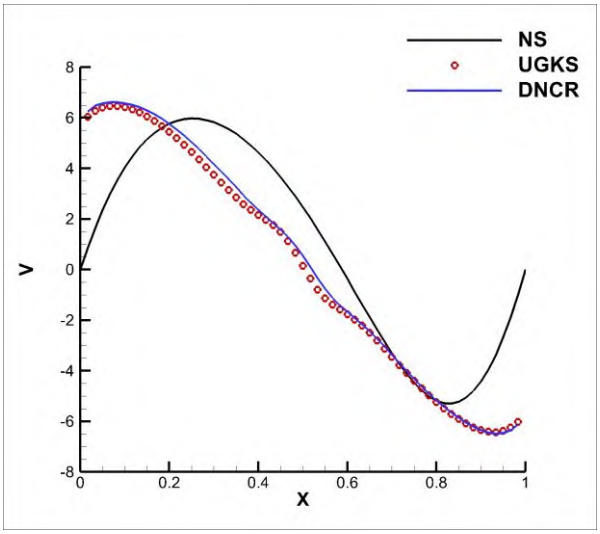
(b)$Y=0.5\mathrm{m}$ 压力分布对比



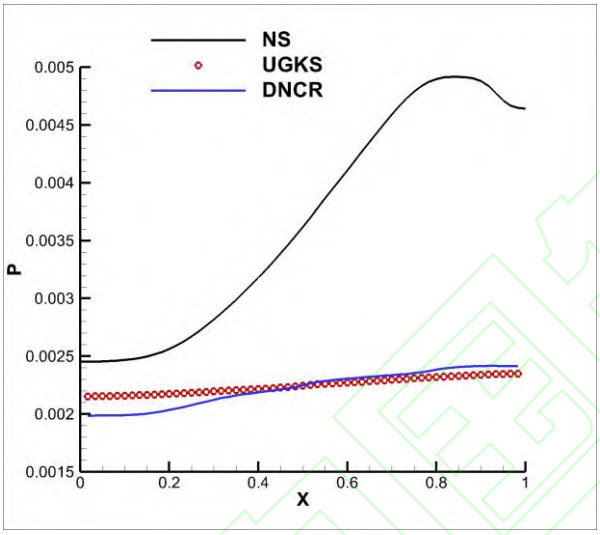
(c) $Y=0.5\mathrm{m}$ 温度分布对比

图 15 $Kn0.5$ 计算结果对比

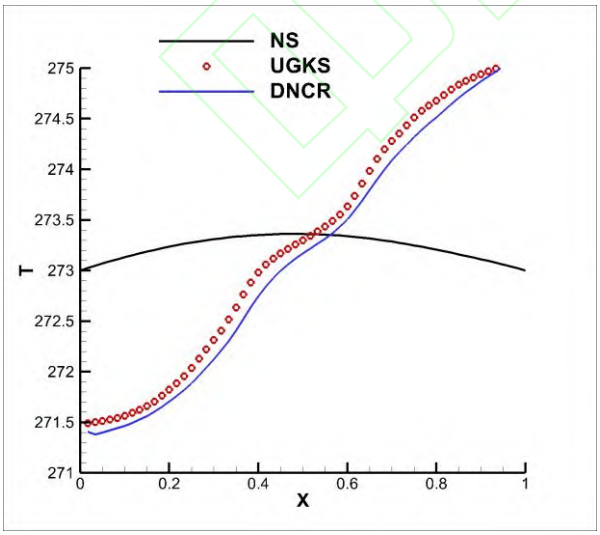
Fig. 15 Comparison of $Kn0.5$ calculation results



(a) $Y=0.5\,\mathrm{m}$ 速度分布对比



(b) $Y=0.5\,\mathrm{m}$ 压力分布对比



(c) $Y=0.5\,\mathrm{m}$ 温度分布对比

图 16 $Kn1.5$ 计算结果对比

Fig. 16 Comparison of $Kn1.5$ calculation results

5 结论与展望

机器学习中的极端随机树模型对基于数据驱动非线性本构关系（DNCR）的复杂高维非线性回归问题具有较好的建模能力，其中特征参数的选择与模型参数的调优是机器学习建模中十分重要的关键问题，直接影响模型性能与精度。本文在 DNCR 方法基础上通过不同努森数条件下二维顶盖驱动方腔流算例对高维非线性建模涉及的特征参数选取、参数调优开展相关验证与研究工作，选取若干典型状态（包括训练集外延状态）对极端随机树模型的泛化性能开展研究，评估了相关模型与方法的计算精度与计算效率。计算结果初步表明通过特征参数筛选与模型调参，DNCR 中的极端随机树模型不仅对训练集努森数上下限范围内中间状态具有较好的预测能力，对训练状态范围以外的外延状态预测效果也较好，体现出了一定的泛化能力。同时，DNCR 方法在模型训练完成后计算效率与 NS 方程基本保持同一量级，能够同时提高现有非平衡流数值方法的计算精度与计算效率，具有较高应用潜力。

然而，作为机器学习方法在物理建模过程中的应用案例，这类方法仍普遍面临（1）与已知物理定律关联困难；（2）公开高可信度数据集匮乏的难题。因此，针对现有 DNCR 方法在机器学习模型泛化性能、训练集数据来源与数据成本及气固边界条件物理适定性方面的研究尚存在较大空白，需要有针对性的开展相关理论与应用创新性研究，进一步提升 DNCR 方法非平衡流动描述能力。

参 考 文 献

[1] 沈青. 稀薄气体动力学[M]. 北京:国防工业出版社, 2003.

[2] Xu K, Huang J. A unified gas-kinetic scheme for continuum and rarefied flows[J]. Journal of Computational Physics, 2010,229(20):7747-7764.

[3] 刘沙,王勇,袁瑞峰,张瑞,陈健锋,朱亚军,卓丛山,钟诚文.统一气体动理学方法研究进展[J].气体物理,2019,4(4):1-13).

[4] Bird G A. Molecular gas dynamics and the direct simulation of gas flows[M]. Oxford: Clarendon Press, 1994:

458.

[5] Jeong S, Chiba K, Obayashi S. Data mining for aerodynamic design space[C]. AIAA-2005-5079, 2005.

[6] Kumano T, Jeong S, Obayashi S, et al. Multidisciplinary design optimization of wing shape for a small jet aircraft using Kriging model[C].AIAA-2006-932 , 2006.

[7] Chiba K, Obayashi S. Data mining for multidisciplinary design space of regional jet wing[J]. Journal of Aerospace Computing, Information, and Communication, 2007, 4:1019-1036.

[8] 陈杰, 孙刚. 基于SOM神经网络的超临界翼型设计[J] .力学季刊, 2011, 32(3) : 411-417.
Chen Jie, Sun Gang. Supercritical airfoil design based on SOM neural network[J]. Chinese Quarterly of Mechanics, 2011, 32(3): 411-417. (in Chinese)

[9] 司景喆,孙刚.基于神经网络的风机叶片叶尖翼型设计[J].力学季刊, 2012, 33(4) : 672-678.
Si Jingzhe, Sun Gang. Design of wind turbine blade based on SOM [J]. Chinese Quarterly of Mechanics, 2012, 33(4): 672-678. (in Chinese)

[10] Milano Michele, Koumoutsakos Petros. Neural Network Modeling for Near Wall Turbulent Flow[J]. Journal of Computational Physics, 2002, 182(1): 1-26.

[11] Yarlanki S.-Rajendran-B.-Hamann-H. Estimation of turbulence closure coefficients for data centers using machine learning algorithms[C]//13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, USA: 13th InterSociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems, 2012: 38-42.

[12] Tracey B,Duraisamy K, Alonso JJ. A Machine Learning Strategy to Assist Turbulence Model Development[C]//53rd AIAA Aerospace Sciences Meeting, [S.l.]: American Institute of Aeronautics and Astronautics, 2015.

[13] Ling J, Kurzawski A, Templeton J. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance[J]. Journal of Fluid Mechanics, 2016, 807: 155-166.

[14] Xiao H, Wu J, Wang J, et al. Quantifying and Reducing Model-form Uncertainties in Reynolds Averaged Navier-stokes Equations: a Data-driven, physics-informed Bayesian Approach[J]. Journal of Computational Physics, 2016, 324: 115-136.

[15] Edeling W N, Cinnella P, Dwight R P, et al. Bayesian estimates of parameter variability in the k–ε turbulence model[J]. Journal of Computational Physics,2014 258: 73-94.

[16] Wang Z, Lan C E, Brandon J M. Fuzzy logic modeling of nonlinear unsteady aerodynamics [R]. AIAA-98-4351 ,1998.

[17] Wang Z, Lan C E, Brandon J M. Fuzzy logic modeling of lateral-directional unsteady aerodynamics[R]. AIAA-1999-4012, 1998.

[18] Wang Z, LiJ, Lan C E, etal. Estimation of unsteady aerodynamic models from flight test data [R]. AIAA-2001-4017, 2001.

[19] Wang Z, Lan C E, Brandon J M. Estimation of lateral-directional unsteady aerodynamic models from flight test data[R]. AIAA-2002-4626, 2002.

[20] Lan C E, Li J, Yau W, et al. Longitudinal and lateral-directional coupling effects on nonlinear unsteady aerodynamic modeling from flight data[R]. AIAA-2002-4804, 2002.

[21] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees[J]. Machine Learning, 2006,63(1):3-42.

[22] 王爱平,万国伟,程志全,et al.支持在线学习的增量式极端随机森林分类器[J].软件学报,2011,22(9):2019-2074.

[23] 于普兵. 基于DSMC和气体动理学统一格式的激波结构模拟[A]. 中国力学学会物理力学专业委员会. 第十二届全国物理力学学术会议论文摘要集[C].中国力学学会物理力学专业委员会:中国力学学会,2012:1.

[24] 赵文文. 高超声速流动Burnett方程稳定性与数值计算方法研究[D].浙江大学,2014.

[25] 李航.统计学习方法[M].北京:清华大学出版社,2012.03

[26] 张振亚,王进,程红梅,王煦法.基于余弦相似度的文本空间索引方法研究[J].计算机科学,2005,32(9):160-163.

[27] 贾怀勤. 应用统计[M]：对外经济贸易大学出版社，2010.03.

Study of machine learning method in the correction of rarefied nonlinear constitutive relations

LI Tingwei[1,3], ZHANG Mang[2], ZHAO Wenwen[1,*], CHEN Weifang[1], JIANG Lijian[1]

1. *College of Aeronautics and Astronautics, Zhejiang University, Hangzhou 310027, China*
2. *Research & Development Center, China Academy of Launch Vehicle Technology, Beijing 100076, China*
3. *Department of Aerospace Systems and Applications, The 54th Research Institute of CETC, Shijiazhuang 050081, China*

Abstract: The continuum medium hypothesis in rarefied non-equilibrium flow field has been invalid. The rarefied non-equilibrium flow is mainly researched around the Boltzmann equation and the Unified Gas-kinetic Scheme (UGKS) is a representative method. About numerical simulation of rarefied non-equilibrium flow, the NS equation has high efficiency but low accuracy and the UGKS method has high accuracy but low efficiency. In this paper, a data-driven method for solving the nonlinear constitutive relations of rarefied non-equilibrium flow based on the NS equation and the UGKS method is proposed. The flow field numerical simulation results of the NS solver and the UGKS solver are used as the data set. Based on the characteristic parameters of the flow field, an extremely randomized trees algorithm is used to nonlinearly correct the linear viscous stress term and heat flux term of the NS equation. The numerical solution of the rarefied non-equilibrium flow is obtained by solving the NS macro-conservation equation by coupling nonlinear constitutive relations. Using a two-dimensional lid-driven cavity case to carry out related work which is selecting characteristic parameters and tuning parameters involved in high-dimensional non-linear modeling. Several states are selected to study the generalization ability of the extremely randomized trees model. Finally, the evaluation of the calculation accuracy and calculation efficiency shows the superiority of the method proposed in this paper.

Keywords: rarefied non-equilibrium flow; constitutive relations; data-driven; extremely randomized trees; model parameters

---

Received：2020-06-09；Revised：2020-07-02；Accepted：2020-08-30；Published online:
URL:
Foundation item：National Numerical Wind Tunnel Project (NNW2019ZT3-A08); National Natural Science Foundation of China (6162790014)
*Corresponding author. E-mail: wwzhao@zju.edu.cn