

基于分级神经网络的柱形代数分解变元序选择

朱章鹏^{1,2} 陈长波²

1 重庆邮电大学 重庆 400065

2 中国科学院重庆绿色智能技术研究院 重庆 400714

(z.zhangpeng@qq.com)

摘要 柱形代数分解是广泛应用于求多项式系统实数解的一种计算方法。不同的变元序对其计算时间有显著影响。已有选序算法多基于启发式的经验算法,准确率不高。少数基于机器学习的方法使用的数据集较小,且基于复杂人工特征。文中在随机生成大量多项式系统与所有序计算时间标注的数据基础上,提出一类新的多项式显性表示特征和一种新的分级神经网络。首先根据最差序计算时间将数据集划分成4个不同计算难度的子集并分别建立预测最优序的分类模型,其次建立预测最长计算时间的回归模型,最后根据回归模型预测最长计算时间并据其自动选择相应难度分类模型预测最优变元序。实验结果表明,显性特征的性能优于复杂人工特征,且在困难问题上分级神经网络所预测最优序的性能约为经验选序算法的3倍。

关键词: 分级神经网络;柱形代数分解;变元序;回归;特征选择

中图法分类号 TP391

Variable Ordering Selection for Cylindrical Algebraic Decomposition Based on Hierarchical Neural Network

ZHU Zhang-peng^{1,2} and CHEN Chang-bo²

1 Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2 Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China

Abstract Cylindrical algebraic decomposition (CAD) is a widely used approach for computing the real solutions of polynomial systems. The choice of variable ordering has a significant impact on its computation time. Most of existing ordering selection algorithms are based on heuristic empirical algorithms, whose accuracy are not high. A few approaches based on machine learning use small data sets and are based on complex human characteristics. In this paper, on the basis of randomly generating a large set of polynomial systems, which are tagged with timings obtained by applying different orderings for computing CAD, a new kind of explicit representation feature and a new hierarchical neural network are proposed. Firstly, according to the computation time of CAD with the worst ordering, the data set is divided into four subsets with different computation difficulties, and the classification models are established respectively. Secondly, a regression model for predicting the longest computation time is built. Finally, the longest computation time is predicted according to the regression model, based on which a classification model with right computation difficulty is automatically selected to predict the optimal variable ordering. Experimental results show that the performance of explicit features is better than that of complex handcrafted features, and the performance of the optimal ordering predicted by hierarchical neural network on difficult problems is about two times better than that of an empirical algorithm.

Keywords Hierarchical neural network, Cylindrical algebraic decomposition, Variable ordering, Regression, Feature selection

很多科学和工程问题都可以转化为代数方程求解^[1-3]。因此,研究出高效的计算机代数方程求解工具具有重要意义。柱形代数分解(Cylindrical Algebraic Decomposition, CAD)是计算机代数中求解半代数系统的一种重要方法^[4]。在使用柱形代数分解方法时,需要输入一个变元序。研究表明,变元序的选择对使用柱形代数分解求解半代数系统的时间有重要影响^[5]。

机器学习是人工智能的一个分支。机器学习通过大量数

据的训练,发现数据的内在规律,从而提高分类和预测的准确性。数据的表示对机器学习的性能有重要影响。当数据集足够大时,深度学习可以从数据中学习数据的表示,从而提升模型的性能^[6]。目前,机器学习被广泛应用于推荐算法、计算机视觉、物联网等计算机应用领域,同时也与医学、材料、经济学等领域形成交叉应用学科。在计算机数学中,机器学习应用也非常广泛,如 Mall 等^[7]提出用单层 Chebyshev 神经网络求解椭圆偏微分方程, Lample 等^[8]用机器学习求解符号微积

基金项目:国家自然科学基金面上项目(11771421, 1671377, 61572024);中国科学院“西部之光”;重庆市院士牵头科技创新引导专项(cstc2018jcyj-yszxX0002)

This work was supported by the National Natural Science Foundation of China (11771421, 1671377, 61572024), Chinese Academy of Sciences “Light of the West”, Chongqing Academician-led Science and Technology Innovation Guidance Project (cstc2018jcyj-yszxX0002).

通信作者:陈长波(chenchangbo@cigit.ac.cn)

分问题,并且模型性能明显优于符号数学工具。

机器学习的发展为解决柱形代数分解变元序提供了新的解决方案。在 QEPCAD、Reduce、Mathematica、Maple 等中实现的柱形代数分解软件包基于传统经验算法实现变元序,这些方法往往对某一类问题效果较好,而很难适用于所有问题。基于 Chen 等^[9]在 Maple 软件包 RegularChains 中实现的柱形代数分解,Huang 等^[10]提出使用机器学习选择启发式算法,England 等^[11]用机器学习算法根据计算时间直接预测变元序,我们在之前的研究中^[12]使用支持向量机的方法根据胞腔个数直接预测变元序。实验表明,在给定数据集上,机器学习方法在解决变元序问题上比传统经验序算法效果更好。

不同于已有的研究,本文的数据集(282 802)远大于以往研究的数据集(6 117, 32 680)。对比已有的研究发现,当数据集较小时手工设计的特征明显优于数据的自然特征。一种自然的想法是当数据集足够大时,使用深度学习的方法直接从数据中学习特征。但本文第 3 节实验表明,直接使用多项式系数得到的特征效果并不理想。对已有研究中使用的手工特征进一步分析发现,这些特征和系数大小无关,只依赖于多项式项的信息。受此启发,本文在多项式的显性特征(多项式系数特征)的基础上衍生出基于多项式项的特征,即不考虑系数大小,只考虑系数零或非零。

同时,对数据集进一步分析发现,大部分样例的最坏序计算时间小于 10s,属于相对容易的问题,而我们更关心困难样例最佳序的选择。为了解决样本均衡问题,本文提出了一种分级神经网络:首先根据最坏变元序(最长计算时间)将数据集划分成不同子类,对每一子类建立预测最佳变元序(最短计算时间)的分类模型,其次建立回归模型预测最大计算时间。预测时,利用回归模型划分问题难度,自动选择相应分类模型,预测最优变元序。实验结果表明,在不同问题难度的数据集上分级神经网络上都明显优于 maple 命令 SuggestVariableOrder(SVO) 实现的经验选序算法。分级神经网络在预测问题难度的同时能有效预测最优变元序,在困难数据集上所选序计算 CAD 的时间只有传统选序算法 SVO 的 1/3。

本文第 1 节简要介绍了柱形代数分解序问题;第 2 节介绍本文实验数据的生成、特征选择和数据预处理;第 3 节介绍分类神经网络对比实验,包括不同特征选择方法的实验结果及对比分析;第 4 节介绍分级神经网络的结果及对比实验。

1 柱形代数分解简介

1973 年,G. E. Collins 提出了柱形代数分解方法^[13],随后大量学者对其进行了改进^[14-20]。柱形代数分解是将有序的空空间 R^n 分解成有限个连通的半代数子集,并且所有子集呈柱形排列,其对任意的 $1 \leq k < n$, R^n 上的两个子集 A 和 B 在 R^k 上的投影完全相同或不相交^[13]。

在使用 CAD 时,变元序的选择会影响 CAD 的计算时间和效率^[20]。例如,给定多项式系统:

$$f = \begin{cases} 23x_2^3 - 36x_3^3 - 51x_1x_2 + 5 \\ 57x_1x_3 + 33x_2^2 + 17x_2x_3 - 91x_3^2 + 29 \end{cases} \quad (1)$$

使用 Chen 等^[9,21]在 Maple 软件包 RegularChains 中实现的 CylindricalAlgebraicDecompose,当输入的序为 $x_1 > x_3 > x_2$

时,胞腔个数为 155,计算时间为 0.52 s;当输入的序为 $x_2 > x_3 > x_1$ 时,胞腔个数为 2 147,计算时间为 137.53 s。

图 1 对本文数据集中最佳变元序和最差变元序的计算时间做了比较分析。

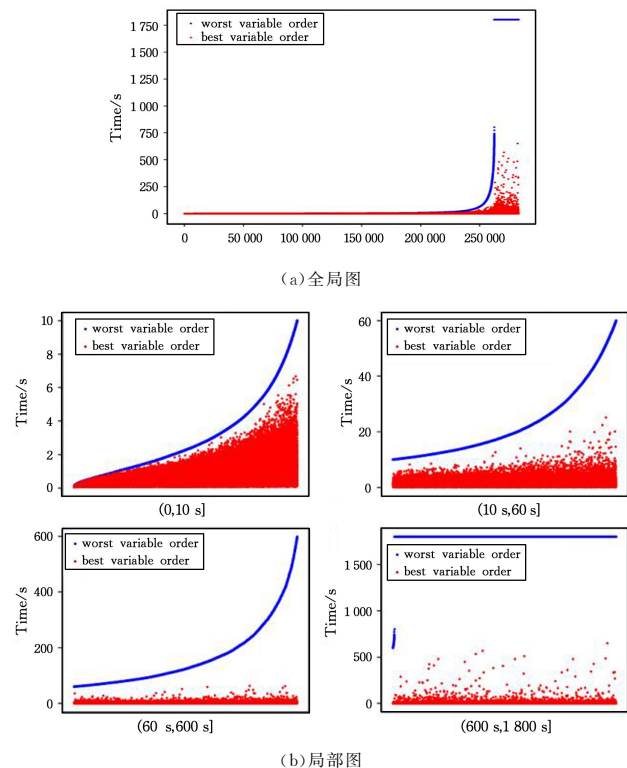


图 1 变元序计算时间

Fig. 1 Computing time of variable orders

从全局图中可以看出,大部分数据最优序与最差序的时间差较小;从局部图可以看出,随着最差序计算时间的增长,最佳变元序与最差变元序的时间差也变大。

2 实验数据

2.1 数据生成

本文的数据集是通过数学软件 Maple 生成,其中用 RegularChains 工具包用来计算多项式系统调用 CAD 计算的 actual 计算时间^[21]。多项式系统的结构如下:

$$F = \begin{cases} \sum_{m=1}^4 Cx_1^a x_2^b x_3^c + D \\ \sum_{m=1}^4 Cx_1^a x_2^b x_3^c + D \end{cases} \quad (2)$$

$$0 \leq a + b + c \leq 3, -100 \leq C \leq 100, -200 \leq D \leq 200$$

为在有限时间内得到大量数据,将每个变元序的最大计算时间限制为 900 s。三变元多项式的变元序为 (x_1, x_2, x_3) , (x_1, x_3, x_2) , (x_2, x_1, x_3) , (x_2, x_3, x_1) , (x_3, x_1, x_2) , (x_3, x_2, x_1) 。

在一个具有 5 个节点、每个节点配置 2 个 Intel E5-2620 CPU(6 核)、64 GB 内存的计算机集群上,15 个进程并行计算,运行 17 天 9 小时后总共得到 282 802 条有效数据。

2.2 特征选择

在机器学习实验中,数据特征的选择会直接影响机器学习模型的上限,因此,特征选择的方法在机器学习中尤为重要。在文献^[9]中,手工设计的基于多项式图结构的特征的效

果优于多项式系数特征的效果,本文实验中引用了这两种特征。实验过程中,手工设计的特征都是基于多项式的项,因此本文在多项式系数的基础上衍生出多项式项的 01 特征。

(1) 基于多项式图结构的特征

在文献[9]中提出了一种基于多项式图结构的特征提取方法,该方法一共提取了多项式的 10 种特征:

- 1) 和 $\{x_i | i=1, 2, 3\}$ 出现在同一多项式中 $\{x_j | j \neq i\}$ 的个数;
- 2) $\{x_i | i=1, 2, 3\}$ 出现的多项式的个数;
- 3) $\{x_i | i=1, 2, 3\}$ 在所有多项式中次数 (degree) 的最大值;
- 4) $\{x_i | i=1, 2, 3\}$ 在每个多项式中最大次数之和;
- 5) $\{x_i | i=1, 2, 3\}$ 作为主变元在每个多项式中的首项系数次数的最大值;
- 6) $\{x_i | i=1, 2, 3\}$ 在每个多项式中所属单项个数的最大值;
- 7) $\{x_i | i=1, 2, 3\}$ 在所有多项式中所属单项次数的最大值;
- 8) $\{x_i | i=1, 2, 3\}$ 在所有多项式中的次数之和;
- 9) $\{x_i | i=1, 2, 3\}$ 作为主变元在每个单项式中的首项系数次数之和;
- 10) $\{x_i | i=1, 2, 3\}$ 在所有多项式中所属单项次数之和。

基于此特征选择方法,本文数据集提取的特征一共 30 个,后文将使用此特征选择的方法称为 Graph Features(GF)。

(2) 基于多项式系数的特征

根据本文实验数据的多项式结构,补入多项式中的缺项,并根据字典序排列,将所有单项的系数作为特征。本文多项式项的系数(稠密表示)一共有 40 个,因为是显性特征,因此称为 Coefficients Features(CF)。

(3) 多项式项的 01 特征

由于多项式项的系数差异性较大,在一定程度上过于突出单项的重要性,因此在多项式系数特征的基础上衍生出了基于多项式项的 01 特征。将多项式项的系数用 01 表示,其中系数不为 0 的表示为 1,其余的表示为 0。相对于依赖经验而人为设计、手动提取的 30 个特征,01 表示的项的特征能更好地表述多项式结构。由于特征表示为 0、1,因此称其为 One-hot Features(OHF)。

给定多项式:

$$f = \begin{cases} 61x_1^3 + 94x_1x_2^2 + 81 \\ -93x_1^3 - 58x_1 + 21x_3 - 61 \end{cases} \quad (3)$$

在上述多项式的基础上,根据上述 3 种特征选择方法提取的特征分别如下:

Graph Feature: 2, 1, 1, 2, 1, 1, 3, 2, 1, 6, 2, 1, 0, 1, 0, 2, 1, 1, 3, 3, 1, 8, 2, 1, 0, 1, 0, 4, 1, 1。

Coefficients Features: 81, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 94, 0, 0, 0, 61, -61, 21, 0, 0, 0, 0, 0, 0, 0, 0, -58, 0, 0, 0, 0, 0, 0, 0, 0, -93。

One-hot Features: 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1。

2.3 数据预处理

由于实验数据中多项式变元序的个数固定为 6 个,因此

将问题划分为多分类问题。根据不同变元序的计算时间,将计算时间最短的变元序标记为最优变元序,并只标记一个最优变元序。由于部分数据中存在计算时间超时的变元序,因此将这部分变元序的计算时间设定为两倍时间上限 1800s。

为提高神经网络的训练速度,在训练之前,先将数据用 z-score 方法进行归一化(OHF 不需要归一化),将其中 4/6 的数据作为训练集,1/6 的数据作为验证集,1/6 的数据作为测试集。实验中,根据验证集的效果对网络模型和参数进行调整,以获取最优模型,最后用测试集测试模型泛化效果。本文的实验结果均基于测试集。

3 分类神经网络

分级神经网络分为非线性回归网络和分类神经网络两部分。文中首先直接使用分类神经网络对不同特征选择方法做了对比实验,得到其中最优的特征选择方法,然后用该特征选择方法进行分级神经网络实验。

3.1 分类神经网络结构

BP 神经网络是一种逆误差方向传播算法,核心是利用梯度搜索技术,通过信号的向前输出和误差的反向输出不断调整网络参数,以达到最小误差^[22]。在 BP 神经网络中,隐含层的层数和节点个数的选取会影响神经网络的训练效率和训练精度。一般来说,隐含层过于简单会降低网络容错性,造成“欠拟合”;隐含层过于复杂虽然会提高网络精度,但容易造成“过拟合”。

本文用于分类的 BP 人工神经网络结构如图 2 所示。

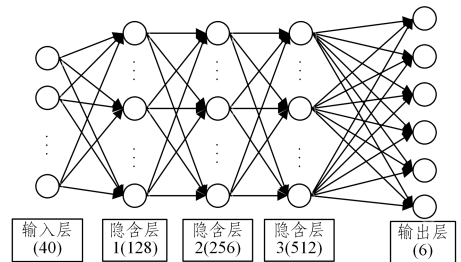


图 2 分类神经网络结构

Fig. 2 Classification neural network structure

网络包含三层隐含层,隐含层节点个数用试凑法进行测试。在实验过程中需要根据具体模型在验证集上的准确率作相应调整,最终发现大部分模型使用 $128 \times 256 \times 512$ 节点个数的三层隐藏层网络结构效果最好。

3.2 模型评估标准

在使用神经网络的方法直接预测最优变元序时,对于一个多项式系统模型预测结果只有一个最优序,但在实际数据中一个多项式系统可能有多个最优序。为了有效评估模型的实际预测效果,实验中测试准确率的计算方法如下:

$$Accuracy = \frac{\sum_{n=1}^D \frac{|R \cap P|}{|R|}}{D} \quad (4)$$

其中, D 为全部测试集大小, R 为实际标签, P 为预测标签。当预测序为多个实际最优序中的一个时就将模型预测结果标记为 +1。将测试集中所有预测为正确的样例个数除以测试集大小便得到最终测试集中模型预测的准确率。

3.3 分类模型的训练与测试

在实验中,首先使用 DF 进行实验,获取最优模型(Model for Coefficients Feature-s, M-CF)。然后在测试集上将测试结果与实际最优时间(best variable order time, BVO-T)、SVO 所选最优序的计算时间作比较,结果如表 1 所列。

表 1 原始数据集对比实验

Table 1 Comparison experimentation of original dataset

	BVO-T	SVO	M-CF
Average time/s	0.998	4.33	5.59
Accuracy/%	100	47.30	58.31

实验结果表明,用 DF 直接进行实验时,最终在测试集上预测变元序的平均计算时间多于 SVO 方法的平均计算时间,说明直接使用数据自然特征的效果较差。

在实验中通过多数数据进行分析观察后发现数据的计算时间的分布情况很不均衡,大量数据的最大计算时间较短,因此根据多项式的最大计算时间进行分段,划分为(0,10s],[10s,60s],[60s,600s],[600s,1800s] 4 个时间区间,划分后的数据分布如表 2 所列。

表 2 数据分布

Table 2 Data distribution

Time/s	(0,10]	(10,60]	(60,600]	(600,1800]
Count	210561	38540	13061	20640

根据 4 个时间区间,将数据集重新划分为 4 个部分,并重新划分训练集、验证集、测试集。

对比分析不同特征表示的数据发现,当使用 GF 表示数据时存在大量特征相同的数据。因此在基于 OHF 表示的基础上对数据进行清洗。清洗完成后,数据集大小由 282802 减少为 233145。

根据 4 个时间区间,重新将用 OHF 表示的数据集进行划分,在保证数据集相同的条件下,使用 OHF,CF 和 GF 分别建立机器学习模型进行对比(Model for One-hot features, M-OHF; Model for Coefficients Features, M-DF; Model for Graph Features, M-GF),对不同特征选择方法和实际平均计算时间、SVO 的平均计算时间进行比较。除此之外,文中还用全部数据集(233145)的 OHF 建立模型(M-OHF-Total)进行比较。实验结果如图 3 所示。

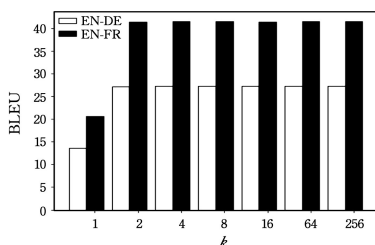


图 3 基于不同特征的手动分类神经网络

Fig. 3 Manual neural network classification based on different features

与经验择序算法 SVO 相比,分类神经网络的预测序计算 CAD 的平均计算时间均明显更短。从整体来看,在划分时间区间后,M-OHF、M-CF、M-GF 3 种方法在(0,600s]区间内的平均时间相差不大,但是在(600s,1800s]区间内 M-OHF 的

平均计算时间明显更短,说明基于多项式项的特征(OHF)优于手工提取的多项式的特征(GF)。

当特征选择方法为 OHF 时,相比于用全部数据集直接实验(M-OHF-Total),在(600s,1800s]区间内,划分时间区间的方法(M-OHF)的平均计算时间减少 10s,说明划分时间区间的方法在预测难度较大的问题时性能有极大提升。

4 分级神经网络

上述实验结果表明,OHF 特征选择方法优于其他两种特征选择方法,同时应当根据不同的时间区间进行分别建模。本节提出一种分级神经网络,通过第一级的回归模型自动选择针对不同时间区间的分类模型;然后通过二级分类模型预测最佳变元序。

4.1 非线性回归模型

在分级神经网络结构中首先通过非线性回归算法建立回归模型预测多项式的最大计算时间。非线性回归模型结构如图 4 所示。

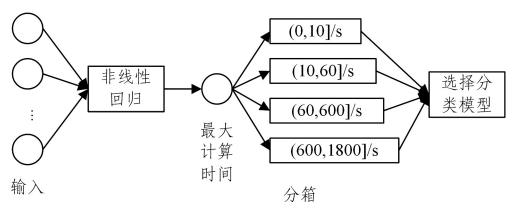


图 4 非线性回归网络结构

Fig. 4 Nonlinear regression network structure

首先根据最大计算时间建立非线性回归模型,预测多项式的最大计算时间,然后根据 4 个时间区间进行分箱处理,得到多项式系统的计算难度,最后就可以根据计算难度选择相应的分类模型。

为了和不采用分级神经网络的 OHF 实验进行对比,在分级神经网络的实验中,使用与 OHF 实验相同的训练集、验证集和测试集。在使用线性回归模型预测多项式系统的最大计算时间后,将预测时间进行分箱处理,分别划分到 4 个时间区间内,分箱的准确率为 78.85%。

4.2 分级神经网络结构

在分类神经网络和非线性回归网络的基础上,本文提出了能同时预测多项式计算难度和最优变元序的分级神经网络。网络结构如图 5 所示。

分级神经网络的训练与测试的完整过程如下:1)生成多项式组,计算 CAD 使用不同序的计算时间,得到初始数据集;2)根据最大计算时间划分 4 个时间区间,将数据集分成 4 部分;3)利用特征工程,得到数据的特征表示;4)得到最大计算时间的完整数据集;5)建立预测最大计算时间的回归模型;6)根据第 2)步划分的 4 个数据集和数据的特征表示,以最短计算时间为标准标记数据标签,得到 4 个子数据集;7)用 BP 神经网络分别建立预测的最有序的分类模型;8)根据特征工程,得到新多项式组的特征表示;9)用回归模型预测新多项式组的最大计算时间;10)根据最大计算时间,得到新多项式组应选择的分类模型;11)用对应的分类模型,预测新多项式组的最优变元序;12)用 CAD 和新多项式组的最优变元序,得到对应的计算结果。

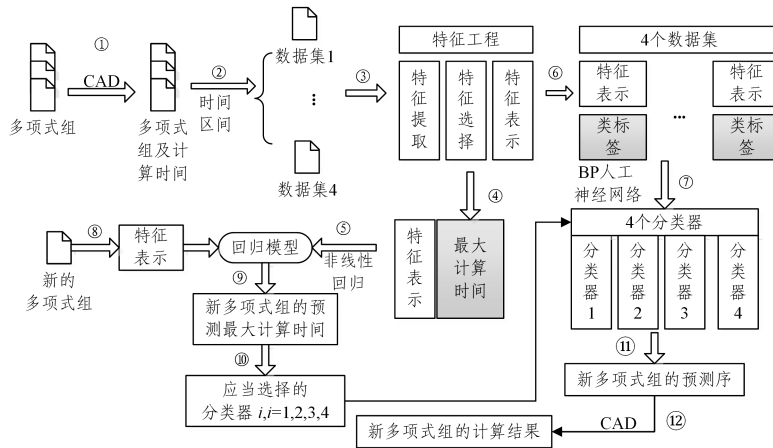


图5 分级神经网络结构

Fig. 5 Hierarchical neural network structure

4.3 分级神经网络训练与测试

采用分级神经网络可以预测多项式系统的计算难度,并根据难度自动选择最佳的分类神经网络。分类神经网络(Multi-class classification neural network, MNN)与分级神经网络(Hi-erarchical neural network, HNN)及 M-OHF-Total 的实验对比如图6所示。

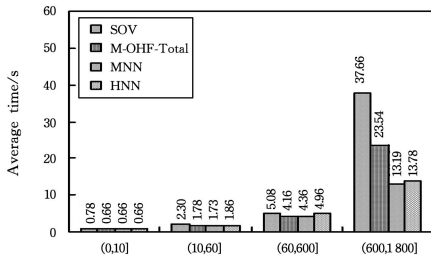


图6 分级神经网络对比实验

Fig. 6 Hierarchical neural network comparison experiment

上述4个实验预测序中计算时间超时的样例个数如表3所列。

表3 计算超时的样例个数

Table 3 Number of timeout examples

方法	SVO	M-OHF-Total	MNN	HNN
数量/个	58	24	13	14

在4个时间区间内,基于神经网络的方法所用时间均明显比SVO方法时间短。对于较困难问题(600s以上),根据不同时间区间建模的MNN和HNN方法在困难问题上明显好于直接使用所有数据建模的M-OHF-Total方法,并且预测序中计算时间超时的样例个数明显少于SVO和M-OHF-Total。采用分级神经网络HNN的优点在于可以实现自动划分问题难度并预测最优变元序,而且性能和手动划分时间区间建模的MNN方法不相上下。当最大计算时间低于600s时,最优变元序的平均计算时间均低于5s;当最大计算时间大于600s时,平均计算时间达到13s左右。

在实际应用中,当计算时间低于5s时所需要的计算资源较少,等待时间较短;当计算时间平均大于13s时,计算资源增多,且等待时间增长,而人们往往无法接受长时间的等待。

结束语 目前,柱形代数分解变元择序多采用传统的经验选择算法。本文随机生成大量数据,利用神经网络建立机器学习模型预测变元序。通过对数据集和已有特征选择方法

进行分析比较,发现了数据中隐藏的一些关键信息,并据此提出一种新的衍生显示特征,在此基础上提出了一种新的分级神经网络预测变元序。实验结果表明该方法解决困难问题的速度远快于传统的经验算法。未来,我们拟将该方法推广到三变元以上的多项式系统,并将其应用于半代数系统的求解中。

参考文献

- [1] LLIAS S, KOTSIREAS, EDGAR M. Applications of computer algebra[M]. Kalamata: Springer International Publishing, 2017.
- [2] BLUMLEIN J, MAIER A, MARQUARD P, et al. From momentum expansions to post-Minkowskian Hamiltonians by computer algebra algorithms[J]. Physics Letters B, 2020, 801.
- [3] JOSE L G G, GABRIEL A V, PEDRO R C, et al. SFOPDES: A Stepwise First Order Partial Differential Equations Solver with a Computer Algebra System [J]. Computers and Mathematics with Applications, 2019, 78(9): 3152-3164.
- [4] WANG D M, XIA B C, LI Z M. Computer Algebra[M]. Beijing: Tsinghua University Press, 2007.
- [5] BROWN C W, DAVENPORT J H. The complexity of quantifier elimination and cylindrical algebraic decomposition[C]// International Symposium on Symbolic and Algebraic Computation, 2007: 54-60.
- [6] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning [M]. London: Massachusetts Institute of Technology Press, 2016.
- [7] MALL S, CHAKRAVERTY S. Single Layer C-hebyshev Neural Network Model for Solving Elliptic Partial Differential Equations[J]. Neural Processing Letters, 2017, 45(3): 1-16.
- [8] LAMPLE G, CHARTON F. Deep Learning for Symbolic Mathematics [EB/OL]. (2019-10-02) [2019-12-23]. <https://arxiv.org/abs/1912.01412>.
- [9] CHEN C B, MAZA M M, XIA B C, et al. Computing cylindrical algebraic decomposition via triangular decomposition[C]// International Symposium on Symbolic and Algebraic Computation, 2009: 95-102.
- [10] HUANG Z Y, ENGLAND M, WILSON D J, et al. Applying machine learning to the problem of choosing a heuristic to select the variable ordering for cylindrical algebraic decomposition [C]// Intelligent Computer Mathematics, 2014: 92-107.