

基于有序编码的核极限学习顺序回归模型

李佩佳^{①②③} 石勇^{*②③④} 汪华东^{②③} 牛凌峰^{②③④}

^①(中国科学院大学计算机与控制学院 北京 101408)

^②(中国科学院大数据挖掘与知识管理重点实验室 北京 100190)

^③(中国科学院虚拟经济与数据科学研究中心 北京 100190)

^④(中国科学院大学经济与管理学院 北京 100190)

摘要: 顺序回归是机器学习领域中介于分类和回归之间的有监督问题。在实际中,许多带有序关系标签的问题都可以被建模成顺序回归问题,因此顺序回归受到众多学者的关注。基于极限学习机(ELM)的算法能有效避免因迭代过程陷入的局部最优解,减少训练时间,但基于极限学习机的算法在顺序回归问题上的研究较少。该文将核极限学习机与纠错输出编码相结合,提出了一种基于有序编码的核极限学习顺序回归模型。该模型有效解决了如何在顺序回归中取得良好的特征映射以及如何避免传统极限学习机中隐层节点个数依赖于人工设置的问题。为验证提出模型的有效性,该文在多个顺序回归数据集上进行了测试,测试结果表明,相比于传统ELM模型,该文提出的模型在准确率上平均提升了10.8%,在数据集上预测表现最优,而且获得了最短的训练时间,从而验证了模型的有效性。

关键词: 纠错输出编码; 顺序回归; 极限学习机; 核函数

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2018)06-1287-07

DOI: 10.11999/JEIT170765

Ordered Code-based Kernel Extreme Learning Machine for Ordinal Regression

LI Peijia^{①②③} SHI Yong^{②③④} WANG Huadong^{②③} NIU Lingfeng^{②③④}

^①(School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 101408, China)

^②(Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

^③(Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing 100190, China)

^④(School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China)

Abstract: Ordinal regression is one of the supervised learning issues, which resides between classification and regression in machine learning fields. There exist many real problems in practice, which can be modeled as ordinal regression problems due to the ordering information between labels. Therefore ordinal regression has received increasing interest by many researchers recently. The Extreme Learning Machine (ELM)-based algorithms are easy to train without iterative algorithm and they can avoid the local optimal solution; meanwhile they reduce the training time compared with other learning algorithms. However, the ELM-based algorithms which are applied to ordinal regression have not been exploited much. This paper proposes a new ordered code-based kernel extreme learning ordinal regression machine to fill this gap, which combines the kernel ELM and error correcting output codes effectively. The model overcomes the problems of how to get high quality feature mappings in ordinal regression and how to avoid setting the number of hidden nodes by manual. To validate the effectiveness of this model, numerous experiments are conducted on a lot of datasets. The experimental results show that the model can improve the accuracy by 10.8% on average compared with traditional ELM-based algorithms and achieve the state-of-the-art performance with the least time.

Key words: Error correct output code; Ordinal regression; Extreme learning machine; Kernel function

收稿日期: 2017-07-28; 改回日期: 2018-01-22; 网络出版: 2018-03-14

*通信作者: 石勇 yshi@ucas.ac.cn

基金项目: 国家自然科学基金(71110107026, 71331005, 91546201, 11671379, 111331012), 中国科学院大学资助项目(Y55202LY00)

Foundation Items: The National Natural Science Foundation of China (71110107026, 71331005, 91546201, 11671379, 111331012), The Grant of University of Chinese Academy of Sciences (Y55202LY00)

1 引言

顺序回归(ordinal regression)介于机器学习领域中分类和回归两大类问题之间。与分类不同,顺序回归的标签不仅属于某有限集合而且具有顺序信息,如{非常积极,积极,中性,消极,非常消极};不同于普通回归,顺序回归的顺序标签不具有数值含义。现实中许多应用都被看作是顺序回归问题,如:客户评论打分、情感分析、信用评级、图像分类、排序^[1-3]等。在这些应用中,样本被分到不同的错误类别,代价往往不同。如在信用评估领域中,一个“信用极差”的公司被错分到“中性”或“信用极好”的类别,影响大有不同。所以顺序回归模型在解决许多应用问题上扮演着不可替代的角色。当处理顺序回归问题时需考虑两个事实,一是预测的类别应与实际类别尽可能接近,即错分到不同类别带来的损失不同,二是有效利用标签的顺序信息有助于构建更加合理及高效的模型。

目前,顺序回归领域已有很多成功的模型^[4]。第1类是将顺序回归当作成普通的数值回归或多分类的朴素模型,如支持向量机,支持向量回归机,极限学习机(ELM)^[5,6],该类方法忽略了标签之间的顺序关系。第2类是基于阈值的模型,如支持向量顺序回归机^[7]等,该类方法将序关系用于构建模型即假设具有顺序结构的标签上存在一个潜在的连续变量,通过寻找一个实值预测函数将样本映射到1维的实数轴上,然后寻找一些连续分界点将该实数轴划分为连续的区间,每个区间代表一个类别。第3类是基于顺序分区的模型,该方法是将顺序回归问题根据标签的序关系进行有序的二元分解,将其转化为多个二元分类问题,然后采用一个或者多个模型进行处理。这类方法将标签间的顺序信息与第1类普通多分类模型相结合,使顺序信息有效嵌入到普通多分类模型中,这不仅继承和保留了普通多分类模型的优点,而且还能利用标签的顺序信息。文献[4]通过实验比较了多个顺序回归模型,实现性能显示基于ELM的顺序回归模型在训练时间上具有明显优势。一个好的顺序回归模型不仅需要性能好还应具备训练速度快的优点。所以本文提出基于有序编码的核极限学习顺序回归模型,将顺序信息和核函数嵌入到普通多分类极限学习机中,该模型属于第3类顺序回归模型。

极限学习机由学者Huang等人^[5]提出,常被看作为单层前馈神经网络训练算法并用于分类或回归^[8],主要特点包括:(1)连接输入层与隐层之间的权重矩阵和隐藏层偏置向量直接被随机初始化;(2)不需要基于梯度更新的迭代算法防止陷入局部最优

解。这两个特点使ELM不仅具备统计学习模型无需较多训练数据的优点,而且拥有神经网络学习模型的自动学习有效特征的优势。ELM相比于神经网络算法简化了众多参数学习、缩短了训练时间,另外还提高了学习效率及泛化性能^[5]。因此,极限学习机引起学者广泛关注并在分类和回归的多个领域中得到了成功应用,如在线学习^[9]、噪声预测^[10]等。另外,极限学习机也为深度学习领域中模型压缩问题提供了新的解决思路^[11]。

ELM可被应用在顺序回归领域。Deng等人^[12]将ELM算法应用在顺序回归问题中,其中ELM的良好表现证明了ELM在顺序回归中的有效性。Riccardi等人^[13]提出应用在顺序回归上的集成ELM模型。最近一项关于顺序回归的全面综述^[4]显示基于ELM的顺序回归模型^[12]在性能可比较前提下,在训练时间上拥有不俗的表现。但据我们所知目前基于ELM的顺序回归模型相对较少,ELM算法在顺序回归领域研究还不充分。具有核函数的ELM将样本从低维空间映射到高维特征空间,使得许多低维不可分的问题在高维空间中变得可分,另外,ELM需要随机选择隐藏节点个数,即对这些参数的选择只能通过试凑法获得,影响了ELM的学习效率和鲁棒性,核函数的引入恰好解决了这些问题。该类模型目前在顺序回归领域尚无公开报道。鉴于此,本文基于有序编码和核极限学习机算法,提出一种新型的核极限学习顺序回归模型,该模型有效解决了如何在顺序回归问题中取得良好的特征映射以及如何避免传统ELM隐层节点个数依赖于人工设置的问题。为了验证该模型的有效性,本文在多个数据集上对该模型进行了测试,并与传统基于ELM的模型进行了分析比较。

本文结构如下:第2节介绍ELM算法原理,第3节介绍本文提出的基于有序编码的核极限学习顺序回归模型,第4节将展示数值实验并进行结果分析,第5节给出总结。

2 极限学习机算法原理

极限学习机由学者Huang等人^[5]提出,其发展具有深厚的理论基础。通用近似定理^[14]说明具有非线性激活函数(如Logistic、Sigmoid函数)的单隐层前馈神经网络可以逼近任意Borel可测的函数^[11]。Huang与Babri^[15]证明了在一个有限训练集中,一个最多具有 N 个隐层节点和任意非线性激活函数的单隐层前馈神经网络可学习 N 个显著样本。该理论指出了ELM所需最多的隐藏节点个数的上界。文献[16]继续证明了ELM具有上述的通用近似能力。另外,文

献[17]证明了随机选择具有 N 个隐层节点的单隐层前馈神经网络的输入权重和隐藏层偏置仍可以有效学习 N 个显著样本。换句话说,输入权重和隐藏层偏置项在实际应用中不需要训练可直接随机获得。Huang等人在文献[5]中继续证明了如果那些激活函数是无限次可微,需要的隐层节点个数小于或等于 N ,其中 N 为样本点个数,并提出了ELM。关于ELM模型的构建,根据Bartlett前馈神经网络的泛化表现

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \mathbf{h}(\mathbf{x}_2) \\ \vdots \\ \mathbf{h}(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \cdots & h_L(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & \cdots & h_L(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_N) & h_2(\mathbf{x}_N) & \cdots & h_L(\mathbf{x}_N) \end{bmatrix} = \begin{bmatrix} g(w_1\mathbf{x}_1 + b_1) & g(w_2\mathbf{x}_1 + b_2) & \cdots & g(w_L\mathbf{x}_1 + b_L) \\ g(w_1\mathbf{x}_2 + b_1) & g(w_2\mathbf{x}_2 + b_2) & \cdots & g(w_L\mathbf{x}_2 + b_L) \\ \vdots & \vdots & \ddots & \vdots \\ g(w_1\mathbf{x}_N + b_1) & g(w_2\mathbf{x}_N + b_2) & \cdots & g(w_L\mathbf{x}_N + b_L) \end{bmatrix}_{N \times L} \quad (2)$$

L 是隐层节点个数, $g(\mathbf{x})$ 是激活函数, β 是连接隐藏层和输出层的权重矩阵, \mathbf{T} 是训练数据真实标签矩阵:

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{L1} & \beta_{L2} & \cdots & \beta_{LM} \end{bmatrix}_{L \times M}, \quad \mathbf{T} = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & \cdots & t_{1M} \\ t_{21} & t_{22} & \cdots & t_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ t_{N1} & t_{N2} & \cdots & t_{NM} \end{bmatrix}_{N \times M} \quad (3)$$

故ELM算法可总结为:

(1)随机初始化隐层节点个数 L ,输入权重 \mathbf{w} 和隐层偏置权重 \mathbf{b} ,激活函数类型 $g(\mathbf{x})$;

(2)计算隐层输出矩阵 \mathbf{H} ;

(3)根据优化目标函数得到输出权重矩阵 $\beta = \mathbf{H}^+ \mathbf{T}$ 其中 \mathbf{H}^+ 是Moore-Penrose广义逆矩阵。

根据脊回归理论^[20]在求广义逆矩阵过程中对 $\mathbf{H}\mathbf{H}^T$ 或 $\mathbf{H}^T\mathbf{H}$ 的对角加入 $1/C$ 可以得到更优的泛化性能,故得 $\beta = \mathbf{H}^T \left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T}$ (当 $N < L$)或

$\beta = \left(\frac{1}{C} + \mathbf{H}^T\mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T}$ (当 $N \geq L$),则相应ELM可

表示为 $f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\beta = \mathbf{h}(\mathbf{x})\mathbf{H}^T \left(\frac{1}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{T}$ 或

$f(\mathbf{x}) = \mathbf{h}(\mathbf{x})\beta = \mathbf{h}(\mathbf{x}) \left(\frac{1}{C} + \mathbf{H}^T\mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T}$ 。

3 基于有序编码的核极限学习顺序回归模型

顺序回归问题可以用以下数学语言描述:给定

理论^[18]:训练误差越小,权重范数越小,则网络的泛化性能越好。基于上述理论保障^[19],ELM用数学描述如下:对任意 N 个样本 (\mathbf{x}_i, t_i) 中 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ 表示 n 维输入向量, $t_i \in R^M$ 是该样本的标签,则ELM优化目标函数可表示为

$$\text{Min: } \|\beta\| + C \|\mathbf{H}\beta - \mathbf{T}\|^2 \quad (1)$$

其中,“ $\|\cdot\|$ ”代表L2范数, \mathbf{H} 是隐层输出矩阵:

训练样本 $(\mathbf{x}_i, t_i) \in R^n \times Y$, $i = 1, 2, \dots, N$,其中 $\mathbf{x}_i \in R^n$ 是第 i 个样本输入, N 是样本个数, n 是样本维度, t_i 是对应样本的标签。假设有 M 类样本,为不失一般性我们令 $Y = \{1, 2, \dots, M\}$ 表示其具有顺序信息的标签集合。顺序回归的目标是找到一个从输入空间到输出空间的映射函数使模型分配的标签尽可能接近样本的真实值。如第1节所述普通多分类模型在不考虑顺序信息前提下可直接应用于顺序回归中。在标准ELM中,ELM将样本标签 t_i 自动编码为标签向量 $t_i \in R^M$ 即所属类别一列为“1”,其他默认为“-1”或“0”,如表1所示(本文以5类为例);而本文提出的考虑顺序信息的基于有序编码的核极限学习顺序回归模型需要以下3个步骤:

(1)编码。首先将具有顺序信息标签的样本重新进行编码。

(2)核极限学习机的学习。将重新编码后的样本送入核极限学习机进行学习。

(3)解码预测。给定一个新样本,根据loss-based距离预测新样本,即选择与真实标签loss-based距离最小的作为预测标签。

有效的编码标签能有效表达标签中的顺序信息,是连接ELM和顺序回归问题的桥梁;合适的

表1 普通多分类编码,行表示其不同类别,列表示不同类别对应的编码位的编码

标签类别	编码位				
	1	2	3	4	5
1	1	-1	-1	-1	-1
2	-1	1	-1	-1	-1
3	-1	-1	1	-1	-1
4	-1	-1	-1	1	-1
5	-1	-1	-1	-1	1

学习算法能保证模型的有效运行；合理的解码方式保证了在解码产生错误码时提供纠错的能力。

3.1 编码和解码

在众多的编码方案中，纠错输出编码(error correct output code)可有效刻画标签间有序关系。纠错输出编码除了能将多分类问题转化为二分类问题，为解决多分类问题提供一种有效的思路，而且能将标签信息重新编码，并在解码产生错误码时提供纠错的能力，正因如此，纠错输出编码成为了多分类领域的研究热点^[21,22]。更重要的是，纠错输出编码策略可用于表示标签间的有序关系，成为连接普通多分类模型和顺序回归模型的桥梁。

在纠错输出编码策略中常常使用编码矩阵将类别信息表示成二元码(-1, +1)矩阵，使多类问题转化为多个二分类问题，其中“-1”，“+1”分别表示两种不同类别。但在顺序回归问题中使用纠错输出编码策略与普通的多分类编码策略往往不同。普通多分类编码策略并没有将标签的顺序信息表示出来，具体来说只是将所属类别一列标为“1”，其他默认为“-1”，如表1所示；而顺序回归中使用的编码策略则将所属类别以及相比该类别更小的类别编码位均标为“1”，其他标为“-1”，如表2所示。这样的编码方式能有效刻画标签间的有序关系。

关于解码策略的研究工作主要是对现有解码方法的改进。在解码策略中常常将输出编码和目标编码的距离如汉明距离、欧式距离作为解码策略。文献[21]提出一种基于目标损失函数(loss-based)的解码策略并证明了该解码策略优于基于汉明距离的解码策略，该解码方式在计算预测编码向量与实际编码向量差异时，采用指数函数计算两者的距离即 $D_{\text{LOSS}} = e^{-\alpha t_i}$, $i = 1, 2, \dots, N$ 。 α_i 与 t_i 分别表示第 i 个样本的预测编码向量及实际编码向量。所以本文在设计模型时拟采用基于目标损失函数的解码策略。

3.2 核极限学习机的学习

核函数的引入有利于提高特征映射质量，有利

表2 顺序回归编码，行表示其不同类别，列表示不同类别对应的编码位的编码

标签类别	编码位				
	1	2	3	4	5
1	1	-1	-1	-1	-1
2	1	1	-1	-1	-1
3	1	1	1	-1	-1
4	1	1	1	1	-1
5	1	1	1	1	1

于求解输出权重并提高 ELM 泛化能力^[23]，从而使 ELM 在顺序回归中得到更好的应用。文献[8]将核函数引入到极限学习机中，解决了 ELM 中隐层节点个数依赖于人为设置的问题，同时满足了特征映射的需求。核极限学习机(KELM)的发展同样具有相应的理论保证。文献[24-26]证明了 ELM 与 SVM 之间的关系，受支持向量机中核函数启发：核方法仅涉及到样本点之间的点积运算，与特征空间的维数无关，所以文献[8]提出了引入核函数的 ELM。即

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) \quad (4)$$

其中， $\mathbf{h}(\mathbf{x}_i) = [g(w_1 \mathbf{x}_i + b_1), \dots, g(w_L \mathbf{x}_i + b_L)]$ 这样有效地避免了“维数灾难”问题。待选的核函数包括高斯 RBF 核函数，多项式核函数，Sigmoid 核函数等。KELM 可表示为

$$f(\mathbf{x}) = \mathbf{h}(\mathbf{x}) \mathbf{H}^T \left(\frac{1}{C} + \mathbf{H} \mathbf{H}^T \right)^{-1} \mathbf{T} \\ = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left(\frac{1}{C} + \boldsymbol{\Omega}_{\text{ELM}} \right)^{-1} \mathbf{T} \quad (5)$$

其中， $\boldsymbol{\Omega}_{\text{ELM}} = \mathbf{H} \mathbf{H}^T$, $\boldsymbol{\Omega}_{\text{ELM}, i, j} = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j)$ 是核矩阵。

4 数值实验

为了验证本文提出的核极限学习顺序回归模型(KELMOR)模型的有效性，本节利用文献[27]提供的顺序回归数据，将本文提出的 KELMOR 模型与其他基于 ELM 的模型进行对比，包括：ELM (Extreme Learning Machine)模型^[5]、KELM (Kernel Extreme Learning Machine)模型^[8]及 ELMOR (Ordinal Extreme Learning Machine)模型^[12]，这里 ELM 和 KELM 模型采用表1的编码方式把顺序回归看作普通多分类问题。除此之外，我们还比较了其顺序回归领域中较经典的支持向量回归机(SVOR)^[7]模型在性能上的差异。本节将依次介绍使用的评价标准、数据集、实验参数设定和实验结果。

4.1 评价标准

在顺序回归领域中常使用的评价标准是平均0-1误差(Mean Zero-One Error, MZE)和平均绝对值误差(Mean Absolute Error, MAE)^[4,28]。MZE 与正确率相反，代表分错的样本数量占总样本数量的比例，定义为

$$\text{MZE} = \frac{1}{N} \sum_{i=1}^N [\mathbf{o}_i \neq \mathbf{t}_i] = 1 - \text{Acc} \quad (6)$$

其中， $\mathbf{o}_i = (o_{i1}, o_{i2}, \dots, o_{iM})$ 代表模型预测标签， $\mathbf{t}_i =$

$(t_{i1}, t_{i2}, \dots, t_{iM})$ 表示该样本真实标签, N 代表样本数量, M 代表类别数, Acc 表示模型的准确率。MZE 的值位于区间 $[0,1]$, 衡量模型整体的表现, 其值越低代表模型性能越好。与 MZE 相比, MAE 表示预测样本偏离真实样本的平均程度, MAE 的值位于区间 $[0, M - 1]$, MAE 定义为

$$MAE = \frac{1}{N} \sum_{i=1}^N |o_i - t_i| \quad (7)$$

此外, 本文中还将比较各个模型的训练时间。

4.2 实验数据集

文献[27]中提供的数据是顺序回归领域中常用的数据集¹⁾, 其中包括多个不同的基准数据集。表 3 详细展示了 5 个基准数据集的统计情况, 包括样本数量、特征数量、类别以及类别分布。第 1, 第 2, 第 3, 第 4, 第 5 列分别表示数据集名称, 该数据的总样本数, 特征维度, 类别数。为防止随机性带来的误差, 本文将直接采用文献[27]提供的已随机分好的 20 折数据。

4.3 实验参数

在该实验中, 模型 ELM 和 ELMOR 隐层节点选择范围为 $[100,1000]$, 模型 SVOR, KELM 和 KELMOR 采用 RBF 高斯核, 其中 C 参数选择范围为 $[10^{-3}, 10^3]$, K 参数选择范围为 $[10^{-3}, 10^3]$, 激活函

数均采用 Sigmoid 函数。所有模型在 4.1 节描述的数据集上运行, 以在训练集上 5 折交叉验证后获得最低 MAE 的参数为最优参数, 之后将在测试集上直接使用该参数测试。模型在一个数据上的最终性能表现将取在 20 折数据运行后的平均值。

4.4 实验结果及分析

表 4 展示了各个模型在上述数据集上的性能表现, 黑色加粗字体部分为最优性能。从表 4 中可以看出, KELMOR 模型在大多数数据集上关于 MZE, MAE 比其他所有基于 ELM 的传统模型表现优越。相比 ELM 模型, KELMOR 在准确率上最高提高了 23.8%(CC5 数据集), 平均提高了 10.8%。此外, KELM 和 ELMOR 模型相对比 ELM 几乎在所有数据集上都有相应提高, 对比 KELM 和 ELMOR 模型, 在所有数据集上 KELM 模型优于 ELMOR 模型, 这一现象强调了在这些数据集上加入核函数的必要性; 而对比 KELM 和 KELMOR 模型, KELMOR 几乎在所有的数据集上性能优于 KELM, 故考虑标签的顺序信息和加入核函数有助于提高模型的性能。此外, 对比 KELMOR 和 SVOR, 我们可以看出 KELMOR 在 MZE 及 MAE 上达到与 SVOR 可比较的性能, 在多个数据上甚至略优于 SVOR。

除此之外, 我们对比了各个模型在上述数据集上运行的时间如表 5 所示, 黑色加粗字体部分表示时间消耗最短。如上所述 KELMOR 与 SVOR 在 MZE 和 MAE 上性能较接近, 但从表 5 看来, 所有基于 ELM 的模型相比 SVOR 在时间性能上有着较大的优势, 尤其在数据量相对较大的数据集上如 Cal5 Census5, Census*5, 而基于 ELM 的模型在时间性能上表现比较接近。通过对比时间, 我们可以看出, KELMOR 模型在大多数数据集上的时间效

表 3 实验数据集相关描述

数据集	总样本数	特征维度	类别数	样本/类
Pyrim5	74	27	5	≈ 15
Computer*5(CC5)	8192	21	5	≈ 1639
Cal.housing5(Cal5)	20640	8	5	4128
Census5	22784	8	5	≈ 4557
Census*5	22784	16	5	≈ 4557

表 4 KELMOR 与模型 SVOR, ELM, KELM 和 ELMOR 在多个数据集上的 MZE, MAE

数据集	MZE					MAE				
	SVOR	ELM	KELM	ELMOR	KELMOR	SVOR	ELM	KELM	ELMOR	KELMOR
Pyrim5	0.498	0.634	0.567	0.576	0.495	0.630	1.026	0.812	0.926	0.602
	±0.102	±0.085	±0.061	±0.094	±0.068	±0.125	±0.233	±0.166	±0.277	±0.101
CC5	0.330	0.623	0.442	0.535	0.385	0.359	1.095	0.527	0.858	0.435
	±0.014	±0.050	±0.019	±0.051	±0.016	±0.017	±0.155	±0.035	±0.159	±0.021
Cal5	0.519	0.668	0.538	0.634	0.515	0.628	1.052	0.739	1.024	0.616
	±0.013	±0.030	±0.019	±0.033	±0.021	±0.021	±0.078	±0.046	±0.122	±0.018
Census5	0.566	0.683	0.589	0.640	0.563	0.773	1.130	0.875	1.037	0.740
	±0.012	±0.036	±0.021	±0.025	±0.014	±0.020	±0.102	±0.045	±0.091	±0.020
Census*5	0.562	0.666	0.576	0.631	0.556	0.723	1.062	0.845	0.998	0.741
	±0.022	±0.011	±0.015	±0.016	±0.013	±0.029	±0.037	±0.033	±0.056	±0.017

¹⁾<http://www.gatsby.ucl.ac.uk/~chuwei/ordinalregression>

表 5 KELMOR 与模型 SVOR, ELM, KELM 和 ELMOR
在多个数据集上的训练时间(s)

数据集	SVOR	ELM	KELM	ELMOR	KELMOR
Pyrim5	0.0019	0.0064	0.0002	0.0061	0.0004
CC5	0.1207	0.0264	0.0016	0.0303	0.0011
Cal5	0.2502	0.0269	0.0016	0.0240	0.0011
Census5	0.3206	0.0086	0.0024	0.0066	0.0018
Census*5	0.4532	0.0450	0.0028	0.0377	0.0020

率表现最优。从而从实验中更加验证了加入核函数和表顺序信息的有效性。我们分析, 模型 KELMOR 比 KELM 多考虑了标签顺序, 比 ELMOR 多考虑了核, 之所以能取得较好的时间表现, 得益于 KELMOR 中的核函数和有序编码对顺序信息的特殊处理有助于输入数据和标签的有效表示, 使得模型更好更快地学习输入特征及映射函数。

5 结论

本文提出了一种基于有序编码的核极限学习顺序回归模型。该模型将核函数引入到顺序回归中, 使原始数据映射到更易可分的特征空间, 提高了模型的准确率; 将输出纠错编码引入表示标签的顺序信息, 使核极限学习机有效应用在顺序回归上。该文提出的新模型对比多个基于 ELM 的传统模型, 在顺序回归领域众多数据集上取得了优越的实验性能, 突出了引入核函数和标签顺序信息的重要性。另一方面, 基于 ELM 的算法本身由于避免了迭代过程, 训练时间已较快, 而 KELMOR 模型在训练时间上更加优于其他基于 ELM 的算法。因此, 该模型相比其他基于 ELM 的模型和神经网络等算法更加适用于对训练时间要求较高的实际顺序回归应用场景如: 在线学习, 情感分析、用户推荐等。

参 考 文 献

- [1] NAKOV P, RITTER A, ROSENTHAL S, *et al.* SemEval-2016 task 4: Sentiment analysis in Twitter[C]. International Workshop on Semantic Evaluation, San Diego, USA, 2016: 1-18. doi: 10.18653/v1/S16-1028.
- [2] TIAN Q, CHEN S, and TAN X. Comparative study among three strategies of incorporating spatial structures to ordinal image regression[J]. *Neurocomputing*, 2014, 136: 152-161. doi: 10.1016/j.neucom.2014.01.017.
- [3] CORRENTE S, DOUMPOS M, GRECO S, *et al.* Multiple criteria hierarchy process for sorting problems based on ordinal regression with additive value functions[J]. *Annals of Operations Research*, 2017, 251(1/2): 117-139. doi: 10.1007/s10479-015-1898-1.
- [4] GUTIÉRREZ P A, PÉREZ-ORTIZ M, SANCHEZ-MONEDERO J, *et al.* Ordinal regression methods: Survey and experimental study[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2016, 28(1): 127-146. doi: 10.1109/TKDE.2015.2457911.
- [5] HUANG G B, ZHU Q Y, and SIEW C K. Extreme learning machine: Theory and applications[J]. *Neurocomputing*, 2006, 70(1): 489-501. doi: 10.1016/j.neucom.2005.12.126.
- [6] RAJASEKARAN S and PAI G A V. Neural Networks, Fuzzy Systems and Evolutionary Algorithms: Synthesis and Applications[M]. Haryana, India: Rajkamal Electric Press, 2017: 151-168.
- [7] CHU W and KEERTHI S S. Support vector ordinal regression[J]. *Neural Computation*, 2007, 19(3): 792-815. doi: 10.1162/neco.2007.19.3.792.
- [8] HUANG G B, ZHOU H, DING X, *et al.* Extreme learning machine for regression and multiclass classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 513-529. doi: 10.1109/TSMCB.2011.2168604.
- [9] UCAR A, DEMIR Y, and GÜZELI C. A new facial expression recognition based on curvelet transform and online sequential extreme learning machine initialized with spherical clustering[J]. *Neural Computing and Applications*, 2016, 27(1): 131-142. doi: 10.1007/s00521-014-1569-1.
- [10] 徐涛, 郭威, 吕宗磊. 基于快速极限学习机和差分进化的机场噪声预测模型[J]. 电子与信息学报, 2016, 38(6): 1512-1518. doi: 10.11999/JEIT150986.
XU Tao, GUO Wei, and LÜ Zonglei. Prediction model of airport noise based on fast extreme learning machine and differential evolution[J]. *Journal of Electronics & Information Technology*, 2016, 38(6): 1512-1518. doi: 10.11999/JEIT150986.
- [11] GOODFELLOW I, BENGIO Y, and COURVILLE A. Deep Learning[M]. Massachusetts, USA, MIT Press, 2016: 165-480. doi: 10.1038/nature14539.
- [12] DENG W Y, ZHENG Q H, LIAN S, *et al.* Ordinal extreme learning machine[J]. *Neurocomputing*, 2010, 74(1): 447-456. doi: 10.1016/j.neucom.2010.08.022.
- [13] RICCARDI A, FERNÁNDEZ-NAVARRO F, and CARLONI S. Cost-sensitive AdaBoost algorithm for ordinal regression based on extreme learning machine[J]. *IEEE Transactions on Cybernetics*, 2014, 44(10): 1898-1909. doi: 10.1109/TCYB.2014.2299291.
- [14] HORNIK K, STINCHCOMBE M, and WHITE H. Multilayer feedforward networks are universal approximators[J]. *Neural Networks*, 1989, 2(5): 359-366. doi: 10.1016/0893-6080(89)90020-8.

- [15] HUANG G B and BABRI H A. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions[J]. *IEEE Transactions on Neural Networks*, 1998, 9(1): 224–229. doi: 10.1109/72.655045.
- [16] HUANG G B, CHEN L, and SIEW C K. Universal approximation using incremental constructive feedforward networks with random hidden nodes[J]. *IEEE Transactions on Neural Networks*, 2006, 17(4): 879–892. doi: 10.1109/TNN.2006.875977.
- [17] HUANG G B. Learning capability and storage capacity of two-hidden-layer feedforward networks[J]. *IEEE Transactions on Neural Networks*, 2003, 14(2): 274–281. doi: 10.1109/TNN.2003.809401.
- [18] BARTLETT P L. The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network[J]. *IEEE Transactions on Information Theory*, 1998, 44(2): 525–536. doi: 10.1109/18.661502.
- [19] TANG J, DENG C, and HUANG G B. Extreme learning machine for multilayer perceptron[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2016, 27(4): 809–821. doi: 10.1109/TNNLS.2015.2424995.
- [20] HOERL A E and KENNARD R W. Ridge regression: Biased estimation for nonorthogonal problems[J]. *Technometrics*, 1970, 12(1): 55–67. doi: 10.1080/00401706.1970.10488634.
- [21] ALLWEIN E L, SCHAPIRE R E, and SINGER Y. Reducing multiclass to binary: A unifying approach for margin classifiers[J]. *Journal of Machine Learning Research*, 2000, 1(12): 113–141. doi: 10.1162/15324430152733133.
- [22] 雷蕾, 王晓丹, 罗玺, 等. ECOC 多类分类研究综述[J]. *电子学报*, 2014, 42(9): 1794–1800. doi: 10.3969/j.issn.0372-2112.2014.09.020.
- LEI Lei, WANG Xiaodan, LUO Xi, *et al.* An overview of multi-classification based on error-correcting output codes[J]. *Acta Electronica Sinica*, 2014, 42(9): 1794–1800. doi: 10.3969/j.issn.0372-2112.2014.09.020.
- [23] HUANG G, HUANG G B, SONG S, *et al.* Trends in extreme learning machines: A review[J]. *Neural Networks*, 2015, 61: 32–48. doi: 10.1016/j.neunet.2014.10.001.
- [24] LIU Q, HE Q, and SHI Z. Extreme support vector machine classifier[C]. 12th Pacific-Asia Conference on Knowledge Discovery and Data Mining, Osaka, Japan, 2008: 222–233. doi: 10.1007/978-3-540-68125-0_21.
- [25] FRÉNAVY B and VERLEYSEN M. Using SVMs with randomised feature spaces: an extreme learning approach[C]. European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 2010: 315–320.
- [26] HUANG G B, DING X, and ZHOU H. Optimization method based extreme learning machine for classification[J]. *Neurocomputing*, 2010, 74(1): 155–163. doi: 10.1016/j.neucom.2010.02.019.
- [27] CHU W and GHARAMANI Z. Gaussian processes for ordinal regression[J]. *Journal of Machine Learning Research*, 2005, 6(7): 1019–1041.
- [28] BACCIANELLA S, ESULI A, and SEBASTIANI F. Evaluation measures for ordinal regression[C]. The Ninth International Conference on Intelligent Systems Design and Applications, Pisa, Italy, 2009: 283–287. doi: 10.1109/ISDA.2009.230.
- 李佩佳: 女, 1990 年生, 博士生, 研究方向为机器学习、数据挖掘.
- 石 勇: 男, 1956 年生, 教授, 研究方向为数据挖掘、知识管理.
- 汪华东: 男, 1988 年生, 博士, 研究方向为机器学习、最优化理论.
- 牛凌峰: 女, 1981 年生, 副教授, 研究方向为最优化理论、数据挖掘.