基于神经切线核的学件 RKME 规约

谭志豪 史浩宇 陈梓轩 姜 远
(南京大学计算机软件新技术国家重点实验室 南京 210023)

摘 要 当前机器学习技术已经在大量领域得到广泛应用,然而仍面临许多亟待解决的问题:依赖大量的训练数据和训练技巧、难以适应环境变化、数据隐私/所有权的保护、灾难性遗忘等等.最近,学件范式使得上述问题同时得到系统性地解决成为可能.在该范式下,用户面临新的机器学习任务时可以通过学件基座系统方便地复用他人的结果,而不必从头开始.学件范式的核心在于规约,规约使得学件基座系统在不接触原始数据的情况下,可以根据用户的需求快速识别出对用户任务有帮助的学件.近期研究均通过缩略核均值嵌入(Reduced Kernel Mean Embedding,RKME)为模型构造规约,并通过构建学件原型系统验证了范式的有效性.在实际中,学件基座系统中往往包含在各种领域任务、数据类型上构建的机器学习模型,而传统的 RKME 规约面临维度灾难的问题,难以适用于高维数据,例如图像场景.为了拓展 RKME 规约的适用范围,本文引入神经切线核进行 RKME 规约构造.为提升方法的高效性,本文进一步通过神经网络高斯过程与随机特征近似,快速为各种模型生成 RKME 规约.最后,本文在真实数据构建的销量预测、图像分类场景的学件基座系统中进行大量实验验证了所提出方法的有效性和高效性,所提出方法相比于传统 RKME 规约查搜准确率显著提升近 9%,且实验结果表明改进后的规约在图像任务上具有良好的隐私保护性质.代码见:https://github.com/tanzh-lamda/RKME-NTK.

关键词 学件;学件基座系统;规约;神经切线核;缩略核均值嵌入
中图法分类号 TP181 **DOI 号** 10.11897/SP.J.1016.2024.01232

Learnware Reduced Kernel Mean Embedding Specification Based on Neural Tangent Kernel

TAN Zhi-Hao SHI Hao-Yu CHEN Zi-Xuan JIANG Yuan
(*State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing* 210023)

Abstract Machine learning technology has been successfully applied in various fields. Nevertheless, several challenges still need to be addressed. In classic machine learning paradigm, developing a high-quality model for a new task from scratch requires a substantial amount of labeled data, expertise, and computational resources, making the process difficult and costly. Moreover, although source data is crucial for transferring and reusing existing efforts, concerns over data privacy and proprietary generally hinder the sharing of experience among developers. Recently, the learnware paradigm has been developed to systematically tackle these challenges. This paradigm enables users to utilize the learnware dock system and leverage numerous existing high-performing models when faced with new machine learning tasks, instead of building machine learning models from scratch. For high-performing models of any structure from various tasks, a learnware consists of the model itself and a specification which captures the model's specialty, like its statistical properties. In this paradigm, developers worldwide can submit their well-trained models spontaneously into a learnware dock system (formerly known as learnware market). The system uniformly generates a

收稿日期:2023-06-05;在线发布日期:2024-04-10.本课题得到国家自然科学基金(62250069,62176117)和江苏省研究生科研与实践创新计划项目(KYCX23_0159)资助.**谭志豪**,博士研究生,主要研究方向为机器学习与数据挖掘.E-mail:tanzh@lamda.nju.edu.cn.**史浩宇**,本科生,主要研究方向为机器学习与数据挖掘.**陈梓轩**,硕士,主要研究方向为机器学习与数据挖掘.**姜 远**(通信作者),博士,教授,博士生导师,主要研究领域为机器学习与数据挖掘.E-mail:jiangy@lamda.nju.edu.cn.

specification for each model to form a learnware and accommodates it. The core of this paradigm lies in the specification, which enables the learnware dock system to identify and assemble existing helpful learnwares to solve new machine learning tasks according to the user requirement. Note that the learnware dock system should be able to preserve the raw data of model developers and users. Recent studies have demonstrated the efficacy of the learnware paradigm with reduced kernel mean embedding (RKME) specification, which makes a good approximation for the distribution of training data used by the model without revealing the raw data. In practice, the learnware dock system comprises machine learning models from different domains and various data types, whereas the traditional RKME specification faces curse of dimensionality, which makes it hard to be applied in high dimensional scenarios like images. In order to broaden the scope of applicability of RKME specification for all these scenarios, this paper explores the construction of RKME specification based on neural tangent kernels (NTK), and improves its efficiency through neural network Gaussian process (NNGP) and random feature approximation. This approach enables the efficient and accurate generation of the RKME specification for various models in practice. Finally, the experiments on real-world data in sales forecasting and image classification scenarios validate the effectiveness of the proposed algorithm. Compared to traditional RKME specification, our method has improved the identification accuracy by nearly 9% in image classification scenarios. Benefiting from the random feature approximation, our method is even more efficient while maintaining the accuracy improvement. And the experimental results show that the improved specification has good privacy-preserving properties on image data, making it even impossible for the human eye to discern any information about the original real data. Our proposed neural tangent kernel-based specification is well suited as a specification implementation for image data, which serves as a foundation for subsequent research on the learnware paradigm. Our method implementation and experimental code have been open sourced and can be found here: https://github.com/tanzh-lamda/RKME-NTK.

Keywords　learnware; learnware dock system; specification; neural tangent kernel; reduced kernel mean embedding

1　引　言

机器学习技术在大量领域得到了成功应用，包括自然语言处理、图像、视频、语音等[1-3]. 然而，当下的机器学习仍面临许多亟待解决的问题：想要从头开始构建一个优秀的机器学习模型仍然非常困难，往往需要大量的标记数据、训练资源以及专业知识与技巧；已有的优质模型往往难以应对环境的变化，在旧环境中表现优异的模型面对新的环境可能表现很差甚至失效；模型在适应新环境的过程中容易产生灾难性遗忘[4]；此外由于数据隐私与所有权的问题，阻碍了数据的分享，导致难以复用他人的结果. 虽然已经有大量研究分别针对以上问题开展，然而在实际中，这些问题往往同时出现并相互影响.

针对上述机器学习面临的问题，近期提出的学件(learnware)范式[5-6]进行了统一的考虑，使得同时解决这些问题成为可能. 学件由模型和规约(specification)构成，其中规约描述了模型的效用，使其可以被高效准确地识别和查搜. 在学件范式中，任意模型的开发者都可以自发地将模型提交给学件基座系统(之前称为学件市场)进行售卖或分享，如果学件基座系统接受了该模型，会为其分配合适的规约组成学件，并将其安放在规约空间的合适位置. 当新用户需要解决机器学习任务时，它可以将需求提交给学件基座系统，然后学件基座系统会基于规约识别出对用户任务有帮助的学件. 用户可以直接使用学件或基于其数据进一步“打磨”以获得更好的表现，而不必从头开始构建模型. 值得注意的是，学件基座系统不会接触到开发者和用户的原始数据，从而保护各方的数据隐私.

基于上述学件范式，核心问题是：在开发者与用

户数据隐私不泄露给学件基座系统的前提下,市场如何能够高效地从大量模型中为用户任务识别出有帮助的模型?解决这一问题的关键在于规约的设计.最新的研究基于缩略核均值嵌入(Reduced Kernel Mean Embedding,RKME)为模型构造规约[6],实现了在不泄露原始数据的同时,对模型的能力进行刻画.具体而言,RKME 规约通过优化数据分布映射到再生核希尔伯特空间后的距离,为原始任务生成缩略集,实现对分布信息的近似表示,从而保证原始数据不被泄露.RKME 规约使得在保护隐私的前提下为用户任务识别有帮助的学件成为可能.基于 RKME 规约,已有工作[6-8]设计了相应的学件查搜与复用算法,并在一些场景中验证了算法的有效性.

虽然基于高斯核的 RKME 规约具有良好的理论性质[6],然而在许多高维数据的应用场景中,例如图像数据等,高斯核会面临维度灾难的问题[9],样本复杂度高[10],导致实际表现不佳.然而,在学件范式中,由于学件基座系统的开放性,市场中往往包含开发者提交的来自各种领域的不同任务上的机器学习模型,如何使 RKME 规约适应各种类型的数据是当下亟待解决的问题.此外,虽然各种特定结构的神经网络往往具备处理不同数据类型的能力,但由于神经网络训练过程的黑盒特性,使用神经网络替代基于再生核希尔伯特空间中分布距离的计算会丧失规约的理论保障[11].

为了拓展 RKME 规约的适用范围,本文提出引入神经切线核(Neural Tangent Kernel,NTK)[12]实现为各种任务的模型构造准确的 RKME 规约.神经切线核刻画了各种神经网络结构在无限宽度下准确的学习收敛过程,代表了各种神经网络结构在梯度下降过程中对应的隐式先验.本文通过使用神经切线核进行核均值嵌入,并通过多种近似技术进一步解决其效率问题,实现为各类机器学习任务上的模型构造准确的 RKME 规约,从而更准确地识别对用户任务有帮助的模型.

本文的主要贡献包括 4 个方面:

(1)本文通过引入神经切线核进行 RKME 规约构造,在延续 RKME 规约良好理论性质的基础上,充分利用了神经网络结构带来的隐式先验,从而能够为各种任务上的模型构造准确的 RKME 规约;

(2)为解决神经切线核 RKME 规约构造的效率问题,以及 NTK 对缩略集难以求导的问题,我们提出了RKME-NTK-RF算法:基于神经网络高斯过程(Neural Network Gaussian Process,NNGP)[13]与随机特征(Random Feature,RF)近似技术[14],实现对神经切线核矩阵、缩略集梯度优化的近似计算,从而高效生成神经切线核 RKME 规约;

(3)基于神经切线核 RKME 规约,我们设计了相应的学件上传与部署流程;

(4)在基于真实数据集构建的销量预测、图像分类任务的学件基座系统中,验证了提出的算法可以高效准确地查搜到对用户任务有帮助的模型,相比高斯核基线取得显著提升.此外,本文算法在图像任务上具备良好的隐私保护性质.

2 相关工作

本节分别介绍学件范式和神经切线核两个方面的相关工作.

2.1 学件范式

在学件范式中,来自各种领域的大量模型可以被开发者提交到学件基座系统,学件基座系统基于规约对大量学件进行组织管理,从而当用户面临新的机器学习任务时,可以通过学件基座系统部分复用他人结果,而不必从头开始.这与传统的迁移学习[15-16]、领域自适应[17-18]等领域的研究有着本质不同.首先,它们往往假定用来迁移、适应的源域任务数量较少且均对目标域任务潜在有用,直接基于少量源域任务来帮助目标域任务的学习.而学件范式下,由于学件基座系统的开放性,市场内的大量模型中可能仅有极少量对用户任务有帮助,所以如何根据用户需求准确高效地识别出有帮助的一个或多个学件是学件范式中的一个重要问题.其次,它们往往假定可以使用源域任务的原始数据.而在现实中,由于数据隐私、所有权等问题,学件基座系统通常无法接触到开发者或者用户的原始数据.此外,模型复用[19-20]、假设迁移学习[21-22]、无源域自适应[23]等研究仅需接触到源域任务上训练的模型,但依然假定源域任务上的所有模型都对目标任务有帮助.

学件范式由周志华教授提出并探索发展[5-6].文献[6]提出通过语义规约与统计规约的方式构建学件规约,最新的研究主要建立在缩略核均值嵌入(RKME)作为统计规约的基础上.例如,针对用户任务中包含学件基座系统未见过的任务的情形,Zhang 等人[8]基于 RKME 规约进行混合比例估计,识别出用户任务中无法被学件基座系统解决的部分,将剩

余部分分配给合适的学件进行预测，且提供了用户任务上的泛化错误率保障；针对现实中同类任务的模型训练自各种不同特征空间的问题，Tan 等人[24-25]通过子空间学习，在不使用额外辅助数据的情况下，为异构特征空间的学件构建统一的规约空间，从而可以识别并复用不同特征空间上的学件来帮助用户任务；此外，Xie 等人[26]通过在市场中构建锚位学件，实现了仅需评价学件基座系统中少量学件即可为用户匹配到最有帮助的多个学件的任务，极大提升了学件查搜的效率；Liu 等人[27]通过利用学件坞中不断增长的学件规约包含的信息构建学件规约索引，使得学件查搜更精准高效. 近期，基于 RKME 规约，首个开源学件基座系统北冥坞[28]初步实现了对学件范式全流程的支持，为学件相关研究提供了一个初步的科研平台. 已有的研究均基于高斯核函数计算 RKME 规约，本文提出基于神经切线核的 RKME 规约可以适用于更多任务类型的模型，且维持其理论上可解释的优势. 更多关于学件范式的介绍，我们推荐读者参考文献[6].

2.2　神经切线核

在深度神经网络的理论研究中，神经切线核最早被发现可以刻画无限宽的全连接神经网络的训练过程[12]. 具体来说，无限宽的全连接神经网络的梯度下降学习过程等价于基于神经切线核进行泛函空间中的核梯度下降. 后来，该结论逐渐被拓展到卷积神经网络(CNTK)[29]、图神经网络(GNTK)[30]、循环神经网络(RNTK)[31]，直至推广到任意结构的神经网络[32-34]. 理论研究也给出了对网络宽度非渐进的要求[29]. 于是，神经切线核成为核方法与各种神经网络之间的理论桥梁，通过使用神经切线核搭配基础的核方法学习器，便可能实现一个复杂网络的学习过程.

除了理论上的结果，诸多实验中神经切线核也被广泛验证取得优异表现，例如在 UCI 数据集上 NTK SVM 打败包含随机森林在内的 SOTA 及经典方法[35]；在图像数据上，通过利用无限宽卷积神经网络的先验，在所有核方法算法中取得最优，并具有良好的稳定性[36]；此外，NTK 被证明同样具备表示学习的能力[37]. 值得注意的是，在数据集蒸馏[38]任务上，基于核方法的算法通过使用神经切线核取得了在各种数据集上的最优表现，且相比于之前的最优算法取得了大幅提升[39].

由于核方法固有的核矩阵的计算与存储开销，以及神经切线核的计算开销，在实际任务中通常需要依赖一些加速方法. 除了对核方法通用的近似方法之外(例如 Nyström 近似[40]、随机特征近似[14]等)，近期也有许多研究针对神经切线核的计算设计相应的加速方法[41-42]. 此外，Loo 等人[43]使用神经网络高斯过程的核函数近似神经切线核，进而通过有限宽随机网络进行随机特征近似，实现低成本地快速生成核矩阵，并在数据集蒸馏任务上取得了优异效果. 本文在规约构造算法部分沿用了该近似方法，然而缩略集生成方式、目的等完全不同.

3　预备知识

3.1　核均值嵌入

核均值嵌入(Kernel Mean Embedding，KME)[44]技术的核心思想是将定义在$\mathcal{X}$上的概率分布 P 映射到一个核函数对应的再生核希尔伯特空间，表示为该空间中均值函数的形式：

$$\mu_P := \int_{\mathcal{X}} k(\boldsymbol{x}, \cdot)\mathrm{d}P(\boldsymbol{x}),$$

其中$k:\mathcal{X}\times\mathcal{X}\to\mathbb{R}$是正定的核函数，其对应的再生核希尔伯特空间记为$\mathcal{H}_k$. 当核函数$k$符合$\mathbb{E}_{x\sim P}[\sqrt{k(\boldsymbol{x},\boldsymbol{x})}]<\infty$，KME 存在且满足 $\mu_P\in\mathcal{H}_k$.

核均值嵌入作为概率分布的表示，具有优良的理论性质. 当核函数为特征核(Characteristic Kernel)[45]时，KME 刻画了分布的所有信息，即原始分布与其 KME 表示互为一一映射.

在学习问题中，我们往往只能获得从分布 P 中独立采样得到的数据集 $D=\{(\boldsymbol{x}_i)\}_{i=1}^n$. 此时，可以通过计算如下的经验 KME $\hat{\mu}_p$作为原始分布 P 所对应 KME 的近似表示：

$$\hat{\mu}_p = \frac{1}{n}\sum_{i=1}^{n} k(\boldsymbol{x}_i, \cdot).$$

Smola 等人[44]证明，使用再生核希尔伯特空间$\mathcal{H}_k$中范数作为距离度量，经验 KME 以 $O(1/\sqrt{n})$的速率收敛至原始分布 KME.

3.2　缩略核均值嵌入规约

学件范式的核心在于规约的设计. 规约的基本要求为能够刻画模型的能力，同时不泄露模型的原始训练数据. 虽然上述经验 KME 能够刻画模型原始任务分布并具有良好的理论性质，然而其包含全部原始训练数据，不满足规约所需的隐私性质.

近期，研究提出的缩略核均值嵌入(Reduced

Kernel Mean Embedding,RKME)规约[6]通过优化生成缩略核均值嵌入近似原始任务数据的经验KME的思想,实现了对原始数据分布近似表示的同时,不暴露模型的原始训练数据.具体而言,RKME旨在求解带权重的缩略集$\{(\beta_i,\boldsymbol{z}_i)\}_{i=1}^m$来最小化与原始数据$\{(\boldsymbol{x}_i)\}_{i=1}^n$的最大均值差异,从而得到在再生核希尔伯特空间中与原始数据经验KME的距离最小的RKME $\tilde{\mu}_P=\sum_{i=1}^m\beta_i k(\boldsymbol{z}_i,\cdot)$:

$$\min_{\boldsymbol{\beta},\boldsymbol{Z}}\left\|\frac{1}{n}\sum_{i=1}^n k(\boldsymbol{x}_i,\cdot)-\sum_{j=1}^m\beta_j k(\boldsymbol{z}_j,\cdot)\right\|_{\mathcal{H}_k}^2 \tag{1}$$

其中k为使用的核函数,$\mathcal{H}_k$为核函数对应的再生核希尔伯特空间.展开上式可得:

$$F(\boldsymbol{\beta},\boldsymbol{Z})=\sum_{i,j=1}^n\frac{1}{n^2}k(\boldsymbol{x}_i,\boldsymbol{x}_j)+\sum_{i,j=1}^m\beta_i\beta_j k(\boldsymbol{z}_i,\boldsymbol{z}_j)-2\sum_{i=1}^n\sum_{j=1}^m\frac{\beta_j}{n}k(\boldsymbol{x}_i,\boldsymbol{z}_j) \tag{2}$$

由于其中第一项仅与模型开发者原始数据集有关,于是在进行缩略集迭代更新时可以不考虑.缩略集的优化过程可以采用交替优化的方式.当固定集合$\boldsymbol{Z}$更新权重向量$\boldsymbol{\beta}$时,可以直接求解得到$\boldsymbol{\beta}$的闭式解.由于$\boldsymbol{\beta}$向量的维度为缩略集的大小(远小于原始样本数n),所以计算非常高效.求解$\boldsymbol{\beta}$向量后,对于缩略集$\boldsymbol{Z}$中的每个元素$\{(\boldsymbol{z}_i)\}_{i=1}^m$,均可以通过梯度下降来迭代优化如下:

$$\boldsymbol{z}_i^t=\boldsymbol{z}_i^{t-1}-\eta\frac{\partial F(\boldsymbol{\beta}^t,\boldsymbol{Z})}{\partial\boldsymbol{z}_i} \tag{3}$$

对于高斯核函数,我们可以显式地写出其梯度.

基于RKME规约,已有工作[6-8]设计了相应的学件查搜与复用算法,分析了利用RKME规约在学件基座系统查搜到的模型复用于用户任务上的泛化风险上界,并在大量实验中验证了学件范式下RKME规约的有效性.

3.3 神经切线核

记f_θ为参数为θ的任意深度神经网络,对于任意的两个输入$\boldsymbol{x},\boldsymbol{x}'\in\mathcal{X}$,其神经切线核(Neural Tangent Kernel,NTK)定义为

$$\hat{\Theta}(\boldsymbol{x},\boldsymbol{x}'):=\langle\nabla_\theta f_\theta(\boldsymbol{x}),\nabla_\theta f_\theta(\boldsymbol{x}')\rangle.$$

直觉上,该核函数定量刻画了网络在对一个新的观测$\boldsymbol{x}'$处做梯度下降时,其函数f_θ在泛函空间的变化.基于对神经网络所对应的NTK核函数的分析,Jacot等人[12]在神经网络理论方面取得重要突破.具体而言,若f_θ为一个随机初始化的多层感知机网络,当其网络宽度趋于无穷时,其NTK依概率收敛至一个固定的核函数,且该核函数在网络训练过程中保持不变.记$\hat{\Theta}_t(\boldsymbol{x},\boldsymbol{x}')$为网络$f_\theta$在训练时间$t$时的NTK,则对于任意训练时间$t\geq0$,可得

$$\lim_{width\to\infty}\hat{\Theta}_t(\boldsymbol{x},\boldsymbol{x}')=\Theta(\boldsymbol{x},\boldsymbol{x}').$$

基于该极限情况下的NTK,可以完美地刻画无限宽网络的训练过程,即当f为一个无限宽神经网络时,其泛函空间中的梯度下降可以被基于NTK的核梯度下降刻画:

$$\partial_t f_t=-\eta\Theta\cdot\nabla_f\mathcal{L}(f_t),$$

其中$\mathcal{L}$为损失函数.以上结论已被证明适用于任意结构的深度神经网络[32-33].

4 本文方法

本节介绍提出的RKME-NTK-RF算法.首先,我们介绍了在RKME规约中引入NTK的根本意义与困难;为解决计算开销方面带来的挑战,我们进而介绍了本文提出的加速后的RKME-NTK规约构造算法;最后,基于该规约,我们给出了部署阶段的学件查搜算法.

4.1 基于神经切线核的RKME规约

神经切线核通过刻画无限宽深度神经网络的梯度下降优化过程,以核函数的形式代表了无限宽的神经网络结构与梯度下降优化方法所带来的隐式先验.相比于直接使用深度神经网络面临的不可解释、黑盒调参等问题,NTK在利用神经网络表达能力的同时,取得更稳定的效果和严谨的理论保障[46].在监督学习任务中,已有研究表明[29]直接使用NTK作为核函数,结合核回归算法,即可等价于使用无限宽网络进行梯度下降得到的解.然而在RKME规约构建的过程中,该如何合适地纳入NTK的能力?是否可以直接以NTK作为核均值嵌入与再生核希尔伯特空间中距离计算的核函数?

记基于NTK核函数Θ的RKME规约为$\tilde{\mu}_P=\sum_{i=1}^m\beta_i\Theta(\boldsymbol{z}_i,\cdot)$,缩略集对应的经验分布为$\hat{q}:=\sum_{i=1}^m\beta_i\delta_{\boldsymbol{z}_i}$,其中$\delta$为狄拉克$\delta$函数,且$\sum_{i=1}^m\beta_i=1$;对于原始训练数据$\{(\boldsymbol{x}_i)\}_{i=1}^n$,其经验分布记为$\hat{p}:=\frac{1}{n}\sum_{i=1}^n\delta_{\boldsymbol{x}_i}$,于是NTK的再生核希尔伯特空间中二者的距离为

$$Dist_\Theta^2=\int_{\mathcal{X}}\int_{\mathcal{X}}\Theta(\boldsymbol{x},\boldsymbol{y})(\hat{p}-\hat{q})(\boldsymbol{x})(\hat{p}-\hat{q})(\boldsymbol{y})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{y}.$$

幸运的是,对于该距离,近期有关最大均值差异的研

究[47]揭示了其与原始神经网络之间的关系. 具体而言，该研究证明了上述 RKME 规约计算中我们关心的经验分布距离，可以通过以下基于原始神经网络的表示方式近似表示：

$$Dist_{\Theta}^{2} \approx \frac{1}{t}(L(\theta_t)-L(\theta_0)),$$

此处，$L(\theta)$是精心构造的网络优化目标：

$$L(\theta)=-\int_{\mathcal{X}} f_\theta(\boldsymbol{x})(\hat{p}-\hat{q})(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$$
$$=-\frac{1}{n}\sum_{i=1}^{n} f_\theta(\boldsymbol{x}_i)+\sum_{i=1}^{m}\beta_i f_\theta(\boldsymbol{z}_i),$$

其中，t 是对优化目标 $L(\theta)$进行梯度下降的训练时长，t 时刻的网络参数为 θ_t；上式在训练时长较小时成立，近似误差有关训练时长的上界为$\mathcal{O}(t)$.

由上述理论结果可知，作为 RKME 规约生成过程的核心，分布距离的计算如果使用神经切线核 NTK，可以充分利用各种网络结构的先验. 具体而言，基于上述结果推导，可知使用NTK作为最大均值差异距离的核函数，其证据函数(Witness Function)等价于使用 NTK 相应的网络结构进行短时间梯度下降的一阶变化率. 在网络结构的特征表示先验下，样例带来的一阶变化率越小，则证据函数的值越小，从而此处的样本密度差距的权重越小.

虽然使用 NTK 可以为分布距离计算纳入神经网络结构的先验，但已有研究表明，基于 NTK 的准确核矩阵计算[43]以及 NTK 核函数值对于样本的准确梯度计算[39]都极其昂贵，效率难以接受. 而在学件范式下，为了保护数据隐私，RKME 规约生成需要在开发者本地完成，需要满足高效性.

4.2　神经切线核 RKME 规约高效计算

为了提高神经切线核 RKME 规约的可扩展性，受神经切线核高效计算方面研究[43]的启发，本文采用神经网络高斯过程核替代神经切线核的准确计算，并进一步使用有限宽网络随机特征进行核矩阵元素的近似计算.

首先，使用神经网络高斯过程(NNGP)替代 NTK 的计算. 相比于 NTK 刻画了无限宽网络的训练过程，NNGP 则仅刻画了无限宽网络对应的前向计算过程，于是可以被更高效地计算. 具体而言，在一定假设下，一个随机初始化的无限宽任意结构神经网络，得到的网络函数 f_θ 服从相应的由 NNGP 核 K 刻画的高斯过程：

$$f_\theta \sim \mathcal{GP}(0,K).$$

事实上，NNGP 可以作为 NTK 的粗略近似. 例如，对于 l 层深度的 ReLU 激活函数的无限宽全连接网络，其 NTK Θ^l 与 NNGP 核函数的关系为

$$\Theta^l = K^l + \dot{K}^l \odot \Theta^{l-1}$$
$$= K^l + \dot{K}^l \odot (K^{l-1} + \dot{K}^{l-1} \odot (K^{l-2}\cdots))$$
$$= K^l + \dot{K}^l \odot K^{l-1} + \dot{K}^l \odot \dot{K}^{l-1} \odot K^{l-2} + \cdots,$$

其中 $\dot{K}$ 即是反向传播过程中对激活函数求导得到的矩阵，元素值均小于 1，于是 NNGP 核 $\dot{K}^l$ 可以作为 NTK 的粗略近似.

其次，由于 NNGP 核 K 的定义来源于无限宽网络最后一层输出 $\boldsymbol{u}^l$的期望内积，即：

$$K(\boldsymbol{x},\boldsymbol{x}') = K^l(\boldsymbol{x},\boldsymbol{x}') := \mathbb{E}[u_i^l(\boldsymbol{x})u_i^l(\boldsymbol{x}')].$$

于是可以进一步通过有限宽网络输出特征的内积进行高效的随机特征近似. 具体而言，对于模型开发者本地数据集 $\boldsymbol{X}=\{(\boldsymbol{x}_i)\}_{i=1}^{n}$，可以通过采样生成 N 个随机网络，输入维度为 M，从而对原始数据集进行随机特征表示如下：

$$\hat{\Phi}(\boldsymbol{X}) \leftarrow \frac{1}{\sqrt{NM}}[f_{\theta_1}(\boldsymbol{X}),\cdots,f_{\theta_N}(\boldsymbol{X})]^{\mathrm{T}} \in \mathbb{R}^{|NM|\times n} \quad (4)$$

其中 θ_i代表第 i 个随机生成的神经网络的参数. 对于后续生成的缩略集，也可以通过同样的方式计算其随机特征表示，然后通过简单的矩阵乘法得到所需的近似核矩阵. 同时，因为不涉及反向求导过程，且均用有限宽网络表示，规约构造过程中对于缩略集的梯度信息可以依赖深度学习框架高效完成. 最终的规约构造算法伪代码如算法 1 所示.

算法 1.　随机特征神经切线核 RKME 规约构造算法.

输入：模型提交者本地数据集$\mathcal{D}=\{(\boldsymbol{x}_i,\boldsymbol{y}_i)\}_{i=1}^{n}$，随机网络初始化参数分布 $p(\theta)$，随机网络个数 N，随机网络输出维度 M，学习率 η，迭代轮数 T

输出：神经切线核 RKME 缩略集规约$\{\boldsymbol{\beta},\boldsymbol{Z}\}$

1. 随机初始化缩略集 $\boldsymbol{Z}=\{\boldsymbol{z}_i\}_{i=1}^{m}$；
2. FOR $t=1:T$
3. 　根据市场提供的分布 $p(\theta)$参数采样生成 N 个输出维度为 M 的随机网络；
4. 　根据随机网络计算数据集随机特征表示：
$\hat{\Phi}(\boldsymbol{X}) \leftarrow \frac{1}{\sqrt{NM}}[f_{\theta_1}(\boldsymbol{X}),\cdots,f_{\theta_N}(\boldsymbol{X})]^{\mathrm{T}} \in \mathbb{R}^{|NM|\times n}$；
5. 　根据随机网络计算缩略集随机特征表示：
$\hat{\Phi}(\boldsymbol{Z}) \leftarrow \frac{1}{\sqrt{NM}}[f_{\theta_1}(\boldsymbol{Z}),\cdots,f_{\theta_N}(\boldsymbol{Z})]^{\mathrm{T}} \in \mathbb{R}^{|NM|\times m}$；
6. 　计算核矩阵：
$\hat{\boldsymbol{K}}_{\boldsymbol{XZ}} \leftarrow \hat{\Phi}(\boldsymbol{X})^{\mathrm{T}}\hat{\Phi}(\boldsymbol{Z})$；$\hat{\boldsymbol{K}}_{\boldsymbol{ZZ}} \leftarrow \hat{\Phi}(\boldsymbol{Z})^{\mathrm{T}}\hat{\Phi}(\boldsymbol{Z})$；
7. 　求解权重向量 $\boldsymbol{\beta}$；
8. 　对缩略集 $\boldsymbol{Z}$ 进行梯度更新；
9. END FOR

4.3　部署阶段

在学件部署阶段，学件基座系统通过规约识别

对用户有帮助的学件推荐给用户，从而帮助用户快速解决其学习任务. 学件基座系统可以使用上述算法为用户生成神经切线核 RKME 缩略集，进而与学件的 RKME 规约进行匹配. 具体而言，假设学件基座系统中包含 C 个学件 $\{(\boldsymbol{f}_c,\tilde{\boldsymbol{\mu}}_c)\}_{c=1}^{C}$，其中 $\tilde{\boldsymbol{\mu}}_c$ 为第 c 个学件的神经切线核 RKME 规约，$\boldsymbol{f}_c$ 为第 c 个学件的模型，则学件基座系统可以通过计算 RKME 之间的 RKHS 距离，从而匹配最合适的学件推荐给用户，用户可以复用学件模型解决其学习任务. 具体学件匹配算法伪代码如算法 2 所示.

算法 2. 基于神经切线核 RKME 规约的学件匹配.

输入：用户本地数据 $\mathcal{D}_u=\{(\boldsymbol{x}_i)\}_{i=1}^{n_u}$，包含 C 个学件的学件基座系统 $\{(\boldsymbol{f}_c,\tilde{\boldsymbol{\mu}}_c)\}_{c=1}^{C}$

输出：学件基座系统推荐学件 $(\boldsymbol{f}_c,\tilde{\boldsymbol{\mu}}_c)$ 给用户

1. 通过算法 1 为用户数据生成神经切线核 RKME 缩略集 $\{\boldsymbol{\beta}_u,\boldsymbol{Z}_u\}$；
2. 用户将神经切线核 RKME 缩略集 $\{\boldsymbol{\beta}_u,\boldsymbol{Z}_u\}$ 发送至学件基座系统；
3. 学件基座系统计算用户 RKME 与学件 RKME 规约之间的 RKHS 距离；
4. 学件基座系统推荐 RKHS 距离最近的第 j 个学件给用户：

$$j=\arg\min_{c=1,\cdots,C}\left\|\tilde{\boldsymbol{\mu}}_c-\sum_{i=1}^{m}\beta_{ui}k(\boldsymbol{z}_{ui},\cdot)\right\|_{\mathcal{H}_k};$$

5. 复用学件模型在用户数据上进行预测.

值得注意的是，由于 RKME 缩略集规模远小于原始训练数据集，所以采用准确计算的 NTK 核矩阵来计算 RKME 之间的 RKHS 距离，而不使用随机近似来进行学件匹配，避免引入额外的随机性，从而可以准确地匹配对用户有帮助的学件.

本节中，我们采取了最基础的 RKME 规约的匹配方法，用于验证所提出方法的有效性. 同时，本文所提出的规约构造方式也可以被自然地拓展到锚位学件构建[26]、异构规约构建[25]等场景.

5 实验结果与分析

本节中，我们构建了包含大量学件的学件市场，学件模型来自销量预测、图像分类的机器学习任务. 基于该市场，我们针对提出的 RKME-NTK-RF 算法在学件范式中的实际表现进行了实验验证，包括在大规模数据集上规约的生成效率，以及算法构造的神经切线核 RKME 规约对于为用户需求匹配学件的有效性.

5.1 规约构造效率

为保护数据隐私，RKME 规约的生成需要在开发者本地完成，所以规约生成过程应尽可能地高效. 在本实验中，我们从核矩阵计算效率、RKME 规约构造效率两个方面分别进行我们的算法与以往算法的效率对比.

RKME 规约的构造算法中，开销最大的步骤为核矩阵的计算，我们使用随机生成的数据，对核矩阵计算效率进行对比. 对比方法为使用传统的高斯核函数（RBF），以及使用 Neural Tangents 加速框架[48]计算的未近似的准确 NTK 核函数，包括 NTK-MLP 与 NTK-CNN 两种核函数，进行 RKME 规约的构建，其中 NTK 对应的网络结构如下：NTK-MLP 对应多层全连接模型，使用 3 层隐层结构，每层包含 256 维特征，并使用 ReLU 激活函数，且不包含任何 Dropout 操作；对于 NTK-CNN，使用 3 层卷积进行特征表示，每层卷积输出 256 维特征的张量，并在每层卷积后通过（2×2）的池化操作进行降维. 将最后输出的张量进行拼接，通过一层全连接.

实验结果如表 1 所示，可以看出，本文提出的规约构造算法的核矩阵计算相比于使用 Neural Tangents 框架准确计算 NTK，在计算效率上有明显的提升，甚至在模拟图像数据上，准确的 NTK 计算由于内存不足无法得到结果. 我们的算法甚至可以超越具有显式表达的传统高斯核函数的计算效率.

表 1 随机数据核矩阵生成耗时

数据维度	核函数	时间/s
20k×80	RBF	1.03±0.05
	NTK-MLP	5.19±0.74
	NTK-RF-MLP	**0.29±0.70**
20k×3×32×32	RBF	2.28±0.12
	NTK-MLP	6.34±2.39
	NTK-CNN	NaN
	NTK-RF-MLP	**0.32±0.22**
	NTK-RF-CNN	1.75±0.74

此外，在规约生成过程中，需要对缩略集进行迭代更新. 这一过程也会带来额外的时间消耗. 我们使用随机生成的模拟数据测试提出算法的完整规约生成过程中的时间消耗. 由于精确计算的 NTK 难以对缩略集计算梯度信息，于是我们仅与高斯核进行比较. 实验结果如表 2 所示，可以看出本文算法能够以接近传统显式核函数的计算效率近似构建出基于神经切线核的 RKME 规约. 同时，可以看出随着数据维度增加，将默认的缩略集规模从 50 提高到 100，我们的方法仍然维持极高的效率.

表 2　随机数据规约生成耗时

数据维度	缩略集	核函数	时间/s
200k×80	50	RBF	**0.38±0.02**
		NTK-RF-MLP	1.53±0.84
200k×500	50	RBF	1.82±0.05
		NTK-RF-MLP	**1.76±0.83**
	100	RBF	5.74±0.05
		NTK-RF-MLP	**2.03±0.97**
20k×3×32×32	100	RBF	1.38±0.13
		NTK-RF-MLP	**1.02±0.13**
		NTK-RF-CNN	3.51±0.87

5.2　表格数据实验

我们构建了大量来自销量预测任务的学件作为学件市场，基于本文提出的规约构造算法，验证从市场中进行学件查搜的准确率. 市场中的模型构建方式如下：基于在 Kaggle 上公开的经典销量预测任务 Predict Future Sales(PFS)[49]，该数据集中的销量数据来自 52 家商店. 按照销量信息的日期划分训练集和验证集，基于每家商店的训练数据训练模型，并为其构建 RKME 规约，构成了学件市场中的学件. 我们分别使用岭回归模型及 LightGBM 模型[50]作为学件市场中的模型.

我们采用留一验证的方法验证我们的方法的查搜效率，即使用每家商店的验证集作为目标任务，在学件市场中基于规约进行查搜，选择除本商店的学件之外，其他学件中再生核希尔伯特空间中的距离最小的学件进行直接复用. 在这里，我们对比了不同的规约算法选择的模型相比于平均误差带来的提升. 实验结果如表 3 所示.

表 3　销量预测任务基于规约查搜模型准确率（相比于市场模型平均表现）

规约类型	提升比例/%	
	LightGBM	Ridge
RKME-RBF	24.3±27.7	20.6±20.1
RKME-NTK	**24.7±27.3**	**21.6±21.6**

由表 3 中可以看出我们的方法即使在常见的表格数据任务上，相比高斯核规约仍有一定的提升，且相比随机从市场中选择模型，均能够带来超过 20% 的提升. 这说明了学件范式可以帮助用户利用少量数据识别对其有帮助的学件.

5.3　图像数据实验

为了验证我们的方法在表格数据以外的数据类型上的表现，我们基于 CIFAR-10[51]图像分类任务构建学件市场，以验证我们的方法在图像数据上的性能. 在 CIFAR-10 数据集中，共有 10 个类别的图像数据，其中，训练集每个类别有 5000 张图片，共 50000 张训练样本；测试集中每个类别有 1000 张图片，共 10000 张测试样本. 所有样本均是大小为 32×32 的 3 通道 RGB 彩色图像.

对于 CIFAR-10 数据集，我们按类别对训练集进行不均匀采样，为 50 个学件构建了只包含部分类别的不均衡训练数据集. 这使得学件市场中不存在任何一个学件可以准确处理所有类别的数据；只有训练数据与目标任务数据分布最接近的学件，才可能在目标任务上取得良好表现. 具体地，每个类别被采样的概率服从一个随机多项分布，仅有 4 个类别上采样概率为正，采样比例为 0.4∶0.4∶0.1∶0.1. 最终，每个学件的训练集包含 12000 个样本，覆盖 CIFAR-10 中 4 个类别的数据.

我们基于选中的数据，为每个学件训练模型，使用 3 层卷积＋1 层全连接的结构，每层卷积输出维度为 64 维. 我们使用带动量的随机梯度下降算法对模型进行训练，为了防止模型在对应任务上出现过拟合，对每个模型训练 35 个 epoch. 最终，所有模型在与其训练集同分布采样的测试集上的准确率达到约 98%.

此外，我们使用 CIFAR-10 的测试集数据，构建了 50 个目标任务. 与构建学件训练集类似，为使得任务之间存在一定的差异，对测试集进行不均匀采样. 具体地，每个类别被采样的概率服从一个随机多项分布，仅有 6 个类别上采样概率为正，采样比例为 0.3∶0.3∶0.1∶0.1∶0.1∶0.1. 最终，每个目标任务包含 3000 个样本，覆盖 CIFAR-10 中 6 个类别的数据.

对于所有学件和目标任务，缩略集样本点数量均为 $m=48$，在此实验中，我们对比不同的规约算法选择的模型在目标任务上进行投票复用得到的分类准确率. 在计算缩略集过程中，使用 NTK-RF-CNN 核函数进行核矩阵计算；在查搜学件过程中，使用 Neural Tangents 库进行核矩阵计算. 与现有的 RBF 核进行比较，CIFAR-10 数据上的结果如表 4 所示.

表 4　图像分类任务基于规约查搜模型准确率

规约类型	复用准确率/%
RKME-RBF(Top1)	47.7±2.6
RKME-NTK-CNN(Top1)	**55.4±2.4**
RKME-RBF(Top3)	50.5±2.8
RKME-NTK-CNN(Top3)	**59.4±1.8**
Oracle	62.7±0.1
Market Average	34.5±0.0

其中，Oracle 对应采样所服从的多项分布参数已知，直接选择采样分布最相似学件的表现，可以作为采样参数未知条件（正常情景）下的准确率上界；Average 对应市场中所有学件的平均表现，即随机挑选学件. Top1 和 Top3 分别对应选择相似度最高的一个或三个学件投票集成结果.

从结果中可以看出，对于大量目标任务上的平均表现，本文提出的 NTK-RF-CNN 核方法在使用图像数据建立的学件市场上的效果明显优于 RBF 核函数，提高约 7%～9%；且挑选出的模型的复用准确率相比从市场中随机挑选学件，准确率的提升接近 25%. 同时，我们的方法选出学件的准确率距离准确率上界 Oracle-Top1 仅有约 3%的差距.

从图 1 中可以看出，在绝大部分用户任务上，我们的方法所选出的学件的准确率（实线）显著高于 RBF 规约核函数选出的学件的准确率（浅色虚线，深色区域下边缘），在图中被标注为浅色区域；在所有任务上，我们的方法所选出来的学件均优于随机选择学件（深色虚线）；在部分任务上，我们的方法所选学件投票复用后准确率超过市场中的最优学件

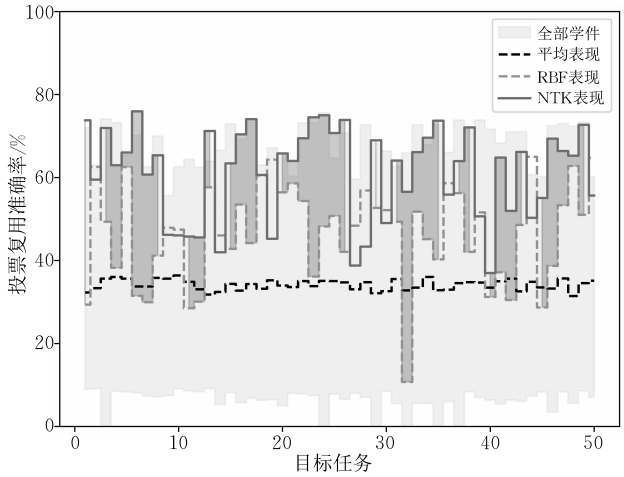

图 1 在所有用户任务上各种方法的表现对比

（浅色区域上边缘）.

我们又将同一学件对应 RBF 和 NTK-RF-CNN 规约缩略集中权重最大的 8 个缩略点分别绘制在图 2 的第 1 行和第 3 行，其中，β 代表了该缩略点在规约中所占权重；将训练数据集中与 RBF 规约中缩略点欧氏距离最近的真实数据绘制在第 2 行的对应位置，在下方标注了二者间的欧氏距离.

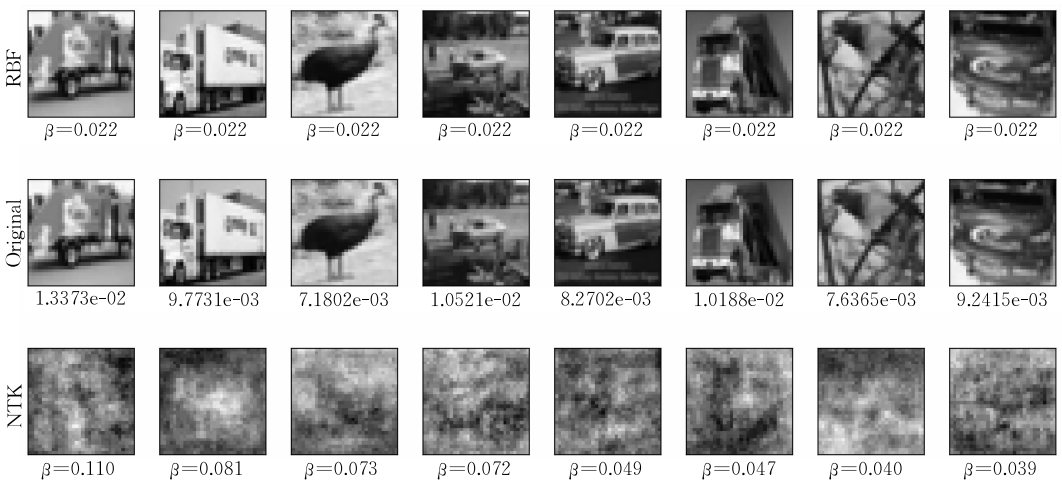

图 2 图像数据 RKME 缩略集规约可视化

我们可以发现，使用 RBF 作为规约核函数，在处理图像类的高维数据时，几乎不再存在有价值的簇结构，而是近似随机收敛，所以 β 的值均相等且为 0.022. 这导致高维情况下，使用基于欧氏距离的 RBF 核存在着潜在的隐私泄露风险，人眼可以从中轻易获取到真实数据的信息. 而相比之下，本文提出的 NTK-RF-CNN 方法，可以在表示空间中合理地度量图像之间的关系，从而正常地生成有价值的规约. 且人眼难以从缩略集中获得真实数据信息，缩略集的权重也被充分利用. 这同样验证了相较 RBF 规约核函数，本文提出的方法能够在图像类的高维数据上，更高效安全地帮助用户选择对其有帮助的学件.

6 总结及展望

学件范式以规约为核心，通过构建学件基座系统，使得用户在解决新的机器学习任务时可以部分复用他人结果，不必从头开始. 规约需要在刻画模型特性与效用的同时，不会泄露开发者的数据隐私，并

且可以针对用户的需求高效准确地识别有帮助的模型.本文通过引入神经切线核,缓解了 RKME 规约在高维数据时面临的维度灾难的问题,拓展了 RKME 规约的表示能力与适用范围,同时保持了 RKME 规约优良的理论性质.此外,本文通过神经网络高斯过程与随机特征近似对 NTK 核矩阵进行了高效计算,从而在提升规约能力的同时,可以更高效地完成规约构造.通过真实数据集上的实验验证了所提出算法的有效性.学件范式为同时解决机器学习中许多关键问题提供了系统性框架.如何从理论上证明规约对开发者数据的隐私保护能力是亟待进一步研究的问题.

参　考　文　献

[1] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020: 1877-1901

[2] He Kaiming, Zhang Xiangyu, Ren Shaoqing, et al. Deep residual learning for image recognition//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778

[3] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436-444

[4] De Lange M, Aljundi R, Masana M, et al. A continual learning survey: Defying forgetting in classification tasks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(7): 3366-3385

[5] Zhou Zhi-Hua. Learnware: On the future of machine learning. Frontiers of Computer Science, 2016, 10(4): 589-590

[6] Zhou Zhi-Hua, Tan Zhi-Hao. Learnware: Small models do big. Science China Information Sciences, 2024, 67(1): 112102

[7] Wu Xi-Zhu, Xu Wen-Kai, Liu Song, et al. Model reuse with reduced kernel mean embedding specification. IEEE Transactions on Knowledge and Data Engineering, 2023, 35(1): 699-710

[8] Zhang Yu-Jie, Yan Yu-Hu, Zhao Peng, et al. Towards enabling learnware to handle unseen jobs//Proceedings of the AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2021, 35(12): 10964-10972

[9] Ramdas A, Reddi S J, Póczos B, et al. On the decreasing power of kernel and distance based nonparametric hypothesis tests in high dimensions//Proceedings of the AAAI Conference on Artificial Intelligence. Texas, USA, 2015, 29(1): 3571-3577

[10] Suzuki T. Adaptivity of deep ReLU network for learning in Besov and mixed smooth Besov spaces: Optimal rate and curse of dimensionality//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018: 1-26

[11] Lopez-Paz D, Oquab M. Revisiting classifier two-sample tests//Proceedings of the International Conference on Learning Representations. Toulon, France, 2017: 1-27

[12] Jacot A, Gabriel F, Hongler C. Neural tangent kernel: Convergence and generalization in neural networks//Proceedings of the Advances in Neural Information Processing Systems. Montréal, Canada, 2018: 8580-8589

[13] Lee J, Bahri Y, Novak R, et al. Deep neural networks as Gaussian processes//Proceedings of the International Conference on Learning Representations. Vancouver, Canada, 2018

[14] Rahimi A, Recht B. Random features for large-scale kernel machines//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2007: 1177-1184

[15] Pan S J, Yang Qiang. A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359

[16] Zhuang Fu-Zhen, Qi Zhi-Yuan, Duan Ke-Yu, et al. A comprehensive survey on transfer learning. Proceedings of the IEEE, 2020, 109(1): 43-76

[17] Mansour Y, Mohri M, Rostamizadeh A. Domain adaptation: Learning bounds and algorithms//Proceedings of the 22nd Annual Conference on Learning Theory. Montreal, Canada, 2009: 20-35

[18] Ganin Y, Lempitsky V. Unsupervised domain adaptation by backpropagation//Proceedings of the 32nd International Conference on Machine Learning. San Diego, USA, 2015: 1180-1189

[19] Yang Yang, Zhan De-Chuan, Fan Ying, et al. Deep learning for fixed model reuse//Proceedings of the AAAI Conference on Artificial Intelligence. San Francisco, USA, 2017: 2831-2837

[20] Zhao Peng, Cai Le-Wen, Zhou Zhi-Hua. Handling concept drift via model reuse. Machine Learning, 2020, 109: 533-568

[21] Kuzborskij I, Orabona F. Stability and hypothesis transfer learning//Proceedings of the 30th International Conference on Machine Learning. Scottsdale, USA, 2013: 942-950

[22] Du S S, Koushik J, Singh A, et al. Hypothesis transfer learning via transformation functions//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 574-584

[23] Kundu J N, Venkat N, Babu R V, et al. Universal source-free domain adaptation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA, 2020: 4544-4553

[24] Tan Peng, Tan Zhi-Hao, Jiang Yuan, et al. Towards enabling learnware to handle heterogeneous feature spaces. Machine Learning, 2024, 113(4): 1839-1860

[25] Tan Peng, Tan Zhi-Hao, Jiang Yuan, et al. Handling learnwares developed from heterogeneous feature spaces without auxiliary data//Proceedings of the 32nd International Joint Conference on Artificial Intelligence. Washington, USA, 2023: 4235-4243

[26] Xie Yi, Tan Zhi-Hao, Jiang Yuan, et al. Identifying helpful learnwares without examining the whole market//Proceedings of the 26th European Conference on Artificial Intelligence. Kraków, Poland, 2023: 2752-2759

[27] Liu Jian-Dong, Tan Zhi-Hao, Zhou Zhi-Hua. Towards making learnware specification and market evolvable//Proceedings of the 26th AAAI Conference on Artificial Intelligence. Vancouver, Canada, 2024, 38(12): 13909-13917

[28] Tan Zhi-Hao, Liu Jian-Dong, Bi Xiao-Dong, et al. Beimingwu: A learnware dock system. arXiv preprint arXiv:2401.14427, 2024

[29] Arora S, Du S S, Hu Wei, et al. On exact computation with an infinitely wide neural net//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 8139-8148

[30] Du S S, Hou Kang-Cheng, Salakhutdinov R R, et al. Graph neural tangent kernel: Fusing graph neural networks with graph kernels//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 5724-5734

[31] Alemohammad S, Wang Zichao, Balestriero R, et al. The recurrent neural tangent kernel//Proceedings of the International Conference on Learning Representations. Vienna, Austria, 2021

[32] Yang G. Tensor programs II: Neural tangent kernel for any architecture. arXiv preprint arXiv:2006.14548, 2020

[33] Yang G, Littwin E. Tensor programs IIb: Architectural universality of neural tangent kernel training dynamics//Proceedings of the 38th International Conference on Machine Learning. 2021: 11762-11772

[34] Tan Zhi-Hao, Xie Yi, Jiang Yuan, et al. Real-valued backpropagation is unsuitable for complex-valued neural networks//Proceedings of the Advances in Neural Information Processing Systems. New Orleans, USA, 2022, 35: 34052-34063

[35] Arora S, Du S S, Li Zhiyuan, et al. Harnessing the power of infinitely wide deep nets on small-data tasks//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020

[36] Lee J, Schoenholz S, Pennington J, et al. Finite versus infinite neural networks: An empirical study//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2020, 33: 15156-15172

[37] Yang G, Hu E J. Tensor programs IV: Feature learning in infinite-width neural networks//Proceedings of the 38th International Conference on Machine Learning. 2021: 11727-11737

[38] Nguyen T, Chen Zhou-Rong, Lee J. Dataset meta-learning from kernel ridge-regression//Proceedings of the International Conference on Learning Representations. Vienna, Austria, 2021

[39] Nguyen T, Novak R, Xiao Lechao, et al. Dataset distillation with infinitely wide convolutional networks//Proceedings of the Advances in Neural Information Processing Systems. 2021: 5186-5198

[40] Williams C, Seeger M. Using the Nyström method to speed up kernel machines//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2000: 682-688

[41] Zandieh A, Han I, Avron H, et al. Scaling neural tangent kernels via sketching and random features//Proceedings of the Advances in Neural Information Processing Systems. 2021: 1062-1073

[42] Novak R, Sohl-Dickstein J, Schoenholz S S. Fast finite width neural tangent kernel//Proceedings of the 39th International Conference on Machine Learning. Baltimore, USA, 2022: 17018-17044

[43] Loo N, Hasani R, Amini A, et al. Efficient dataset distillation using random feature approximation//Proceedings of the Advances in Neural Information Processing Systems. New Orleans, USA, 2022: 13877-13891

[44] Smola A, Gretton A, Song Le, et al. A Hilbert space embedding for distributions//Proceedings of the International Conference on Algorithmic Learning Theory. Sendai, Japan, 2007: 13-31

[45] Fukumizu K, Gretton A, Sun Xiaohai, et al. Kernel measures of conditional dependence//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2007: 489-496

[46] Bietti A, Mairal J. On the inductive bias of neural tangent kernels//Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada, 2019: 12873-12884

[47] Cheng Xiu-Yuan, Xie Yao. Neural tangent kernel maximum mean discrepancy//Proceedings of the Advances in Neural Information Processing Systems. 2021, 34: 6658-6670

[48] Novak R, Xiao Le-Chao, Hron J, et al. Neural tangents: Fast and easy infinite neural networks in python//Proceedings of the International Conference on Learning Representations. Addis Ababa, Ethiopia, 2020

[49] Guschin A, Ulyanov D, Trofimov M, et al. Predict future sales. Kaggle, 2018. https://kaggle.com/competitions/competitive-data-science-predict-future-sales

[50] Ke Guo-Lin, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree//Proceedings of the Advances in Neural Information Processing Systems. Long Beach, USA, 2017: 3146-3154

[51] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Technical Report, 2009

TAN Zhi-Hao, Ph. D. candidate. His main research interests include machine learning and data mining.

SHI Hao-Yu, undergraduate student. His main research interests include machine learning and data mining.

CHEN Zi-Xuan, M. S. His main research interests include machine learning and data mining.

JIANG Yuan, Ph. D., professor. Her main research interests include machine learning and data mining.

Background

Machine learning has achieved significant success in various real-world applications, including medicine, robotics, and ecology. However, developing a well-performed model necessitates several essential conditions, such as sufficient labeled data, adequate computational resources, and proficient training skills. Therefore, most ordinary users can hardly produce high-quality models starting from scratch. Furthermore, it is difficult to identify and reuse beneficial models among different users due to data privacy and proprietary concerns. To tackle the above issues simultaneously, the learnware paradigm was proposed and developed by Prof. Zhi-Hua Zhou to establish a learnware dock system containing numerous machine learning models, enabling users to build models by reusing existing efforts instead of starting from scratch. A learnware is a well-performed model with a specification representing its specialty and utility, enabling the model to be adequately identified for subsequent user tasks. Developers can spontaneously submit their trained models on various tasks to the dock system, and the dock system assigns specifications to accepted models. Given a new user task, the dock system can identify helpful learnwares based on the submitted user requirements. Note that the learnware dock system has no access to the raw data of developers and users.

The core of learnware paradigm lies in the specification, which plays a crucial role in model characterization and identification without leaking raw data. Recently the Reduced Kernel Mean Embedding (RKME) specification was proposed, which concisely represents the model's training data distribution without exposure of raw data. Based on this specification, a recent line of studies about the learnware paradigm succeeded in identifying helpful learnwares by measuring distribution similarity between RKME specifications and user tasks. However, in practice, the learnware dock system comprises machine learning models from different domains and various data types, whereas the RKME specification based on traditional kernel methods faces curse of dimensionality, which makes it hard to be applied in high dimensional scenarios like images. In this paper in order to alleviate the curse of dimensionality, we make the first attempt to explore the construction of RKME specification based on neural tangent kernels (NTK), and improves its efficiency through neural network Gaussian process (NNGP) and random feature approximation. This approach enables the efficient and accurate generation of the RKME specification for various models. Compared to traditional RKME specification, our method has improved the identification accuracy significantly by nearly 9%.

This work is supported by the National Natural Science Foundation of China (Nos. 62250069, 62176117) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (No. KYCX23_0159).