



# 领域知识指导的模型重用

吴西竹, 周志华\*

南京大学计算机软件新技术国家重点实验室, 南京 210023

\* 通信作者. E-mail: zhouzh@lamda.nju.edu.cn

收稿日期: 2017-05-16; 接受日期: 2017-05-22; 网络出版日期: 2017-11-14

国家自然科学基金 (批准号: 61333014) 资助项目

**摘要** 机器学习模型的生命周期通常比较短暂, 大量的机器学习模型针对特定任务设计, 在完成任务之后即失去使用价值. 然而, 一个精心设计和训练的模型通常更精炼地概括了训练数据中蕴含的知识. 更进一步地, 当无法获取原始训练数据时, 已有的预训练模型就是仅剩的信息来源. 本文提出了一种重用已有的预训练机器学习模型来辅助当前任务学习的框架, 尤其适用于当前任务数据量不足的情形. 该方法把已有模型视为黑盒, 不需要已有模型的结构信息, 并使用领域知识对已有模型进行筛选和结合. 实验表明, 在该框架下, 通过对已有模型的重用, 多个实际任务上的性能可以得到显著提升.

**关键词** 机器学习, 模型重用, 领域知识, 环境变化, 学件

## 1 引言

机器学习是目前处理许多实际应用的重要方法, 包括计算机视觉、自然语言处理、生物信息学等. 通常情况下, 我们需要足够多的数据用于训练, 并由机器学习专家针对具体问题设计学习算法.

一方面, 数据, 特别是高质量的数据是宝贵的资源. 尽管在学术领域已有许多公开数据集, 由于版权、隐私、经济利益等原因, 许多数据集不可公开. 另一方面, 目前针对不同的任务, 已有许多训练好的机器学习模型. 这些模型有的使用了大量训练数据, 有的使用了巧妙的学习和优化算法, 有的在目前最先进的计算设备上仍需要数周甚至数月的训练时间完成训练. 这些模型在指定的任务上通常具有较好的性能, 但由于只能处理预先设置的任务, 一旦任务发生改变, 就只能被抛弃, 导致我们必须重新在当前任务上训练新的模型.

如何使用更少的数据, 重用已有模型, 适应环境变化, 是 Zhou<sup>[1,2]</sup> 提出的学件 (learnware) 概念所需要达成的目标之一. 本文针对实际问题中的常见场景, 细化了两个需要注意的限制条件: (1) 已有模型的训练数据无法获得. 因此, 传统的迁移学习<sup>[3]</sup> 方法不能直接应用; (2) 已有的模型可能使用不同的未知算法进行训练, 我们只能使用它们的输入输出, 而不了解模型的结构.

**引用格式:** 吴西竹, 周志华. 领域知识指导的模型重用. 中国科学: 信息科学, 2017, 47: 1483-1492, doi: 10.1360/N112017-00106  
Wu X Z, Zhou Z H. Model reuse with domain knowledge (in Chinese). Sci Sin Inform, 2017, 47: 1483-1492, doi: 10.1360/N112017-00106

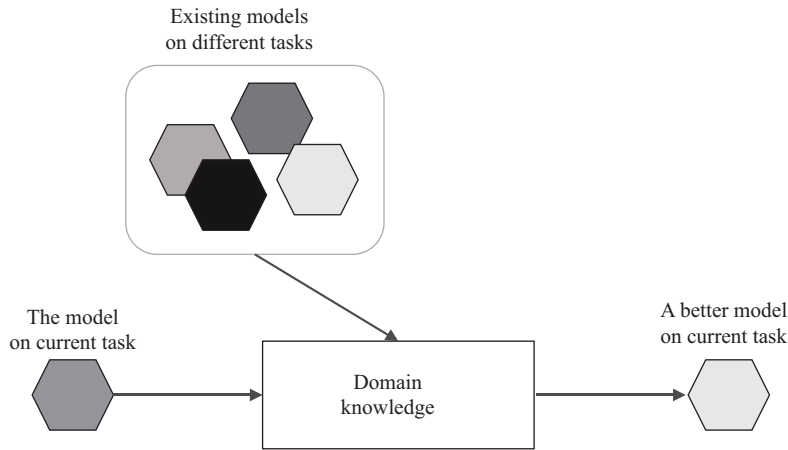


图 1 MRDK 框架示意图

Figure 1 The illustration of the MRDK framework

在这两个限制条件之外, 我们还注意到, 实际的机器学习模型通常都会提供其输出的含义, 而不仅是数值. 例如, 当一个训练好的沙漠图片分类器输出它对一张图片的预测时, 输出不仅是数值, 还是该分类器对这一图片是否是沙漠的观点. 基于“沙漠中缺水”这一领域知识, 我们在设计一个预测图片中是否有水的分类器时, 该沙漠分类器的预测很可能可以提供帮助. 如果我们有更多的领域知识, 已有分类器的输出就会更有意义, 从而帮助当前任务的学习.

本文提出了名为 MRDK (model reuse with domain knowledge) 的机器学习框架, 利用领域知识结合已有模型, 在环境发生变化的情况下, 仍能帮助当前任务的学习. 该框架的示意图见图 1. 具体而言, 在本文提出的框架中, 已有模型被视为黑盒, 并且无法获得训练这些模型所用的训练数据. 而在当前任务中, 训练数据极少, 无法直接得到一个性能较好的模型. 在领域知识的指导下, 已有模型可以辅助当前任务的学习, 以得到性能更佳在当前任务模型.

本文第 2 节对适应环境变化和利用领域知识两方面的工作进行了简要总结. 第 3 节提出并详细阐释了 MRDK 框架. 第 4 节通过在具体问题上的实例化实验, 验证了 MRDK 框架的作用. 最后总结全文, 并对未来值得研究的问题进行了初步探讨.

## 2 相关工作

本节分别介绍目前机器学习领域中, 适应环境变化和利用领域知识两方面的工作.

### 2.1 适应环境变化

传统机器学习模型只能完成预先设置好的任务, 难以适应环境变化. 已有的领域自适应 (domain adaptation) 方法<sup>[4~6]</sup>和迁移学习方法<sup>[3,7]</sup>依赖于以往任务数据, 需要同时使用以往数据和当前任务数据, 训练新的模型, 而无法使用已有模型. 在实际应用中, 以往任务的数据多数不能公开或无法取得, 给使用领域自适应方法和迁移学习方法带来了困难.

最近, Zhou<sup>[1,2]</sup>提出了学件的概念, 强调利用已有机器学习模型及其规约, 以适应环境变化的重要性. 并指出环境变化的适应主要有两种基本情形: (1) 任务不断变化, 模型需要能自动感知环境的变

化、自适应地对自己进行调整; (2) 任务与以往研究过的任务不同, 直接处理当前任务存在困难, 用户需要找寻以往非相同环境下的相关模型并设计合理的方法加以使用.

在学习过程中可能出现新的类别, 学习系统如何能够自动感知到新类别的出现, 并适当地对自身进行调整, 这属于第 1 种情形的任务之一. Da 等<sup>[8]</sup> 研究了一种方法, 能够利用未标记样本中蕴含的结构信息, 使得在新类别出现之后, 仍能保持在以往类别上的分类性能, 并且准确区分新增类别和已有类别. Mu 等<sup>[9]</sup> 利用矩阵素描 (matrix sketching) 法研究了在流式数据上任务类别发生变化时, 模型如何自适应的问题. Zhu 等<sup>[10,11]</sup> 提出了在多标记及多示例多标记问题中, 处理任务类别变化的方法. 在这些方法中提出的机器学习模型能够感知环境变化, 自我演进, 以取得较好性能.

本文主要研究了第 2 种情形下, 如何设计一种方法, 让用户能够利用已有的相关模型帮助当前任务. 在这一情形下, Li 等<sup>[12]</sup> 提出了 CAPO 方法, 能使用已有的非线性模型, 在任务评测指标变化的情形下, 高效地优化新的指标, 但不能处理任务类别的变化. Yang 等<sup>[13]</sup> 提出的方法与本工作最为相关, 在他们的工作中, 已有模型作为监督信息使用, 可以参与当前任务模型的学习过程, 与当前模型共同训练, 能够减少当前任务所需的数据量. 这一方法不需要使用已有模型的训练数据, 对任务类别变化有良好的适应性. 但是, 在 Yang 等<sup>[13]</sup> 的工作中, 已有模型必须能够参与共同训练, 这需要了解模型结构, 给使用者造成了较大限制. 而本文的工作不需要了解已有模型的结构信息, 有利于使用复杂的已有模型和同时使用多个不同类型的模型来适应任务变化.

## 2.2 利用领域知识

在机器学习研究中, 领域知识通常指训练数据之外的较高层次的知识, 这些知识一般无法刻画为特征放入训练数据中. 在学习过程中使用领域知识通常能够起到促进作用. 例如, Zhao 等<sup>[14]</sup> 利用领域知识指导聚类, Dai 等<sup>[15]</sup> 利用领域知识学习通用特征表示, 另一些工作<sup>[16~18]</sup> 用领域知识从较少的图片中学习如何检测新增目标物体.

利用领域知识的方法根据算法和应用的的不同也各不相同. 一种多视图聚类的工作<sup>[14]</sup> 通过利用“必连 (must-link)”和“勿连 (cannot-link)”关系, 补全示例关系矩阵, 实现聚类. 一种主题模型的工作<sup>[19]</sup> 在后验分布中加入一阶逻辑规则来利用领域知识. Dai 等<sup>[15]</sup> 将一阶逻辑知识用于缩小特征学习的假设空间, 再使用统计学习方法从较小的假设空间学习, 既能提高效率, 又能生成符合人类认知的特征表示. 这些方法尽管在特定领域取得了较好的效果, 但大多需要在模型的训练阶段引入领域知识, 无法适应环境变化. 另一些计算机视觉领域针对小样本学习 (one-shot learning, zero-shot learning) 的工作<sup>[18,20]</sup> 具有较好的适应环境变化的能力, 但依赖于视觉领域的一些领域知识, 无法直接推广到其他领域知识的利用. 本文提出的 MRDK 框架设计了一种较为通用的利用方式, 可以使用多种不同的领域知识形式指导已有模型的重用, 以适应变化后的当前任务.

## 3 MRDK 框架

本节形式化描述 MRDK 框架. 使用小写字母表示数值, 大写字母表示矩阵, 花体大写字母表示空间.

假设已有完成不同分类任务的  $n$  个模型  $\{f_j\}_{j=1}^n$ , 每个  $f_j: \mathcal{X} \rightarrow \mathcal{Y}_j$ . 现在面临新的任务, 需要训练机器学习模型  $f: \mathcal{X} \rightarrow \mathcal{Y}_0$ . 但当前任务只有一个较小的数据集  $D = (X, Y) = \{x_i, y_i\}_{i=1}^m$ , 其中  $x_i \in \mathcal{X}$ ,  $y_i \in \mathcal{Y}_0$ ,  $m$  较小. 如果直接在数据集  $D$  上训练, 由于数据不充足, 得到的  $f_0$  性能可能比较低, 不足

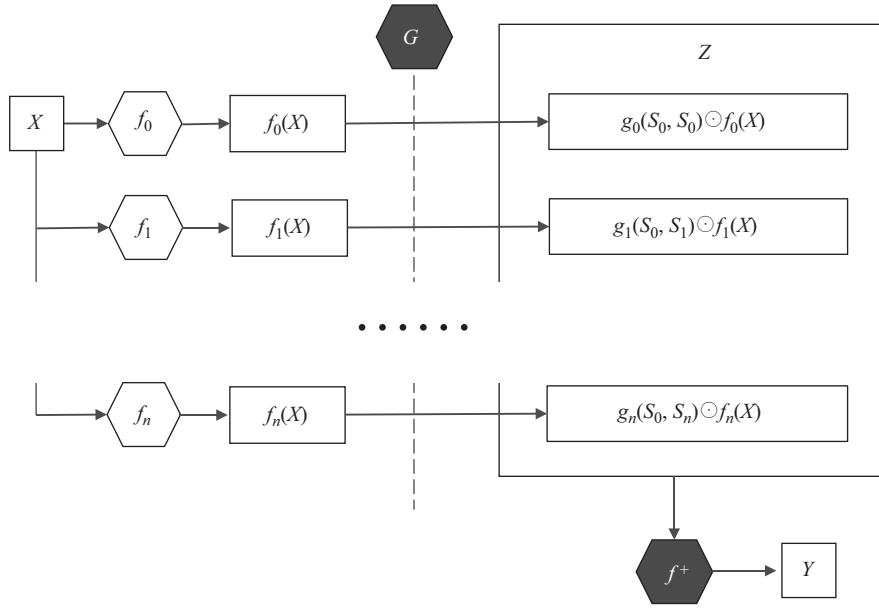


图 2 MRDK 框架中数据和模型的流程关系  
Figure 2 The data and model flow of MRDK

以实用.

可以注意到, 尽管已有模型的输出空间并不与当前任务所需的  $\mathcal{Y}_0$  一致, 但它们的输入空间与当前任务是相同的. 另外, 还知道已有分类器和当前任务输出的含义 (如输出的类别名称、输出类别的描述等), 这些含义信息被刻画在  $S_j$  中. 通过利用领域知识, 建立一个选择函数  $g_j : S_0 \times S_j \rightarrow \{0, 1\}^{|\mathcal{S}_j|}$ , 用于选择与  $\mathcal{Y}_0$  空间相关的  $y_j$ . 对每一个  $f_j$ , 其输出空间通过  $g_j$  的筛选成为  $g_j(S_0, S_j) \odot f_j$ . 例如, 如果  $\mathcal{Y}_1 = (y_{11}, y_{12}, y_{13})$  且  $g_1(S_0, S_1) = (1, 0, 1)$ , 则  $g_1(S_0, S_1) \odot \mathcal{Y}_1 = (y_{11}, y_{13})$ , 第 2 个类别的输出被筛选. 用  $G = \{g_j\}_{j=1}^n$  表示所有选择函数的集合, 之后会更详细地讨论如何从领域知识的不同形式来建立函数  $G$ .

借助已有模型的帮助, 分类器  $f^+$  可以通过以下结构风险最小化 (structural risk minimization) 的方式定义:

$$\min_{f^+} \mathcal{L}(Y, f^+(Z)) + \Omega(f^+) \quad \text{s.t.} \quad Z = \Xi_{j=0}^n (g_j(S_0, S_i) \odot f_j(X)), \quad (1)$$

其中  $\mathcal{L}$  为损失函数,  $\Omega$  为正则项, 这两项的具体定义由应用决定. 符号  $\Xi$  是拼接符号, 水平拼接所有输入矩阵, 形如  $\Xi_{j=0}^n (A_j) = [A_0; A_1; \dots; A_n]$ . 图 2 展示了在 MRDK 框架下, 数据和模型之间的关系. 其中六边形表示模型, 矩形表示数据, 灰底六边形表示该模型需要指定或训练得到, 白底六边形则是已有模型. 通过使用 MRDK 框架, 预期  $f^+$  比  $f_0$  有更好的性能, 并在第 4 节的实验中加以验证.

由于对已有模型没有任何算法、结构假设, 因此可以采用极为复杂的高性能模型. 同时, 损失函数、正则函数和  $f^+$  的形式都是灵活可变的, 因此可将 MRDK 方法作为一种通用框架, 在不同的应用中使用. 在第 4 节中会进一步体现这一点.

从实践角度出发,  $f^+$  本身不需要特别复杂. 一个原因是  $f^+$  的输入是已有模型筛选后的输出, 已具有较高层次的抽象含义. 另一个原因是在当前任务中, 训练数据较少, 过高的复杂度会带来过拟合的风险.

MRDK 中的关键组件是选择函数  $g_j : S_0 \times S_j \rightarrow \{0, 1\}^{|S_j|}$ , 其具体形式可由不同的领域知识的形式确定. 例如, 如果领域知识可以刻画为相似度矩阵  $M^j \in \mathbb{R}^{|S_0| \times |S_j|}$ , 其中  $M_{ab}^j$  为当前任务的第  $a$  项输出与已有模型  $f_j$  的第  $b$  项输出的相关程度, 那么选择函数可以通过阈值判断定义为

$$g_j(S_0, S_j) = ((M^j - T)_{\dagger}^T \mathbf{1})_+, \quad \text{其中 } T = [t]^{|S_0| \times |S_j|}, \quad \mathbf{1} = [1]^{|S_0| \times 1}, \quad (2)$$

其中  $(\cdot)_+$  表示括号中大于 0 的元素置为 1, 其余元素置为 0 的函数. 这一选择函数表示, 已有模型输出与当前某项任务的相关程度高于阈值, 则保留这一输出值, 否则筛去这一输出值. 如果领域知识以非负有权无向图的形式给出, 则可以先将其转化为邻接矩阵的表示方式, 再使用 Floyd-Warshall 算法计算出距离矩阵  $W$ , 最后根据距离矩阵  $W$  计算出相似度矩阵  $M$ . 较为简单直接的做法是取  $M_{ij} = e^{-W_{ij}}$  或是取  $M_{ij} = 1 - W_{ij} / \max(W)$ , 再利用式 (2) 构造选择函数. 对于必连 (must-link) 和勿连 (cannot-link) 关系, 由于两种关系均对最终预测有帮助, 应同时选择与当前任务有连接的已有模型输出. 对于一阶逻辑 (first-order logic) 表达式形式的领域知识, 可以从当前任务的输出变量出发, 将包含这些变量的表达式中的其他变量纳入覆盖集合, 作为选择函数. 在第 4 节中会以具体实例说明如何从领域知识建立选择函数.

## 4 实验

本节将 MRDK 框架运用到两个具体应用中: 生物领域的蛋白质功能预测和计算机视觉领域的图像分类. 以验证通过运用该方法可以在当前任务数据量较小的情况下, 通过对已有模型的利用, 提升模型性能.

### 4.1 蛋白质功能预测

蛋白质测序技术现已比较成熟, 已知的不同生物的蛋白质序列种类超过七千万种. 然而, 对于这些蛋白质的功能人们知之甚少, 目前绝大多数的蛋白质功能只能由计算生物学的方法给出可能的预测. 一份近期对 54 种计算生物学方法的比较研究<sup>[21]</sup>显示, 目前主流的大多数方法均使用了机器学习技术.

使用包括细菌 (bacteria)、古细菌 (archaea)、真核生物 (eukaryota) 三域系统的 5 种不同生物蛋白质组 (proteome, 蛋白质组为该生物的全部蛋白质) 进行了实验, 这 5 种生物的学名为 *Geobacter sulfurreducens*, *Azotobacter vinelandii*, *Haloarcula marismortui*, *Pyrococcus furiosus*, *Saccharomyces cerevisiae*. 为方便起见, 在文中使用对应的助记名 GEOSL, AZOVD, HALMA, PYRFU, YEAST 代替学名. 这些蛋白质的数据从 UniProt 数据库<sup>[22]</sup>取得, 其分子功能由人工实验给出, 按基因本体<sup>[23]</sup>分类体系标注. 由于蛋白质一般同时具有多种功能, 因此每个样本有多个标记. 这些基因组的样本个数、标记个数等信息如表 1 所示.

在实验中, 使用蛋白质特征工具包 ProFET<sup>[24]</sup>提取特征, 从而保证不同生物的蛋白质特征统一. 进行 5 次实验, 每次实验时, 某一生物的蛋白质分类任务被视为待研究的当前任务, 该生物的蛋白质功能只有很少的标注 (每个类别仅有一个样本有标记), 而其他 4 种生物视为已经研究过, 具有所有的样本和标记. 在已经研究过的生物上, 分类模型可以得到充分训练, 并作为已有模型纳入 MRDK 框架, 帮助待研究任务. 这与实际研究中的场景一致, 在某一生物刚被发现时, 能够获得的有标记蛋白质是极少的, 但我们已经拥有了许多研究过的生物的蛋白质功能信息.

表 1 蛋白质功能数据集信息  
Table 1 Proteome datasets information

Domain	Proteome	#Instance	#Class	Label cardinality
Bacteria	GEOSL	378	319	3.143
	AZOVD	406	340	3.993
Archaea	HALMA	304	234	3.247
	PYRFU	425	321	4.480
Eukaryota	YEAST	3507	1566	5.887

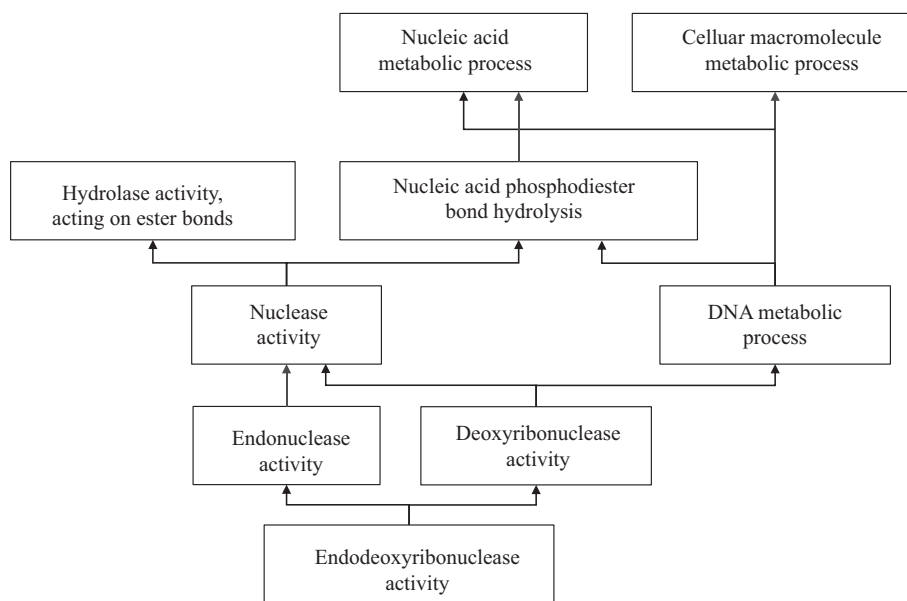


图 3 脱氧核糖核酸内切酶活性功能涉及的部分层级图

Figure 3 Part of the ancestor chart of endodeoxyribonuclease activity

本实验使用的领域知识是蛋白质功能层级图, 这是一张极为庞大的有向无环图, 体现了功能分类的不同层次, 每一种功能是一个节点, 某种功能隶属于更高层次的功能则两点之间有单向边. 图 3 展示了一种名为脱氧核糖核酸内切酶活性 (endodeoxyribonuclease activity) 的功能涉及的部分层级图.

利用这一关系, 构造选择函数  $G$ , 筛除在功能层级图中距离 2 以上的蛋白质功能的模型, 只保留较为相关的模型输出. 使用多标记学习中的常用指标 Hamming loss 和 F-measure 作为评测标准, 以 Logistic Regression 作为模型  $f^+$  的具体形式, 进行 10 次实验, 记录均值. 实验结果见表 2, 其中 Hamming loss 这项指标数值越低越好 (表 2 中用  $\downarrow$  表示), F-measure 这项指标数值越高越好 (表 2 中用  $\uparrow$  表示).

本文进行了  $p$  值为 0.01 的 t-test 显著性检验, 加粗数值表示具有显著性. 可以看出, 在所有测试生物的蛋白质组上, 训练数据不足的原模型  $f_0$ , 经过 MRDK 框架对已有其他模型的利用, 都可以得到性能提升的新模型  $f^+$ , 其中仅 PYRFU 这一蛋白质组的提升没有显著性, 其余生物上均具有显著提升.

表 2 蛋白质功能预测实验结果

Table 2 Results of protein function prediction

Proteome	Hamming loss ↓		F-measure ↑	
	$f_0$	$f^+$	$f_0$	$f^+$
GEOSL	0.070	<b>0.023</b>	0.064	<b>0.124</b>
AZOVD	0.096	<b>0.024</b>	0.035	<b>0.071</b>
HALMA	0.035	<b>0.017</b>	0.097	<b>0.175</b>
PYRFU	0.022	0.017	0.173	0.183
YEAST	0.108	<b>0.009</b>	0.012	<b>0.080</b>

表 3 图像数据集信息

Table 3 Information of image datasets

Dataset	#Instance	#Class	Label cardinality
Scene	2000	5	1.236
VOC07	9963	20	1.437
MS-COCO	122218	80	2.926
NUS-WIDE	133441	81	1.761

## 4.2 图像分类

图像分类是计算机视觉中的常见任务,成熟的分类系统目前已经能够达到接近人类的识别精度.然而,这些系统通常都依赖于已知类别上大量的训练数据,当训练数据不足时,直接使用已有的算法和模型效果不佳.这一场景非常适合使用本文 MRDK 框架,一是已经具有很多成熟的图像分类模型,二是这些图像分类模型通常训练自大量的数据,直接从原始数据出发,帮助当前任务训练新模型代价很大.本文采用了 4 个图像数据集,分别是 Scene<sup>[25]</sup>, VOC07<sup>1)</sup>, NUS-WIDE<sup>[26]</sup> 和 MS-COCO<sup>[27]</sup>, 这些数据集的基本信息见表 3. 其中 MS-COCO 和 NUS-WIDE 数据量巨大,均超过了 10 万张图片,即使在 NVIDIA 的高性能 GPU K40 加速下,仍需要数天时间进行完整训练.

和生物数据集时的实验类似,我们进行了 4 次,每次取一个数据集作为当前任务的数据集进行实验.为了模拟实际场景下数据不足的情形,在当前任务中每个类只取 10 张图像作为有标记样本.其余 4 个数据集样本充足,取所有样本训练作为已有模型,帮助当前任务.已有模型均使用端到端的深度神经网络,按 Gao 等<sup>[28]</sup>的方法训练.对当前任务,由于数据不足以训练完整的深度神经网络,使用 VGG-Net<sup>[29]</sup>中 FC8 层的输出作为深度特征,再使用传统的 Logistic Regression 学习得到分类模型.本实验中使用的领域知识是词汇的语义知识.在图像分类任务上,已有模型的输出均对应着某一具体的概念,如“沙漠”、“海滩”、“杯子”、“夏威夷”等.使用在 Wikipedia 和 Gigaword 语料库上训练得到的词嵌入模型 GloVe<sup>[30]</sup>作为词的向量表示,通过计算向量之间的距离来刻画两个词之间的语义相似度.使用计算得到的相似度值(范围在 0~1 之间)构建相似度矩阵  $M$ ,令阈值  $T$  为所有元素为 0.8 的矩阵,通过式 (2) 构建选择函数,将相似度小于 0.8 的模型输出筛除.

本文仍然使用 Hamming loss 和 F-measure 作为评测标准,进行 10 次实验,进行 t-test 显著性检验,记录均值.实验结果见表 4. 其中 Hamming loss 这项指标数值越低越好(表 4 中用 ↓ 表示), F-measure 这项指标数值越高越好(表 4 中用 ↑ 表示).

1) <http://www.pascal-network.org/challenges/VOC>.

表 4 图像分类实验结果  
Table 4 Results of image classification

Dataset	Hamming loss ↓		F-measure ↑	
	$f_0$	$f^+$	$f_0$	$f^+$
Scene	0.160	0.152	0.685	<b>0.703</b>
VOC07	0.056	<b>0.047</b>	0.623	<b>0.665</b>
MS-COCO	0.040	<b>0.035</b>	0.441	<b>0.454</b>
NUS-WIDE	0.035	<b>0.029</b>	0.198	0.196

从表 4 中可以看出, 通过 MRDK 框架对已有模型的利用, 在大多数情形下  $f^+$  均比  $f_0$  有显著提升. 仅在 NUS-WIDE 上的 F-measure 有不显著的轻微下降 (约 0.002). 这可能与 NUS-WIDE 中许多类别涉及抽象概念, 无法从其他模型获取有效的帮助有关.

## 5 结语

一个实用的机器学习模型通常需要耗费大量人力物力进行数据采集、算法设计、参数调优等繁杂工作, 但仅限于完成训练时指定的任务, 一旦环境发生变化, 只能抛弃原有模型, 重新学习. 本文提出了 MRDK 框架, 尝试通过利用领域知识来重用已有的模型, 在数据匮乏的情形下, 帮助当前任务的学习. 在蛋白质功能预测和图像分类两个具体应用上的实验表明, 使用该方法可以显著提升模型性能, 在当前任务上取得较好的效果.

如何让已有模型能适应环境变化是一个重要问题, 对于精炼和重用过去数据中蕴含的知识, 加速或是更好地学习当前任务具有重要意义. 本工作要求已有模型具备相同的输入特征才可重用, 如何放松这一要求, 实现更广泛的模型重用, 仍是一个有待研究的课题.

## 参考文献

- 1 Zhou Z H. Learnware: on the future of machine learning. *Front Comput Sci*, 2016, 10: 589–590
- 2 周志华. 机器学习: 发展与未来. *中国计算机学会通讯*, 2017, 13: 44–51
- 3 Pan S J, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng*, 2010, 22: 1345–1359
- 4 Jiang J. A literature survey on domain adaptation of statistical classifiers. 2008. [http://sifaka.cs.uiuc.edu/jiang4/domain\\_adaptation/survey/da\\_survey.pdf](http://sifaka.cs.uiuc.edu/jiang4/domain_adaptation/survey/da_survey.pdf)
- 5 Blitzer J, McDonald R, Pereira F. Domain adaptation with structural correspondence learning. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Sydney, 2006. 120–128
- 6 Glorot X, Bordes A, Bengio Y. Domain adaptation for large-scale sentiment classification: a deep learning approach. In: *Proceedings of the 28th International Conference on Machine Learning*, Bellevue, 2011. 513–520
- 7 Luo J, Tommasi T, Caputo B. Multiclass transfer learning from unconstrained priors. In: *Proceedings of the 2011 International Conference on Computer Vision*, Washington, 2011. 1863–1870
- 8 Da Q, Yu Y, Zhou Z H. Learning with augmented class by exploiting unlabeled data. In: *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, Québec City, 2014. 1760–1766
- 9 Mu X, Zhu F D, Du J, et al. Streaming classification with emerging new class by class matrix sketching. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 2373–2379
- 10 Zhu Y, Ting K M, Zhou Z H. Multi-label learning with emerging new labels. In: *Proceedings of the 16th International Conference on Data Mining*, Barcelona, 2016. 1371–1376
- 11 Zhu Y, Ting K M, Zhou Z H. Discover multiple novel labels in multi-instance multi-label learning. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 2977–2984



- 12 Li N, Tsang I W, Zhou Z H. Efficient optimization of performance measures by classifier adaptation. *IEEE Trans Pattern Analy Mach Intell*, 2013, 35: 1370–1382
- 13 Yang Y, Zhan D C, Fan Y, et al. Deep learning for fixed model reuse. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 2831–2837
- 14 Zhao P, Jiang Y, Zhou Z H. Multi-view matrix completion for clustering with side information. In: *Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Jeju, 2017. 403–415
- 15 Dai W Z, Zhou Z H. Combining logical abduction and statistical induction: discovering written primitives with human knowledge. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, San Francisco, 2017. 4392–4398
- 16 Li F F, Fergus R, Perona P. One-shot learning of object categories. *IEEE Trans Pattern Analy Mach Intell*, 2006, 28: 594–611
- 17 Palatucci M, Pomerleau D, Hinton G E, et al. Zero-shot learning with semantic output codes. In: *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, Vancouver, 2009. 1410–1418
- 18 Lampert C H, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization. *IEEE Trans Pattern Analy Mach Intell*, 2014, 36: 453–465
- 19 Mei S, Zhu J, Zhu J. Robust regbayes: selectively incorporating first-order logic domain knowledge into bayesian models. In: *Proceedings of the 31st International Conference on Machine Learning*, Beijing, 2014. 253–261
- 20 Fu Y W, Sigal L. Semi-supervised vocabulary-informed learning. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016. 5337–5346
- 21 Radivojac P, Clark W T, Oron T R, et al. A large-scale evaluation of computational protein function prediction. *Nat Methods*, 2013, 10: 221–227
- 22 The UniProt Consortium. Uniprot: a hub for protein information. *Nucleic Acids Res*, 2014, 43: D204
- 23 Ashburner M, Ball C A, Blake J A, et al. Gene ontology: tool for the unification of biology. *Nat Genet*, 2000, 25: 25–29
- 24 Ofer D, Linial M. Profet: feature engineering captures high-level protein functions. *Bioinformatics*, 2015, 31: 3429–3436
- 25 Zhou Z H, Zhang M L. Multi-instance multi-label learning with application to scene classification. In: *Proceedings of the 19th International Conference on Neural Information Processing Systems*. Cambridge: MIT Press, 2006. 1609–1616
- 26 Chua T S, Tang J H, Hong R C, et al. Nus-wide: a real-world web image database from national university of singapore. In: *Proceedings of the ACM International Conference on Image and Video Retrieval*, Santorini, 2009
- 27 Lin T Y, Maire M, Belongie S J, et al. Microsoft COCO: common objects in context. In: *Proceedings of the 13th European Conference on Computer Vision*, Zürich, 2014. 740–755
- 28 Gao B B, Xing C, Xie C W, et al. Deep label distribution learning with label ambiguity. *IEEE Trans Image Process*, 2017, 26: 2825–2838
- 29 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *Comput Vision Recogn*, 2014. arXiv:1409.1556
- 30 Pennington J, Socher R, Manning C D. Glove: global vectors for word representation. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, Doha, 2014. 1532–1543

## Model reuse with domain knowledge

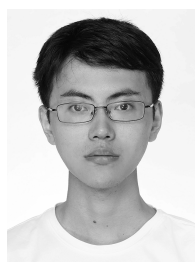
Xizhu WU & Zhihua ZHOU\*

*National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*

\* Corresponding author. E-mail: zhouzh@lamda.nju.edu.cn

**Abstract** The life spans of machine learning models are often short and a large number of models are wasted because they can only be applied to a specific task. However, a well-designed, carefully trained model contains learned knowledge from its task, which may be more concise than training data. Furthermore, when we have no access to training data, the trained model is the last remaining source of information. This study introduces a framework to reuse existing models trained in other tasks and help improve the model for the current task, especially when limited data is available for the current task. This framework incorporates high-level domain knowledge to combine existing models and treat them as black boxes, in order for them to be universal for complex models. Experiments on applying the framework to practical problems demonstrate that we can improve the performance on the current task by reusing existing models.

**Keywords** machine learning, model reuse, domain knowledge, environment change, learnware



**Xi-Zhu WU** was born in 1991. He received his BSc degree in Computer Science in Kuangyaming Honor School of Nanjing University in 2014. He was admitted to study for a MSc degree at Nanjing University without entrance examination in 2014 and transferred for a Ph.D. degree in 2016. His research interests include machine learning and data mining.



**Zhi-Hua ZHOU** was born in 1973. He received his Ph.D. degree in Computer Science from Nanjing University, China, in 2000. He is currently a professor at Nanjing University. His research interests include artificial intelligence, machine learning, and data mining.