

范畴表示机器学习算法

徐晓祥 李凡长 张 莉 张 召
(苏州大学计算机科学与技术学院 江苏苏州 215006)
(苏州大学机器学习与类脑计算国际合作联合实验室 江苏苏州 215006)
(20154027004@stu.suda.edu.cn)

The Category Representation of Machine Learning Algorithm

Xu Xiaoxiang, Li Fanzhang, Zhang Li, and Zhang Zhao
(School of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006)
(Joint International Research Laboratory of Machine Learning and Neuromorphic Computing, Soochow University, Suzhou, Jiangsu 215006)

Abstract For a long time, it is thought that the representation is one of the bottleneck problems in the field of machine learning. The performance of machine learning methods is heavily dependent on the choice of data representation. The rapidly developing field of representation learning is concerned with questions surrounding how we can best learn meaningful and useful representations of data. We take a broad view of the field and include topics such as deep learning, feature learning, metric learning, compositional modeling, structured prediction, reinforcement learning, etc. The range of domains to which these techniques apply is also very broad, from vision to speech recognition, text understanding, etc. Thus, the research on new representation methods for machine learning is a piece of work which is long-term, explorative and meaningful. Based on this, we propose several basic concepts of category representation of machine learning methods via the category theory. We analyze the decision tree, support vector machine, principal component analysis and deep neural network with category representation and give the corresponding category representation for each algorithms; the category representation of decision tree, slice category representation of support vector machine, and functor representation of the neural network. We also give the corresponding theoretical proof and feasibility analysis. According to further reach of category representation of machine learning algorithms, we find the essential relationship between support vector machine and principal component analysis. Finally, we confirm the feasibility of the category representation method using the simulation experiments.

Key words category representation; machine learning; representation of machine learning; category representation learning; category representation learning algorithm

摘 要 长期以来,人们认为表示问题是机器学习领域的瓶颈问题之一.机器学习方法的性能在很大程度上依赖于数据表示的选择.数据表示领域的主要问题是更好地学习到有意义和有用的数据表示.

收稿日期:2016-05-24;修回日期:2017-04-07
基金项目:国家自然科学基金项目(61373093,61402310,61672364,61672365);苏州大学东吴学者计划项目;江苏省普通高校研究生科研创新计划项目
This work was supported by the National Natural Science Foundation of China (61373093, 61402310, 61672364, 61672365), the Soochow Scholar Program of Soochow University, and the Graduate Innovation and Practice Program of Colleges and Universities in Jiangsu Province.
通信作者:李凡长(lfz@suda.edu.cn)

宽泛来看数据表示领域有深度学习、特征学习、度量学习、成分建模、结构化预测和强化学习等. 这些技术应用的范围也非常广泛, 包括图像、语音识别和文字理解等. 因此, 研究机器学习表示方法是一件长期且具有探索意义的工作. 基于此, 利用范畴理论来研究机器学习方法的表示, 提出了范畴表示机器学习方法的基本概念. 对决策树、支持向量机、深度神经网络等方法进行研究分析, 提出了范畴表示分类算法、范畴表示决策树算法、切片范畴表示主成分分析和支持向量机算法、范畴函子表示深度学习方法, 给出相应的理论证明及可行性分析. 并对这 5 种算法做了深入分析, 找到了主成分分析和支持向量机之间的本质联系, 最后通过仿真实验论证范畴表示方法的可行性.

关键词 范畴表示; 机器学习; 机器学习表示; 范畴表示学习; 范畴表示学习算法

中图法分类号 TP391

众所周知, 机器学习方法的性能很大程度上依赖于数据的表示. 表示作为机器学习基本问题之一, 一直倍受研究者的关注. 表示学习领域的发展速度决定更好的数据表示出现的时间. 表示具有广泛的应用领域, 主要涉及深度学习^[1-3]、特征学习^[4-5]、度量学习^[6]、核学习^[7]、成分模型^[8-9]、非线性结构预测^[10-11]、非凸优化问题^[12-13]等. 在 2013—2015 年举办的学习表示国际会议^[14-23]上出现了许多为解决这一问题的思路与方法. 这些方法仅是从某类学习方法出发探讨表示问题, 而没有给出具有普遍意义的表示方式. 如文献[1]从深度学习角度研究文本识别, 文献[7]用基于高斯核密度的贝叶斯网络做分类等. 并不是从整体普遍的角度来考虑, 就分类而言, 我们考虑所有的分类算法, 如何寻找这些算法的相同点、相异点; 寻找到了又如何表达, 又如何利用这些相同点、相异点来提高算法效率或者寻找新的算法. 为此本文利用范畴理论来研究机器学习方法的一般表示, 将现有的以及可能有的机器学习方法包含进范畴表示学习的框架里来给出理论证明, 并在文中对几种机器学习方法做了比较详细的研究分析.

综上所述, 本文用范畴理论来描述机器学习方法的表示问题有两大优势: 1) 机器学习表示问题可以转换成输入输出以及它们的关系问题来解决, 这符合了范畴的基本思想; 2) 范畴理论是一个强大且富有兼容性的数学理论, 可以将我们现有的以及将有的机器学习方法都用范畴表示, 从而进一步利用范畴良好的数学性质来建立机器学习表示理论体系.

1 机器学习方法范畴表示定理

为了便于理解范畴表示学习理论框架, 在此给出范畴的基本概念.

定义 1. 范畴(category) $\mathcal{C}$. 由 2 部分组成:

1) 一族对象(object) obj , 包括样例和实例;

2) 任意一对对象 A, B , 对应映射的集合 $\mathcal{C}(A, B)$. 其元素称为态射(morphism), 使得当 $A \neq A'$ 或者 $B \neq B'$ 时, $\mathcal{C}(A, B)$ 与 $\mathcal{C}(A', B')$ 不交. 并满足下面条件:

① 复合运算律(composition law). 若 $f \in \mathcal{C}(A, B), g \in \mathcal{C}(B, C), A, B, C \in obj$, 则存在唯一的 $gf \in \mathcal{C}(A, C)$, 称为 f 与 g 的复合.

② 结合律(associativity). 若 $f \in \mathcal{C}(A, B), g \in \mathcal{C}(B, C), h \in \mathcal{C}(C, D), A, B, C, D \in obj$, 则有 $(hg)f = h(gf)$.

③ 单位态射(identity morphism). 每一个对象 A 存在一个态射 $I_A \in \mathcal{C}(A, A)$ 使得对任意的 $f \in \mathcal{C}(A, B)$ 及 $g \in \mathcal{C}(C, A)$ 有 $fI_A = f$ 和 $I_{Ag} = g$.

例 1. 以集合 $A = \{1, 2, 3\}, B = \{a, b, c\}$ 为对象, 则映射共有 $3^3 = 27$ 个构成集合 $\mathcal{C}(A, B)$. 在此对于每一个映射就是一个态射. 更一般地给出一些常见范畴的对象和态射. 如表 1 所示:

Table 1 Example of Category

表 1 范畴实例

Category	Object	Morphism
Set	Set	Function
Top	Top Space	Continuous Function
Vector	Vector Space	Linear Transformation
Group	Group	Group Homomorphism

根据机器学习方法的一般定义: $F: X \rightarrow Y$, 其中, F 表示机器学习方法, X 表示样例集合, Y 表示实例集合. 结合范畴的基本概念, 下面给出机器学习方法范畴(category of machine learning algorithm, CMLA)的定义.

定义 2. 机器学习方法范畴 $\mathcal{C}$. 由下面 2 部分组成:

- 1) 对象族 obj ;
- 2) 任意一对对象 A, B 对应映射的集合 $\dot{C}(A, B)$. 其元素称为态射,使得当 $A=A'$ 或者 $B=B'$ 时, $\dot{C}(A, B)$ 与 $\dot{C}(A', B')$ 不交.

说明:1)这里的对象对应学习问题中的各种训练集、测试集、属性和项的集合,所有的对象组成对象族 obj ;2)态射是2个对象之间保持特性的过程的一种抽象,这里对应样例集合到实例集合之间的过程抽象.对于不同算法之间的区别在于样例集合、实例集合不同或为空,以及过程抽象不同或者过程选择不同.从而我们将单个学习问题中的各种输入样例集合表示为 $A:A \in obj$,输出实例集合表示为 $B:B \in obj$ 和其他看成一个或几个对象,并将所有学习问题的所有对象组成对象族;将单个学习问题中的输入样例输出特性之间保持特性的过程抽象为态射,从而所有的态射就构成了对象对应映射的集合 $\dot{C}(A, B)$,这就是一个学习状态.

定理 1. 机器学习方法范畴表示定理(category representation of machine learning algorithm, CRMLA):机器学习方法构成一个范畴.即机器学习方法可用范畴表示.

证明. 根据定义 2 描述的机器学习方法范畴的构成可知,每一个输入样例都有对应的输出特性,即样例集与特性集的映射为满射.整个机器学习方法系统的各种训练集、测试集、属性和项的集合构成对象族 obj .在 obj 中,任何一对对象 A, B ,我们都可以找到一个集合 $\dot{C}(A, B)$,使它们形成一个态射.比如对象 A 定义为一个算法的输入集合,对象 B 定义为一个算法的输出集合,在任何系统中都存在一个从输入到输出的一条路径,这条路径我们可以用 $\dot{C}(A, B)$ 表示,即从对象 A 到对象 B 的一个态射.

同时在单个机器学习方法中,输入对象 A ,经过一个或多个模块的处理即多个态射的运算,可以得到输出对象 B ,在此过程中满足每对不同对象之间的态射是不相交的,即 $A=A'$ 或者 $B=B'$ 时, $\dot{C}(A, B)$ 与 $\dot{C}(A', B')$ 不交,满足范畴的定义.同时容易验证存在单位态射,由定义 2 可知在算法中单个态射为满射,从而满足复合运算率以及结合律.综上所述,机器学习方法构成一个范畴. 证毕.

2 范畴表示分类算法

已知集合: $C = \{y_1, y_2, \dots, y_n\}$ 和 $I = \{x_1,$

$x_2, \dots, x_m, \dots\}$,确定映射规则 $y = f(x)$,使得任意 $x_i \in I$ 有且仅有一个 $y_i \in C$ 使得 $y_i = f(x_i)$ 成立(这里不考虑模糊数学里的模糊集情况).其中, C 叫作类别集合,其中每一个元素是一个类别;而 I 叫作项集合,其中每一个元素是一个待分类项; f 叫作分类器.分类算法的任务就是构造分类器 f .这就是分类算法的基本思路.

- 定义 3.** 分类算法范畴.由 2 部分组成:
- 1) 对象族 $obj = \{C', I, \dots\}$. 其中 C' 为对应的 x_i 的集合(在实际操作中一般 $C=C'$).
- 2) 任意一对对象 A, B 对应映射的集合 $\dot{C}(A, B)$. 其元素称为态射 $f:C' \rightarrow I$,使得当 $A=A'$ 或者 $B=B'$ 时, $\dot{C}(A, B)$ 与 $\dot{C}(A', B')$ 不交.

说明:1)这里的对象是分类算法中的各种类别集合与项集合,所有的对象组成对象族;2)态射是对应的分类构成的集合.

证明. 显然存在单位态射,由定义 3 可知对于任意 $x_i \in I$ 有且仅有一个 $y_i \in C'$ 从而可知:1)每对不同对象之间的态射是不相交的;2)所有态射都为满射,从而满足复合运算率以及结合律. 证毕.

- 算法 1.** 范畴表示朴素贝叶斯分类算法.
- 输入:训练数据 $\mathbf{X} = (x_1, x_2, \dots, x_n)$ 、类别信息 $\mathbf{Y} = (y_1, y_2, \dots, y_m)$ 、测试数据 $\mathbf{x}$;
- 输出:类别信息 y .

步骤 1. 计算每个类别条件下各个特征属性划分的频率: $P(x_1 | y_1), P(x_2 | y_1), \dots, P(x_n | y_1); P(x_1 | y_2), P(x_2 | y_2), \dots, P(x_n | y_2); \dots; P(x_1 | y_m), P(x_2 | y_m), \dots, P(x_n | y_m)$;

步骤 2. 计算训练样本中每个类别的频率:

$$P(x | y_i)P(y_i) = P(x_1 | y_i)P(x_2 | y_i) \cdots P(x_n | y_i)P(y_i);$$

步骤 3. $y = y_{i: \max(P(x | y_i)P(y_i))}$.
结合分类算法范畴以及朴素贝叶斯分类算法,我们不难发现,设 $f = y_{i: \max(P(x | y_i)P(y_i))}$,那么定义朴素贝叶斯算法范畴.在朴素贝叶斯算法范畴中,对象为训练数据 $\mathbf{X}$ 、类别信息 $\mathbf{Y}$,那么 $f:\mathbf{X} \rightarrow \mathbf{Y}$ 就是朴素贝叶斯算法范畴的一个态射.

- 算法 2.** 范畴表示决策树算法.
- 输入: $E = D_1 \times D_2 \times \dots \times D_m$ 是 m 维有穷向量空间,其中 D_i 是有穷离散符号集, E 中的元素 $e = (v_1, v_2, \dots, v_m)$ 叫作例子,其中 $v_j \in D_j, j = 1, 2, \dots, m$. 设 PE 和 NE 是 E 的 2 个子集,分别叫作正例集和反例集,向量空间 E 中的 PE 和 NE 的大小分别定义为 p 和 n ;

输出:决策树 T .

步骤 1. 节点 S 信息熵: $Entrop(S)=-p/(p+n)\times\lg((p/(p+n))-n/(p+n)\times\lg(n/(p+n))$;

步骤 2. 以属性 A 作为决策树的根所需的期望信息:

$$Entrop(S,A)=\sum_{i=1}^m(p_i+n_i)/(p+n)\times Entrop(S_i),$$

其中 p_i 和 n_i 分别是 E_i 的正实例和负实例的值;

步骤 3. 以 A 为根的信息增益:

$$Gain(S,A)=Entrop(S)-Entrop(S_i,A);$$

步骤 4. 选择使 $Gain(S,A)$ 最大的属性 A^* , 即 $Entrop(S_i,A)$ 最小, 作为根节点;

步骤 5. 对 A^* 的不同取值对应的 E 的 k 个子集 E_i , 递归调用步骤 1~4 过程生成 A^* 的子节点 $B_1, B_2, \cdots, B_k$, 最终生成决策树.

对比算法 1 与算法 2, 我们不难发现, 最终生成的决策树是决策树算法范畴中的一个态射, 并且定义也类似, 在此就不进行赘述. 但是我们发现: 如果在解决同一问题时, 那么分别由朴素贝叶斯算法生成的态射和决策树方法生成的态射都属于同一训练数据到类别信息映射集合的一个元素, 但是明显可以发现决策树算法的态射比朴素贝叶斯分类算法的态射要复杂得多. 这就体现了范畴表示算法的多样性.

算法 3. 范畴表示支持向量机算法.

输入: 训练数据 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m)$ 、类别信息 $\mathbf{Y}=(y_1, y_2, \cdots, y_n)$ 、测试数据 $\mathbf{x}$;

输出: 决策函数 g .

在此仅从宏观的分类算法角度来研究支持向量机算法的范畴表示, 第 3 节还会从本质核心的角度来研究, 所以仅展示算法的输入输出, 具体过程在第 3 节阐述.

之前已经讲述了 2 个算法实例, 举这 2 个算法是为了说明范畴表示学习方法的多样性. 在此给出第 3 个算法实例的原因是与第 3 节用切片范畴表示支持向量机算法做对比, 更进一步说明范畴表示学习方法的多样性. 由于已经给出了 2 个算法实例, 在此用一句话简略说明支持向量机算法范畴: 以训练数据 $\mathbf{X}$ 和类别信息 $\mathbf{Y}$ 作为对象, 以决策函数 $g:\mathbf{X}\rightarrow\mathbf{Y}$ 作为其态射的范畴. 纵观这 3 个算法实例, 我们可以发现范畴表示学习方法十分简洁精炼, 但又可以说明问题.

3 切片范畴表示 PCA&SVM 算法

为了说明如何用切片范畴表示主成分分析(principal component analysis, PCA)和支持向量机(support vector machine, SVM)算法, 在此给出切片范畴的基本概念.

定义 4. 切片范畴(slice category). 设 B 为范畴 $\mathcal{C}$ 中的一个对象, 考虑范畴中以 B 为值域的态射 $\varphi:T\rightarrow B$, 记作 (T, φ) . 则可以构造一个以所有 (T, φ) 为对象的新范畴, 任意 2 个以 B 为值域的态射 $f:A\rightarrow B$ 和 $g:C\rightarrow B$, 构造态射 $h:(A, f)\rightarrow(C, g)$ 是指 $h:A\rightarrow C$ 满足 $f=gh$. 如图 1 所示:

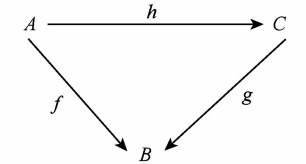


Fig. 1 Slice category over B
图 1 B 上切片范畴

称这样得到的范畴就为 B 上的切片范畴(slice category over B), 记作 $\mathcal{C}/B$. 对偶可以得到 B 下的余切片范畴(coslice category under B), 记作 $B/\mathcal{C}$.

例 2. 对象 $A=\{a, b, c\}, B=\{1, 2, 3\}, C=\{\blacksquare, \bullet, \blacktriangle\}$ 对应态射和切片范畴如图 2 所示:

$f:$	a	b	c
	1	2	3

morphism $g: B\rightarrow C$:

$g:$	$\blacksquare$	$\bullet$	$\blacktriangle$
	1	2	3

Slice Category over B $h:A\rightarrow C$:

$h:$	a	b	c
	$\blacksquare$	$\bullet$	$\blacktriangle$

Coslice Category under B $h':C\rightarrow A$:

$h':$	$\blacksquare$	$\bullet$	$\blacktriangle$
	a	b	c

Fig. 2 Morphism and slice category
图 2 态射和切片范畴

主成分分析是一种统计分析方法, 它可以从多元事物中解析出主要影响因素, 主成分分析的目的在于将高维数据投影到较低维空间.

算法 4. 主成分分析.

输入:原始数据—— n 行 m 列矩阵 $\mathbf{X}$;

输出:降维数据—— k 行 m 列矩阵 $\mathbf{Z}$.

步骤 1. 将 $\mathbf{X}$ 的每一行(代表一个属性字段)进行零均值化,即减去这一行的均值 $\bar{\mathbf{X}}$;

步骤 2. 求出协方差矩阵 $\mathbf{W}=1/m\mathbf{X}'\mathbf{X}'^T$, $\mathbf{X}'=\mathbf{X}-\bar{\mathbf{X}}$ 及协方差矩阵的特征 λ 值和对应的特征向量;

步骤 3. 将特征向量按对应特征值大小从上到下按行排列成矩阵 $\mathbf{W}'$,取前 k 行组成矩阵 $\mathbf{P}$, $\mathbf{Z}=\mathbf{P}\mathbf{X}$ 即为降维到 k 维后的数据.

设数据原始空间及其内部关系构成范畴 $\dot{N}$,不妨设 $f:\mathbf{X}\rightarrow\mathbf{O}$ 为范畴 $\dot{N}$ 中的一个态射, $\mathbf{O}$ 为数据本质. 在数据的低维空间中其以及其内部关系构成范畴 $\dot{U}$,不妨设 $g:\mathbf{Z}\rightarrow\mathbf{O}$ 为范畴 $\dot{U}$ 中的一个态射,其中 $\mathbf{Z}$ 为低维数据. 构造 $\mathbf{O}$ 上的切片范畴,不妨设为 $\dot{S}$,有 $h:\mathbf{X}\rightarrow\mathbf{Z}$,满足 $f=hg$. 其中的 f ,我们在大多数情况下不知道或者在可计算时间内无法求解,但是可以通过构造切片范畴找到合适的 $\mathbf{Z}$,满足在 $\mathbf{Z}$ 上比 $\mathbf{X}$ 上更容易更快速地寻找到理想的 g . 在此虽然我们不知道 f 和 g 的具体表达式,但是我们可以给出 $\mathbf{Z}$ 上的切片范畴 $\dot{S}$ 的态射表达式 $h=\mathbf{P}\mathbf{X}$. 此时在 $\mathbf{Z}$ 上,数据就呈现出一种以尽可能少的维度表示最多的数据本质信息.

支持向量机是一个有监督的学习模型,通常用来进行模式识别、分类以及回归分析.

算法 5. 支持向量机.

输入:训练数据 $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n)$ 、类别信息 $\mathbf{Y}=(y_1, y_2, \cdots, y_m)$ 、测试数据 $\mathbf{x}$;

输出:决策函数 $g=\mathbf{w}\mathbf{x}+b$.

步骤 1. 原规划 $\max 1/\|\mathbf{w}\|$,使得 $y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1$,其中 $i=1,2,\cdots,n$,即 $\min \|\mathbf{w}\|^2/2$ 使得 $y_i(\mathbf{w}^T\mathbf{x}_i+b)\geq 1$,其中 $i=1,2,\cdots,n$;

步骤 2. 拉格朗日对偶 $L(\mathbf{w},b,\mathbf{a})=\|\mathbf{w}\|^2/2-\sum_{i=1}^n a_i(y_i(\mathbf{w}^T\mathbf{x}_i+b)-1)$,其中 $\mathbf{a}=(a_1, a_2, \cdots, a_n)$ 为拉格朗日参数;

步骤 3. 由拉格朗日对偶求出 a_i ,从而:

$$\mathbf{w}=\sum_{i=1}^m a_i y_i \psi(\mathbf{x}_i),$$

$$b=-\left(\max_{i: y_i=-1} \mathbf{w}\psi(\mathbf{x}_i)+\min_{i: y_i=1} \mathbf{w}\psi(\mathbf{x}_i)\right)/2,$$

$$g=\sum_{i=1}^m a_i y_i < \psi(\mathbf{x}_i), \psi(\mathbf{x}) > +b,$$

其中, $\psi(\mathbf{x})$ 是核函数.

类似地,设数据原始空间及其内部关系构成范畴 $\dot{N}$,不妨设 $f:\mathbf{X}\rightarrow\mathbf{Y}$ 为范畴 $\dot{N}$ 中的一个态射, $\mathbf{Y}$ 为数据标签在此可类比数据本质. 在数据的高维甚至无限维空间中其以及其内部关系构成范畴 $\dot{U}$,不妨设 $g:\mathbf{H}\rightarrow\mathbf{Y}$ 为范畴 $\dot{U}$ 中的一个态射,其中 $\mathbf{H}$ 为高维甚至无限维数据. 构造 $\mathbf{H}$ 上的切片范畴,不妨设为 $\dot{S}$,有 $h:\mathbf{X}\rightarrow\mathbf{H}$,满足 $f=hg$. 其中的 f ,我们在大多数情况下不知道或者在可计算时间内无法求解,但是可以通过构造切片范畴找到合适的 $\mathbf{H}$,满足在 $\mathbf{H}$ 上比 $\mathbf{X}$ 上更容易更快速地寻找到理想的 g . 在此对应的 $\mathbf{H}$ 上的切片范畴 $\dot{S}$ 的态射表达式 $h=\phi(\mathbf{x})$. 结合第 2 节 SVM 算法的宏观范畴表示,我们可以发现对同一算法从不同范畴角度观察分析会得到不一样的表达但是它们都是同一个算法.

SVM 与 PCA 算法的本质核心都是将数据投影到其他空间,然后利用其他空间的特性来进行处理. 但是 SVM 与 PCA 不同的是:1)SVM 寻找的空间不是低维空间而是高维甚至无限维空间,无法轻易直观地对数据进行表示;2)在 PCA 中 f 和 g 都是不知道具体表达式的,但在 SVM 中 g 是已表达的. 下面给出切片范畴表示 PCA&SVM 算法示意图,如图 3 所示:

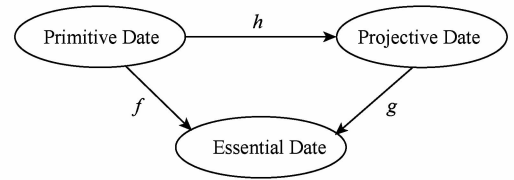


Fig. 3 Slice category representation of PCA&SVM

图 3 切片范畴表示 PCA&SVM 算法

4 范畴函子表示 Deep Learning 算法

为了说明如何用范畴函子表示 Deep Learning 算法,在此给出范畴函子的基本概念.

定义 5. 函子(functor). 设 $\dot{C}$ 和 $\dot{U}$ 是 2 个范畴. 一个函子 $F:\dot{C}\rightarrow\dot{U}$ 是由 2 个映射组成:

$$objn\rightarrow obju:A\mapsto F(A),$$

$$morn\rightarrow moru:f\mapsto F(f),$$

满足 $dom(F(f))=F(dom(f))$, $cod(F(f))=F(cod(f))$, $F(I_A)=I_{F(A)}$. 并且若 $dom(g)=cod(f)$,则 $F(gf)=F(g)F(f)$. 其中, dom 为定义域, cod 为值域.

Deep Learning 算法常见有 3 种类型:有监督深度学习、无监督深度学习和混合深度学习. 监督学习主要用于处理回归和分类问题;无监督学习主要用于处理聚类和密度估计问题;混合深度网络则综合了前 2 种方式. 为了呼应第 2 节,在此仅以有监督深度学习中的分类算法为例.

算法 6. 深度学习算法.

输入:训练数据 $\mathbf{X}=\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_n\}$ 、标签数据 $\mathbf{Y}=\{y_1,y_2,\cdots,y_m\}$ 、测试样本 $\mathbf{x}$ 、 k 层深度网络;
输出:样本标签 y .
步骤 1. 预训练,通过最小化损失函数:

$$J(\mathbf{W},b)=1/(2m)\times\sum_{i=1}^n\|h(\mathbf{x}_i)-\mathbf{x}_i\|^2+\gamma/2\times\|\mathbf{W}\|^2,$$

求得 $\mathbf{W},b$,其中 $\mathbf{x}_i$ 表示第 i 个样本, γ 为超参数,

$$h(\mathbf{x})=f(\mathbf{W}^T\mathbf{x}+b)=f(\sum_{i=1}^n\mathbf{W}_i^T\mathbf{x}_i+b);$$

步骤 2. 通过最小化交叉熵函数,对 $\mathbf{W},b$ 进行微调 $J(\mathbf{W})=-1/m\times\sum_{i=1}^n\sum_{j=1}^kbl(y_i=j)\times\ln(e^{\mathbf{w}_i^T\mathbf{x}_i}/\sum_{j=1}^ke^{\mathbf{w}_j^T\mathbf{x}_i})$,其中 $bl(y_i=j)=\begin{cases}1, & y_i=j, \\ 0, & y_i\neq j;\end{cases}$

步骤 3. 求出属于第 i 类的概率 $\mathbf{P}_i$:

$$\mathbf{P}_i=p(y=1|\mathbf{x},\mathbf{W})=e^{\mathbf{w}_i^T\mathbf{x}_i}/\sum_{j=1}^ke^{\mathbf{w}_j^T\mathbf{x}_j},$$

从而 $y=y_{i:\max P_i}$.

设 Deep Learning 算法(n 层)的每一层为一个范畴,记为 $\dot{C}_1,\dot{C}_2,\cdots,\dot{C}_n$. 第 i 层,其对应对象族为 $objd_i$ 关系为 $f_i:subs(objd_i)\rightarrow subs(objd_i)$,其中 $subs$ 为子集. 从而算法的目标就是构建函子串,使得在兼顾鲁棒性的情况下提升精确度. 函子串定义如下:第 i 层向第 $i+1$ 层的函子 $F_i:\dot{C}_i\rightarrow\dot{C}_{i+1}$ 对应 2 个映射为 $objd_i\rightarrow objd_{i+1}:A\mapsto F_i(A);mord_i\rightarrow mord_{i+1}:f\mapsto F_i(f)$. 从而我们就得到了 Deep Learning 算法的范畴函子表示.

5 实验仿真

本文进行实验仿真的目的是为了验证范畴表示机器学习方法的可行性,第 2~4 节中我们对决策树算法、支持向量机算法、深度学习方法进行了理论阐述以及证明. 在本节主要针对这 3 个算法进行可行性验证.

5.1 决策树算法实验仿真

在第 2 节中我们将决策树算法进行了范畴表示,下面就从范畴表示角度进行计算验证算法的可行性. 表 2 为 10 位志愿者的房产、婚姻、收入、债务情况. 我们将以此为例,通过房产、婚姻、收入来预测能否还清债务,并预测第 11 位志愿者(无房产、单身、收入 5.5 万元)的债务情况.

Table 2 Situation of Volunteers

表 2 志愿者情况

ID	Property (Y/N)	Marriage (Single, Married,Divorce)	Income (Unit: 10 000 CNY)	Debt (Y/N)
1	Y	Single	12.5	N
2	N	Married	10	N
3	N	Single	7	N
4	Y	Married	12	N
5	N	Divorce	9.5	Y
6	N	Married	6	N
7	Y	Divorce	22	N
8	N	Single	8.5	Y
9	N	Married	7.5	N
10	N	Single	9	Y

由第 2 节决策树范畴中的态射构造方法我们可以计算(不妨将收入分为大于等于 8 和小于 8 这 2 类):

$$Entrop(S)=-0.3\lg 0.3-0.7\lg 0.7=0.811.$$

$$Entrop(Property_Y)=0,$$

$$Entrop(Property_N)=0.985.$$

$$Entrop(Marriage_{Single})=1,$$

$$Entrop(Marriage_{Married})=0,$$

$$Entrop(Marriage_{Divorce})=1.$$

$$Entrop(Income_{\geq 8})=0.985,$$

$$Entrop(Income_{<8})=0.$$

$$Entrop(S,Property)=$$

$$0.3\times 0+0.7\times 0.985=0.69,$$

$$Entrop(S,Marriage)=$$

$$0.4\times 0+0.4\times 0+0.2\times 1=0.6,$$

$$Entrop(S,Income)=0.7\times 0.985+0.3\times 0=0.69.$$

$$Gain(S,Property)=0.811-0.69=0.121,$$

$$Gain(S,Marriage)=0.811-0.6=0.211,$$

$$Gain(S,Income)=0.811-0.69=0.121.$$

从而我们的根节点就是婚姻,重复此过程得到

其他分支节点及叶子节点,得到最终态射 f 如图 4 所示:

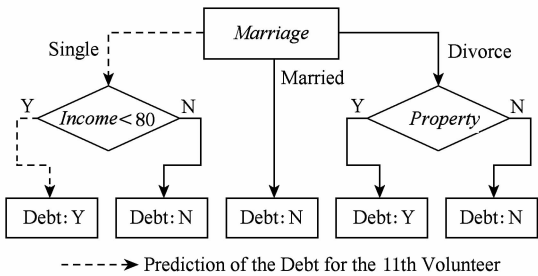


Fig. 4 The morphism f
图 4 态射 f

从而我们的预测第 11 位志愿者的债务情况为 Y(图 4 中虚线箭头)。

5.2 SVM 算法实验仿真

在第 3 节中我们将 SVM 算法进行了范畴表示,下面就从范畴表示角度进行计算验证算法的可行性.实验从人工数据集和 ORL 人脸数据集上进行验证.

人工数据集(采用线性核)采用服从高斯分布的 2 组点,分别以 $(0,0)$ $(2.5,2.5)$ 为中心且标准差为 1 的 2 组点,每组 100 个,进行 10 次实验取平均准确率为 96.57%.其中单次实验效果如图 5 所示:

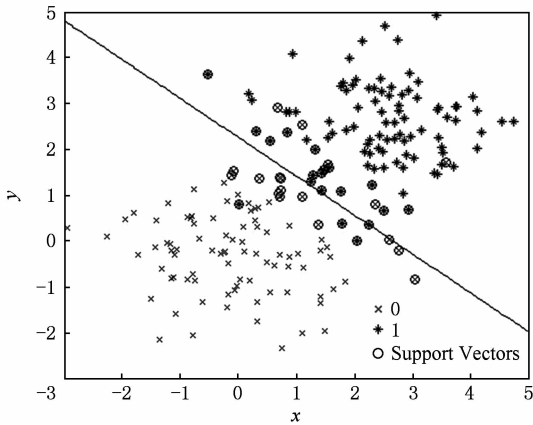


Fig. 5 Experimental results on artificial data sets
图 5 人工数据集实验效果

ORL 人脸数据集(采用高斯核)包含 400 幅人脸图像 $(40 \text{ 人} \times 10)$,其中每幅图像大小为 112×92 .其中一半测试一半训练,高斯核参数选取 $(0.01, 128)$.多分类构建 $39 \times 40/2$ 个分类器投票决定类别,10 次实验平均准确率为 93.68%.

由这 2 组实验不仅验证范畴表示机器学习方法的可行性,同时第 1 组实验又构成了第 2 节范畴表示分类算法的一个例子,第 1 组实验验证了切片范畴表示 SVM 算法的可行性.结合第 1,2 组实验发

现 SVM 可用范畴理论来表示体现了范畴表示机器学习方法的多样性.

5.3 神经网络算法实验仿真

MNIST 数据集^[24]是由 0~9 的手写体数字组成(即 10 类数据),包含了 60 000 张训练图像和 10 000 张测试图像,每张图像为 28×28 的灰度图,图 6 给出了 MNIST 数据集的一些示例样本.图 7 为在手写体数据集上 196 个隐单元的可视化效果.

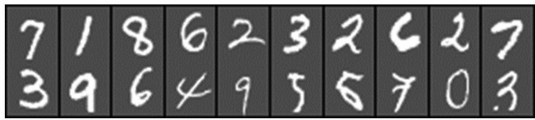


Fig. 6 Sample for handwritten data set
图 6 手写体数据集的示例样本

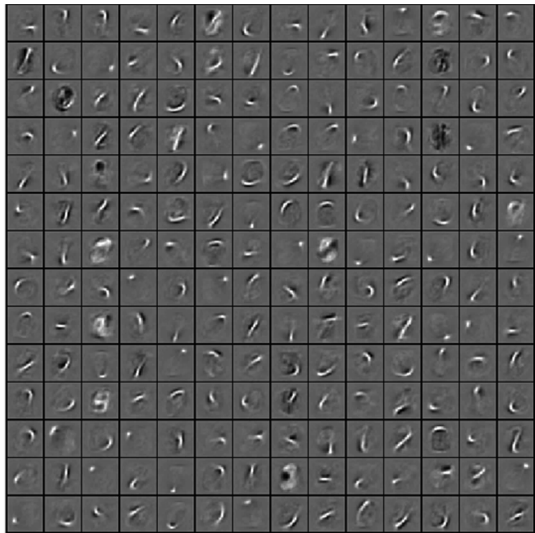


Fig. 7 Visualization of 196 hidden units on handwritten data set

图 7 在手写体数据集上 196 个隐单元的可视化效果

COIL 数据集^[25]包含的样本为日常生活中的一些瓶子、罐子以及盒子之类的物体.该数据集有 7 200 张灰度图像,共 100 类数据,每个类别 72 张图像,每张图像的大小为 32×32 .图 8 给出了该数据集的一些示例样本,有 100 个样本,分别属于不同的类别.图 9 COIL 数据集上训练所得 196 个隐单元的可视化效果

我们从每个类别中随机地挑选出 50 张图像,共计 5 000 张图像作为训练集,剩余的 1 200 张图像作为测试集.实验中,MNIST 和 COIL 数据均被归一化到了 $[0,1]$ 之间.

本次实验在 2 个数据集上分别采用深度学习方法,网络的隐单元数均被设为 196. MNIST 训练

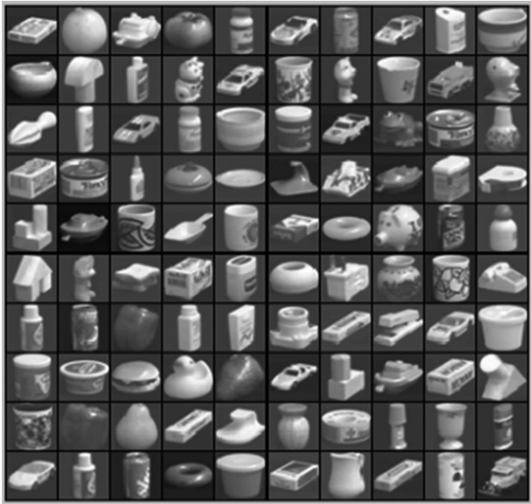


Fig. 8 Sample for COIL data set
图 8 COIL 数据集的示例样本

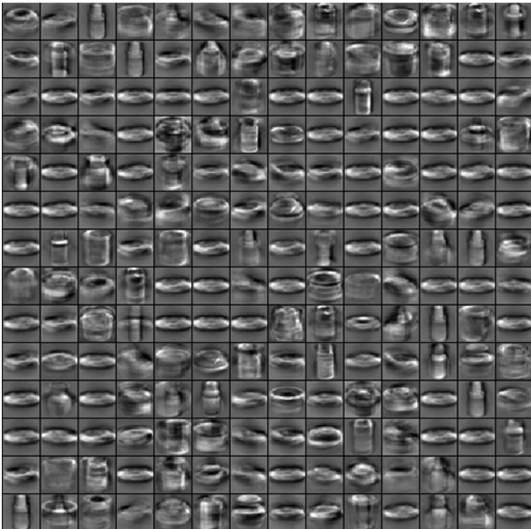


Fig. 9 Visualization of 196 hidden units on COIL data set
图 9 COIL 数据集上训练所得 196 个隐单元的可视化效果

样本被随机划分成 600 批, 每批 100 个样本; COIL 训练样本被随机划分成 50 批, 每批 100 个样本 (每一类包含一个数据). 最后通过 softmax 分类器对测试数据进行分类 10 次实验求平均准确率: 在 MNIST 数据集上为 93.03%, 在 COIL 数据集上为 98.08%. 这不仅验证了范畴表示方法的可行性, 同时由可视化效果(图 7 和图 9)可以发现范畴表示方法十分符合人类认知习惯.

6 结束语

本文提出了一种机器学习表示新方法, 即范畴表示机器学习方法, 并且结合多种机器学习方法来

实例验证了机器学习方法表示针对统一性问题的一种可能解法. 这种方法包容了所有的机器学习方法, 并能对部分方法的本质原理给出依据. 因此, 本文的创新点主要体现在:

- 1) 机器学习方法可用范畴表示. 不同算法可用同一范畴表示, 同一算法可用不同范畴理论表示. 这体现了范畴表示机器学习方法的统一性和多样性.
- 2) 范畴表示机器学习方法能够从表面以及本质上看清算法, 同时能够寻找符合人们的认知习惯的表示特征.

范畴表示学习刚刚开始, 还有许多工作要做. 例如范畴表示学习的拓扑问题、范畴表示学习的流形问题、范畴表示学习的优化问题等.

参 考 文 献

[1] Belharbi S, Chatelain C, Herault R, et al. Deep structured output learning for unconstrained text recognition [OL]. 2015 [2016-02-06]. <https://arxiv.org/pdf/1412.5903.pdf>

[2] Mao Junhua, Xu Wei, Yang Yi, et al. Deep captioning with multimodal recurrent neural networks [OL]. 2014[2016-02-06]. <https://arxiv.org/pdf/1412.6632v3.pdf>

[3] Goodfellow I J, Vinyals O, Saxe A M. Qualitatively characterizing neural network optimization problems [OL]. 2014 [2016-02-06]. <https://arxiv.org/pdf/1412.6544.pdf>

[4] Liu Tong, Si Yunjuan, Wen Dunwei, et al. Dictionary learning for VQ feature extraction in ECG beats classification [J]. Expert Systems with Applications, 2016, 53: 129-137

[5] Tao Chao, Pan Hongbo, Li Yansheng, et al. Unsupervised spectral-spatial feature learning with stacked sparse autoencoder for hyperspectral imagery classification [J]. IEEE Geoscience & Remote Sensing Letters, 2015, 12(12): 2438-2442

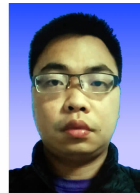
[6] Wang Wei, Hu Baogang, Wang Zengfu. Efficient and scalable information geometry metric learning [C] //Proc of the 13th IEEE Int Conf on Data Mining. Los Alamitos, CA: IEEE Computer Society, 2013: 1217-1222

[7] Wang Shuangcheng, Gao Rui, Wang Limin. Bayesian network classifiers based on Gaussian kernel density [J]. Expert Systems with Applications, 2016, 51(C): 207-217

[8] Crouzen P, Hermanns H. Aggregation ordering for massively compositional models [C] //Proc of the 10th Int Conf on Application of Concurrency to System Design. Los Alamitos, CA: IEEE Computer Society, 2010: 171-180

[9] Virtanen T, Gemmeke J F, Raj B, et al. Compositional models for audio processing: Uncovering the structure of sound mixtures [J]. IEEE Signal Processing Magazine, 2015, 32(2): 125-144

- [10] Michel O, Hero A. Tree structured non-linear signal modeling and prediction [J]. IEEE Trans on Signal Processing, 1995, 3(2): 267-270
- [11] Patel M, Shah H. Protein secondary structure prediction using support vector machines [C] //Proc of the 1st Int Conf on Machine Intelligence and Research Advancement. Piscataway, NJ: IEEE, 2013: 233-244
- [12] Parks D, Truszczyński M. Routing non-convex grids without holes [C] //Proc of the 1st Great Lakes Symp on IEEE. Piscataway, NJ: IEEE, 1991: 157-162
- [13] Ding Yin, Selesnick I W. Artifact-free wavelet denoising : Non-convex sparse regularization, convex optimization [J]. IEEE Signal Processing Letters, 2015, 22(9): 1364-1368
- [14] Bribiesca E, Bribiesca-Contreras G. 2D tree object representation via the slope chain code [J]. Pattern Recognition, 2014, 47(10): 3242-3253
- [15] Chen Lichi, Hsieh J W, Yan Yilin, et al. Vehicle make and model recognition using sparse representation and symmetrical SURFs [J]. Pattern Recognition, 2013, 48(6): 1143-1148
- [16] Fraz M, Edirisinghe E A, Sarfraz M S. Mid-level representation based lexicon for vehicle make and model recognition [C] //Proc of the 22nd Int Conf on Pattern Recognition. Piscataway, NJ: IEEE, 2014: 393-398
- [17] Liu Haifeng, Yang Zheng, Yang Ji, et al. Local coordinate concept factorization for image representation [J]. IEEE Trans on Neural Networks & Learning Systems, 2014, 25(25): 1071-1082
- [18] Healy M J. Category theory applied to neural modeling and graphical representations [C] //Proc of the 1st Int Joint Conf on Neural Networks. Piscataway, NJ: IEEE, 2000: 30-45
- [19] Jadoon W, Zhang Haixian. Locality features encoding in regularized linear representation learning for face recognition [C] //Proc of the 2nd Int Conf on Frontiers of Information Technology. Los Alamitos, CA: IEEE Computer Society, 2013: 189-194
- [20] Lei Hao, Mei Kuizhi, Zheng Nanning, et al. Learning group-based dictionaries for discriminative image representation [J]. Pattern Recognition, 2014, 47(2): 899-913
- [21] Lemus E, Bribiesca E, Garduño E. Representation of enclosing surfaces from simple voxelized objects by means of a chain code [J]. Pattern Recognition, 2014, 47(4): 1721-1730
- [22] Liang Yan, Lai Jianhuang, Yuen P C, et al. Face hallucination with imprecise-alignment using iterative sparse representation [J]. Pattern Recognition, 2014, 47(10): 3327-3342
- [23] Wang Lingfeng, Yan Hongping. Visual tracking via kernel sparse representation with multikernel fusion [J]. IEEE Trans on Circuits & Systems for Video Technology, 2014, 24(7): 1132-1141
- [24] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition [J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324
- [25] Nene S A, Nayar S K, Murase H. Columbia object image library (COIL-100), CUCS-006-96 [R]. New York: Department of Computer Science, Columbia University, 1996



Xu Xiaoxiang, born in 1991. Doctoral candidate. His main research interests include Lie group machine learning, category theory and homology theory.



Li Fanzhang, born in 1964. Professor and PhD supervisor. Senior member of CCF. His main research interests include Lie group machine learning and dynamic fuzzy logic.



Zhang Li, born in 1975. Professor and PhD supervisor. Senior member of CCF. His main research interests include pattern recognition, machine learning and image processing (zhangliml@suda.edu.cn).



Zhang Zhao, born in 1984. Professor and Master supervisor. Senior member of CCF. His main research interests include pattern recognition, machine learning, data mining and computer vision (cszzhang@suda.edu.cn).