智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

从随机集落影到随机点落影——隶属函数用于机器学习

汪培庄, 鲁晨光

引用本文:
汪培庄, 鲁晨光. 从随机集落影到随机点落影——隶属函数用于机器学习[J]. 智能系统学报, 2025, 20(2): 305-315.
WANG Peizhuang, LU Chenguang. From random set falling shadows to a random point falling shadow: membership functions for machine learning[J]. *CAAI Transactions on Intelligent Systems*, 2025, 20(2): 305-315.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202309028

DOI: 10.11992/tis.202309028
网络出版地址: https://link.cnki.net/urlid/23.1538.TP.20241223.1247.006

从随机集落影到随机点落影——隶属函数用于机器学习

汪培庄，鲁晨光
（辽宁工程技术大学 智能工程与数学研究院，辽宁 阜新 123000）

摘　要：从样本分布求得隶属函数是重要的也是困难的。汪培庄的随机集落影理论使用集值统计得到隶属函数，从而在统计和模糊逻辑之间架起桥梁。但是，通常的样本并不包含集值，所以该理论不够实用。鲁晨光使用语义信息方法推导出用样本分布优化隶属函数的 2 个公式，它们和集值统计结果一致，可谓随机点落影方法。该方法可以用于多标签分类、最大互信息分类、混合模型、贝叶斯确证等。深度学习最新潮流中用的相似函数和估计互信息就是隶属函数和语义互信息的特例。因为最大语义信息准则和最大似然准则以及正则化最小误差平方准则兼容，并且隶属函数比似然函数迁移性更好，比反概率函数更容易构造，隶属函数有希望被广泛用于机器学习。

关键词：模糊集合；隶属函数；样本分布；语义信息测度；机器学习；多标签分类；最大互信息分类；混合模型；贝叶斯确证
中图分类号：TP3; O21; O23　　**文献标志码**：A　　**文章编号**：1673−4785(2025)02−0305−11

中文引用格式：汪培庄，鲁晨光．从随机集落影到随机点落影——隶属函数用于机器学习 [J]．智能系统学报，2025，20(2)：305–315.
英文引用格式：WANG Peizhuang, LU Chenguang. From random set falling shadows to a random point falling shadow: membership functions for machine learning[J]. CAAI transactions on intelligent systems, 2025, 20(2): 305–315.

From random set falling shadows to a random point falling shadow: membership functions for machine learning

WANG Peizhuang, LU Chenguang
(Intelligence Engineering and Mathematics Institute, Liaoning Technical University, Fuxin 123000, China)

Abstract: Obtaining membership functions from sample distributions is essential and challenging. Wang Peizhuang's random set falling shadow theory uses set-valued statistics to derive membership functions, bridging the gap between statistics and fuzzy logic. However, traditional samples do not include set values, limiting the practical applicability of this theory. Lu Chenguang addressed this issue by using a semantic information method to derive two formulas for optimizing membership functions based on sample distributions. This method, known as the random point falling shadow method, is compatible with set-valued statistics. The resulting membership functions have applications in multilabel classification, maximum mutual information classification, mixed models, and Bayesian confirmation. Furthermore, the similarity function and estimated mutual information in modern deep learning techniques are special cases of the membership function and semantic mutual information. The maximum semantic information criterion is compatible with the maximum likelihood criterion, and the regularized least square error criterion, and the membership function is more transferable and easier to construct than likelihood functions or inverse probability functions. Thus, the membership function and the semantic information method hold considerable potential for widespread use in machine learning.

Keywords: fuzzy set; membership function; sampling distribution; semantic information measure; machine learning; multilabel classification; maximum mutual information classification; mixed model; Bayesian confirmation

学习函数是为了分类和聚类。隶属函数[1]和相似函数本该是最合适的学习函数，但是由于它们过去一直很难通过统计从样本分布得到，流行

收稿日期：2023−09−15.　**网络出版日期**：2024−12−23.
基金项目：国家自然科学基金重大项目 (9688007-1).
通信作者：鲁晨光. E-mail: lcguang@foxmail.com.

的学习函数是似然函数、反概率函数[2-3](比如 Logisitic 函数)。似然函数的主要缺点是：当先验概率分布改变后，以前学到的似然函数就不再适用。而 Logistic 函数作为学习函数在类别大于 2 时用于分类是困难的[4]，所以仍然需要用隶属函数和相似函数作为学习函数。

求解隶属函数有多种方法，包括：专家经验法、二元对比排序法[5]、模糊统计法以及拟合模糊分布法[6]。本文方法属模糊统计和拟合模糊分布法。根据样本分布构造隶属函数用于机器学习，已有很多研究[6-10]，但是本文希望提供一个更简洁且应用更广的方法。

本文作者之一汪培庄于 80 年代初提出隶属函数的统计解释——随机集落影理论[11-12]，它后来成为人工智能基础理论——因素空间理论[13]的重要部分。根据随机集落影理论，一个模糊集合的隶属函数可以通过一个随机集合的无数取值 (简称集值) 的统计得到。这一理论的重要意义是在模糊逻辑和统计之间架起桥梁。然而，通常的用于机器学习的样本并不包含集值，以至于集值统计方法很难推广。本文另一作者鲁晨光于 90 年代初使用基于统计的隶属函数建立语义信息测度[14-15]，后来把它发展为语义信息理论[16-19]，并用语义信息方法推导出从样本分布求解隶属函数的 2 个公式[19-20]。因为它们和集值统计公式兼容，语义信息方法可谓随机点落影方法。最近几年，鲁晨光把这样的隶属函数 (又叫真值函数) 用于机器学习的多个领域[19-20]。最近几年出现的以互信息神经估计 (mutual information neural estimation, MINE)[21] 和信息噪声对比估计 (information noise contrastive estimation, InfoNCE)[22] 为代表的深度学习新潮流也和语义信息方法相互支持。

本文将回顾从随机集落影到随机点落影的历史；简单介绍用样本分布优化隶属函数的语义信息方法以及隶属函数在机器学习领域的应用；讨论潜在应用和挑战。

1　背景知识

1.1　隶属函数、真值函数和相似函数之间的关系

令 X 是一随机变量，表示一个实例，取值 $x\in U=\{x_1, x_2, \cdots\}$；$Y$ 是一随机变量，表示一个标签或假设，取值 $y\in V=\{y_1, y_2, \cdots\}$。设 U 中使 y_j 为真的元素构成一个模糊子集 θ_j(即 y_j=“x 在 θ_j 中”)，则 x 在 θ_j 中的隶属度，记为 $m(\theta_j|x)$，就是命题函数 $y_j(x)$ 的真值函数 $T(y_j|x)$，都用 $T(\theta_j|x)$ 表示，即

$$T(\theta_j|x)=T(y_j|x)=m(\theta_j|x)$$

式中：θ_j 也可以理解为模型参数，y_j 的逻辑概率就是 Zadeh 定义的模糊事件的概率[23]：

$$T(y_j)=T(\theta_j)=\sum_i P(x_i)T(\theta_j|x_i)$$

假设对于每个 y_j 存在一个典型或柏拉图的理念 x_j，它使 $T(\theta_j|x_j)$=1，则隶属度就是 x_i 和 x_j 的相似度。如果 $U=V$，y_j 就变成估计，即 $y_j=\hat{x}_j$=“估计 x 是 x_j”，x 和 x_j 之间的相似度就是两者的混淆概率。比如：全球定位系统 (global positioning systems, GPS, 非特指美国的 GPS) 指示的位置和实际位置的相似度就是两者的混淆概率。估计的真值函数通常可以表示为失真函数的函数，即 $T(\theta_j|x)=\exp[-d(x, x_j)]$(本文假设 exp 是 log 的反函数)。如果失真取决于距离的平方，则有

$$T(\theta_j|x)=\exp[-d(x,x_j)]=\exp\left[-\frac{(x-x_j)^2}{2\sigma^2}\right] \tag{1}$$

设和 x_j 相混淆的所有 x 构成一个模糊集合 θ_j，则相似函数就是隶属函数，也是分辨率函数。GPS 的精度，平均误差平方的开方 (root mean square, RMS) 就反映其分辨率。相似函数或分辨率函数是不能用概率或条件概率表示的，因为它独立于 $P(x)$ 和 $P(y)$[20]。

1.2　学习函数的进化——从似然函数到隶属函数

机器学习的任务是：先用样本优化学习函数，再把学习函数用于概率预测或分类。优化学习函数有多种准则，比如最大正确率、最大似然度、最大互信息、最大估计互信息、最小失真、最小交叉熵等准则。主要的 3 种学习函数及其和样本分布的关系是：

1) 似然函数 $P(x|\theta_j)$(或写作 $P(x|y_j, \theta)$)，它逼近 x 的后验概率分布 $P(x|y_j)$；

2) 反概率函数 $P(\theta_j|x)$[3]，它逼近转移概率函数 $P(y_j|x)$。当标签数目 n=2 时，可用 Logistic 函数表示它。

3) 隶属函数 (包括相似函数)，它正比于关联函数 $m(x, y_j)=P(x, y_j)/[P(x)P(y_j)]$[20]，即

$$T(\theta_j|x)\propto m(x,y_j)=\frac{P(x,y_j)}{P(x)P(y_j)}=\frac{P(x|y_j)}{P(x)}=\frac{P(y_j|x)}{P(y_j)}\propto P(y_j|x)$$

因为似然函数在先验概率分布 $P(x)$ 改变后会失效，也因为有时候需要用最小误差 (或失真) 准则优化模型参数，所以似然函数不能满足要求。反概率函数能弥补似然函数的上述缺点。因为在 $P(x)$ 变为 $P'(x)$ 后，能用贝叶斯公式从 $P'(x)$ 和 $P(\theta_j|x)$ 得到新的概率预测 $P'(x|\theta_j)$。当标签数目

n=2 时, 反概率能作为正确率或误判率。但是, 当 n>2 时, 构造反概率函数是困难的, 因为有限制条件:

$$\sum_j P(\theta_j|x_i)=1, i=1,2,\cdots$$

一个权宜的方法是二元关联法 (binary relevance)[4]: 把 n 个标签学习转换为 n 对标签学习。这种方法是不经济的。另外 n>2 时, 用 $P(\theta_j|x)$ 表示正确率或误差率也是有问题的[20]。

为了克服上述 2 种学习函数的缺点, 鲁晨光提出用隶属 (或真值) 函数作为学习函数, 以便简化多标签学习[19]。不谋而合的是: MINE 和 InfoNCE 用相似函数作为学习函数, 并且用估计互信息作为目标函数。随后出现的深度信息最大化 (deep infomax, DIM)[24]、简单对比学习表示 (simple contrastive learning representations, SimCLR)[25]、动量对比 (momentum contrast, MoCo)[26] 等 (用于自监督学习) 都使用了相似函数和估计互信息, 表现出强大学习能力。而相似函数就是隶属函数的特例, 估计互信息是鲁晨光 30 年前提出的语义互信息[14] 的特例。

1.3　隶属函数的统计解释: 随机集落影

按照汪培庄的随机集落影理论, 隶属函数可看作一个随机集合 S_j 的无数取值 S_{jk}, k=1, 2, … (它们是清晰集合) 投影在 x 的论域 U 上产生的 (见图 1)。

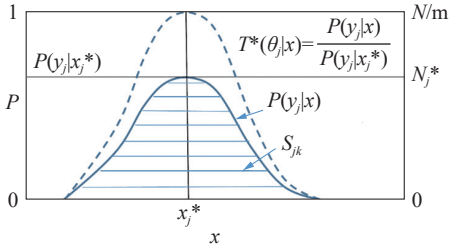


图 1　随机集落影理论中的集值统计方法

Fig. 1　Set-value statistic method in the random set falling shadow theory

用随机集落影方法得到的隶属函数如图 1 中虚线所示。S_{jk} 是一个集合值 (细横线), N 是样例总数, x_j^*是落影最厚的点, N_j^*是覆盖 x_j^*的集值的数目。图中等式表明, 用随机点落影方法得到的隶属函数和用随机集落影方法得到的隶属函数相同。根据随机集落影理论, x 在模糊集 θ_j 中的隶属度, 就是 x 落在随机集 S_j 中的概率, 即

$$T(\theta_j|x)=P(x\in S_j)=\frac{1}{N_j^*}\sum_k P(x\in S_{jk}) \qquad (2)$$

张南纶等[27] 以年龄 x 和属性“年轻人”“中年人”……为例验证了这个理论, 即通过集值统计表明, 这样得到的隶属函数随样本增大而稳定收敛。

2　用语义信息方法求解隶属函数

2.1　P-T 概率框架和语义信道

鲁晨光的语义信息论使用了兼用统计概率和逻辑概率的 P-T 概率框架[28]。

一个假设或标签 y_j 有 2 种概率: 统计概率 $P(y_j)$ 和逻辑概率 $T(y_j)$。统计概率是归一化的 (相加等于 1), 而逻辑概率不是归一化的。比如 y 表示 3 个可能的标签中的一个: y_1=“非成年人”, y_2=“成年人”, y_3=“年轻人”。3 个标签被选择的概率 (统计概率) 之和是 1, 但是它们为真的概率 (逻辑概率) 之和大于 1, 因为 y_1 和 y_2 的逻辑概率之和就等于 1。

贝叶斯定理可以推广到逻辑概率, 即

$$P(x|\theta_j)=T(\theta_j|x)P(x)/T(\theta_j)$$

$$T(\theta_j)=\sum_i P(x_i)T(\theta_j|x_i)$$

式中 $P(x|\theta_j)$ 是似然函数。假设隶属函数的最大值是 1, 从 $P(x)$ 和 $P(x|\theta_j)$ 也能得到隶属函数[19]:

$$T(\theta_j|x)=T(\theta_j)P(x|\theta_j)/P(x)$$

$$T(\theta_j)=1/\max[P(x|\theta_j)/P(x)]$$

一组转移概率函数构成一个香农信道, 一组隶属函数组成一个语义信道。当后者匹配前者时, 即 $T(\theta_j|x)\propto P(y_j|x)$ (j=1,2,…), 有 $P(x|\theta_j)$=$P(x|y_j)$ (j=1,2,…)。这时, 语义信息量达到最大。

2.2　兼容 Popper 思想的语义信息测度

鲁晨光提出的语义信息量公式为

$$I(x_i;\theta_j)=\log\frac{P(x_i|\theta_j)}{P(x_i)}=\log\frac{T(\theta_j|x_i)}{T(\theta_j)}$$

语义信息公式图解见图 2。它意味着: 先验逻辑概率越小, 后验逻辑概率越大, 语义信息量越大; 如果偏差过大, 语义信息量是负的。它符合 Popper 思想。当隶属函数总是 1 的时候, 上面语义信息公式就变成 Carnap 和 Bar-Hillel 的语义信息公式[29]。

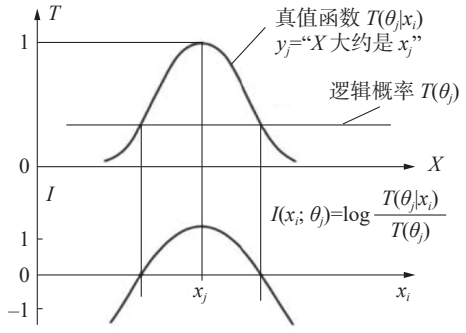


图 2　语义信息量公式

Fig. 2　Semantic information amount formula

因为 $I(x_i;\theta_j)$ 是后验真除以先验真的对数, 也可谓为 y_j 反映 x_i 的逼真度。对 $I(x_i;\theta_j)$ 求平均, 就得到语义 Kullback-Leibler (KL) 信息 $I(X;\theta_j)$ 和语

义互信息 $I(X;Y_\theta)$：

$$I(X;\theta_j)=\sum_i P(x_i|y_j)\log\frac{P(x_i|\theta_j)}{P(x_i)}=\sum_i P(x_i|y_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)} \tag{3}$$

$$I(X;Y_\theta)=\sum_i\sum_j P(x_i,y_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)} \tag{4}$$

当 $P(x|\theta_j)=P(x|\theta_j)(j=1,2,\cdots)$ 时，语义互信息就变成香农互信息[30]。从式 (3) 可以看出，最大语义信息准则等价于最小交叉熵准则和最大似然准则。把式 (1) 代入式 (4)，就可以看出最大互信息准则兼容正则化最小误差平方准则 (regularized least squares, RLS)。

2.3　信息率逼真函数 $R(G)$

推广信息率失真函数 $R(D)$，即用 $I(x_i;\theta_j)$ 代替失真 $d(x_i,y_j)$，并用语义互信息下限 G 代替平均失真上限 D，就得到保真度信息率函数 $R(G)$[16,31]。所有 $R(G)$ 函数都是碗状的 (见图 3)，并且其中有个重要的点：$s=\mathrm{d}R/\mathrm{d}G=1$，$R=G$。表示在这一点语义信道匹配香农信道，信息效率 $r=G/R$ 最大，为 1。关于 $R(G)$ 的更多讨论见文献 [31]。

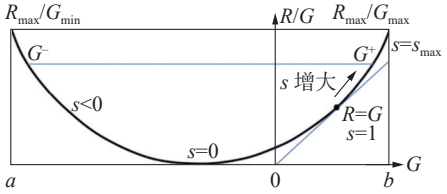


图 3　信息率逼真函数 $R(G)$

Fig. 3　The rate fidelity function $R(G)$

2.4　优化隶属函数的 2 个公式——随机点落影

从式 (3) 可知，当 $P(x|\theta_j)=P(x|y_j)$ 或 $T(\theta_j|x)\propto P(y_j|x)$ 时，平均语义信息量达最大。如果样本很大，令 $T(\theta_j|x)$ 的最大值是 1，可得

$$T(\theta_j|x)=\frac{\dfrac{P(x|y_j)}{P(x)}}{\max\left[\dfrac{P(x|y_j)}{P(x)}\right]}=\frac{P(y_j|x)}{\max[P(y_j|x)]} \tag{5}$$

如果样本不够大，则最大化 $I(X;\theta_j)$ 可得

$$T*(\theta_j|x)=\arg\max_{T(\theta_j|x)}\sum_i P(x_i|y_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)} \tag{6}$$

式 (5) 和式 (2) 完全兼容。因为

$$m_{\theta_j}(x)=P(x\in S_j)=\frac{1}{N_j^*}\sum_k P(x\in S_{jk})=\frac{1/N}{N_j^*/N}\sum_k P(x\in S_{jk})=\frac{P(y_j|x)}{P(y_j|x_j^*)}=\frac{P(y_j|x)}{\max[P(y_j|x)]}=T(\theta_j|x)$$

这说明随机点落影的结果等价于随机集落影的结果。

可以用机器学习语言来解释式 (5)：一个集值就是一个含有单标签多实例的样例 $(x_{j1},x_{j2},\cdots;y_j)$，隶属度 $T(\theta_j|x_i)$ 就是所有含有 x_i 的样例中 y_j 出现的概率。但是，单标签多实例的样例只有通过实验得到，从现实文本中只能得到含有单标签单实例的样例 $(x_i;y_j)$ 和样本分布 $P(x,y)$。现在，可以在想象中用多个单标签单实例的样例拼凑一个单标签多实例的样例或集值 S_{jk}，然后求 x_i 在 S_j 中的概率。这样，随机集落影就变成随机点落影 (见图 1)。

3　隶属函数用于机器学习

3.1　用于多标签学习和分类

考虑多标签学习——一种有监督学习。从样本 $\{(x_k,y_k),k=1,2,\cdots,N\}$ 能得到样本分布 $P(x,y)$。然后用式 (5) 得到优化的隶属函数——这是模糊统计法；还可以用式 (6) 求带参数的优化的隶属函数——这是拟合模糊分布法。多标签学习和分类的 2 个步骤如图 4 所示。

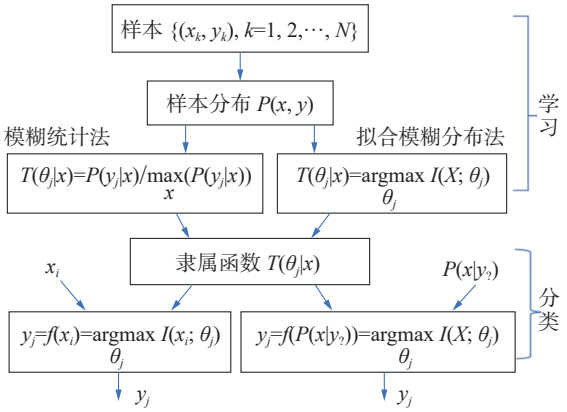


图 4　多标签学习和分类

Fig. 4　Multi-label learning and classification

假设真值函数是高斯函数，因为语义信道匹配香农信道时，应有

$$T(\theta_j|x)\propto\frac{P(x|y_j)}{P(x)}\propto P(y_j|x)$$

所以可以用 $P(x|y_j)/P(x)$ 或 $P(y_j|x)$ 的期望和标准偏差作为 $T(\theta_j|x)$ 的期望和标准偏差。

图 5 给出了一个例子。其中，x 表示年龄，y_j=“年轻人”，其真值函数是高斯函数。如果 $P(x|y_j)$ 是由 $P(x)$ 和 $P(y_j|x)$ 产生的，且 $P(y_j|x)$ 和高斯函数越成正比，则优化的真值函数和 $P(y_j|x)$ 的期望和标准差就越是接近。

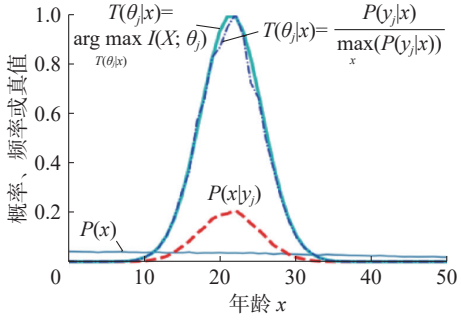


图 5　图解从 $P(x|y_j)$ 和 $P(x)$ 求 $T(\theta_j|x)$ 的 2 个公式

Fig. 5　Illustrating the two formulas for obtaining $T(\theta_j|x)$ from $P(x|y_j)$ 和 $P(x)$

如果分布 $P(x|y_j)/P(x)$ 形状近似于水坝截面, 称参数化的这种形状的函数为水坝函数 (见图 6)。可以采用下面转换从高斯函数 $g(x)$ 得到水坝函数 $b(x)$:

$$b(x)=1-[1-g(x)]^n \tag{7}$$

反函数是 $g(x)=1-[1-b(x)]^{1/n}$。其中 $n>1$。n 越大, 则"坝顶"越宽, "斜坡"越陡。不能用样本分布优化 $b(x)$, 但是能用修正的样本分布 $d(x)=1-[1-P(x|y_j|P(x)]^{1/n}$ 优化 $g(x)$, 即用 $d(x)$ 的期望和标准偏差作为 $g(z)$ 的期望和标准偏差, 然后用式 (7) 得到优化的真值函数。图 6 验证了这个方法, 图中使用 $n=3$。

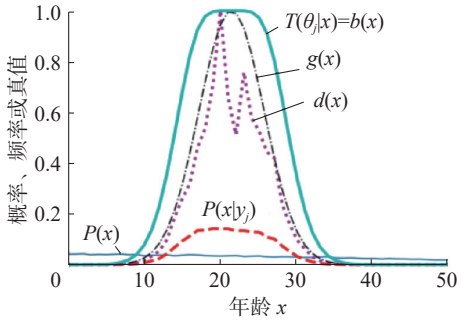


图 6　使用水坝函数 $b(x)$ 作为隶属函数

Fig. 6　Using a dam function $b(x)$ as a membership function

对于"老年人"的隶属函数, 可用 Logistic 函数表示。如果只知道 $P(y_j|x)$ 而不知道 $P(x)$, 可以假设 $P(x)$ 是等概率的, 即 $P(x)=1/|U|$, 然后优化隶属函数:

$$I(X;\theta_j)=\sum_i P(x_i|y_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)}=$$

$$\sum_i\frac{P(y_j|x_i)}{\sum_k P(y_j|x_k)}\log\frac{T(\theta_j|x_i)}{\sum_k T(\theta_j|x_k)}+\log|U|$$

对于多标签分类, 可以用分类器:

$$y_j*=\arg\max_{y_j} I(x;\theta_j)=\arg\max_{y_j}\log\frac{T(\theta_j|x)}{T(\theta_j)}$$

如果 x 是不确定的, 即只知道 $P(x|y_?)$, 则可以用分类器:

$$y_j*=\arg\max_{y_j} I(X;\theta_j)=$$

$$\arg\max_{y_j}\sum_i P(x_i|y_?)\log\frac{T(\theta_j|x)}{T(\theta_j)}$$

如果使用失真准则, 可以用 $-\log T(\theta_j|x)$ 作为失真函数[31], 或用 $T(\theta_j|x)$ 取代式中 $I(x;\theta_j)$。

3.2　用于不可见实例最大互信息分类

这种分类属于半监督学习。下面以医学检验和信号检测为例说明 (见图 7)。

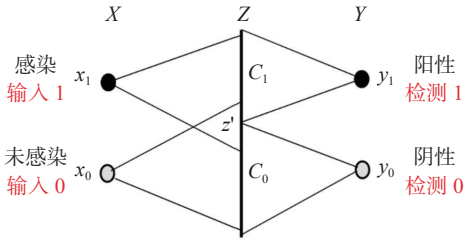


图 7　医学检验和信号检测

Fig. 7　Medical tests and signal detections

图中 Z 是取值 $z\in C$ 的随机变量。概率分布 $P(x)$ 和 $P(z|x)$ 是给定的。分类是根据 $z\in C_j$ 选择 y_j。任务是: 求产生最大香农互信息的分类器 $y=h(z)$。

图 8 是最大互信息分类方法流程。

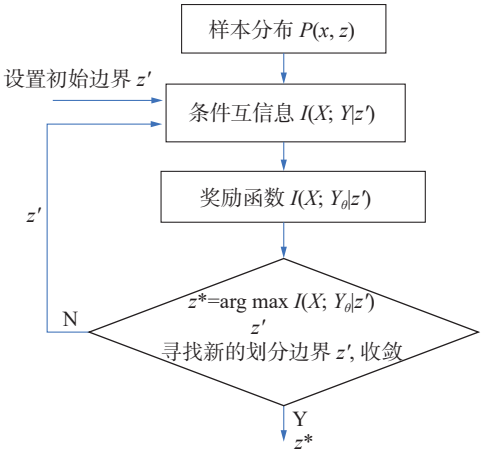


图 8　最大互信息分类流程

Fig. 8　Flow of maximum mutual information classification

方法不限于二分类。设 C_j 是 C 的子集, 并且 $y_j=f(z|z\in C_j)$。$S=\{C_1, C_2, \cdots\}$ 是 C 的一个划分。目的是找到最优划分:

$$S*=\arg\max_S I(X;Y_\theta|S)=$$

$$\arg\max_S\sum_j\sum_i P(C_j)P(x_i|C_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)}$$

先假设一个划分, 然后作下面迭代。

匹配 1　让语义信道匹配香农信道，即从

$$P(y_j|x)=\sum_{z_k\in C_j}P(z_k|x),\ j=1,2,\cdots$$

得到 $T(\theta_j|x)$ 和 $I(x;\theta_j)$。然后，对于每个 z 得到条件信息或奖励函数：

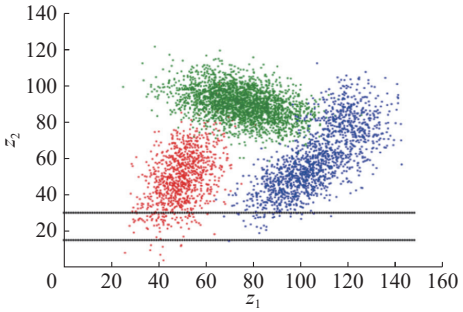
$$I(X;\theta_j|z)=\sum_i P(x_i|z)I(x_i;\theta_j),\ j=1,2,\cdots$$

匹配 2　令香农信道匹配语义信道——通过分类器：

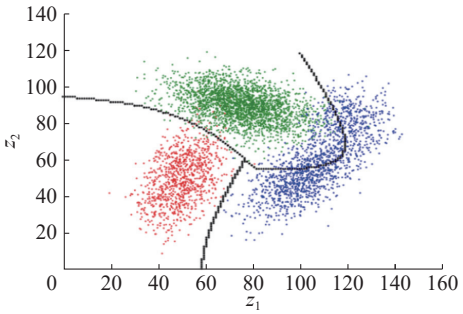
$$y_j{*}=f(z)=\arg\max_{y_j} I(X;\theta_j|z),\ j=1,2,\cdots$$

重复匹配 1 和匹配 2 直至 S 不再变化。收敛的 S 就是要找的 S*。详见文献 [19] 中利用 $R(G)$ 函数的收敛证明。

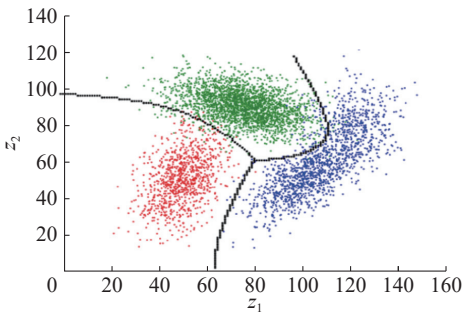
图 9 给出了一个例子。详细数据见文献 [19] (4.2 节)。图 9(a) 中 2 条水平线表示很糟糕的初始划分。图 9(d) 显示了互信息随迭代次数变化 (收敛很快)。z 现在变成二维矢量 $\boldsymbol{z}$，图中 z_1 和 z_2 是 $\boldsymbol{z}$ 的 2 个分量。



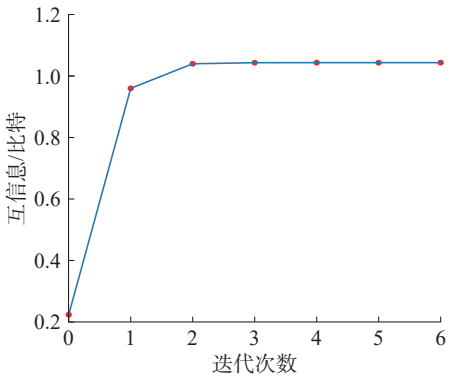
(a) 很差的初始划分(见 2 条水平线)



(b) 1 次迭代后的划分



(c) 2 次迭代后的划分



(d) 互信息随迭代次数变化情况

图 9　最大互信息分类

Fig. 9　Maximum mutual information classification

3.3　用于解释和改进用于混合模型的 EM 算法

期望−最大 (expectation-maximization, EM) 算法常用于混合模型。对于 $P(x)=\sum_j P(y_j)P(x|y_j)$，只知道样本分布 $P(x)$，然后使用混合模型 $P_\theta(x)=\sum_j P(y_j)P(x|\theta_j)$ 逼近 $P(x)$，使得相对熵 $H(P\|P_\theta)$ 接近 0。改进的 EM 算法可谓 EnM 算法，其中 n 意思是重复期望算法 n 次。迭代之前设置初始的 $P(x|\theta_j)$ 和 $P(y_j)$, j=1,2,⋯，每个迭代包含 2 个匹配。

匹配 1　令香农信道 $P(y|x)$ 匹配语义信道 $T(y|x)$，即重复 2 个公式 n(比如 n=3) 或更少次：

$$\begin{aligned}&P(y_j|x)=P(y_j)P(x|\theta_j)/P_\theta(x)\\ \text{其中}\quad &P_\theta(x)=\sum_j P(y_j)P(x|\theta_j)\\ &P^{+1}(y_j)=\sum_x P(x_i)P(y_j|x_i)\end{aligned}\tag{8}$$

式 (8) 就是求解 $R(G)$ 函数用到的最小化香农互信息的公式。在上面所有步骤中，只有在改变 θ 的第一步中，香农互信息 R=$I(X;Y)$ 和语义互信息 G=$I(X;Y_\theta)$ 可能增大也可能减小，其他步骤都不改变 θ，都会减小香农互信息。

匹配 2　令语义信道匹配香农信道，通过

$$\begin{aligned}&P(x|\theta_j^{+1})=P(x)P(x|\theta_j)/P_\theta(x)\\ &P_\theta(x)=\sum_j P(y_j)P(x_i|\theta_j)\end{aligned}$$

直至相对熵或 θ 不能改进为止。

为了证明迭代收敛，能推导出

$$H(P\|P_\theta)=R-G+H(P^{+1}(y)\|P(y))\tag{9}$$

式中 $H(P\|P_\theta)$ 是相对熵或 KL 离散度。因为匹配 2 最大化 G 且匹配 1 最小化 R 和 $H(P^{+1}(y)\|P(y))$，所以 $H(P\|P_\theta)$ 能接近 0。

图 10 是 EnM 算法流程。

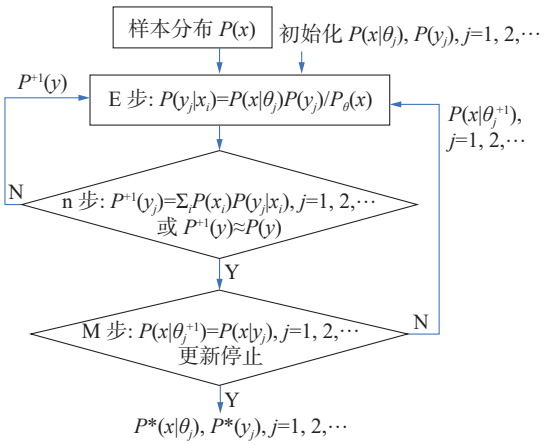
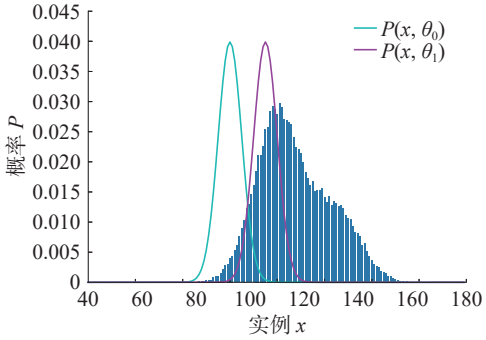


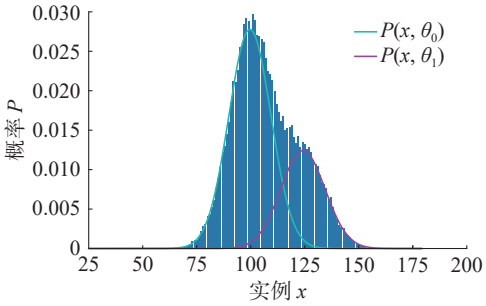
图 10　求混合模型的 EnM 算法流程

Fig. 10　Flow of the EnM algorithm for mixture models

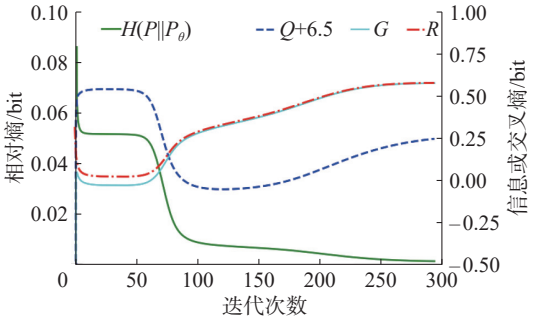
图 11 给出一个高斯混合模型的例子, 用以比较 EM 和 E3M 算法。真实模型参数是 (μ_1, μ_2, σ_1, σ_2, $P(y_1)$) = (100, 125, 10, 10, 0.7), 初始参数是 (μ_1, μ_2, σ_1, σ_2, $P(y_1)$) = (80, 95, 5, 5, 0.5)。



(a) 2 个初始部件



(b) 2 个收敛的部件



(c) Q(完全数据的对数似然度)、R、G 和 $H(P||P_\theta)$随迭代次数变化情况

图 11　用一个不易收敛的例子比较 EM 和 E3M 算法

Fig. 11　Comparing the EM and E3M algorithms by using a example that is hard to converge

结果显示, EM 算法需要大约 340 次迭代, 而 E3M 算法需要大约 240 次迭代。这个例子揭示 EnM(包括 EM 算法) 收敛是因为语义互信息 G 和香农互信息 R 相互靠近, 并不是因为完全数据对数似然度 Q 不断增大 (流行观点)。

3.4　用于贝叶斯确证和因果确证

贝叶斯确证的任务是评价样本分布对大前提的支持。比如对于医学检验 (见图 7), 一个大前提是“如果一个人检验呈阳性 (y_1), 那么他被感染 (x_1)”, 简记为 $y_1 \to x_1$。鲁晨光把确证分为信道确证 (评价检验手段有多好) 和预测确证 (看概率预测 $P(x_1|\theta_1)$ 有多可靠)[32]。对于信道确证, 一个真值 (或隶属) 函数可看作清晰真值函数 $T(y_1|x) \in \{0,1\}$和永真句的真值函数 (总是 1) 的组合:

$$T(\theta_1|x) = b_1' + b_1 T(y_1|x)$$

永真句的比例 b_1'就是不信度。可信度是 b_1, 它和 b_1'的关系是 $b_1'=1-|b_1|$(见图 12)。

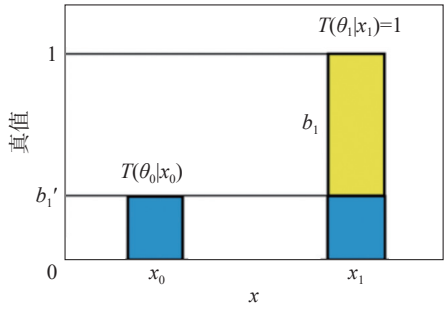


图 12　真值 (或隶属) 函数

Fig. 12　Truth (or membership) function

通过改变 b_1 最大化语义 KL 信息 $I(X; \theta_1)$, 优化的可信度 b_1, 记为 b_1^*, 就是确证度[32]:

$$b_1^* = b^*(y_1 \to x_1) = \frac{P(y_1|x_1) - P(y_1|x_0)}{\max(P(y_1|x_1), P(y_1|x_0))} = \frac{R^+ - 1}{\max(R^+, 1)}$$

式中 $R^+=P(y_1|x_i)/P(y_1|x_0)$ 是阳性似然比, 反映检验阳性有多可靠。这个结论和医学检验理论兼容。

考虑预测确证度, 假设 $P(x|\theta_1)$ 是 0-1 部分和等概率部分的组合, 0-1 部分的比例就是预测的可信度, 优化的可信度就是预测确证度:

$$c_1^* = c^*(y_1 \to x_1) = \frac{P(x_1|y_1) - P(x_0|y_1)}{\max(P(x_1|y_1), P(x_0, y_1))} = \frac{a - c}{\max(a, c)}$$

式中: a 是正例的个数, c 是反例的个数。2 种确证度都可用于概率预测, 即求 $P(x|\theta_1)$。

Hemple 曾提出确证悖论, 即乌鸦悖论。根据经典逻辑中的等价条件, “如果 x 是乌鸦, 则 x 是黑的” (规则 1) 等价于 “如果 x 不是黑的, 则 x 不是乌鸦” (规则 2)。一支白粉笔支持规则 2; 因而

也支持规则 1。但是，根据常识，一只黑乌鸦支持规则 1，一个非黑的乌鸦反对规则 1；一个不是乌鸦的东西，比如一只黑猫或一只白粉笔，和规则 1 不相关。因此，在等价条件和常识之间存在悖论。使用确证度 c_1*，可以肯定常识是对的，等价条件是错的 (对于模糊大前提)，因而可以消除乌鸦悖论。而其他确证测度都不能消除乌鸦悖论[32]。

因果推断理论中使用因果概率[33]：

$$P_d = \max\left[0, \frac{P(y_1|x_1)-P(y_1|x_0)}{P(y_1|x_1)}\right] = \max\left(0, \frac{R^+-1}{R^+}\right)$$

表示原因 x_1 替代 x_0 导致结果 y_1 的必然性。其中 $P(y_1|x)$=$P(y_1|\mathrm{do}(x))$ 是干预 x 导致的 y_1 的后验概率[34]。鲁晨光用语义信息方法得到的因果确证度是[35]

$$Cc(x_1/x_0 => y_1) = b_1* = \frac{P(y_1|x_1)-P(y_1|x_0)}{\max(P(y_1|x_1),P(y_1|x_0))} = \frac{R^+-1}{\max(R^+,1)}$$

它兼容上述因果概率 (式中“=>”表示因果关系)，但是还能表示负的因果关系，比如表示疫苗抑制感染的必然性。

3.5　用于模糊约束控制和强化学习

假设要把羊群赶到一个指定牧场——用隶属函数 $T(\theta_j|x)$ 表示，$P(x)$ 是羊群不加约束时的概率分布，$P(x|c_j)$ 是采用控制手段 c_j 后羊群的概率分布，可以用语义 KL 信息评价 c_j 如何符合目的，合目的信息公式为

$$I(X;\theta_j,c_j) = \sum_i P(x_i|c_j)\log\frac{P(x_i|\theta_j)}{P(x_i)} = \sum_i P(x_i|c_j)\log\frac{T(\theta_j|x_i)}{T(\theta_j)}$$

类似的任务是：

1) 要把粮食产量提高到某个范围 (比如每公顷接近或高于 10 t——可用 Logistic 函数表示)，$P(x)$ 就是当前的粮食产量的概率分布，$P(x|c_j)$ 是采用控制 c_j 后的粮食产量的概率分布；

2) 要把人口死亡年龄控制到某个范围 (比如接近或超过 80 岁)，$P(x)$ 是当前人口死亡年龄分布，$P(x|c_j)$ 是控制后的死亡年龄分布。

容易证明，当 $P(x|c_j)$ 的分布集中在 $T(\theta_j|x)$ 最大点 (为 1) 时，合目的信息最大。但是还要考虑控制成本 (假设它和 KL 信息 $I(X;c_j)$ 成正比) 和控制效率 r=$I(X;\theta_j)/I(X;c_j)$。当 $P(x|c_j)$=$P(x|\theta_j)$ 时，控制效率 r 最大，等于 1。如果要继续增大 $I(X;\theta_j)$ 并希望成本尽可能小，$R(G)$ 函数的参数解[31] 说明可以选择：

$$P(x_i|c_j) = \frac{P(x_i)[T(\theta_j|x_i)]^s}{\sum_k P(x_k)[T(\theta_j|x_i)]^s}$$

式中 s>1 是 $R(G)$ 函数参数解中的参数。该式提供了权衡合目的信息 G 和控制效率 r 的方法[20]。

3.6　用于自监督学习

自监督学习用图像或文本的一部分预测另一部分，后者可能是下一个或掩盖的实例。预测的和真实的实例之间存在相似性。

Belghazi 等[21] 提出 MINE 时使用了学习函数：

$$\exp[T_w(x,y_j)] \propto P(y_j|x)$$

式中 y_j 是预测或估计 $\hat{x}_j$。虽然 $T_w(x,y_j)$ 不是负的，但是可理解为保真度函数。设 $T_{\max}$ 最大保真度，失真函数就是 $d(x,y_j)$=$T_{\max}$−$T_w(x,y_j)$。所以 $\exp[T_w(x,y_j)]$ 和相似函数 $\exp[-d(x,y_j)]$ 成正比，也可以看作是相似函数。

Oord 等[22] 在介绍 InfoNCE 的文章中明确提出：使用一个学习函数，希望它正比于 $m(x,y)$=$P(x|y)/P(x)$。文中的表达式为

$$f_k(\boldsymbol{x}_{t+k},\boldsymbol{c}_t) \propto P(\boldsymbol{x}_{t+k}|\boldsymbol{c}_t)/P(\boldsymbol{x}_{t+k})$$

式中：$\boldsymbol{c}_t$ 是从前面的数据得到的特征矢量；$\boldsymbol{x}_{t+k}$ 是要预测的特征矢量；$f_k(\boldsymbol{x}_{t+k},\boldsymbol{c}_t)$ 就是根据 $\boldsymbol{c}_t$ 预测的 $\boldsymbol{x}_{t+k}$ 和实际的 $\boldsymbol{x}_{t+k}$ 之间的相似函数，用它可以构造估计互信息，然后优化模型参数。这种方法把对比学习中的 N 对 Logistic 函数变成 N 个 Softmax 函数，因而可以直接用于多标签学习。和 MINE 中学习函数不同，InfoNCE 用一个类似于隶属函数的函数作为相似函数，这样更加灵活。

MINE 和 InfoNCE 和鲁晨光的语义信息方法在本质上是一样的，其共同特点是：

1) 用正比于 $P(y_j|x)$ 的隶属函数 $T(\theta_j|x)$ 或相似函数 $S(x,y_j)$ 作为学习函数，它的最大值一般是 1，它的平均是划分函数 Z_j。

2) x 和 y_j 之间的估计信息或语义信息是 $\log[T(\theta_j|x)/Z_j]$ 或 $\log[S(x,y_j)/Z_j]$。

3) 求平均信息仍然用统计概率分布 $P(x,y)$。

4　讨论

4.1　隶属函数来自主观定义还是客观统计？

如何确定隶属函数？历来存在 2 条路径：来自主观定义和来自客观统计。本文认为两者并不矛盾。首先，隶属函数是主观的，可以来自定义。命题函数的语义可以通过真值函数定义，而真值函数就是隶属函数。控制领域和专家系统中的隶属函数大多是主观定义的。

但是，对于概率预测，存在一个最优的隶属函

数, 它和统计相匹配, 如式 (5) 所示, 是客观的。当语义信道匹配香农信道, 即 $T(\theta_j|x) \propto P(y_j|x)$, j=1,2,⋯时, 语义信息量达到最大。这时用隶属函数 $T(\theta_j|x)$ 作贝叶斯预测和用转移概率函数 $P(y_j|x)$ 作贝叶斯预测, 结果相等, 即 $P(x|\theta_j)$=$P(x|y_j)$。

4.2　隶属函数用作学习函数的优点

和似然函数 $P(x|\theta_j)$ 或反概率函数 $P(\theta_j|x)$ 相比, 隶属函数作为学习函数有下面优点:

1) $P(x)$ 改变后, 也可以使用贝叶斯公式从 $P(x)$ 和 $T(\theta_j|x)$ 得到新的概率预测 $P(x|\theta_j)$;

2) 便于用负指数函数表示, 构造多个隶属函数时没有归一化限制;

3) 用于学习和分类时, 便于使用最大语义或估计互信息准则——它兼容最大似然准则和 RLS 准则。

4) 也便于按最小平均失真准则分类, 因为 $-\log T(\theta_j|x)$ 就反映失真或误差。

5) 优化的隶属函数和关联函数 $m(x, y_j)$ 成正比, 它们独立于 $P(x)$ 和 $P(y_j)$, 反映 x 和 y_j 之间的内在联系, 因而有更好的可迁移性。

4.3　随机点落影方法和其他求隶属函数方法比较

随机点落影方法有下面 2 个特点:

1) 适合大样本且多标签学习, 能直接从样本分布 $P(x, y)$ 得到一组隶属函数 $T(\theta_j|x)$ (j=1,2,⋯);

2) 隶属函数不仅像转移概率函数 $P(y_j|x)$ 一样, 能代入贝叶斯公式, 用于概率预测, 还适合作为学习函数和约束函数, 用于度量语义信息和约束控制。

这 2 个特点是其他求隶属函数方法[6-10]不具有的。但是随机点落影方法也有局限性。比如, 和二元对比排序法[5]相比, 后者更适合较小样本; 和专家经验法比, 后者更适合制定控制规则[36-37]。

4.4　解释深度学习

为解释以自动编码器 (autoencoder) 和深度信念网 (deep belief network, DBN) 等深度神经网络的成功, Tishby 等[38]提出信息瓶颈解释, 认为优化深度神经网络时, 需要最大化一些环节之间的香农互信息并同时最小化另外一些环节之间的香农互信息。然而, 从 $R(G)$ 函数的角度看, 自动编码器和 DBN 的每一层都需要最大化语义互信息 G 并最小化香农互信息 R; 预训练就是语义信道和香农信道相互匹配, 使得 G=R 且相对熵 $H(P||P_\theta)$=0 (见式 (9))。微调就通过增大 s(使划分边界变陡) 同时增大 R 和 G(见图 3)。

最近 OpenAI 的研究人员[39-40]用无损数据压缩 (或限失真) 前提下最小化香农互信息或复杂性解释通用人工智能, 这和信道匹配解释类似。不过, 用隶属 (或相似) 函数构造的语义 (或估计) 信息测度表示约束条件, 更具有一般性。因为通常的损失是模糊的且语义相关的。所以, 约束条件不是无损编码, 而是有损编码时的解码逼真度或语义信息量。

4.5　潜在应用: 用隶属函数构造信道混合模型机

到目前为止, 只有用似然函数 (比如高斯似然函数) 构造混合模型。但是, 也能用高斯真值函数构造高斯信道混合模型, 并用 EnM 算法求解它。不同的是, 在匹配 1 中, 需要用先验概率分布 $P_0(x)$ 和 $T(\theta_j|x)$ 产生似然函数:

$$P(x|\theta_j) = P_0(x)T(\theta_j|x) \Big/ \sum_i P_0(x_i)T(\theta_j|x)$$

然后重复式 (8)n 次。$P_0(x)$ 和 $P(x)$ 不同, 前者是先验概率分布而后者不是。可以假设 $P_0(x)$ 是等概率的。在匹配 2 中, 让语义信道匹配香农信道, 即令

$$T(\theta_j^{+1}|x) = \exp\left[\frac{-(x-\mu_j)^2}{2\sigma_j^2}\right] \propto \frac{P(y_j|x)}{P(y_j)} = \frac{P_0(x)T(\theta_j|x)}{T(\theta_j)P_\theta(x)} \propto \frac{P_0(x)T(\theta_j|x)}{P_\theta(x)}$$

这意味着能用 $P(y_j|x)$ 或 $P_0(x)T(\theta_j|x)/P_\theta(x)$ 的期望和标准偏差作为 $T(\theta_j^{+1}|x)$ 的期望和标准偏差。

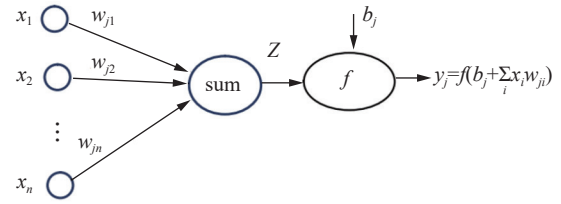
信道混合模型可以用于神经网络的无监督学习或预训练, 从而得到信道混合模型机[20]。其功能类似于有限波尔兹曼机, 但是不需要考虑梯度下降和反向传播。这时候网络权重参数就是隶属度。

4.6　用隶属函数构造神经网络

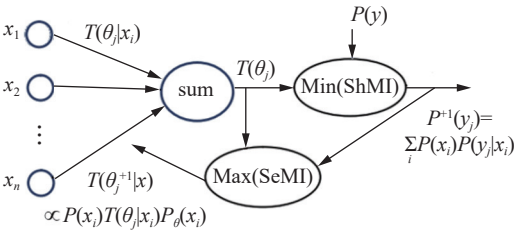
用隶属度或相似度作为神经网络权重参数时, 神经网络就包含语义信道和香农信道。然后可用语义信息方法优化神经网络。比如, 信道混合模型机的一个神经元和一个标准的神经元如图 13 所示。

模糊逻辑——特别是兼容布尔代数的模糊逻辑——看来也能用于神经网络。比如神经网络常用的激活函数 ReLU(x)=max(0, x) 就是鲁晨光建立色觉机制模型[41]用的逻辑差运算 $f(a \wedge \neg b)$=max(0, a–b) 在 b=0 时的特例。隶属函数、模糊逻辑以及语义信息方法用于神经网络, 将使神经网络更加易于理解。

深度学习领域出现了许多令人惊异有效方法, 特别是特征抽取方法。隶属函数和语义信息方法用于解释和改进深度学习刚刚开始, 要赶上深度学习的步伐, 这是严峻挑战。



(a) 标准神经元



(b) 信道混合模型机的神经元

图 13　比较信道混合模型机的神经元和标准神经元

Fig. 13　Comparing a neuron in the channel mixture model machine and a standard neuron

5　结束语

汪培庄的随机集落影理论给隶属函数的统计解释打下理论基础。但是，由于集值统计需要的集值需要通过大量实验才能得到，实践中希望从一般的样本分布得到隶属函数。鲁晨光在汪培庄的隶属函数统计解释的基础上，使用语义信息方法，推导出用样本分布优化隶属函数的 2 个公式。它们和随机集的统计结果完全兼容，可谓随机点落影方法。随机点落影方法得到的隶属函数的特点是：适合大样本和多标签学习；适合概率预测和语义信息度量；不含主观性；适合多种机器学习方法。这样的隶属函数已经在多标签分类、最大互信息分类和混合模型等任务中显示很好效果。最近深度学习领域出现的以 MINE 和 InfoNCE 为代表的新潮流——使用相似函数作为学习函数并使用估计互信息作为目标函数——和鲁晨光的方法异途同归，也支持把隶属函数(或相似函数) 作为重要的学习函数。随机点落影方法应能促使隶属函数在机器学习领域发挥更大作用。

一个有意义的探索是：把隶属函数用作神经网络权重参数，并用信道混合模型方法预训练深度神经网络，从而简化深度学习。深度学习中存在很多巧妙的特征抽取方法，如何从语义信息论角度理解和改进它们，需要更多研究。

参考文献：

[1] ZADEH L A. Fuzzy sets[J]. Information and control, 1965, 8(3): 338−353.

[2] FISHER R A. On the mathematical foundations of theoretical statistics[J]. Philosophical transactions of the royal society of London series A, containing papers of a mathematical or physical character, 1922, 222: 594−604.

[3] FIENBERG S E. When did Bayesian inference become “Bayesian”?[J]. Bayesian analysis, 2006, 1(1): 1−40.

[4] ZHANG Minling, LI Yukun, LIU Xuying, et al. Binary relevance for multi-label learning: an overview[J]. Frontiers of computer science, 2018, 12(2): 191−202.

[5] 赵光荣. 用二元相对比较法确定模糊顺序[J]. 系统工程, 1984, 2(4): 104−106.
ZHAO Guangrong. Determination of fuzzy order by binary relative comparison method[J]. Systems engineering, 1984, 2(4): 104−106.

[6] BHATTACHARYYA R, MUKHERJEE S. Fuzzy membership function evaluation by non-linear regression: an algorithmic approach[J]. Fuzzy information and engineering, 2020, 12(4): 412−434.

[7] DOMBI J, RIGÓ P R. The construction of multidi- mensional membership functions and its application to feasibility problems[J]. Fuzzy sets and systems, 2023, 469: 108634.

[8] REN Yaxue, LYU Jinfeng, LIU Fucai. A novel fuzzy model identification approach based on FCM and Gaussian membership function[C]//2020 39th Chinese Control Conference. Shenyang: IEEE, 2020: 1209−1214.

[9] SHUKLA A K, MUHURI P K. Deep belief network with fuzzy parameters and its membership function sensitivity analysis[J]. Neurocomputing, 2025, 614: 128716.

[10] 范轶博, 赵涛, 解相朋. 广义二型模糊系统的自组织规则生成方法[J]. 智能系统学报, 2024, 19(3): 646−652
FAN Yibo, ZHAO Tao, XIE Xiangpeng. Self-organizing rule generation method for a general type-2 fuzzy system[J]. [J]. CAAI transactions on intelligent systems, 2024, 19(3): 646−652.

[11] WANG Peizhuang. From the fuzzy statistics to the falling random subsets[M]//WANG P P, ed. Advances in Fuzzy Sets, Possibility Theory, and Applications. Boston: Springer, 1983: 81−96.

[12] 汪培庄. 模糊集与随机集落影[M]. 北京: 北京师范大学出版社, 1985.
WANG Peizhuang. Fuzzy set and random set falling shadow[M]. Beijing: Beijing Normal University Publishing House, 1985.

[13] 汪培庄, 刘海涛. 因素空间与人工智能[M]. 北京: 北京邮电大学出版社, 2021.
WANG Peizhuang, LIU Haitao. Factor space and artificial intelligence[M]. Beijing: Beijing University of Posts and Telecommunications Press, 2021.

[14] LU Chenguang. Shannon equations reform and applications[J]. BUSEFAL, 1990, 44: 45−52.

[15] 鲁晨光. 广义熵和广义互信息的编码意义[J]. 通信学报, 1994, 15(6): 37−44.
LU Chenguang. Meanings of generalized entropy and generalized mutual information for coding[J]. Journal on communications, 1994, 15(6): 37−44.

[16] 鲁晨光. 广义信息论[M]. 合肥: 中国科学技术大学出版

社, 1993.
LU Chenguang. Generalized information theory [M]. Hefei: China University of Science and Technology Press, 1993.
[17] LU Chenguang. A generalization of Shannon's information theory[J]. International journal of general systems, 1999, 28(6): 453−490.
[18] LU Chenguang. Channels' matching algorithm for mixture models[C]//International Conference on Intelligence Science. Beijing: Springer, 2017: 321–332.
[19] LU Chenguang. Semantic information G theory and logical Bayesian inference for machine learning[J]. Information, 2019, 10(8): 261.
[20] LU Chenguang. Reviewing evolution of learning functions and semantic information measures for understanding deep learning[J]. Entropy, 2023, 25(5): 802.
[21] BELGHAZI M I, BARATIN A, RAJESWAR S, et al. MINE: mutual information neural estimation[C]// Proceedings of the 35th International Conference on Machine Learning. Stockholm: PMLR, 2018: 1–44.
[22] OORD A V D, LI Y, VINYALS O. Representation learning with contrastive predictive coding[EB/OL]. (2018−07−10)[2023−09−15]. https://arxiv.org/abs/1807.03748.
[23] ZADEH L A. Probability measures of fuzzy events[J]. Journal of mathematical analysis and applications, 1968, 23(2): 421−427.
[24] HJELM R D, FEDOROV A, LAVOIE-MARCHILDON S, et al. Learning deep representations by mutual information estimation and maximization[EB/OL]. (2018−08−20)[2023−09−15]. https://arxiv.org/abs/1808.06670.
[25] CHEN T, KORNBLITH S, NOROUZI M, et al. A simple framework for contrastive learning of visual representations[C]//Proceedings of the 37th International Conference on Machine Learning. Vienna: JMLR, 2020: 1597−1607.
[26] HE Kaiming, FAN Haoqi, WU Yuxin, et al. Momentum contrast for unsupervised visual representation learning[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 9726−9735.
[27] 张南纶. 随机现象的从属特性及概率特性[J]. 武汉建材学院学报, 1981, 3(3): 9−24.
ZHANG Nanlun. The membership and probability characteristics of random appearances[J]. Journal of Wuhan University of technology, 1981, 3(3): 9−24.
[28] LU Chenguang. The P–T probability framework for semantic communication, falsification, confirmation, and Bayesian reasoning[J]. Philosophies, 2020, 5(4): 25.
[29] CARNAP R, BAR-HILLEL Y. An outline of a theory of semantic information, technical report No. 247[R]. Research Laboratory of Electronics, Cambridge: MIT, 1952.
[30] SHANNON C E, WEAVER W. The mathematical theory of communication[M]. Urbana: The University of Illinois Press, 1963.
[31] LU Chenguang. Using the semantic information G measure to explain and extend rate-distortion functions and maximum entropy distributions[J]. Entropy, 2021, 23(8): 1050.
[32] LU Chenguang. Channels' confirmation and predictions' confirmation: from the medical test to the raven paradox[J]. Entropy, 2020, 22(4): 384.
[33] ROBINS J, GREENLAND S. The probability of causation under a stochastic model for individual risk[J]. Biometrics, 1989, 45(4): 1125−1138.
[34] PEARL J. Causal inference in statistics: an overview[J]. Statistics surveys, 2009, 3: 96−146.
[35] LU Chenguang. Causal confirmation measures: from Simpson's paradox to COVID-19[J]. Entropy, 2023, 25(1): 143.
[36] ANBALAGAN P, JOO Y H. Fuzzy membership- function-dependent design of aperiodic sample-data control scheme for nonlinear PMSG-based WECS with quantization measurements via refined looped Lyapunov functional[J]. Information sciences, 2024, 661: 120149.
[37] 李东升, 邵山, 陈军, 等. 不确定隶属函数 T-S 模糊控制器设计与稳定分析[J]. 智能系统学报, 2010, 5(1): 17−23.
LI Dongsheng, SHAO Shan, CHEN Jun, et al. Design and stability analysis of a fuzzy controller with uncertain degrees of membership[J]. CAAI transactions on intelligent systems, 2010, 5(1): 17−23.
[38] TISHBY N, ZASLAVSKY N. Deep learning and the information bottleneck principle[C]//2015 IEEE Information Theory Workshop. Jerusalem: IEEE, 2015: 1−5.
[39] RAE J. Compression for AGI[EB/OL]. (2023−02−08)[2023−09−15]. https://www.nxrte.com/jishu/16893.html.
[40] SUTSKEVER L. An observation on generalization [EB/OL]. (2023−08−14) [2023−09−15]. https://simons.berkeley.edu/talks/ilya-sutskever-openai-2023-08-14.
[41] 鲁晨光. 色觉的译码模型及其验证[J], 光学学报, 1989, 9(2): 158−163.
LU Chenguang. Decoding model of color vision and its verification [J]. Journal of optics, 1989, 9(2): 158−163.

作者简介:

汪培庄, 教授, 博士生导师, 曾任国际模糊系统协会副主席。主要研究方向为模糊数学及其在人工智能中的应用。获得国家级和部委级奖励多项、国际奖 1 项。发表学术论文 200 余篇, 出版学术著作 4 部。E-mail: peizhuangw@126.com。

鲁晨光, 辽宁工程技术大学客座教授。主要研究方向为语义信息论、机器学习、色觉机制、投资组合、美感和进化。发表学术论文 40 余篇, 出版学术专著 4 部。E-mail: lcguang@foxmail.com。