

一类动态模糊机器学习算法研究

谢 琳, 李凡长

(苏州大学计算机科学与技术学院, 江苏苏州 215006)

摘 要: 现有的各种机器学习方法在面对数据的动态模糊性问题时, 面临严峻挑战. 本文主要针对数据的动态模糊性, 给出了一类动态模糊机器学习模型, 及其局部线性嵌入算法和改进的局部线性嵌入算法, 并对算法的稳定性进行分析, 通过实例验证说明算法是有效的.

关键词: 动态模糊集; 动态模糊逻辑; 动态模糊机器学习

中图分类号: TP181 **文献标识码:** A **文章编号:** 0372-2112 (2008) 12A-114-06

Research on the Kind of Dynamic Fuzzy Machine Learning Algorithm

XIE Lin, LI Farr zhang

(College of Computer Science and Technology, Soochow University, Suzhou, Jiangsu 215006, China)

Abstract: The machine learning is an important technology of the data analysis. But facing data dynamic fuzziness questions, many learning algorithms faced with stem challenge. This paper mainly aims at the character of dynamic fuzzy of data, gives the kind of dynamic fuzzy machine learning model, the locally linear embedding algorithm and the improved locally linear embedding algorithm. It analyses the algorithms stability, and confirms their effective through the examples.

Key words: dynamic fuzzy sets; dynamic fuzzy logic; dynamic fuzzy machine learning

1 引言

机器学习作为人工智能领域的一个分支, 近几年, 由于数据性质的复杂性与用户需求的多样性, 使得各种机器学习方法层出不穷, 如文献[1~8]的相关内容. 但是目前人们比较公认的机器学习概念的说法还是 Simon 对学习的阐述: “如果一个系统能够通过执行某种过程而改进它的性能, 这就是学习.” 这个说法的要点是: 其一, 学习是一个过程; 其二, 学习是对一个系统而言的; 其三, 学习能够改变系统的性能. “过程”、“系统”与“改变性能”是学习的三个要点, 显然, 如果将系统考虑为人, 这个说法也是成立的. 根据这种说法, 用现有的机器学习方法对其进行验证就发现其不足了. 其原因是这三个要点具有动态模糊性(character of dynamic fuzzy): 如学习过程, 本质上具有动态模糊性; 系统的变化, 系统的好坏等本质上也是具有动态模糊性的; 改变系统性能, 其改变的结果如何, 改变的性能如何等均具有动态模糊性. 由此说明, 在整个机器学习活动及机器学习系统中动态模糊性是普遍存在的. 利用动态模糊集研究机器学习文献[9, 10]取得了一系列成果. 本文在此基础上进一步研究.

2 动态模糊机器学习模型

一个系统的学习可以认为是它在运行过程中对自身的调整, 体现为系统的结构或参数的一系列的变化, 如果用数学语言来描述, 学习就可以定义为一个集合到另一个集合的映射.

定义 1 动态模糊机器学习空间 由一切动态模糊机器学习要素构成的用于描述学习过程的空间称为动态模糊机器学习空间. 它由{学习样例、学习算法、输入数据、输出数据、表示理论}五个元素组成, 可表示为 $(S, S) = \{(\vec{E}x, \vec{E}x)ER, (\vec{X}, \vec{X}), (Y, Y), ET\}$

定义 2 动态模糊机器学习 DFML(Dynamic Fuzzy Machine Learning) 动态模糊机器学习 (\vec{l}, \vec{l}) 是指在动态模糊机器学习空间 (S, S) 中的一个输入数据集 (\vec{X}, \vec{X}) 到一个输出数据集 (Y, Y) 的映射, 可表示为 $(\vec{l}, \vec{l}): (\vec{X}, \vec{X}) \mapsto (Y, Y)$.

定义 3 动态模糊机器学习系统 动态模糊机器学习空间 (S, S) 中的五要素按照一定的学习机制形成的具有学习能力的计算机系统, 称为动态模糊机器学习系统 DFMLS(Dynamic Fuzzy Machine Learning System).

收稿日期: 2008-05-07; 修回日期: 2008-12-10

基金项目: 国家自然科学基金(No. 60775045); 江苏省自然科学基金(No. BK2005027, No. BK2002040)

* 注: (S, S) 中 \leftarrow, \rightarrow 是 S 状态发展变化趋势的直观表示. 文中其他处相同^[11].

定义 4 动态模糊机器学习模型 DFMLM (Dynamic Fuzzy Machine Learning Model) $DFMLM = \{(\underline{S}, \overline{S}), (\underline{L}, \overline{L}), (\underline{u}, \overline{u}), (\underline{y}, \overline{y}), (\underline{p}, \overline{p}), (\underline{I}, \overline{I}), (\underline{O}, \overline{O})\}$, 其中 $(\underline{S}, \overline{S})$ 是被学习部分(动态环境/动态模糊学习空间), $(\underline{L}, \overline{L})$ 是动态学习部分, $(\underline{u}, \overline{u})$ 是 $(\underline{S}, \overline{S})$ 到 $(\underline{L}, \overline{L})$ 的输出, $(\underline{y}, \overline{y})$ 是 $(\underline{L}, \overline{L})$ 到 $(\underline{S}, \overline{S})$ 的动态反馈, $(\underline{p}, \overline{p})$ 是系统学习性能指标, $(\underline{I}, \overline{I})$ 是外界环境对动态模糊机器学习系统的输入, $(\underline{O}, \overline{O})$ 是本系统对外界的输出。

现在对系统采用离散化处理, $(\underline{S}, \overline{S}), (\underline{L}, \overline{L})$ 用状态空间表示, 于是有:

定义 5 动态模糊机器学习模型 DFMLM 描述为:

$$(\underline{x}, \overline{x})(k+1) = G_1((\underline{x}, \overline{x})(k), (\underline{u}, \overline{u})(k), \zeta_1(k)) \quad (1)$$

$$(\underline{y}, \overline{y})(k) = G_2((\underline{x}, \overline{x})(k), (\underline{u}, \overline{u})(k), \zeta_2(k)) \quad (2)$$

$$(\underline{p}, \overline{p})(k) = \sum_{i=1}^k P((\underline{y}, \overline{y})(i)) \quad (3)$$

其中 $(\underline{x}, \overline{x})$ 是 $(\underline{S}, \overline{S})$ 在时刻 k 的状态变量, $(\underline{u}, \overline{u})(k)$ 是 $(\underline{S}, \overline{S})$ 的动态输出, $\zeta_1(k)$ 是状态方程中的随机干扰, $(\underline{y}, \overline{y})(k)$ 是 $(\underline{L}, \overline{L})$ 的动态反馈输出, $\zeta_2(k)$ 是观察随机误差. k 表示时刻, 只取整数值. 假定其中的向量全部为有限维状态变量. $(\underline{p}, \overline{p})$ 表示系统学习性能指标, $P(i)$ 是一个标量函数, 表示时刻 i 时的系统学习性能^[12].

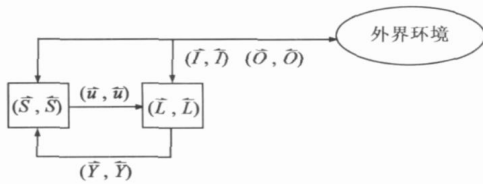


图1 动态模糊机器学习模型框图

3 动态模糊机器学习算法

在机器学习中有许多数据是非线性高维数据, 这给机器学习中的数据预处理工作带来了一定的困难, 因此需要对高维数据进行降维(维数约简)。

3.1 局部线性嵌入法 (locally linear embedding, 简称 LLE)

局部线性嵌入法是 Saul 和 Roweis 于 2000 年提出的一种非线性降维方法^[13], 它的基本思想是将全局非线性转化为局部线性, 而互相重叠的局部邻域能够提供全局结构的信息, 这样对每个局部进行线性降维后, 再按照某种规则将结果组合在一起, 就能得到低维的全局坐标表示。

设在动态模糊高维空间 $(\underline{R}^D, \overline{R}^D)$ 中有动态模糊数据集 $(\underline{X}, \overline{X}) = \{(\underline{x}_1, \overline{x}_1), (\underline{x}_2, \overline{x}_2), \dots, (\underline{x}_n, \overline{x}_n)\}$. 该方法希望通过降维 $(\underline{F}, \overline{F})$ 将 $(\underline{X}, \overline{X})$ 映射到一个相对低维的动态模糊空间 $(\underline{S}^d, \overline{S}^d)$ 中, 求出与其对应的低维动态模

糊向量 $(\underline{Y}, \overline{Y}) = \{(\underline{y}_1, \overline{y}_1), (\underline{y}_2, \overline{y}_2), \dots, (\underline{y}_n, \overline{y}_n)\}$, 同时尽可能地保持原动态模糊数据所携带的信息, 即保持动态模糊数据集的拓扑结构(通过每点的邻域关系确定). 有效的办法是, 在小的局部近似的用线性结构来替代, 也就是在每个局部将非线性降维 $(\underline{F}, \overline{F})$ 用线性降维来近似, 这样整体上的非线性降维问题就转化为局部的线性降维问题。

定义 6 动态模糊降维问题的模型为 $((\underline{X}, \overline{X}), (\underline{F}, \overline{F}))$, 其中: 映射 $(\underline{F}, \overline{F}): (\underline{X}, \overline{X}) \rightarrow (\underline{Y}, \overline{Y})$. 称 $(\underline{F}, \overline{F})$ 为动态模糊数据集 $(\underline{X}, \overline{X})$ 到 $(\underline{Y}, \overline{Y})$ 的降维。

定义 7 称映射 $(\underline{f}, \overline{f}): (\underline{Y}, \overline{Y}) \rightarrow (\underline{T}, \overline{T}) \subset (\underline{R}^D, \overline{R}^D)$ 为嵌入映射. 考虑任意取定的点 $(\underline{x}, \overline{x}) \in (\underline{X}, \overline{X})$, $\{(\underline{x}_l^+, \overline{x}_l^+), l = 1, 2, \dots, k\}$ 表示 $(\underline{x}, \overline{x})$ 的 k 近邻域(即 $(\underline{X}, \overline{X})$ 中离 $(\underline{x}, \overline{x})$ 最近的 k 个点组成的动态模糊集合), 在这个局部 $(\underline{F}, \overline{F})$ 的线性近似为 $(\underline{F}^l, \overline{F}^l)$, 令 $(\underline{y}_b, \overline{y}_b) = (\underline{F}^l, \overline{F}^l)(\underline{x}_b^+, \overline{x}_b^+)$, $(\underline{y}_i^+, \overline{y}_i^+) = (\underline{F}^l, \overline{F}^l)(\underline{x}_i^+, \overline{x}_i^+)$, 则有如下定理:

$$\text{定理 1 若 } (\underline{x}_b^+, \overline{x}_b^+) = \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{x}_l^+, \overline{x}_l^+),$$

$$\text{则 } (\underline{y}_b^+, \overline{y}_b^+) = \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{y}_l^+, \overline{y}_l^+)$$

其中 $(\underline{w}_b^l, \overline{w}_b^l)$ ($l = 1, 2, \dots, n, b = 1, 2, \dots, k$) 是线性组合的动态模糊系数, 满足 $\sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) = (\underline{1}, \overline{1})$.

[证明略]

这说明对局部进行线性降维是不变的. 将 $(\underline{x}_b^+, \overline{x}_b^+)$ 用 $\{(\underline{x}_l^+, \overline{x}_l^+), l = 1, 2, \dots, k\}$ 的线性组合来近似: $(\underline{x}_b^+, \overline{x}_b^+) \sim \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{x}_l^+, \overline{x}_l^+)$, 于是动态模糊系数 $\{(\underline{w}_b^l, \overline{w}_b^l)\}$ 刻画了数据结构在降维下的不变特征, 然后根据 $(\underline{y}_b^+, \overline{y}_b^+) = (\underline{F}^l, \overline{F}^l)(\underline{x}_b^+, \overline{x}_b^+) \sim (\underline{F}^l, \overline{F}^l) \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{x}_l^+, \overline{x}_l^+) = \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{y}_l^+, \overline{y}_l^+)$, 通过 $\{(\underline{w}_b^l, \overline{w}_b^l)\}$ 提供的信息求出 $(\underline{y}_b^+, \overline{y}_b^+)$.

LLE 方法的核心是: 寻求 k 维动态模糊向量 $(\underline{W}_b, \overline{W}_b) = ((\underline{w}_b^1, \overline{w}_b^1), (\underline{w}_b^2, \overline{w}_b^2), \dots, (\underline{w}_b^k, \overline{w}_b^k))$, 使得

$$(\underline{Q}, \overline{Q})(\underline{W}, \overline{W}) = \left\| (\underline{x}_b^+, \overline{x}_b^+) - \sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) (\underline{x}_l^+, \overline{x}_l^+) \right\|^2 \quad (4)$$

最小, 其中: $\sum_{l=1}^k (\underline{w}_b^l, \overline{w}_b^l) = (\underline{1}, \overline{1}), (\underline{w}_b^l, \overline{w}_b^l) = (\underline{0}, \overline{0})$, 以及当 $(\underline{x}_j^+, \overline{x}_j^+) \notin \{(\underline{x}_l^+, \overline{x}_l^+), l = 1, 2, \dots, k\}$ 时, $(\underline{w}_b^l, \overline{w}_b^l) = (\underline{0}, \overline{0})$. $(\underline{W}_b, \overline{W}_b)$ 记录了点 $(\underline{x}_b^+, \overline{x}_b^+)$ 的邻域信息, 即每点的局部拓扑结构, 记它们的全体构成的动态模糊矩阵为 $(\underline{W}, \overline{W})$. 此时选择适当的维数 d , 做 $(\underline{X}, \overline{X})$ 的 d 维嵌入, 即求 $(\underline{Y}, \overline{Y}) = ((\underline{y}_1, \overline{y}_1), (\underline{y}_2, \overline{y}_2), \dots, (\underline{y}_n, \overline{y}_n))$

$\in (\mathbb{R}^{d \times n}, \mathbb{R}^{D \times n})$, 使其满足

$$(\vec{L}, \vec{L}') = ((y_1^*, y_1^*), (y_2^*, y_2^*), \dots, (y_n^*, y_n^*))$$

$$= \min_{(\vec{Y}, \vec{Y}')} \sum_{i=1}^n \|(\vec{y}_i, \vec{y}_i') - (\vec{Y}, \vec{Y}')(\vec{W}_b, \vec{W}_b')^T\|^2$$

$$= \min_{(\vec{Y}, \vec{Y}')} \text{tr}((\vec{Y}, \vec{Y}')(\vec{M}, \vec{M}')(\vec{Y}, \vec{Y}')^T)$$

$$(\vec{M}, \vec{M}') = ((\vec{1}, \vec{1}) - (\vec{W}, \vec{W}')^T)^T((\vec{1}, \vec{1}) - (\vec{W}, \vec{W}')^T)$$

由于 (\vec{W}, \vec{W}') 是由动态模糊数据集 (\vec{X}, \vec{X}') 解出的, 故当动态模糊数据集 (\vec{X}, \vec{X}') 受噪声干扰时, 计算对噪声非常敏感, 特别是当 $(\vec{X}, \vec{X}')^T(\vec{X}, \vec{X}')$ 的特征值较小时, 甚至得不到需要的结果.

令 $(\vec{x}_b^t, \vec{x}_b^{t'})' = (\vec{x}_b^t, \vec{x}_b^{t'}) + (\vec{\epsilon}_i, \vec{\epsilon}_i')$, $(i = 1, 2, \dots, k)$, 代表相应的受噪声干扰的点, 及相应的

$$(\vec{x}_b^t, \vec{x}_b^{t'})' = \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'})' (\vec{x}_b^t, \vec{x}_b^{t'})', \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'})' = (\vec{1}, \vec{1});$$

记 $(\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'})) = ((\vec{x}_b^t, \vec{x}_b^{t'}), (\vec{x}_b^t, \vec{x}_b^{t'}), \dots, (\vec{x}_b^t, \vec{x}_b^{t'}))$ 为 $(\vec{x}_b^t, \vec{x}_b^{t'})$ 的 k 近邻域动态模糊集合,

$$(\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'}))' = ((\vec{x}_b^t, \vec{x}_b^{t'})', (\vec{x}_b^t, \vec{x}_b^{t'}), \dots, (\vec{x}_b^t, \vec{x}_b^{t'}))'$$

为 $(\vec{x}_b^t, \vec{x}_b^{t'})$ 的 k 近邻域动态模糊集合, 及

$$(\vec{W}_b, \vec{W}_b')' = ((\vec{w}_b^1, \vec{w}_b^{1'})', (\vec{w}_b^2, \vec{w}_b^{2'})', \dots, (\vec{w}_b^k, \vec{w}_b^{k'})')$$

$$\text{则 } (\vec{x}_b^t, \vec{x}_b^{t'}) = (\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'}))(\vec{W}_b, \vec{W}_b'),$$

$$(\vec{x}_b^t, \vec{x}_b^{t'})' = (\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'}))'(\vec{W}_b, \vec{W}_b')$$

定理 2 若各点噪声之间, 不同维数之间, 以及 (\vec{W}_b, \vec{W}_b') 与噪声之间是相互独立的, 各点噪声是同均值, 为 $(\vec{0}, \vec{0})'$, 同方差的. 那么对于 $\delta((\vec{W}, \vec{W}') - (\vec{W}_b, \vec{W}_b'))$ 有如下估计:

$$E \|\delta((\vec{W}_b, \vec{W}_b') - (\vec{W}, \vec{W}'))\|^2 \leq \frac{k(k+1)}{(\bar{\lambda}, \bar{\lambda}')_{\min} l} E \|(\vec{W}, \vec{W}')\|^2$$

其中, $\|\cdot\|$ 为欧几里得 (Euclid) 范数;

$$(\vec{\delta}, \vec{\delta}')^2 = \sum_{i=1}^D (\delta_i, \delta_i')^2,$$

$(\delta_i, \delta_i') = \text{Var}(\vec{\epsilon}_i, \vec{\epsilon}_i')$ ($i = 1, 2, \dots, D$), $(\vec{\epsilon}_i, \vec{\epsilon}_i')$ 代表 $(\vec{\epsilon}, \vec{\epsilon}')$ 的第 i 个分量, $l = \text{rank}((\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'})))$, $(\bar{\lambda}, \bar{\lambda}')_{\min}$ 为 $(\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'}))^T, (\vec{U}(\vec{x}_b^t), \vec{U}(\vec{x}_b^{t'}))$ 的最小非零特征值.

[证明略].

由本定理可以得出, 在邻域大小 k 已知的情形之下, (\vec{W}, \vec{W}') 的误差主要由 3 个因素决定:

- ① 噪声干扰的影响, 即 $(\vec{\delta}, \vec{\delta}')$ 的大小;
- ② 邻域的影响, 即 $(\bar{\lambda}, \bar{\lambda}')_{\min}$ 和秩 l 的大小;
- ③ 权重能量的影响, 即 $\|(\vec{W}, \vec{W}')\|$ 的大小.

3.2 改进的局部线性嵌入方法 ILLE/ Improved locally linear embedding

如何确定邻域 k 的大小是个很重要的问题, 如果 k 太小, 映射将会导致子流形脱离连续流形并无法反映任何全局性质, 而如果 k 太大, 整个数据集将被视为局部邻域则映射将会失去非线性特征^[14]. 改进的局部线

性嵌入方法 ILLE 利用文献[15]提出的 DGSOM (Diffusion and Growing Self-Organizing Map) 方法来定义邻域关系, 例如, 如果一个点在 DGSOM 映射中有五个相关联的点, 则这五个数据点均被看作是那个点的邻居, 因此每个样本点的邻域即 k 的大小是不同的, 克服了传统的 LLE 方法中每个点的邻域大小都固定不变的缺点. 另外, 改进的局部线性嵌入方法通过减少噪声干扰和权重能量来减弱噪声的影响, 并且此方法对邻域具有一定的适应性, 可较好地克服对邻域的敏感性问题, 从而较好地解决了上述的 3 个影响因素带来的误差问题.

ILLE 方法的核心是: 寻求 k 维动态模糊向量 $(\vec{W}_b, \vec{W}_b') = ((\vec{w}_b^1, \vec{w}_b^{1'}), (\vec{w}_b^2, \vec{w}_b^{2'}), \dots, (\vec{w}_b^k, \vec{w}_b^{k'}))$, 使得

$$(\vec{Q}, \vec{Q}')(\vec{W}_b, \vec{W}_b') = \left\| (\vec{x}_b^t, \vec{x}_b^{t'}) - \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'}) (\vec{x}_b^t, \vec{x}_b^{t'}) \right\|^2$$

$$+ (\eta, \eta')(\vec{g}, \vec{g}') \left(\left\| \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'}) \right\| \right)$$

$$((\eta, \eta') \geq (\vec{0}, \vec{0}), \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'}) = (\vec{1}, \vec{1})) \quad (5)$$

最小, 再得到相应的动态模糊拓扑结构矩阵 (\vec{W}, \vec{W}') , 最后可得 (\vec{M}, \vec{M}') .

ILLE 通过引入动态模糊正则项 (\vec{g}, \vec{g}') $\left(\left\| \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'}) \right\| \right)$, 避免了 LLE 的病态性, 减弱了噪声干扰的影响, 它的选择要满足两点:

- (1) 使得式(5)的求解稳健;
- (2) 使得式(5)的 $\left\| \sum_{l=1}^k (\vec{w}_b^l, \vec{w}_b^{l'}) \right\|$ 尽可能小, 一般地,

可取 $(\vec{g}, \vec{g}')(\vec{x}_b^t, \vec{x}_b^{t'}) = (\vec{x}_b^t, \vec{x}_b^{t'})^2$. (η, η') 为改进因子, 它的选取与邻域点数 k 、动态模糊数据集的维数 D 及动态模糊数据集的尺度有关, 通过实际计算可知 (η, η') 的选择对结果的影响不是很大.

算法 3.2.1 ILLE 算法

输入:

$$(\vec{X}, \vec{X}') = \{(\vec{x}_1^t, \vec{x}_1^{t'}), (\vec{x}_2^t, \vec{x}_2^{t'}), \dots, (\vec{x}_n^t, \vec{x}_n^{t'})\} \subset (\mathbb{R}^{D \times n}, \mathbb{R}^{D \times n}),$$

$$(\vec{x}_i^t, \vec{x}_i^{t'}) \in (\mathbb{R}^D, \mathbb{R}^D), (i = 1, 2, \dots, n)$$

输出:

$$(\vec{Y}, \vec{Y}') = \{(\vec{y}_1^t, \vec{y}_1^{t'}), (\vec{y}_2^t, \vec{y}_2^{t'}), \dots, (\vec{y}_n^t, \vec{y}_n^{t'})\} \subset (\mathbb{R}^{d \times n}, \mathbb{R}^{d \times n}),$$

$$(\vec{y}_i^t, \vec{y}_i^{t'}) \in (\mathbb{R}^d, \mathbb{R}^d), (i = 1, 2, \dots, n)$$

Begin

For each do $(\vec{x}_i^t, \vec{x}_i^{t'})$

选择邻域: 根据 DGSOM 方法来确定每个点 $(\vec{x}_i^t, \vec{x}_i^{t'})$ 的 k 个邻点 $\{(\vec{x}_b^l, \vec{x}_b^{l'}), l = 1, 2, \dots, k\}$, 则每个 $(\vec{x}_i^t, \vec{x}_i^{t'})$ 的邻域大小 k 可能是不同的. 用 $\{(\vec{x}_b^l, \vec{x}_b^{l'}), l = 1, 2, \dots, k\}$ 的线性组合来表示 $(\vec{x}_i^t, \vec{x}_i^{t'})$, 即用经过 $\{(\vec{x}_b^l, \vec{x}_b^{l'}), l = 1, 2, \dots, k\}$ 的超平面来近似流形在 $(\vec{x}_i^t, \vec{x}_i^{t'})$ 附近的局部块.

(2) 求取权重: 对每个 $(\vec{x}_b^l, \vec{x}_b^{l'})$, 寻求

$(\vec{W}_i, \vec{W}_i) = ((\vec{w}_i^1, \vec{w}_i^1), (\vec{w}_i^2, \vec{w}_i^2), \dots, (\vec{w}_i^k, \vec{w}_i^k))$ 使
 $(\vec{Q}, \vec{Q})(\vec{W}_i, \vec{W}_i) = \left\| (\vec{x}_i, \vec{x}_i) - \sum_{j=1}^k \vec{w}_j^i (\vec{x}_i^j, \vec{x}_i^j) \right\|^2$
 $+ (\vec{r}, \vec{r}) \left\| \sum_{j=1}^k \vec{w}_j^i (\vec{x}_i^j, \vec{x}_i^j) \right\|^2$ 最小, 最后求得动态模糊拓扑
 结构矩阵 (\vec{W}, \vec{W}) .

(3) 降维映射: 最小化损失函数

$(\vec{L}, \vec{L})((\vec{y}_1, \vec{y}_1), (\vec{y}_2, \vec{y}_2), \dots, (\vec{y}_n, \vec{y}_n))$
 $= \min_{(\vec{Y}, \vec{Y})} \sum_{i=1}^n \left\| (\vec{y}_i, \vec{y}_i) - (\vec{Y}, \vec{Y})(\vec{W}_i, \vec{W}_i)^T \right\|^2$
 $= \min_{(\vec{Y}, \vec{Y})} \text{tr}((\vec{Y}, \vec{Y})(\vec{M}, \vec{M})(\vec{Y}, \vec{Y})^T)$
 $(\vec{M}, \vec{M}) = ((\vec{1}, \vec{1}) - (\vec{W}, \vec{W})^T)^T ((\vec{1}, \vec{1}) - (\vec{W}, \vec{W})^T)$
 其中:
 $(\vec{Y}, \vec{Y}) = \{(\vec{y}_1, \vec{y}_1), (\vec{y}_2, \vec{y}_2), \dots, (\vec{y}_n, \vec{y}_n)\} \in (\vec{R}^{l \times n}, \vec{R}^{l \times n})$,
 $(\vec{y}_i, \vec{y}_i) \in (\vec{R}^l, \vec{R}^l)$, $(i = 1, 2, \dots, n)$, 为了求解的唯一
 性, 添加约束条件, 对 $(\vec{y}_i, \vec{y}_i) \in (\vec{R}^l, \vec{R}^l)$, $(i = 1, 2, \dots,$
 $n)$, 加上约束 $\sum_{i=1}^n (\vec{y}_i, \vec{y}_i) = (\vec{0}, \vec{0})$ 及 $(\vec{Y}, \vec{Y})^T (\vec{Y}, \vec{Y}) = (\vec{I},$
 $\vec{I})^{d \times d}$, $(\vec{I}, \vec{I})^{d \times d}$ 为动态模糊单位矩阵.

ILLE 各步的计算复杂度为: 第 (1) 步的计算量为 O
 $(Dn)^2$, 第 (2) 步的计算量为 $O(Dnk^3)$, 第 (3) 步的计算量
 为 $O(dn^2)$. ILLE 的计算复杂度和 LLE 的计算复杂度是
 差不多的, 但对噪声干扰的抵抗能力更强.

4 动态模糊机器学习算法的稳定性分析

考虑动态模糊机器学习系统由如下动态模糊规则
 (dynamic fuzzy rules) 组建^[16]:

R^l : IF $(\vec{x}, \vec{x})(k)$ is $(\vec{A}_1^l, \vec{A}_1^l)$ and $(\vec{x}, \vec{x})(k-1)$ is $(\vec{A}_2^l,$
 $\vec{A}_2^l)$ and...and $(\vec{x}, \vec{x})(k-n+1)$ is $(\vec{A}_n^l, \vec{A}_n^l)$ and (\vec{u}, \vec{u})
 (k) is (\vec{B}_l, \vec{B}_l)

THEN $= (\vec{x}, \vec{x})^l(k+1) = (\vec{A}_p, \vec{A}_p)(\vec{X}, \vec{X})(k) + (\vec{b}_p, \vec{b}_p)$
 $(\vec{u}, \vec{u})(k)$, $l = 1, 2, \dots, m$ (6)

其中:

$(\vec{X}, \vec{X})(k) = ((\vec{x}, \vec{x})(k), (\vec{x}, \vec{x})(k-1), \dots, (\vec{x}, \vec{x})(k-n+1))^T \in (\vec{U}, \vec{U})$ 为系统状态向量; $(\vec{A}_i^l, \vec{A}_i^l) = ((\vec{a}_i^1,$
 $\vec{a}_i^1), (\vec{a}_i^2, \vec{a}_i^2), \dots, (\vec{a}_i^n, \vec{a}_i^n), (\vec{a}_i^1, \vec{a}_i^1), (i = 1, 2, \dots, n)$ 为
 $\Delta V[(\vec{X}, \vec{X})(k)] = (\vec{X}, \vec{X})^T(k+1)(\vec{P}, \vec{P})(\vec{X}, \vec{X})(k+1) - (\vec{X}, \vec{X})^T(k)(\vec{P}, \vec{P})(\vec{X}, \vec{X})(k)$

$$= (\vec{X}, \vec{X})(k) \left[\left(\frac{\sum_{p=1}^m (\vec{A}_p, \vec{A}_p)^T (\vec{v}, \vec{v})^p}{\sum_{p=1}^m (\vec{v}, \vec{v})^p} \right) (\vec{P}, \vec{P}) \left(\frac{\sum_{p=1}^m (\vec{A}_p, \vec{A}_p) (\vec{v}, \vec{v})^p}{\sum_{p=1}^m (\vec{v}, \vec{v})^p} \right) - (\vec{P}, \vec{P}) \right] (\vec{X}, \vec{X})(k)$$

$$= \frac{\sum_{p=1}^m \sum_{q=1}^m \left\{ (\vec{v}, \vec{v})^p (\vec{v}, \vec{v})^q (\vec{X}, \vec{X})^T(k) [(\vec{A}_p, \vec{A}_p)^T (\vec{P}, \vec{P}) (\vec{A}_q, \vec{A}_q) - (\vec{P}, \vec{P})] (\vec{X}, \vec{X})(k) \right\}}{\sum_{p=q=1}^m (\vec{v}, \vec{v})^p (\vec{v}, \vec{v})^q}$$

由式(9)及 $(\vec{v}, \vec{v})^p \geq (\vec{0}, \vec{0})$, 得到 $\Delta V[(\vec{X}, \vec{X})(k)] < (\vec{0}, \vec{0})$. 根据 Lyapunov 稳定性定理, 本定理得证.

相应的动态模糊常数: $(\vec{A}_i^l, \vec{A}_i^l) \in (\vec{R}, \vec{R})^{n \times n}$, $(\vec{B}_i^l, \vec{B}_i^l) \in (\vec{R}, \vec{R})^{n \times n}$ 为动态模糊常数矩阵. 称每条动态模糊规则
 对应的线性状态方程为动态模糊机器学习子系统.

对于学习过程中的噪声干扰问题, 我们采用文献
 [17] 提出的鲁棒因子 (robust factor) 来减少噪声干扰:

$$(\vec{\varphi}, \vec{\varphi})((\vec{e}_i, \vec{e}_i), (\vec{c}_i, \vec{c}_i)) = \frac{\partial (\vec{\phi}, \vec{\phi})((\vec{e}_i, \vec{e}_i), (\vec{c}_i, \vec{c}_i))}{\partial ((\vec{e}_i, \vec{e}_i))} \Big|_{(\vec{e}_i, \vec{e}_i)} \quad (7)$$

其中: $(\vec{e}_i, \vec{e}_i) = (\vec{y}_i, \vec{y}_i) - (\vec{y}_i, \vec{y}_i)$, (\vec{y}_i, \vec{y}_i) 为期望输出
 (expectation output), (\vec{y}_i, \vec{y}_i) 为实际输出, (\vec{c}_i, \vec{c}_i) 为在
 (\vec{x}_i, \vec{x}_i) 附近对偏差 (\vec{e}_i, \vec{e}_i) 作统计估计所得的置信区间的
 中心值. $(\vec{\varphi}, \vec{\varphi})((\vec{e}_i, \vec{e}_i), (\vec{c}_i, \vec{c}_i)) \geq (\vec{0}, \vec{0})$ 是关于 $(\vec{e}_i,$
 $\vec{e}_i)$ 的非负可积函数, 且

$$\max(\vec{\varphi}, \vec{\varphi})((\vec{e}_i, \vec{e}_i), (\vec{c}_i, \vec{c}_i)) = (\vec{\varphi}, \vec{\varphi})((\vec{c}_i, \vec{c}_i), (\vec{c}_i, \vec{c}_i));$$

$$(\vec{\phi}, \vec{\phi})((\vec{0}, \vec{0}), (\vec{c}_i, \vec{c}_i)) = (\vec{0}, \vec{0})$$

记

$$(\vec{X}, \vec{X})(k) = ((\vec{x}, \vec{x})(k), (\vec{x}, \vec{x})(k-1), \dots, (\vec{x}, \vec{x})(k-n+1))^T \in (\vec{U}, \vec{U})$$

及

$$(\vec{A}_p, \vec{A}_p) = \begin{bmatrix} (\vec{a}_p^1, \vec{a}_p^1) & (\vec{a}_p^2, \vec{a}_p^2) & \dots & (\vec{a}_p^{n-1}, \vec{a}_p^{n-1}) & (\vec{a}_p^n, \vec{a}_p^n) \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix}$$

则:

$$(\vec{X}, \vec{X})(k+1) = \frac{\sum_{p=1}^m (\vec{A}_p, \vec{A}_p) (\vec{X}, \vec{X})(k) (\vec{v}, \vec{v})^p}{\sum_{p=1}^m (\vec{v}, \vec{v})^p} \quad (8)$$

定理 3^[18] 动态模糊机器学习系统表达式 (8) 的
 平衡点 $(\vec{0}, \vec{0})$ 是全局渐近稳定的 (asymptotically stable)
 的, 若存在一正定动态模糊矩阵 (\vec{P}, \vec{P}) , 使得

$$(\vec{A}_p, \vec{A}_p)^T (\vec{P}, \vec{P}) (\vec{A}_p, \vec{A}_p) - (\vec{P}, \vec{P}) < (\vec{0}, \vec{0})$$

$$(p = 1, 2, \dots, n) \quad (9)$$

证明 考虑如下的 Lyapunov 函数:

$$V[(\vec{X}, \vec{X})(k)] = (\vec{X}, \vec{X})^T(k) (\vec{P}, \vec{P}) (\vec{X}, \vec{X})(k)$$

其中 (\vec{P}, \vec{P}) 为一正定动态模糊矩阵, 则有:

(10)

5 仿真实验

例 1 太阳黑子问题: 数据取自文献[19], 从 1749 年到 1924 年, 共有 176 个数据, 其中前 100 个数据作为训练样本数据, 其余的用作检验数据.

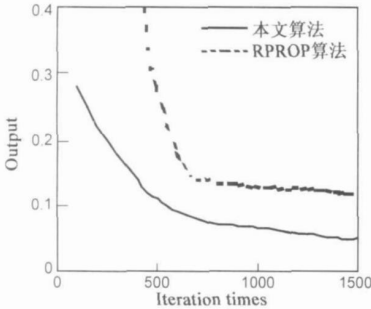


图2 两种算法的误差与迭代步数比较

图 2 是神经网络的弹性 BP 算法 RPROP 和本文算法的误差与迭代步数比较图. RPROP 算法当误差 $(\vec{e}_k, \vec{e}_k) = (\overline{0.4}, \overline{0.4})$ 时陷入局部极小, 所需迭代步数约为 1660; 本文算法当误差 $(\vec{e}_k, \vec{e}_k) < (\overline{0.4}, \overline{0.4})$ 时所需的迭代步数为 455. 当误差 $(\vec{e}_k, \vec{e}_k) = (\overline{0.034}, \overline{0.034})$, 训练基本稳定, 所需迭代步数约为 1050. 图 3 是置初始值 $(\vec{u}_0, \vec{u}_0)(k) = (\vec{0}, \vec{0})$, 增益学习系数 $\alpha = 0.3$, 修正系数 $\beta = 0.2$, 最大容忍误差 $(\vec{\xi}, \vec{\xi}) = (\overline{0.5}, \overline{0.5})$, 可接受误差 $(\vec{\delta}, \vec{\delta}) = (\overline{0.005}, \overline{0.005})$, 当误差 $(\vec{e}_k, \vec{e}_k) = (\overline{0.034}, \overline{0.034})$ 时实际值与预测值的比较图.

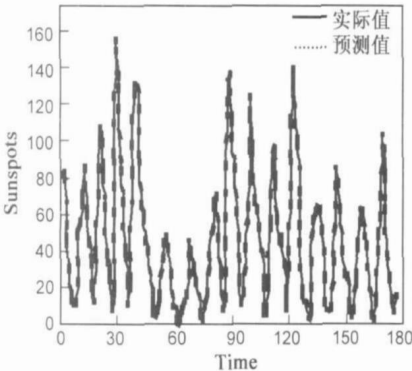


图3 本文算法的预测值与实际值比较

例 2 某公司一段时期的每日股票收盘价时间序列预测本例所用数据取自文献[19], 在所用的 250 个数据中, 前 150 个数据作为训练样本数据, 其余 100 个用作检验数据. 置初始值 $(\vec{u}_0, \vec{u}_0)(t) = (\vec{0}, \vec{0})$, 增益学习系数 $\alpha = 0.3$, 修正系数 $\beta = 0.3$, 最大容忍误差 $(\vec{\xi}, \vec{\xi}) = (\overline{0.5}, \overline{0.5})$, 可接受误差 $(\vec{\delta}, \vec{\delta}) = (\overline{0.005}, \overline{0.005})$.

图 4 是本文算法与神经网络的弹性 BP 算法 RPROP 的误差和迭代步数比较. RPROP 算法当性能指标 $(\vec{p}, \vec{p})(k) = (\overline{0.013}, \overline{0.013})$ 时陷入局部极小, 所需迭

代步数为 146, 本文算法当性能指标小于 $(\overline{0.013}, \overline{0.013})$ 时所需迭代步数为 82; 得到满足精度要求的结果所需迭代步数约为 500. 图 5 是实际值与学习结果的比较, 其中 K 为迭代次数.

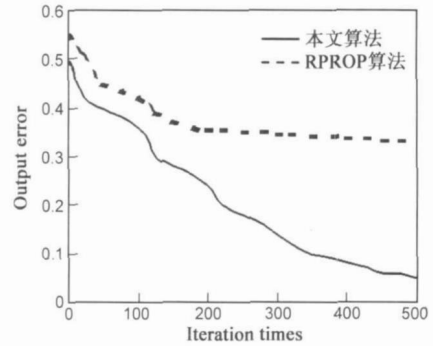


图4 两种算法的误差与迭代步数比较

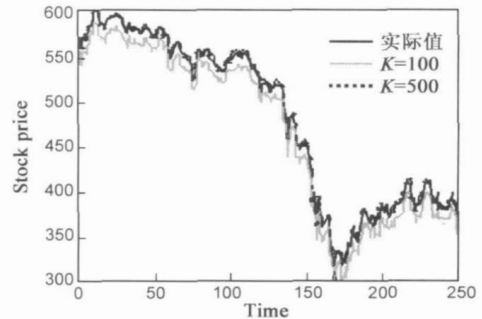


图5 实际值与学习结果的比较

6 总结与展望

由 Simon 对学习的阐述可知, 在整个机器学习活动及机器学习系统中动态模糊性是普遍存在的, 而现有的机器学习方法都不能很好地解决动态模糊性问题, 因此, 本文基于动态模糊集 (DFS) 给出了一类动态模糊机器学习算法, 进一步丰富了动态模糊机器学习的研究内容.

参考文献:

- [1] Samuel L. Some studies in machine learning using the game of checkers, Part II [J]. IBM Journal Research and Development, 1967, 11(4): 601- 618.
- [2] Vapnik V. The Nature of Statistical Learning Theory[M]. New York: Springer Verlag, 1995.
- [3] Sutton R, Barto A. Reinforcement Learning: An Introduction [M]. Cambridge, MA: The MIT Press, 1998.
- [4] Kaelbling L, Littman M, Moore A. Reinforcement learning: A survey[J]. Journal of AI Research, 1996, 4: 237- 285.
- [5] Silva V D, Tenenbaum J B. Global versus local methods in nonlinear dimensionality reduction[J]. Neural Information Processing Systems(NIPS 2002), 2003, 15: 705- 712.
- [6] Blum A, Mitchell T. Combining labeled and unlabeled data with

- co-training [A]. Proc 11th Annual Conf Computational Learning Theory [C]. Madison, WI, 1998. 92– 100.
- [7] Ratsch G, Onoda T, Muller K. Soft margins for AdaBoost [J]. Machine Learning, 2001, 42(3) : 287– 320.
- [8] Dietterich T G, Lathrop R H, Lozano-Perez T. Solving the multiple instance problem with axisparallel rectangles [J]. Artificial Intelligence, 1997, 89(1– 2) : 31– 71.
- [9] LI Far zhang. Dynamic Fuzzy Logic And It Applications [M]. Nova Science Publishers, 2008, 5.
- [10] 张静. 动态模糊机器学习模型及其应用研究 [D]. 江苏苏州: 苏州大学硕士学位论文, 2007. 6.
- [11] 李凡长. 动态模糊逻辑引论 [M]. 云南: 云南科技出版社, 2005.
- [12] 李凡长. 协调机器学习的稳定性研究 [J]. 小型微型计算机系统, 2002, 23(3) : 314– 317.
- [13] Roweis S T, Lawrence K S. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500) : 2323– 2326.
- [14] Xiao Jian, Zhou Zong tan, Hu De wen, Yin Jurr song, Chen Shuang. Self organized locally linear embedding for nonlinear dimensionality reduction [J]. ICNC 2005(3610) : 101– 109.
- [15] Chen S, Zhou Z T, Hu D W. Diffusion and Growing self organizing map: A nitric oxide based neural model [J]. ISNN, 2004, (1) : 199– 204.
- [16] ZHANG Jing, LI Far zhang. Machine learning model based on dynamic fuzzy sets (DFS) and its validation [J]. Journal of Computational Information Systems, 2005, 1(4) : 871– 877.
- [17] 李杰星, 章云, 符曦. 对含噪声数据的一种鲁棒学习算法 [J]. 数值计算与计算机应用, 2000, 6(2) : 112– 120.
- [18] Hassan K Khalil. Nonlinear Systems (Third Edition) [M]. Beijing: Beijing Publishing House of Electronics Industry, 2005.
- [19] 杨叔子, 吴雅. 时间序列分析的工程应用 [M]. 武汉: 华中理工大学出版社, 1994.

作者简介:



谢琳女, 1970 年出生于江苏苏州, 苏州大学计算机科学与技术学院硕士, 主要研究方向为人工智能、机器学习、动态模糊逻辑。

E-mail: xie.lin_2007@126.com



李凡长男, 1964 年出生于云南宣威, 苏州大学教授, 中国人工智能学会理事, 中国人工智能学会的机器学习专委会常务委员, 中国计算机学会高级会员, 中国计算机学会的理论计算机科学专委会委员, 《IEEE Transactions on Fuzzy Systems》等国际国内重要学术期刊的审稿人, 2002 年入选江苏省“333 工程”培养计划, 2008 年被评为江苏省有突出贡献的中青年专家。长期从事计算机科学与技术的教学和科研工作, 主要研究领域: 人工智能、机器学习、动态模糊逻辑、李群在计算机科学中的应用、多 Agent 系统理论及应用等, 先后承担和参与 11 项省部、国家自然科学基金及国家 863 计划项目, 发表期刊论文 150 余篇, 出版专著 4 部, 教材 4 部, 曾获省级科技进步奖二等奖 1 项。他是动态模糊逻辑及李群机器学习的主要提出者。

E-mail: lfzh@suda.edu.cn