

# 基于粒计算的非线性感知机

陈玉明<sup>1,2</sup>, 董建威<sup>1</sup>

(1. 厦门理工学院计算机与信息工程学院, 厦门 361024; 2. 易成功(厦门)信息科技有限公司, 厦门 361024)

**摘要:** 感知机是模式识别领域的二分类判别模型, 具有简单、线性和运算效率高的优点, 也是众多分类器的基础。然而, 感知机无法表达复杂的非线性映射, 难以处理非线性可分数据。针对感知机的固有缺陷, 结合粒计算特性, 本文提出了一种新型的感知机分类模型——粒感知机。通过粒计算理论, 样本在单特征上的粒化形成粒子, 多特征上的粒化构造成粒向量; 进一步定义粒感知机模型, 设计粒感知机策略, 提出粒感知机学习算法。为了求解粒感知机的最优解, 证明了粒感知机损失函数的导数形式, 设计了梯度下降算法, 并从收敛速度、非线性能力与分类精度多方面进行了实验比较, 结果表明所提出的粒感知机模型具有收敛速度快与非线性处理数据的能力。

**关键词:** 感知机; 粒计算; 非线性分类; 粒感知机; 收敛速度

**中图分类号:** TP181      **文献标志码:** A

## Non-linear Perceptron Based on Granular Computing

CHEN Yumin<sup>1,2</sup>, DONG Jianwei<sup>1</sup>

(1. School of Computer and Information Engineering, Xiamen University of Technology, Xiamen 361024, China; 2. E-success (Xiamen) Information Technology Co., Ltd, Xiamen 361024, China)

**Abstract:** Perceptron is a binary classification model in the field of pattern recognition, which has the advantages of simplicity, linearity and high computational efficiency, and it is also the basis of many classifiers. However, the perceptron cannot express complex nonlinear mapping and is difficult to process nonlinear data. Aiming at the inherent defects of perceptron and combining the characteristics of granular computing, we propose a new perceptron classification model—Granular perceptron. According to the theory of granular computing, the granulation of samples on single feature forms granules, and the granulation on multiple features constructs a granular vector. Further, the granular perceptron model is defined, the granular perceptron strategy is designed, and the granular perceptron learning algorithm is proposed. In order to solve the optimal solution of the proposed model, the derivative form of the granular loss function is proved, and its gradient descent algorithm is designed. Finally, some experiments are carried out to compare on the convergence speed, nonlinear ability and classification accuracy. The results show that the proposed model has the ability of fast convergence and nonlinear data processing.

**Key words:** perceptron; granular computing; nonlinear classification; granular perceptron; convergence speed

## 引 言

美国学者 Zadeh 指出现实世界存在大量模糊现象之后,1996 年,香港学者 Lin<sup>[1]</sup>首次提出了粒计算理论。粒计算是一种新的知识表示方法,研究事物的多层次、多粒度表示,处理模糊、不确定的信息,而且是人类认知机理的研究范畴<sup>[2]</sup>。1999 年,加拿大学者 Yao<sup>[3]</sup>提出了基于邻域关系的粒计算,并在知识挖掘领域进行了应用。21 世纪以后,国内众多学者开始关注粒计算领域。Liu 等<sup>[4]</sup>发表了专著论述粒计算理论,并基于粒计算研发了一个医疗诊断专家系统。苗夺谦等<sup>[5-7]</sup>提出知识的不确定性度量,开创性地将信息熵引入粒计算领域。王国胤等<sup>[8]</sup>讨论了多种粒计算模型及其在数据挖掘中的应用。吴伟志等<sup>[9]</sup>定义了一种粒结构,并应用于概念格。粒计算能够分析与处理复杂数据,将复杂的对象细粒度分解而求解问题,并广泛应用于机器学习领域<sup>[10-11]</sup>。机器学习属于人工智能领域,是机器通过学习,期望能模拟人类的思维。它从大量的数据中寻找和学习其背后的隐藏规律,从而达到模拟和预测的目的。机器学习应用广泛,在搜索引擎<sup>[12]</sup>、人脸识别<sup>[13]</sup>、机器翻译<sup>[14]</sup>和语音识别<sup>[15]</sup>等商业领域中均有较好的应用;同时也为学术研究提供了新方法,在生物医学<sup>[16]</sup>、材料科学<sup>[17]</sup>、医疗诊断<sup>[18]</sup>、隐私安全<sup>[19]</sup>和气象预报<sup>[20]</sup>等领域均有重要的学术建树。1957 年 Rosenblatt<sup>[21]</sup>提出了感知机模型,它是机器学习与深度学习的理论基础。Novikoff<sup>[22]</sup>, Minsky 等<sup>[23]</sup>对感知机进行了理论研究,后续改进的感知机模型包括口袋感知机<sup>[24]</sup>、表决感知机<sup>[25]</sup>和带边缘感知机<sup>[26]</sup>。传统感知机模型具有简单、运算效率高的优点,是神经网络与支持向量机的起源算法,也是构建深度学习模型的基础。然而,传统感知机的非线性数据处理能力较弱,其改进算法中通过引入核函数增强了非线性处理能力,但算法复杂度较高。为了克服感知机非线性处理能力弱的缺点,从粒计算的理论出发,从多个特征角度构造粒子与粒向量,建立一种粒分类器。粒的构造本质上是一种集合,则粒计算是一种集合方式运算。信息化是粒计算常用的技术与方法,然而粒化与分类器结合的例子并不多见。以信息粒化为基础,本文提出了一种新型的感知机分类模型—粒感知机。样本在单特征上的粒化形成粒子,多特征上的粒化粒子组成粒向量,从而构造粒感知机模型。进一步设计粒感知机策略,本文提出粒感知机学习模型,讨论粒感知机收敛性。通过设计粒感知机的梯度下降算法,成功实现了基于信息粒形式的数据拟合与预测。最后进行实验分析,验证感知机模型的正确与有效性。

## 1 感知机的粒化

感知机模型简单、参数较少,难以拟合非线性大数据系统。大数据系统具有量大、高维与非线性的特点。为了更好地处理大数据集,本文对样本进行单特征粒化,形成单特征粒子。1 个样本在某个特征上粒化为 1 个粒子,高维特征上进行粒化则构造出这个样本的粒向量。

设感知机系统为  $U=(X, C, \{d\})$ , 其中样本集合为  $X=(x_1, x_2, \dots, x_n)$ , 条件特征集合为  $C=\{c_1, c_2, \dots, c_m\}$ ,  $d$  为决策特征; 给定单样本  $x \in X$ , 对于单特征  $c \in C$ ,  $v(x, c) \in \mathbf{R}$  表示样本  $x$  在特征  $c$  上的值; 该样本  $x$  对应的决策值为  $y \in \{-1, +1\}$ 。

**定义 1** 设感知机系统为  $U=(X, C, \{d\})$ , 对于样本  $x_i, x_j \in X$ , 单特征  $c \in C$ , 则  $x_i$  与  $x_j$  在单特征  $c$  上的曼哈顿距离为

$$s(x_i, x_j) = |v(x_i, c) - v(x_j, c)| \quad (1)$$

**定义 2** 给定感知机系统  $U=(X, C, \{d\})$ , 对于任一样本  $x_i \in X$  和任一条件特征  $c \in C$ , 则  $x_i$  在特征  $c$  上进行粒化, 形成的条件粒子定义为

$$g_c(x_i) = \{g_c(x_i)_j\}_{j=1}^n = \{r_j\}_{j=1}^n = \{r_1, r_2, \dots, r_n\} \quad (2)$$

式中  $r_j = s_c(x_i, x_j)$  表示样本  $x_i, x_j$  在单特征  $c$  上的距离。易知  $s_c(x_i, x_j) \in \mathbf{R}$ , 因此  $r_j \in \mathbf{R}$ 。粒子由粒核组

成,  $g_c(x_i)$ 为粒子,  $g_c(x_i)_j$ 为第  $j$  个粒核。

该样本  $x_i$  对应的决策  $y_i$  在决策特征  $d$  上进行粒化, 形成决策粒子, 定义为

$$g_d(y_i) = \{g_d(y_i)_j\}_{j=1}^n = \{r_j\}_{j=1}^n = \{r_1, r_2, \dots, r_n\} \tag{3}$$

式中  $r_j = 1 - |y_i - y_j|$ 。因为  $y_i, y_j \in \{-1, +1\}$ , 易知  $r_j \in \{-1, +1\}$ 。

感知机系统经过粒化后, 形成条件粒子和决策粒子。条件粒子的值是实数集合, 决策粒子的值是  $\pm 1$  集合。1 个条件粒子和 1 个决策粒子组成 1 条粒规则, 相应的条件粒核和决策粒核则组成粒核规则。

**定义 3** 设  $U = (X, C, \{d\})$  为感知机系统, 对于任一样本  $x_i \in X$ , 任一特征子集  $P \subseteq C$ , 设  $P = \{a_1, a_2, \dots, a_m\}$ , 则  $x_i$  在特征子集  $P$  上的粒向量定义为

$$G_P(x_i) = (g_{a_1}(x_i), g_{a_2}(x_i), \dots, g_{a_m}(x_i))^T \tag{4}$$

式中  $g_{a_m}(x_i)$  为样本  $x_i$  在特征  $a_m$  上的粒子。为方便算计, 特征集  $P = \{a_1, a_2, \dots, a_m\}$  用整数标记, 则粒向量可表示成  $G(x_i) = (g_1(x_i), g_2(x_i), \dots, g_n(x_i))^T$ 。

粒向量由粒子组成, 粒子又由粒核构成。因此粒向量可以是一个矩阵的形式, 表示为

$$G(x_i) = \begin{bmatrix} g_1(x_i)_1, g_1(x_i)_2, \dots, g_1(x_i)_n \\ g_2(x_i)_1, g_2(x_i)_2, \dots, g_2(x_i)_n \\ \vdots \\ g_m(x_i)_1, g_m(x_i)_2, \dots, g_m(x_i)_n \end{bmatrix} = \begin{bmatrix} r_{11}, r_{12}, \dots, r_{1n} \\ r_{21}, r_{22}, \dots, r_{2n} \\ \vdots \\ r_{m1}, r_{m2}, \dots, r_{mn} \end{bmatrix} \tag{5}$$

粒向量也可以用粒核向量表示为

$$G(x_i) = (g(x_i)_1, g(x_i)_2, \dots, g(x_i)_n) \tag{6}$$

式中  $g(x_i)_j = (g_1(x_i)_j, g_2(x_i)_j, \dots, g_m(x_i)_j)^T$  为粒核向量。

感知机的粒向量  $G_P(x_i)$  由粒子组成, 而粒子是 1 个集合的形式。因此, 粒向量的元素是集合, 与传统向量不一样, 传统向量的元素是 1 个实数。

## 2 粒感知机模型

输入样本的特征向量, 感知机输出样本的类别, 它是二分类模型, 分类结果是  $+1$  和  $-1$ 。样本先粒化, 然后进行粒的运算, 从而形成一种粒感知机模型。其输入为粒向量, 输出为只含有  $\pm 1$  的粒子, 进而判别样本的类别。粒感知机模型如图 1 所示。粒向量是由粒子组成的, 粒子是一种含有实数的序列, 表示了样本之间的距离度量与复杂结构, 因此可拟合非线性模型。

**定义 4** 设  $U = (X, C, \{d\})$  为感知机系统, 样本  $\forall x_i \in X$  和特征  $a, b \in C$ , 而且  $x_i$  在  $a, b$  上的粒子分别为  $g_a(x_i) = \{s_j\}_{j=1}^n$  和  $g_b(x_i) = \{t_j\}_{j=1}^n$ , 则两个粒子的加、减、乘、除运算定义为

$$\begin{cases} g_a(x_i) + g_b(x_i) = \{s_j + t_j\}_{j=1}^n \\ g_a(x_i) - g_b(x_i) = \{s_j - t_j\}_{j=1}^n \\ g_a(x_i) * g_b(x_i) = \{s_j * t_j\}_{j=1}^n \\ g_a(x_i) / g_b(x_i) = \{s_j / t_j\}_{j=1}^n \end{cases} \tag{7}$$

**定义 5** 设  $U = (X, C, \{d\})$  为感知机系统, 样本  $\forall x_i, x_k \in X$  和特征  $a \in C$ , 而且  $x_i, x_k$  在  $a$  上的粒子分别为  $g_a(x_i) = \{s_j\}_{j=1}^n$ , 和  $g_a(x_k) = \{t_j\}_{j=1}^n$ , 则两个粒子的加、减、乘、除运算定义为

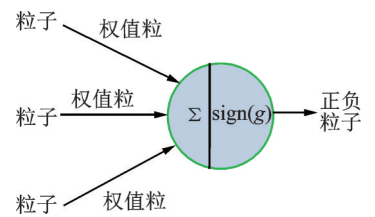


图 1 粒感知机

Fig.1 Granular perceptron

$$\begin{cases} g_a(x_i) + g_a(x_k) = \{s_j + t_j\}_{j=1}^n \\ g_a(x_i) - g_a(x_k) = \{s_j - t_j\}_{j=1}^n \\ g_a(x_i) * g_a(x_k) = \{s_j * t_j\}_{j=1}^n \\ g_a(x_i) / g_a(x_k) = \{s_j / t_j\}_{j=1}^n \end{cases} \quad (8)$$

2个粒子的加减乘除运算结果为1个粒子。定义4的定义是针对同一个样本在不同特征集合上粒化为不同粒子的运算,而定义5则是应用在不同样本在同一特征集合上粒化后粒子的运算。

**定义6** 设  $U=(X, C, \{d\})$  为感知机系统,对于两个样本  $\forall x_i, x_k \in X$ , 存在其相应的粒向量为  $G(x_i)=(g_1(x_i), g_2(x_i), \dots, g_m(x_i))^T$ ,  $G(x_k)=(g_1(x_k), g_2(x_k), \dots, g_m(x_k))^T$ , 则这2个粒向量的点乘为

$$G(x_i) \cdot G(x_k) = G(x_i)^T G(x_k) = g_1(x_i) * g_1(x_k) + g_2(x_i) * g_2(x_k) + \dots + g_m(x_i) * g_m(x_k) \quad (9)$$

2个粒向量的点乘结果为1个粒子。因此,本文可以通过构造1个权值粒向量,让样本的粒向量与权值粒向量进行点乘运算,其结果也为1个粒子。

**定义7** 设  $U=(X, C, \{d\})$  为感知机系统,  $\forall x_i \in X$ , 给定某粒子  $g(x_i) = \{r_j\}_{j=1}^n$ , 则该粒子的 sign 函数定义为

$$\text{sign}(g(x_i)) = \{\text{sign}(r_j)\}_{j=1}^n = \{\text{sign}(r_1), \text{sign}(r_2), \dots, \text{sign}(r_n)\} \quad (10)$$

$$\text{式中 } \text{sign}(r_j) = \begin{cases} +1 & r_j \geq 0 \\ -1 & r_j \leq 0 \end{cases}$$

**定义8** 设  $U=(X, C, \{d\})$  为感知机系统,对于样本  $\forall x_i \in X$ , 粒向量为  $G(x_i) = (g_1(x_i), g_2(x_i), \dots, g_m(x_i))^T$ , 给定权值粒向量  $W = (w_1, w_2, \dots, w_m)^T$  和偏置粒子  $b$ , 则粒感知机表示为

$$f(G(x_i)) = \text{sign}(W \cdot G(x_i) + b) = \text{sign}(w_1 * g_1(x_i) + w_2 * g_2(x_i) + \dots + w_m * g_m(x_i) + b) \quad (11)$$

粒感知机是关于粒子的函数,通过样本条件粒向量与权值粒向量的内积计算与 sign 函数运算,结果为条件粒子,而且其粒核只含有  $\pm 1$ 。同时,感知机系统中决策值进行粒化后形成决策粒子,而决策粒子的粒核也是只含有  $\pm 1$ 。因此,粒感知机系统中的条件粒子与决策粒子进行相乘,其结果为只含有  $\pm 1$  粒子。结果为  $-1$  的粒核则为误分类粒核,将梯度回传修正权值粒核向量,从而形成粒感知机分类器。

**定义9** 设粒子为  $g_c(x) = \{r_j\}_{j=1}^n$ , 其大小定义为

$$q(g_c(x)) = \sum_{j=1}^n r_j \quad (12)$$

若  $q(g_c(x)) > 0$ , 则粒子大小为正;反之,粒子大小为负。

**定义10** 设粒子为  $g_c(x) = \{g_c(x)_j\}_{j=1}^n = \{r_j\}_{j=1}^n$ , 若粒核  $\forall g_c(x)_j > 0$ , 则称  $g_c(x)$  为正粒子;若粒核  $\exists g_c(x)_j \leq 0$ , 则称  $g_c(x)$  为负粒子。

感知机模型的学习训练过程就是优化确定权值粒向量与偏置粒子的值,学习后的感知机模型则可以预测样本的类别。输入样本粒化后,输出粒子,通过度量粒子的大小确定样本的类别。输出结果为正,则判定为正例类别;输出结果为负,则判定为负例类别。

### 3 粒感知机损失函数与学习算法

为了降低大数据系统的复杂性,本文将1个样本按照不同的特征粒化为多个粒子,这些粒子组成1个粒向量。同时,该样本的决策值也粒化为只含有  $\pm 1$  的粒子。为了优化粒感知机模型的参数,需要定

义粒感知机的损失函数,求取损失函数的导数,从而得到粒感知机的梯度下降学习算法。

### 3.1 粒感知机损失函数

传统感知机模型是通过找到正确的分类超平面确定模型参数。利用损失函数,监督感知机的分类性能;利用梯度下降法,最后学习到感知机的分类超平面。经典感知机的损失函数为误分类点到超平面的总距离。

在粒感知器模型中,样本已经粒化为粒子。类似于感知机,粒感知机的损失函数为误分类粒子到超平面的总距离。粒子是粒核的集合,误分类粒子是因为其粒核产生了误分类。本文可以计算粒子中误分类的粒核到超平面的距离,此粒子中误分类粒核个体到超平面的距离之和作为该误分类粒子的距离,因此可得到所有误分类粒子到超平面的总距离。

**定义 11** 给定训练数据集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , 其中  $x_i \in X \subseteq \mathbf{R}^m, y_i \in \{-1, +1\}$ ,  $i = 1, 2, \dots, n$ 。训练集  $T$  粒化为  $GT = \{(G(x_1), g(y_1)), (G(x_2), g(y_2)), \dots, (G(x_n), g(y_n))\}$ , 其中  $G(x_i) = (g_1(x_i), g_2(x_i), \dots, g_m(x_i))^T$  为条件粒向量,  $g(y_i)$  为决策粒子。设  $W = (w_1, w_2, \dots, w_m)^T$  为权值粒向量,  $b$  为偏置粒子。粒感知机的损失函数定义为

$$L(W, b) = - \sum_{G(x_i) \in M} g(y_i) * (W \cdot G(x_i) + b) \quad (13)$$

式中  $M$  为误分类粒向量的集合。

**定理 1** 当粒向量  $G(x_i)$  存在误分类的粒核向量时,  $g(y_i) * (W \cdot G(x_i) + b)$  为负粒子。

**证明** 对于  $\forall x_i, W \cdot G(x_i) + b$  结果为粒子。因粒向量  $G(x_i)$  存在误分类的粒核向量, 所以  $W \cdot G(x_i) + b$  与  $g(y_i)$  必然存在正负相异的粒核。  $g(y_i) * (W \cdot G(x_i) + b)$  的结果仍然为粒子, 也必然存在负的粒核。根据负粒子的定义, 可知  $g(y_i) * (W \cdot G(x_i) + b)$  为负粒子。

### 3.2 粒感知机学习算法

定义粒感知机的损失函数

$$L(W, b) = - \sum_{G(x_i) \in M} g(y_i) * (W \cdot G(x_i) + b) \quad (14)$$

则参数  $W, b$  为下列损失函数极小化问题的解

$$\min_{W, b} L(W, b) = - \sum_{G(x_i) \in M} g(y_i) * (W \cdot G(x_i) + b) \quad (15)$$

式中  $M$  为误分类粒向量的集合。

$W \cdot G(x_i)$  为 2 个粒向量的点乘, 其结果为 1 个粒子, 则  $W \cdot G(x_i) + b$  的结果也为粒子。  $g(y_i) * (W \cdot G(x_i) + b)$  为 2 个粒子相乘, 结果为粒子。因  $G(x_i)$  为误分类粒向量, 则  $W \cdot G(x_i) + b$  与  $g(y_i)$  必然存在正负相异的粒核, 也必然存在  $g(y_i) * (W \cdot G(x_i) + b)$  为负的粒核, 将这些为负的粒核加负号后, 则变为正的粒核。

因此, 粒感知机学习算法由误分类的粒核驱动, 具体实现则采用随机梯度下降法。首先, 参数  $W, b$  进行初始化, 然后利用梯度下降法不断极小化目标函数。其方法是 1 次随机选取 1 个误分类粒子, 按照梯度方向, 修正该粒子中误分类的粒核权值。

**定理 2** 设粒感知机的损失函数为  $L(W, b) = - \sum_{G(x_i) \in M} g(y_i) * (W \cdot G(x_i) + b)$ , 则其对  $W, b$  导数分别为  $\nabla_W L(W, b) = - \sum_{G(x_i) \in M} G(x_i) * g(y_i), \nabla_b L(W, b) = - \sum_{G(x_i) \in M} g(y_i)$ 。

**证明** 因  $L(W, b) = - \sum_{G(x_i) \in M} g(y_i) * (W \cdot G(x_i) + b)$ , 则  $L(W, b) = - \sum_{G(x_i) \in M} (W \cdot G(x_i) * g(y_i) + b * g(y_i))$ 。所以,  $\nabla_W L(W, b) = - \sum_{G(x_i) \in M} G(x_i) * g(y_i)$ ,  $\nabla_b L(W, b) = - \sum_{G(x_i) \in M} g(y_i)$ 。命题得证。

根据上述其导数的相关证明,推导出梯度下降法,进而求解粒感知机的近似最优解。粒感知机算法分为学习算法和分类算法,分别描述如下:

(1) 粒感知机学习算法

输入:训练集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , 其中  $x_i$  为  $m$  维特征向量;其中  $x_i \in X \subseteq \mathbf{R}^m$ ,  $y_i \in \{-1, +1\}, i = 1, 2, \dots, n$ ;学习率为  $\eta (0 < \eta \leq 1)$ ;

输出:  $W, b$ 。

(1)训练集  $T$  粒化为  $GT = \{(G(x_1), g(y_1)), (G(x_2), g(y_2)), \dots, (G(x_n), g(y_n))\}$ ;

(2)随机初始化权值粒向量  $W = (w_1, w_2, \dots, w_m)^T$  和阈值粒子  $b$ ;

(3)从粒化后的训练集中选取  $(G(x_i), g(y_i))$ ;

(4)如果  $g(y_i) * (W \cdot G(x_i) + b)$  为负粒子,则  $W = W + \eta * G(x_i) * g(y_i), b = b + \eta * g(y_i)$ ;

(5)转至(3),直至  $g(y_i) * (W \cdot G(x_i) + b)$  为正粒子;

(6)输出权值粒向量  $W = (w_1, w_2, \dots, w_m)^T$  和偏置粒子  $b$ 。

(2) 粒感知机分类算法

输入:训练集  $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , 权值粒向量  $W = (w_1, w_2, \dots, w_m)^T$  和偏置粒子  $b$ , 测试样本  $x$ ;

输出:测试样本  $x$  的类别。

(1)测试样本  $x$  在训练集中粒为化  $G(x)$ ;

(2)计算  $f(G(x)) = \text{sign}(W \cdot G(x) + b)$ ;

(3)计算粒子的大小  $q(f(G(x)))$ , 为正则判定为正例类别, 为负则判定为负例类别。

## 4 实验分析

实验数据为UCI数据集,具体描述如表1所示。特别说明:Movement\_libras数据集具有15个标签类别,故在该实验中将标签小于8的类别设为负类,标签大于等于8的数据类别设为正类。

表1 数据集描述

Table 1 Dataset description

数据集	样本数	特征数	类别数
Heart	270	13	2
Blood-transfusion	748	4	2
Haberman	306	3	2
Movement_libras	359	91	15
Breast Cancer Coimbra	115	10	2

数据集的量纲存在差异,因此采用最大最小值法进行样本归一化,从而保证数据值在 $[0, 1]$ 区间变化。最大最小值归一化公式为

$$f(x_i) = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{16}$$

为了测试算法的有效性,首先对粒感知机的收敛性进行分析;其次,将选取线性与非线性的数据集,与传统感知机进行比较,测试粒感知机的非线性性能。最后,本文采用多个传统的机器学习分类器算法,比较其分类精度,分析算法的分类性能。

### 4.1 粒感知机的收敛性分析

本次实验利用 Heart 数据集,测试粒感知机误差率是否收敛。首先,从 Heart 数据集随机抽取 9/10 进行训练集,剩下 1/10 的数据用于测试分类效果,迭代次数由 1 递增到 50,实验结果如图 2 所示。然后采样 10 次交叉验证法进行测试。迭代次数由 1 递增到 30,实验结果如图 3 所示。图 2 中,第 1 次迭代误差达到最小,第 3 次迭代后误差则趋于稳定。从图 3 可知,第 2 次迭代误差达到最小值,随后几次迭代小幅波动,迭代到第 10 次,误差趋于稳定。从图 2 和图 3 中可以看出,粒感知机具有快速收敛的特性。因为粒化是整个样本空间进行粒化的,其粒子具备全局的信息,因此收敛迅速,迭代次数很少。

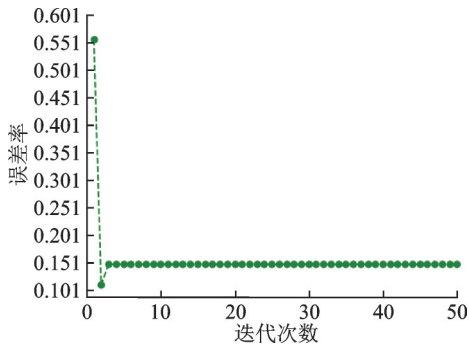


图2 Heart数据集随机训练的误差率

Fig.2 Error rate of random training in Heart dataset

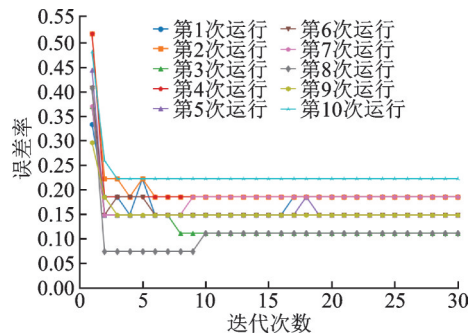


图3 Heart数据集10次交叉验证误差率

Fig.3 Error rate of ten cross-validation of Heart dataset

### 4.2 粒感知机的非线性性能

众所周知,感知机是一种线性模型,难以表达复杂非线性函数,比如异或 XOR。而粒感知机则不同,其参数较多,具备非线性拟合能力。为了验证粒感知机这个优点,本文对 XOR 函数进行模拟实验。用决策表表示 XOR 函数,如表 2 所示。首先本文使用传统的感知机对 XOR 函数进行实验。如图 4 所示,一些特征的输出与函数的输出值不同,说明传统感知机不具备非线性处理能力。用

表 2 XOR 函数  
Table 2 XOR function

$U$	$a$	$b$	$Y$
$x_1$	0	0	0
$x_2$	0	1	1
$x_3$	1	0	1
$x_4$	1	1	0

粒感知机对 XOR 函数以及 Heart 非线性数据集进行实验。由于本文中的定义采用正例负例的情况,故将 XOR 函数的 0 值改为 -1。实验结果如图 5,4 个类型的特征输出均符合 XOR 函数;如图 6 所示,Heart 数据集的准确率也很高,故验证粒感知机具有非线性处理能力。

### 4.3 分类精度比较

经典的机器学习模型有决策树分类、线性回归、K 近邻(K-nearest neighbor, KNN)分类、随机森林分类、Adaboost 分类和梯度提升决策树(Gradient boosting decision tree, GBDT)分类。本文采用精确度来评价算法的分类性能。数据集中使用 10 次交叉验证取平均值的方法。随机森林分类使用 20 个决策

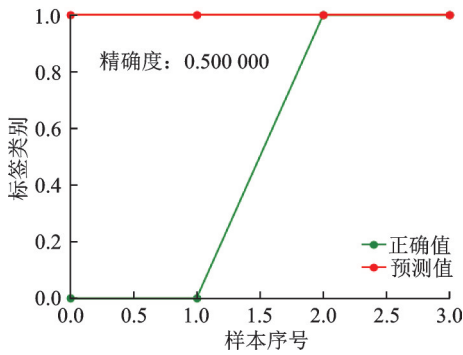


图4 XOR函数感知机预测曲线

Fig.4 Perceptron prediction of XOR function

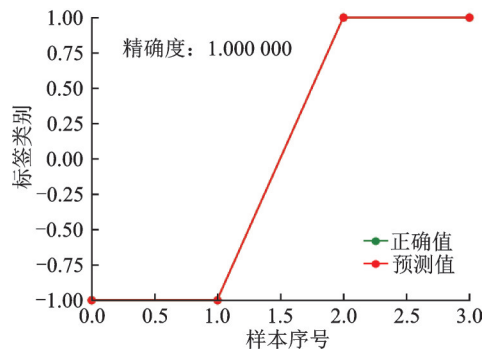


图5 XOR函数粒感知机预测曲线

Fig.5 Granular perceptron prediction of XOR function

树, Adaboost分类使用50个决策树, GBDT分类使用100个决策树。由上节可知, 粒感知机算法在10次迭代就达到收敛稳定的效果, 故这次实验粒感知机使用10次迭代。如表3所示, 结果保留4位小数。

从实验中可以得出, 粒感知机算法在Heart和Haberman数据集上要明显优于其他算法。除了在Movement\_libras数据集上精确度弱一些外, 在其他数据集上与其他算法相比, 大部分情况精确度更高。从这个表中还可以看出, 粒感知机算法明显都要高于线性回归的分类精度, 由此再次印证了粒感知机算法的非线性分类性能。

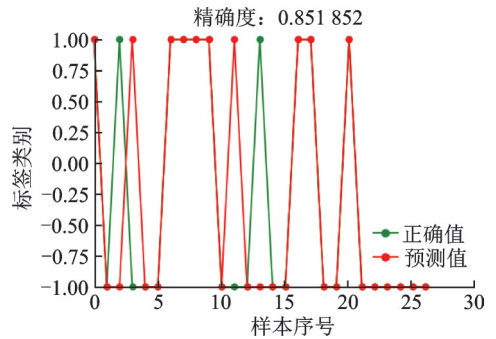


图6 Heart测试集预测曲线

Fig.6 Prediction curve of Heart test set

表3 各数据集分类算法精确度比较

Table 3 Accuracy comparison of classification algorithms for each dataset

数据集	粒感知机	决策树分类	线性回归	KNN分类	随机森林分类	Adaboost分类	GBDT分类
Heart	<b>0.840 7</b>	0.748 1	0.559 2	0.807 4	0.833 3	0.788 9	0.811 1
Movement_libras	0.527 8	0.598 9	0.489 4	0.551 7	0.582 5	<b>0.690 6</b>	0.598 4
Haberman	<b>0.734 8</b>	0.634 7	0.503 4	0.721 8	0.690 0	0.683 4	0.680 2
Blood-transfusion	0.764 5	0.733 9	0.437 3	0.753 9	0.737 8	<b>0.777 9</b>	0.763 2
Breast Cancer Coimbra	0.554 5	0.499 2	0.422 4	0.430 3	0.558 3	<b>0.627 3</b>	0.575 8

### 5 结束语

传统感知机模型不具备非线性分类性能, 故分类效率不高。本文从研究样本的粒化出发, 通过定义粒向量的形式, 本文提出了一种新型的感知机分类模型: 粒感知机。首先, 引入特征粒化方法, 在数据集系统中构建粒子和粒向量, 并定义了粒向量的运算规则。进一步将粒向量进行了应用, 设计粒感知机策略, 提出了粒感知机学习算法。其次, 对粒感知机模型进行参数求解, 论证了粒感知机函数的导数形式, 得到其梯度下降算法。最后进行实验分析, 验证了粒感知机模型的正确性与有效性。在未来



的工作中,研究粒的局部粒化方法,提高粒化速度与粒化性能;研究粒的度量方法,提高粒分类器的分类精度。

#### 参考文献:

- [1] LIN T Y. Granular computing on binary relations I: Data mining and neighborhood systems[J]. *Rough Sets in Knowledge Discovery*, 1998, 2: 165-166.
- [2] 苗夺谦, 张清华, 钱宇华, 等. 从人类智能到机器实现模型——粒计算理论与方法[J]. *智能系统学报*, 2016, 11(6): 743-757.  
MIAO Duoqian, ZHANG Qinghua, QIAN Yuhua, et al. From human intelligence to machine implementation model: Theories and applications based on granular computing[J]. *CAAI Transactions on Intelligent Systems*, 2016, 11(6): 743-757.
- [3] YAO Y Y. Granular computing using neighborhood systems[M]. London: Springer-Verlag, 1999: 539-553.
- [4] LIU Q, JIANG F, DENG D. Design and implement for diagnosis systems of hemorheology on blood viscosity syndrome based on GrC[C]//*Proceedings of 9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*. [S.l.]: Springer-Verlag, 2003, 413-420.
- [5] 苗夺谦. *Rough Set 理论及其在机器学习中的应用研究*[D]. 北京: 中国科学院自动化研究所, 1997.  
MIAO Duoqing. *Research on rough set theory and its application in machine learning*[D]. Beijing: Institute of Automation, Chinese Academy of Sciences, 1997.
- [6] 苗夺谦, 王珏. 粗糙集理论中知识粗糙性与信息熵关系的讨论[J]. *模式识别与人工智能*, 1998 (1): 34-40.  
MIAO Duoqian, WANG Jue. On the relationships between information entropy and roughness of knowledge in rough set theory [J]. *Pattern Recognition and Artificial Intelligence*, 1998 (1): 34-40.
- [7] 苗夺谦, 王珏. 粗糙集理论中概念与运算的信息表示[J]. *软件学报*, 1999, 10(2): 113-116.  
MIAO Duoqian, Wang Jue. An information representation of the concepts and operations in rough set theory[J]. *Journal of Software*, 1999, 10(2): 113-116.
- [8] 王国胤, 张清华, 马希骛, 等. 知识不确定性问题的粒计算模型[J]. *软件学报*, 2011, 22(4): 676-694.  
WANG Guoyin, ZHANG Qinghua, MA Xi'ao, et al. Granular computing models for knowledge uncertainty[J]. *Journal of Software*, 2011, 22(4): 676-694.
- [9] WU W Z, LEUNG Y, MI J S. Granular computing and knowledge reduction in formal contexts[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2009, 21(10): 1461-1474.
- [10] XIA S, LIU Y, XIN D, et al. Granular ball computing classifiers for efficient, scalable and robust learning[J]. *Information Sciences*, 2019, 483: 136-152.
- [11] DICK S, TAPPENDEN A, BADKE C, et al. A granular neural network: Performance analysis and application to re-granulation[J]. *International Journal of Approximate Reasoning*, 2013, 54: 1149-1167.
- [12] 黄际洲, 孙雅铭, 王海峰, 等. 面向搜索引擎的实体推荐综述[J]. *计算机学报*, 2019, 42(7): 1467-1494.  
HUANG Jizhou, SUN Yaming, WANG Haifeng, et al. A survey of entity recommendation in web search[J]. *Chinese Journal of Computers*, 2019, 42(7): 1467-1494.
- [13] 苏煜, 山世光, 陈熙霖, 等. 基于全局和局部特征集成的人脸识别[J]. *软件学报*, 2010, 21(8): 1849-1862.  
SU Yi, SHAN Shiguang, CHEN Xilin, et al. Integration of global and local feature for face recognition[J]. *Journal of Software*, 2010, 21(8): 1849-1862.
- [14] 曹宜超, 高翊, 李淼, 等. 基于单语语料和词向量对齐的蒙汉神经机器翻译研究[J]. *中文信息学报*, 2020, 34(2): 27-32.  
CAO Yichao, GAO Yi, LI Miao, et al. Mongolian-chinese neural machine translation based on monolingual corpora and word embedding alignment[J]. *Journal of Chinese Information Processing*, 2020, 34(2): 27-32.
- [15] 高庆吉, 赵志华, 徐达, 等. 语音情感识别研究综述[J]. *智能系统学报*, 2020, 81(1): 7-19.  
GAO Qingji, ZHAO Zhihua, XU Da, et al. Review on speech emotion recognition research[J]. *CAAI Transactions on Intelligent Systems*, 2020, 81(1): 7-19.
- [16] 徐开琨, 韩明飞, 黄传玺, 等. 基于质谱的蛋白质生物标志物发现中的特征选择与机器学习方法研究进展[J]. *生物工程学*

- 报, 2019, 35(9): 1619-1632.
- XU Kaikun, HANG Mingfei, HUANG Chuanxi, et al. Research progress of feature selection and machine learning methods for mass spectrometry-based protein biomarker discovery[J]. Chinese Journal of Biotechnology, 2019, 35(9): 1619-1632.
- [17] 康靓, 米晓希, 王海莲, 等. 人工神经网络在材料科学中的研究进展[J]. 材料导报, 2020, 34(21): 176-183.
- KANG Jian, MI Xiaoxi, WANG Hailian, et al. Research progress of artificial neural networks in material science[J]. Materials Reports, 2020, 34(21): 176-183.
- [18] 龚军, 杜超, 钟小钢, 等. 基于机器学习算法的原发性高血压并发冠心病的患病风险研究[J]. 解放军医学杂志, 2020, 45(7): 735-741.
- GONG Jun, DU Chao, ZHONG Xiaogang, et al. Researches on the illness risk of essential hypertension complicated with coronary heart disease based on machine learning algorithm[J]. Medical Journal of Chinese People's Liberation Army, 2020, 45(7): 735-741.
- [19] 何英哲, 胡兴波, 何锦雯, 等. 机器学习系统的隐私和安全问题综述[J]. 计算机研究与发展, 2019, 56(10): 2049-2070.
- HE Yingzhe, HU X B, HE J W, et al. Privacy and security issues in machine learning systems: A survey[J]. Journal of Computer Research and Development, 2019, 56(10): 2049-2070.
- [20] 马廷淮, 穆强, 田伟, 等. 气象数据挖掘研究[J]. 武汉理工大学学报, 2010, 32(16): 110-114.
- MA Tinghuai, MU Qiang, TIAN Wei, et al. Research on meteorological data mining[J]. Journal of Wuhan University of Technology, 2010, 32(16): 110-114.
- [21] ROSENBLATT F. The Perceptron: A probabilistic model for information storage and organization in the Brain[J]. Psychological Review, 1958, 65(6): 111-127.
- [22] NOVIKOFF A. On convergence proofs on perceptrons[C]//Proceedings of Symposium on the Mathematical Theory of Automata. New York: Polytechnic Institute of Brooklyn, 1962, 16: 615-622.
- [23] MINSKY M L, PAPER S A. Perceptrons[M]. Cambridge, MA: MIT Press, 1969.
- [24] GALLANT S I. Perception-based learning algorithms[J]. IEEE Transactions on Neural Networks, 1990, 1(2): 179-191.
- [25] FREUND Y, SCHAPIRE R E. Large margin classification using the perceptron algorithm[C]//Proceedings of the 11th Annual Conference on Computational Learning Theory(COLT' 98). [S.l.]: PressACM, 1998.
- [26] LI Y Y, ZARAGOZA H, HERBRICH R, et al. The perceptron algorithm with uneven margins[C]//Proceedings of the 19th International Conference on Machine Learning. Sydney: [s.n.], 2002: 379-386.

#### 作者简介:



陈玉明(1977-),男,教授,研究方向:机器学习,粗糙集和粒计算, E-mail: cym0620@163.com。



董建威(1998-),通信作者,男,硕士研究生,研究方向:机器学习,图像处理,和模式识别, E-mail: 1466781674@qq.com。

(编辑:刘彦东)