

DOI: 10.15923/j.cnki.cn22-1382/t.2018.3.12

# 基于 FUZZY-SVM 的软件质量综合评价

郑 鹏

(莆田学院 信息工程学院, 福建 莆田 351100)

**摘 要:** 提出了模糊支持向量机软件质量评价模型, 利用模糊分布构造隶属度函数解决指标模糊性问题, 克服了过拟合问题, 最终得到全局最优评价点。对 25 个测试软件进行了仿真实验。

**关键词:** 软件质量; ISO/IEC 25010 标准; 模糊数学; 支持向量机; 综合评价

**中图分类号:** TP 311.52   **文献标志码:** A   **文章编号:** 1674-1374(2018)03-0278-04

## FUZZY-SVM based software quality comprehensive evaluation

ZHENG Peng

(College of Information Engineering, Putian University, Putian 351100, China)

**Abstract:** A software quality evaluation model based on fuzzy support vector machine is built to solve index fuzzy problem, overcomes over fitting, and obtain the global optimal evaluation. Simulations for 25 softwares are tested.

**Key words:** software quality; ISO/IEC 25010 standard; Fuzzy mathematics; Support Vector Machine (SVM); comprehensive evaluation.

## 0 引 言

随着国家软件产业进入高速发展阶段, 软件质量评价问题引起了人们的广泛关注。由于传统的评价方法存在很多不足, 支持向量机 (Support Vector Machine, SVM) 算法作为基于统计学的一种新的机器学习算法, 克服了主观赋权法的主观性和维数灾的问题, 其全局寻优也解决了神经网络法中的局部极值问题。文中将模糊数学和 SVM 结合, 提出模糊支持向量机 (Fuzzy Support

Vector Machine, FUZZY-SVM) 模型, 旨在为软件质量的综合评价提供一个新的思路。

## 1 软件质量的度量

软件质量度量指的是对软件质量在开发过程进行质量控制、在验收阶段进行评价评估。为了可以提供一个科学度量的标准, ISO 发布了 ISO/IEC 25010 软件质量度量模型<sup>[1]</sup>。该模型根据最小关联与独立的原则把软件产品划分为 8 个质量特性和 31 个质量子特性。质量特性是代表软件

收稿日期: 2018-03-17

基金项目: 福建省莆田市科技项目(2015G2014)

作者简介: 郑 鹏(1983—), 男, 汉族, 福建仙游人, 莆田学院讲师, 硕士, 主要从事软件工程、人工智能、数据挖掘方向研究, E-mail: 32200267@qq.com.

质量的一个方面,用于描述和评价软件的某方面属性。子特性是质量特性的细分,一个质量特性可以由多个子特性表达,某个子特性也可以属于多个质量特性。ISO/IEC 25010 模型是一个成熟全面的度量模型,文中将选其作为度量模型。

## 2 SVM、FUZZY-SVM 模型

### 2.1 SVM

SVM 算法是基于统计学习理论体系,由 Cortes 等<sup>[2]</sup>提出的一种解决线性不可分问题的新型通用机器学习方法。

假设线性可分样本集为  $(x_i, y_i)$ ,  $i=1, 2, \dots, n, x \in R^c, y \in \{+1, -1\}$ , 其中  $y$  为类别符号。在维度为  $c$  的空间中,将线性判别函数的形式一般表示为:

$$f(x) = \omega x + b$$

分类线性方程表示为:

$$\omega x + b = 0$$

将判别函数进行归一化,使样本满足

$$|f(x)| = 1$$

此时分类间隔等于  $\|\omega\|/2$ 。要使分类间隔最大,即  $\|\omega\|$  最小,需要满足:

$$y_i[(\omega x) + b] - 1 \geq 0, i = 1, 2, \dots, n$$

当  $\|\omega\|$  最小且满足上述条件时的分类面即为最优分类面,把支持最优分类面的  $H_1, H_2$  这两类训练样本点称为支持向量。

对于线性分类,SVM 将其转化为一个不等式约束下的二次函数极值问题。对于非线性的情况,假设有非线性映射  $\varphi: R^d \rightarrow H$ ,要将样本映射到高维特征空间  $H$  中,当要构造最优超平面时,仅需使用空间中的点积  $\varphi(x_i) \cdot \varphi(x_j)$ 。此时,寻找函数  $K$ ,使得

$$K(x_i \cdot x_j) = \varphi(x_i) \cdot \varphi(x_j)$$

这样,在高维空间只需进行内积运算,而且可利用原有空间中的函数来实现。

选用适合的内积函数  $K(x_i \cdot x_j)$  进行非线性变换,可以使核函数满足 Mercer 条件,这样在变幻的空间中的内积得到对应,也得到了线性分类<sup>[3]</sup>。

SVM 算法这种通过向量输入、支持向量的内积处理以及中间节点的线性组合输出,结构类似一个神经网络<sup>[4]</sup>。相当于把  $x = (x_1, x_2, \dots, x_n)$  为输入向量,中间节点  $K(x \cdot x_i)$  是基于支持向量的内积,权值是  $\omega_i = \alpha_i y_i$ ,输出是中间节点的线

性组合:

$$y = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i y_i K(x \cdot x_i) + b \right\} \quad (1)$$

常见的核函数有三类:多项式核函数、RBF 径向基函数和 S 形核函数<sup>[4]</sup>。考虑到软件质量评价指标与结果之间的非线性映射关系,以及径向基 RBF 核函数的参数较少,不会过多影响模型的精度。此外,以往的评价方法对评价结果都是用线性赋权法进行计算的,这对于非线性赋权问题还是有缺陷的,而径向基核函数解决了这些问题,因此选取径向基 RBF 函数作为核函数<sup>[5]</sup>:

$$K(x \cdot x_i) = \exp \left\{ -\frac{|x - x_i|^2}{\sigma^2} \right\} \quad (2)$$

### 2.2 构造模糊隶属度函数

软件质量评价指标中,有些指标是带模糊性的,因此利用模糊数学来进行量化。对于可量化的指标,也存在一些问题:比如不符合软件质量的连续和渐变的特点、部分评价的信息丢失、量化程度不好等,因此需要构造隶属度函数<sup>[6]</sup>可定量指标的评价标准进行处理。

在实际评价应用中,通常可以选择适当的模糊分布<sup>[7]</sup>,再根据实际问题的特点确定各个参数,进而确定隶属函数。根据软件质量渐变连续的特点,梯形分布与半梯形分布的模糊分布比较符合。因此,文中采用梯形分布与半梯形分布构造隶属度函数,在进行 SVM 输入训练之前对原始数据进行标准化。

### 2.3 FUZZY-SVM 模型

FUZZY-SVM 模型是一个模式分类的过程,是指对软件质量特征、因子进行描述辨识和分类解释的过程,如图 1 所示。

首先,模型选择指标体系及评价分级标准。然后对原始数据进行采集,采用模糊隶属度函数对指标数据进行归一化预处理。模型在面临不同类型的软件时,可以根据问题的实际数据特征来选择不同的 SVM 模型和参数。在求解样本数的二次规划问题过程中,还需要设计特殊的训练策略来解决复杂程度与样本数成指数递增关系的问题<sup>[8]</sup>,具体包括:训练集的选择策略以及增量迭代如何进行等。在模型择优的过程中,可循环迭代、多次优化,以最终达到整体上的最优效果,并进行综合评价。如此一来,FUZZY-SVM 模型可有效排除主观性及干扰、容错能力强、操作简单便捷。

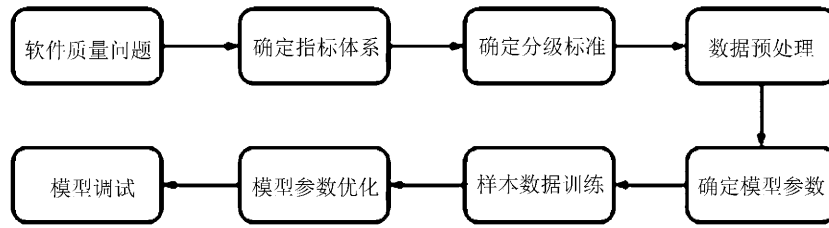


图 1 FUZZY-SVM 模型

### 3 实 例

#### 3.1 指标体系及评价分级标准

假设要对某类的软件进行综合评价。依据 ISO/IEC 25010 标准选取全部特性作为评价指标,用  $I_1 \sim I_{31}$  表示,见表 1。

表 1 软件质量评价指标体系

$I$	评价指标	$I$	评价指标
$I_1$	功能完整性	$I_{17}$	完整性
$I_2$	功能正确性	$I_{18}$	抗抵赖性
$I_3$	功能适合性	$I_{19}$	可追踪性
$I_4$	稳定性	$I_{20}$	真实性
$I_5$	可用性	$I_{21}$	共存性
$I_6$	容错性	$I_{22}$	互操作性
$I_7$	易恢复性	$I_{23}$	易学性
$I_8$	时间特性	$I_{24}$	可被识别适当性
$I_9$	资源特性	$I_{25}$	易用性
$I_{10}$	容量	$I_{26}$	用户错误保护
$I_{11}$	易分析性	$I_{27}$	用户界面美观
$I_{12}$	易重用性	$I_{28}$	易接近性
$I_{13}$	易修改性	$I_{29}$	适应性
$I_{14}$	易测试性	$I_{30}$	易安装性
$I_{15}$	模块性	$I_{31}$	易替代性
$I_{16}$	保密性		

#### 3.2 数据采集和预处理

在指标体系中定量指标包括  $I_1, I_2, I_8, I_9, I_{10}, I_{17}$ , 其余为定性指标。将定性指标的隶属度分为 5 个等级:(差、较差、合格、良好、优秀) = (0, 0.3, 0.6, 0.8, 1.0), 隶属度越大, 该指标对软件质量的贡献程度越大。定量指标中指标  $I_1, I_2, I_{17}$

是极大型指标, 而  $I_8, I_9, I_{10}$  是极小型指标。利用隶属度函数进行训练:

$$I_1 = \begin{cases} 0, & x < 10\% \\ \frac{x - 10}{89}, & 10\% \leq x \leq 99\% \\ 1, & x = 100\% \end{cases}$$

$$I_2 = \begin{cases} 0, & x < 30\% \\ \frac{x - 3}{6.5}, & 30\% \leq x \leq 95\% \\ 1, & 95\% < x \end{cases}$$

$$I_8 = \begin{cases} 1, & x < 2 \text{ s} \\ \frac{10 - x}{8}, & 2 \text{ s} \leq x \leq 10 \text{ s} \\ 0, & 10 \text{ s} < x \end{cases}$$

$$I_9 = \begin{cases} 1, & x < 1c \\ \frac{3 - x}{2}, & 1c \leq x \leq 3c \\ 0, & 3c < x \end{cases}$$

$$I_{10} = \begin{cases} 1, & x < 1 \text{ M} \\ \frac{500 - x}{499}, & 1 \text{ M} \leq x \leq 500 \text{ M} \\ 0, & 500 \text{ M} < x \end{cases}$$

$$I_{17} = \begin{cases} 0, & x < 10\% \\ \frac{x - 10}{89}, & 10\% \leq x \leq 99\% \\ 1, & x = 100\% \end{cases}$$

把软件质量根据评价结果分为 4 个等级, 即优秀、良好、合格、不合格, 并用 1, 2, 3, 4 分别表示。将样本的各项软件质量指标隶属度值作为评价指标输入, 将软件质量综合评价结果作为支持向量机输出。选取收集的 50 份不同等级的数据作为模型训练和测试样本, 其中前 25 个作为训练样本, 后 25 个作为测试样本, 根据隶属度等级和函数归一化样本数据。

### 3.3 模型选择及参数确定

在 CPU Intel Celeron G530 2.4 GHz, RAM 4 GB, windows 7 操作系统, MATLAB 2013a 环境下利用 LIBSVM 工具包, 采用 5 折交叉验证法对训练样本进行测试, 选取了误差最小、模型最优的参数:  $C=8, \sigma=0.5$ 。

## 4 结 语

针对软件质量评价的特殊性, 提出了一种基于 FUZZY-SVM 的软件质量评价模型。模型对不同软件质量等级的样本数据进行训练, 并检验样本数据, 结果基本一致。最后, 再与模糊综合评价方法和神经网络评价方法进行了比较, 结果表明, 该方法正确率和预测精度最高。因此, FUZZY-SVM 在实际中可有效地进行软件质量等级分类、评价, 为决策者在科学客观评价软件质量方面提供了一个新的评价思路。

### 参考文献:

[1] 孙洋, 袁玉宇. ISO/IEC 25010 质量模型标准现状

[J]. 信息技术与标准化, 2008(11): 33-36.

- [2] 赵洪波, 冯夏庭. 非线性位移时间序列预测的进化支持向量机方法及应用[J]. 岩土工程学报, 2003, 25(4): 468-471.
- [3] 杜栋, 庞庆华, 吴炎. 现代综合评价方法与案例精选[M]. 北京: 清华大学出版社, 2015.
- [4] 苏怀智, 温志萍, 吴中如. 基于 SVM 理论的大坝安全预警模型研究[J]. 应用基础与工程科学学报, 2009, 17(1): 40-47.
- [5] 孟洛明, 朱杰辉, 杨杨. 支持向量机回归预测在网络故障检测中的应用[J]. 北京邮电大学学报, 2014, 37: 23-29.
- [6] 杨纶标, 高英仪. 模糊数学原理与应用[M]. 广州: 华南理工大学出版社, 2005.
- [7] 李荣钧. 模糊多准则决策理论与应用[M]. 北京: 科学出版社, 2002.
- [8] 金菊良, 魏一鸣. 复杂系统广义智能评价方法与应用[M]. 北京: 科学出版社, 2008.
- [9] 黄明, 李宪章, 刘杨. 软件质量评审技术[J]. 吉林工学院学报, 1990, 11(3): 25-29.