

基于支持向量机的嫌疑人特征预测

李荣岗, 孙春华, 姬建睿

(合肥工业大学 管理学院, 合肥 230009)

摘 要: 针对大数据环境下, 公安机关计算机核心技术应用不足、备选嫌疑人众多而预测方法相对落后的问题, 提出运用支持向量机(SVM)预测犯罪嫌疑人的模型。根据历史犯罪记录进行特征选择, 训练基于 SVM 的嫌疑人特征预测模型, 通过此模型对案件嫌疑人的各个特征进行预测, 将预测出的特征与备选嫌疑人库中人员特征进行相似度计算, 进而预测出最有可能的嫌疑人。实验结果表明, 与应用分类和回归算法的模型相比, 该模型对预测结果具有较好的解释性, 能够缩小排查范围。

关键词: 大数据; 支持向量机; 特征选择; 分类器; 犯罪预测; 数据挖掘

中文引用格式: 李荣岗, 孙春华, 姬建睿. 基于支持向量机的嫌疑人特征预测[J]. 计算机工程, 2017, 43(11): 198-203.

英文引用格式: LI Ronggang, SUN Chunhua, JI Jianrui. Suspect Characteristics Prediction Based on Support Vector Machine[J]. Computer Engineering, 2017, 43(11): 198-203.

Suspect Characteristics Prediction Based on Support Vector Machine

LI Ronggang, SUN Chunhua, JI Jianrui

(School of Management, Hefei University of Technology, Hefei 230009, China)

[Abstract] In the big data environment, the computer core technology of public security organs is insufficient, many alternative suspects and the forecast method is relatively backward problem, aiming at these problems, this paper proposes the model of using the Support Vector Machine(SVM) to predict the suspect. According to the historical crime record, the model carries on the feature selection firstly, based on SVM training the suspect characteristic prediction model. Through this model, it can predict the various characteristics of the suspect, and calculate the similarity between the characteristics of prediction and the staff characteristics of the suspect database, then can predict the most possible suspects. Experimental results show that compare to the previous models which using classification and regression algorithm, this model has a good explanation for the prediction results; Besides, it can narrow the scope of the investigation.

[Key words] big data; Support Vector Machine(SVM); feature selection; classifier; crime prediction; data mining

DOI: 10.3969/j.issn.1000-3428.2017.11.032

0 概述

犯罪事件每年以一定的比率增长, 而且日趋复杂化。随着信息科技的发展, 数据仓库技术和新型数据源技术得到了不断的应用和发展, 犯罪数据呈指数形式增长。虽然收集存储的犯罪数据非常可观, 但是大量数据信息并没有得到充分的利用, 一般将这些数据作为档案信息进行保存或者进行一些统计分析作为一些业务系统的基础资料。简单来说就是用于存储、统计和共享^[1]。犯罪数据的利用很少, 并且是低效率的, 这与我国犯罪数据的隐蔽性导致数据不易获得有关, 这一局限性限制了犯罪预测研究的发展。另外, 在大数据时代下, 公安机关计算机

核心技术应用不足, 预测方法落后; 公安机关尚未真正重视具体化的犯罪防控工作, 而偏向一般性的, 与具体实务工作相距较远的犯罪“大趋势”“大走向”^[2]; 这些造成了我国的犯罪分析和预测领域发展相对国外比较滞后、不均衡的局面。数据挖掘技术是一种海量数据中寻找有用知识的方法, 已经应用到多个行业、多个领域, 在国内的犯罪预测领域也得到了应用。文献[3]运用数据驱动的可动态优化组合预测方法对犯罪量进行预测; 文献[4]运用随机森林分类器的方法解决犯罪预测的问题; 文献[5]提出了一种基于 Probit 模型的犯罪嫌疑人判定技术。国外对于犯罪预测的研究主要采取实证的方法, 通过调查研究、证据收集、分析和归纳, 从而发现和揭示

基金项目: 国家科技支撑计划项目(2015BAH26F00); 教育部人文社会科学研究一般项目(15YJC630111)。

作者简介: 李荣岗(1989—), 男, 硕士研究生, 主研方向为数据挖掘; 孙春华, 副教授; 姬建睿, 博士研究生。

收稿日期: 2016-08-31 **修回日期:** 2016-11-11 **E-mail:** lrghktk@163.com

犯罪发生的规律。例如利用可视化方法可视化罪犯的社交关系^[6]、犯罪模式的时间和空间可视化^[7];利用统计方法预测青少年暴力犯罪的趋势^[8]和累犯预测^[9];利用机器学习方法分析犯罪行为、模式识别^[10]等,数据挖掘技术得到了广泛运用。同时,嫌疑人预测是犯罪分析的一个主要任务,也是热点问题^[11-12]。文献[12]利用随机森林在患有精神障碍的人群中进行累犯预测;文献[13]利用朴素贝叶斯进行嫌疑人预测等。

目前来看,国内的犯罪预测大多数为宏观预测。在嫌疑人预测问题中,国内与国外的研究大多数运用分类或回归的方法对嫌疑人的可能性进行预测,这可能会造成误判或者漏判问题。为此,本文针对嫌疑人的特征,提出一种新的基于支持向量机(Support Vector Machine, SVM)^[14]的嫌疑人预测模型。

1 犯罪嫌疑人特征预测原理

犯罪案件的特征是一种复杂的非线性数据,包括案件的基本特征、受害者特征和嫌疑人特征等,难以建立精确的数据学模型。支持向量机是专门针对非线性系统的预测方法,拟合精度高、泛化推广能力强,因此,可以运用 SVM 进行建模与预测,同时能够避免出现“维数灾难”“过学习”问题,克服了传统预测方法的缺点,对犯罪嫌疑人的预测问题能够达到很好的效果。本文对犯罪嫌疑人的特征与案件的基本特征和受害者的特征之间的关联性进行研究,对训练样本学习,实现案件特征与犯罪嫌疑人特征的非线性映射关系。根据这种非线性映射关系,建立嫌疑人特征预测模型,通过此模型对案件的嫌疑人的特征进行预测,进而与嫌疑人库中人员进行特征比对,从而缩小嫌疑人的排查范围,加快公安机关的破案速度。

本文研究提出的模型与先前的研究的区别在于:先前的嫌疑人预测问题主要是分析嫌疑人的作案可能性,例如利用随机森林分类的方法^[3],输出结果是 0(不是嫌疑人)或 1(是嫌疑人),存在错误判断的可能性;利用 probit 模型预测嫌疑人可能性^[4],但是不能针对嫌疑人的某一特征具体分析。本文提出的特征预测模型主要有 2 点优势:1)模型直接预测该案件嫌疑人的各个特征,再将特征组合与备选嫌疑人进行比对,而不是直接判定是否是嫌疑人,避免了错判,并能够针对某一具体特征分析,提高预测的可解释性;2)模型的输出特征可以直接与嫌疑人库中的嫌疑人全部特征进行相似度计算,对排名靠前的嫌疑人进行重点侦查,加快办案速度,提高侦破效率。

预测过程如图 1 所示,首先对案件数据进行预处理,预处理包括对数据进行清理、变换、规约和离

散化等;然后利用信息增益特征选择方法进行特征选择;之后利用粒子群优化(Particle swarm Optimization, PSO)算法对模型的参数进行优化处理;最后利用 PSO 算法得到的最优的 SVM 参数对数据进行训练学习,构建特征预测模型。在 SVM 学习的过程中,数据集分成训练集和测试集两部分,训练集用来训练特征预测模型,测试集用来对模型的预测精度进行评价。

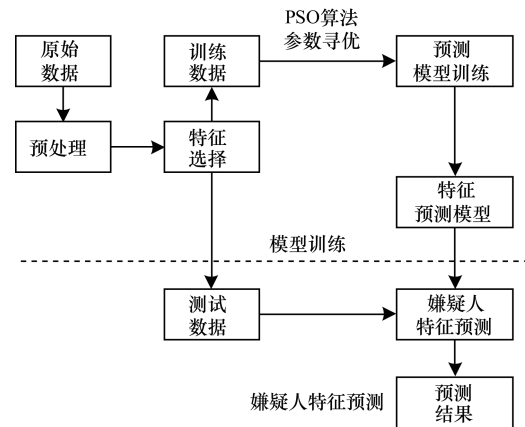


图 1 基于 SVM 的犯罪嫌疑人预测原理

2 基于支持向量机的犯罪嫌疑人预测模型

2.1 支持向量机

支持向量机是近年来发展起来的一种有效的非线性问题处理工具,具有完备的统计学习理论和出色的学习性能。支持向量机是针对结构风险最小化原则提出的,具有很好的泛化能力。支持向量机算法是一个凸优化问题,局部最优解一定是全局最优解。这些特点使支持向量机广泛应用于分类和回归问题。SVM 在面对小样本、高维模式识别和非线性的问题中表现出许多特点与优势。SVM 在模式识别^[15]、回归分析^[16]、时间序列预测^[17]等领域都得到了很好的发展与广泛的应用。基于上述优点,本文运用支持向量机研究嫌疑人预测问题,具有可行性。

SVM 的核函数选择比较重要,选择合适的核函数能提高 SVM 的预测精度,本文选择径向基核函数(Radial Basis Function, RBF)作为 SVM 的核函数。SVM 预测模型精度的主要影响因素包括惩罚参数 C 和核函数参数 g 。如果采用传统的梯度下降算法或者穷举法对 SVM 模型参数进行优化,不仅耗时长,而且效果差。本文通过粒子群优化算法^[18]的全局搜索能力,对 SVM 的参数进行寻优,提高 SVM 预测模型的精度。

2.2 嫌疑人特征预测模型构建

2.2.1 基于信息增益的特征选择

本文基于信息增益的方法进行特征选择。在信息增益中,特征重要性的衡量标准是能够给分类系

统带来多少信息,带来的信息量越多,说明特征越重要^[19]。假定样本集合 D 中第 k 类样本所占的比例为 $p_k (k=1,2,\dots,|y|)$, 则 D 的信息熵定义为:

$$Ent(D) = - \sum_{k=1}^{|y|} p_k \lg p_k \quad (1)$$

假定特征 a 有 V 个取值 $\{a^1, a^2, \dots, a^V\}$, 如果使用 a 对 D 进行划分, 那么会产生 V 个分支结点, 其中, 第 v 个分支结点包含了 D 中所有在特征 a 上取值为 a^v 的样本, 记为 D^v 。根据式(1)计算出 D^v 的信息熵, 考虑到不同的分支结点所包含的样本数不同, 给分支结点赋予权重 $|D^v| / |D|$, 则特征 a 的信息增益为:

$$Gain(D, a) = Ent(D) - \sum_{v=1}^V \frac{|D^v|}{|D|} Ent(D^v) \quad (2)$$

例如, 假定案件数据集 D 中包括需要预测的嫌疑人的性别特征, 根据式(1)可算出信息熵 $Ent(D)$ 。需要选择的特征集中包含时间特征, 时间有 3 个可能的取值: $\{\text{上午}, \text{下午}, \text{夜间}\}$, 若使用时间特征对 D 进行划分, 则可以得到 3 个子集, 分别记为: D^1 (时间 = 上午), D^2 (时间 = 下午), D^3 (时间 = 夜间), 分别计算每个子集中包含男(正例)和女(反例)的比例, 然后根据式(1)可计算出用“时间”划分之后获得的 3 个分支结点的信息熵: $Ent(D^1)$, $Ent(D^2)$, $Ent(D^3)$ 。然后根据式(2)可计算出时间特征的信息增益 $Gain(D, \text{时间})$ 。类似地, 本文可以计算其他特征的信息增益。特征选择中, 一般选取信息增益比较大的特征。

2.2.2 模型构建过程

假定所需要预测的嫌疑人有 k 个特征, 即 k 个类别特征, 对每个类别特征运用信息增益的方法进行特征选择, 则选取了 k 组类别特征, 对 k 组类别特征取交集, 得到特征预测模型的输入特征集合 T , 模型输出为嫌疑人特征向量。

假设每个案件有 m 个案件特征, k 个嫌疑人特征, n 个训练样本, 则构成样本集 S :

$$S = \{X_i, (y_1, y_2, \dots, y_k)\}, X_i \in \mathbb{R}^m, i = 1, 2, \dots, n \quad (3)$$

其中, X_i 为案件特征向量, y_k 为嫌疑人特征。模型的输出为 k 个嫌疑人特征, 因为 SVM 分类器只能单输出, 即输出一个特征, 所以根据 k 个嫌疑人特征构造 k 个分类器, 即将特征预测模型分解为 k 个分类器模型(简称分模型), 则每个分模型构成样本集 s_j :

$$s_j = \{X_i, y_j\}, X_i \in \mathbb{R}^m, y_j \in \{1, 2, \dots, f\} \\ j \in \{1, 2, \dots, k\}, i = 1, 2, \dots, n \quad (4)$$

每个分模型的输入为特征集合 T , 输出为所需要预测的某一特征。这样, 特征预测模型包括 k 个分模型, 每个分模型预测一个嫌疑人特征, 基于每个分类器的结果, 得到嫌疑人的 k 个特征组合成嫌疑人的特征向量, 即为特征预测模型的最终输出。

嫌疑人特征预测模型构建过程图 2 所示, 基于

k 个嫌疑人特征, 将原始数据集 D 分解为 k 个分模型数据集 D_k , 分别训练得到 k 个分模型, 每个分模型的输出为嫌疑人的一个特征 y_k , k 个分模型的输出结果汇总为嫌疑人特征向量 $(y_1, y_2, \dots, y_{k-1}, y_k)$, 即为特征预测模型的输出。其中, 分模型的训练过程: 基于 PSO 算法参数寻优得到 SVM 的最优参数 C 和 g , 然后利用 SVM 对训练数据训练得到分模型。

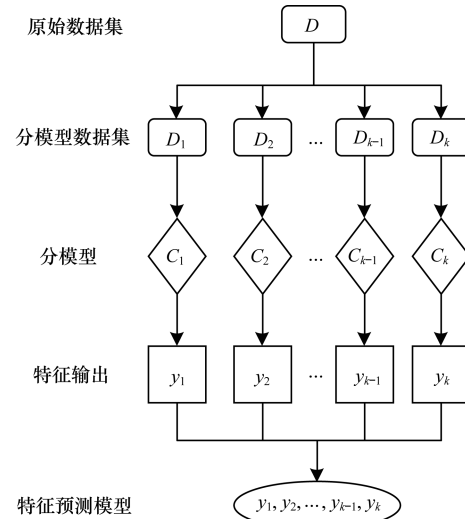


图 2 嫌疑人特征预测模型构建过程

3 实验结果与分析

3.1 预测模型的特征选择

由于犯罪案件数据的隐私性与保密性, 难以从公安机关的数据仓库中获得实际的犯罪数据。本文的实验数据集来源是美国司法统计局的 1992 年—2005 年国家犯罪事件受害者调查级联文件。SVM 的最终决策函数只由支持向量决定, 复杂程度取决于支持向量的数目, 而不是样本空间的维数。根据上述数据集, 选择的嫌疑人特征为年龄、性别、种族、是否黑帮成员、是否吸毒, 作为类别特征, 即模型的输出变量。对于每一个类别特征, 运用信息增益方法进行特征选择, 这样, 基于 5 个类别特征选取了 5 组特征, 对 5 组特征取交集, 得到 11 个特征, 作为模型的输入特征。这 11 个特征分为两部分:

1) 案件基本信息特征: 犯罪事件发生的月份、时间、地点、案件发生室内或者室外、案件种类等 5 个特征。犯罪事件发生的月份为 12 个月中某一个月; 时间是案件发生在某一天的具体时段, 例如上午或者下午; 地点是案发的具体地点, 例如住处、宾馆和车库等; 案件发生室内或者室外表示案件是在室内发生还是室外发生; 案件种类是案件发生的类别, 分为 3 类, 即暴力犯罪、盗窃和家庭犯罪。每一大类别根据案件发生的轻重细分, 例如暴力犯罪可以细分为有伤害严重攻击行为的抢劫、无伤害轻度攻击行为的抢劫等, 3 类别的案件类型细分为 41 个小类型。

2)受害者特征:年龄、性别、种族、婚姻状况、职业、受教育程度。婚姻状况包括已婚、寡、离婚、分居和未婚;种族划分为黑种人、白种人和其他;职业是案件发生时受害人所从事的职业,细分为医生、护士和技师等 21 种职业类型;受教育程度细分为小学及以下、初中、高中、学士、硕士、博士等类别。

嫌疑人包括 5 个特征(年龄、性别、种族、是否黑帮成员和是否吸毒),则需要构建 5 个分模型,即构建 5 个 SVM 分类器,分模型的输入为案件基本信息特征和受害者特征,输出为嫌疑人特征中的某一特征值,分模型输出结果汇总为嫌疑人的特征向量,即为预测模型的最终输出结果。

3.2 实验环境

本文的实验环境包括以下方面:

1)软件条件:Matlab R2015b, Weka 3.6。

2)硬件条件: Intel (R) Core (TM) 2 Duo CPU E7500 @ 2.93 GHz, 2 GB 内存, 500 GB 硬盘, Windows 7 操作系统。

本文中的 PSO 算法寻优过程是以 Matlab R2015b 为实验平台。SVM 算法是通过 WEKA 平台中的 LIBSVM 软件包实现。

3.3 实验过程

实验过程如下:

1)数据处理。数据集中去除有缺失值的案件,选取 3 091 个案件数据进行实验。将案件的特征值量化。例如时间特征的量化,一天 24 h 划分为 6 个时间段:6 a. m—12 noon, 12 noon—3 p. m, 3 p. m—6 p. m, 6 p. m—9 p. m, 9 p. m—12 midnight, 12 midnight—6 a. m, 相对应的特征值为 1~6。量化后的部分数据如表 1 和表 2 所示。嫌疑人的性别、是否黑帮成员和是否吸毒特征取值只有 2 个,属于二分类问题,年龄和种族特征取值有多个,属于多分类问题。SVM 处理多分类问题的算法主要有 2 种:一对多(one-against-rest)和一对一(one-against-one)。其中,已经证明一对一模式性能优于一对多模式^[20]。因此,本文采用一对一的模式。

表 1 案件基本特征和受害者特征量化结果

案件 ID	月份	时间	地点	受害者年龄	婚姻状况	性别	受教育程度	种族	职业	案发室内外	案件类型
1	3	4	8	2	3	2	3	1	27	1	5
2	1	2	5	3	1	2	4	2	3	2	13
3	6	5	9	2	5	1	4	1	27	2	5
4	5	4	21	2	1	2	3	1	27	2	12
5	4	4	23	3	1	2	4	1	27	1	3

表 2 嫌疑人特征量化结果

案件 ID	性别	年龄	黑帮成员	吸毒情况	种族
1	1	5	2	1	1
2	1	4	2	2	2
3	1	5	2	2	1
4	1	5	2	1	1
5	1	6	2	2	1

2 991 条数据用于训练数据,随机抽取的 100 条数据用于测试。为消除各影响因子数据由于量纲和单位不同造成的影响,模型运行前,要将样本的输入作归一化处理,使处理后的数据在 [0,1] 区间。归一化公式为:

$$\hat{x}_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

其中, x_i 为实际值, x_{\max} 和 x_{\min} 分别为最大值和最小值, \hat{x}_i 为归一化后的值。

2)在 Matlab 平台下,利用 PSO 算法分别对 5 个分模型进行 SVM 参数寻优,确定惩罚参数 C 和核函数参数 g 。

3)由 PSO 算法寻优得到的惩罚参数和核函数参数,利用 SVM 训练得到 5 个分模型,分模型组合为最终的嫌疑人特征预测模型。

4)通过 100 条测试数据对模型进行检验,对模型的预测精度进行评价。

3.4 模型评价指标

模型通过两方面指标评价模型的性能优劣:

1)分模型的精确率 P_c 和召回率 R_c :

$$P_c = \frac{N_1}{N_2}$$

$$R_c = \frac{N_1}{N_3}$$

$$F_c = \frac{2P_c R_c}{P_c + R_c} \quad (6)$$

其中, N_1 表示正确分类的样本数, N_2 表示模型判定的总样本数, N_3 表示样本总数, F_c 是加权调和平均值, F_c 值较高时则能说明实验方法比较有效。

2)预测模型输出的特征向量与嫌疑人真实特征向量的相似度大小 Sim :

$$Sim(X, Y) = \frac{M_1}{M_2} \times 100\% \quad (7)$$

其中, X 为模型输出的特征向量, Y 为真实的特征向量, M_1 为特征向量 X 包含正确特征的个数, M_2 为特征向量包含的总特征数。 Sim 值越接近 1 说明模型预测效果越好。

3.5 结果与分析

3.5.1 PSO 参数设置和寻优结果

在 Matlab 平台下,利用 PSO 算法分别对分模型 SVM 参数寻优。粒子群算法参数设置为:初始种群数量为 20,加速常数 $c_1 = 1.5, c_2 = 1.7$,惯性权重 $\omega = 1$,最大迭代次数为 200,适应度函数值取 3 重交叉验证的分类正确率。实验结果得到的 5 个分模型的最优参数 C 和 g 如表 3 所示。其中, C_1 为性别预测模型; C_2 为种族预测模型; C_3 为是否黑帮成员预测模型; C_4 为吸毒情况预测模型; C_5 为年龄预测模型。然后利用得到的 SVM 最优参数训练数据集构造 5 个分模型,分模型输出汇总即为特征预测模型的输出。

表 3 分模型惩罚参数 C 和核函数参数 g

分模型	C	g
C_1	1.009 5	618.734 7
C_2	0.986 2	828.934 9
C_3	14.211 7	0.010 0
C_4	1.896 6	0.010 0
C_5	70.970 5	401.233 6

3.5.2 分模型的精确率和召回率

在 Weka 平台下,利用 PSO 算法参数寻优得到的最优参数 C 和 g ,采用 10 折交叉验证的方法,得到分模型的 P_c 与 R_c 如表 4 所示。性别和是否黑帮成员 2 个特征的 F_c 值分别达到了 82.3% 和 92.8%,模型的预测效果比较好;种族和吸毒情况模型的 F_c 值为 75.2% 和 71.5%,也达到了良好效果;年龄预测模型的 F_c 值最低为 60.1%,这与年龄时间段的划分太细、输入变量选择(数据集限制)有一定关系。总体来讲,分模型的预测效果良好。

表 4 分模型精确率和召回率 %

分模型	P_c	R_c	F_c
C_1	87.9	86.8	82.3
C_2	83.9	80.4	75.2
C_3	94.3	94.3	92.8
C_4	71.5	71.5	71.5
C_5	65.5	65.1	60.1

3.5.3 相似度测试结果

通过 100 条案件数据对模型进行检验,部分结果如表 5 所示。

表 5 特征模型的测试结果 %

案件编号	真实特征向量	模型输出向量	相似度
1	1,1,2,2,6	1,1,2,2,4	80
2	2,2,2,1,6	2,1,2,1,3	60
3	1,1,2,2,2	1,1,2,1,6	60
4	1,1,1,1,4	1,1,1,1,6	80
⋮	⋮	⋮	⋮
99	1,1,2,2,6	1,1,2,2,6	100
100	2,1,2,1,6	1,1,2,1,4	60

实验结果显示,特征预测模型预测的特征与真实特征的平均相似度达到了 78.8%,进而言之,5 个嫌疑人特征能够预测将近 4 个特征,预测效果良好。对于 5 个嫌疑人特征,即年龄、性别、种族、是否黑帮成员、是否吸毒,其中性别、种族、是否黑帮成员 3 个特征预测准确率较高,分别达到 88.2%、80.4% 和 94.6%,是否吸毒和年龄准确率相对较低,达到了 70.8% 和 67.2%。实验证明,运用支持向量机实现的嫌疑人特征预测模型具有可行性。通过此模型,可以预测新发生案件的嫌疑人所具有的可能特征,然后与嫌疑人库中的嫌疑人进行特征比对,特征相似度高的嫌疑人具有较高的作案可能性,重点侦查,从而加快办案速度。

为了说明 SVM 的优势所在,在 Weka 平台下,采用 J48 决策树算法与朴素贝叶斯算法与其进行比较,J48 算法和朴素贝叶斯算法的参数设置为 Weka 默认参数,进行 7 次对比实验,每次实验每个方法分别进行 50 次实验,对 Sim 取平均值。第一次输入变量只有 5 个特征,即月份、时间、地点、受害者年龄和婚姻状况,之后 6 次对比实验分别加入性别、受教育程度、种族、职业、案发室内外、案件类型等特征。其结果见图 3。

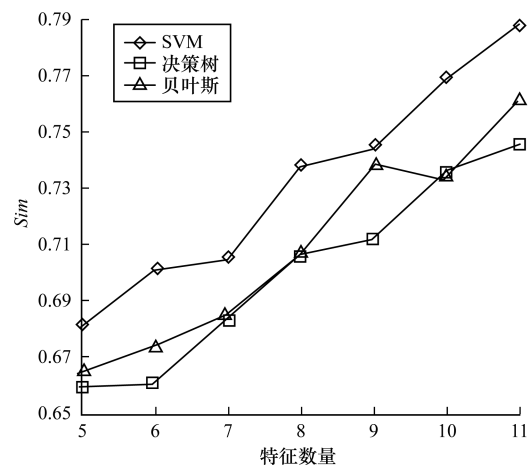


图 3 不同模型预测效果比较

从图 3 可以看出,3 种模型中,SVM 分类方法得出的特征相似度为最好,其次为朴素贝叶斯和决策树,体现出了 SVM 分类准确率高特点。从图 3 中还可以看出,随着输入特征变量的增多,输出的特征与实际的特征相似度越高,说明在一定范围内,模型的输入变量越多,预测效果越好,在以后的研究中选择更多的案件特征应用到模型中,提高预测精度。

4 结束语

针对犯罪案件嫌疑人预测问题,本文基于 SVM 构建嫌疑人特征预测模型。通过对嫌疑人的多个特征建立多个 SVM 分类器,将各分类器输出组合为特征预测模型的输出,即嫌疑人特征向量,从而实现了对嫌疑人的预测。实验结果表明,该模型选取年龄、性别、种族、是否黑帮成员、是否吸毒等 5 个特征对嫌疑人进行预测,达到了较好的效果。由于实验缺少合理的实际应用数据,因此后期需要通过海量的实际数据做进一步研究,以提高模型的预测精度和可靠性。

参考文献

- [1] 邓建锋. 基于 R 语言的罪犯数据聚类研究 [D]. 广州: 中山大学, 2014.
- [2] 赵 军. 我国犯罪预测及其研究的现状、问题与发展趋势——对“中国知网”的内容分析 [J]. 湖南大学学报 (社会科学版), 2011, 25 (3): 155-160.
- [3] 李 明, 薛安荣, 王富强, 等. 犯罪量动态优化组合预测方法 [J]. 计算机工程, 2011, 37 (17): 274-275.
- [4] 孙菲菲, 曹 卓, 肖晓雷. 基于随机森林的分类器在犯罪预测中的应用研究 [J]. 情报杂志, 2014 (10): 148-152.
- [5] 罗森林, 刘 峥, 郭 亮, 等. 基于 Probit 的犯罪嫌疑人判定方法研究 [J]. 北京理工大学学报, 2011, 31 (11): 1337-1341.
- [6] XIANG Yang, CHAU M, ATABAKHSH H, et al. Visualizing Criminal Relationships: Comparison of a Hyperbolic Tree and a Hierarchical List [J]. Decision Support Systems, 2005, 41 (1): 69-83.
- [7] BRUNSDON C, CORCORAN J, HIGGS G. Visualising Space and Time in Crime Patterns: A Comparison of Methods [J]. Computers Environment & Urban Systems, 2007, 31 (1): 52-75.
- [8] ENZMANN D, PODANA Z. Official Crime Statistics and Survey Data: Comparing Trends of Youth Violence Between 2000 and 2006 in Cities of the Czech Republic, Germany, Poland, Russia, and Slovenia [J]. European Journal on Criminal Policy & Research, 2010, 16 (3): 191-205.
- [9] TOLLENAAR N, HEIJDEN P G M V D. Which Method Predicts Recidivism Best?: A Comparison of Statistical, Machine Learning and Data Mining Predictive Models [J]. Journal of the Royal Statistical Society, 2013, 176 (2): 565-584.
- [10] USHA D, RAMESHKUMAR K. A Complete Survey on Application of Frequent Pattern Mining and Association Rule Mining on Crime Pattern Mining [J]. International Journal of Advances in Computer Science and Technology, 2014, 3 (4): 264-275.
- [11] AYAT S, FARAHANI H A, AGHAMOHAMADI M, et al. A Comparison of Artificial Neural Networks Learning Algorithms in Predicting Tendency for Suicide [J]. Neural Computing & Applications, 2013, 23 (5): 1381-1386.
- [12] PFLUEGER M O, FRANKE I, GRAF M, et al. Predicting General Criminal Recidivism in Mentally Disordered Offenders Using a Random Forest Approach [J]. BMC Psychiatry, 2015, 15 (1): 1-10.
- [13] VURAL M S, GOK M. Criminal Prediction Using Naive Bayes Theory [J]. Neural Computing & Applications, 2016: 1-12.
- [14] 罗 瑜. 支持向量机在机器学习中的应用研究 [D]. 成都: 西南交通大学, 2007.
- [15] 闫志勇, 关 欣, 李 铨. 基于 SVM 和增强型 PCP 特征的和弦识别 [J]. 计算机工程, 2014, 40 (7): 170-173.
- [16] 安 欣, 徐 硕, 张录达, 等. 多因变量 LS-SVM 回归算法及其在近红外光谱定量分析中的应用 [J]. 光谱学与光谱分析, 2009, 29 (1): 127-130.
- [17] 丁 芳. 基于 Rijke 管的热声振动涡脱落现象研究及 SVM 时间序列预测模型 [D]. 杭州: 浙江大学, 2016.
- [18] TRELEA I C. The Particle Swarm Optimization Algorithm: Convergence Analysis and Parameter Selection [J]. Information Processing Letters, 2003, 85 (6): 317-325.
- [19] 郑 伟, 吕建新, 马艳丽. 一种基于扩展互信息算法的特征选择方法 [J]. 微计算机信息, 2010, 26 (24): 223-224.
- [20] HSU C W, LIN C J. A Comparison of Methods for Multiclass Support Vector Machines [J]. IEEE Transactions on Neural Networks, 2002, 13 (2): 415-425.

编辑 刘 冰