

基于优化 GridSearch-SVM 算法的服装版型分类预测研究

郑文靖

(西安工程大学 计算机科学学院, 陕西 西安 710048)

摘要:服装定制推荐在服装定制过程中有着重要意义,要实现定制推荐的过程,首先根据其量体数据预测服装版型。根据某服装企业的量体数据与版型数据之间的组合关系,采用 SVM 分类算法,训练分类模型,根据量体数据预测出服装版型,并提出一种二度网格搜索算法(GridSearch),进行支持向量机的 (C, g) 参数寻优,实现最优的分类效果,并缩短运算时间。

关键词:服装定制;量体数据;服装版型;SVM;GridSearch

中图分类号:TS941.26

文献标识码:A

文章编号:1673-0356(2020)04-0048-04

当前时代下,随着消费者对个性化穿着以及对服装尺寸适宜度要求的扩大,服装定制模式渐渐在国内占据了重要的市场份额。定制化是能够满足消费者对个性化追求的重要途径,在“互联网+”时代中,定制行为模式正发生着重要的变革^[1-2]。

随着消费者需求的扩大,传统的定制模式可能对于众多的消费者来说,无法及时满足所有消费者的需求,因此实现服装定制推荐是当前服装定制市场的一种趋势。针对服装定制领域,目前国内主要聚焦于三维人体建模参与定制,并且多数强调的是款式、面料、部件样式的选择,忽略了消费者历史数据中有用的信息,因此,目前国内外服装定制推荐领域研究较少。文献^[3]通过交互日志挖掘,采用有限状态机技术描述客户需求的会话交互行为模型,并运用约束满足理论和多属性效用理论,开发了一种基于约束满足和情感效用的个性化西服定制推荐系统;文献^[4]基于 Kinect 描述了一种人体及服装重构的方法,由多个 Kinect 捕获人体模型,对服装进行 3D 建模,将其变形以适合人体,然后应用缝合将这些模式缝合在一起;文献^[5]设计和开发了基于专家知识的个性化服装搭配系统,提供了个性化的服装搭配服务;文献^[6]采用 BDEU 决策树算法,构筑了用户类别偏好模型,向用户提供了个性化的推荐服务。以上研究对于服装定制推荐领域的涉及有限,因此,本文致力于研究服装定制领域中的量体数据与服装版型。

在大数据时代下,为方便消费者快速选择适合自

己的定制服装版型以及面料绣花搭配,通过 SVM 分类算法对众多量体数据进行挖掘分析,根据用户自身的量体数据进行计算后推荐给用户适合的版型,从而根据版型进一步实现用户的个性化定制搭配推荐。个性化的定制推荐与传统的定制选择相结合,才能为用户提供更好的定制服务。

1 量体特征值与版型类型的 SVM 数据模型

1.1 量体数据特征参数集初始化

设量体数据为输入样本 X ,以上衣量体数据为例, $x_1, x_2, x_3, \dots, x_n$ 分别为代表不同部位的量体特征值,如衣长、胸围、下摆、中腰、肩宽、袖长等部位,输入 n 个特征值作为样本值 $X(X_i)$

建立量体数据参数集 $X = \{x_1, x_2, x_3, \dots, x_n\}$

由于量体特征值之间存在较大差异,如胸围与肩宽之间的差异。需对量体数据集进行归一化处理,使得参数空间在各维度分布均匀,模型具有较优的迭代速度和分类预测效果。归一化方法为:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

1.2 构造最优超平面

以上衣为例,设 $X = \{YC, XW, ZY, XB, JK, XC\}$, YC, XW, ZY, XB, JK, XC 分别代表衣长、胸围、中腰、下摆、肩宽、袖长,设 $Y =$ 版型分类值, X 与 Y 可表示为:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), x_t \in R^d, y_t \in \{y_1, y_2, \dots, y_n\}\}$$

此关系模式下必然存在 $n - 1$ 个超平面,将这些样本分别两两区分,最终实现分类效果,超平面可表示

收稿日期:2020-01-17

基金项目:山东省科技厅项目(2015ZDZX02003)

作者简介:郑文靖(1994-),女,硕士研究生,研究方向为智能信息处理与并行计算,E-mail:997415179@qq.com。

为:

$$w^T x + b = 0$$

式中, $w \in R^d$ 为该平面的权值法向量, $b \in R$ 为偏置量。

则样本到超平面的分类间隔为: $2/\|w\|$, 由于原始样本空间数据的线性不可分性, 因此, 引入惩罚参数 C 来调整支持向量机对已知训练样本构建最优分类超平面时分类错误容忍程度^[7]。所以要建立最优超平面的问题变成在有约束的条件下求:

$$\min \varphi = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i = \frac{1}{2} (w' \cdot w) + C \sum_{i=1}^n \xi_i$$

其约束条件为:

$$y_i(w^T x_i + b) \geq 1, i \in \{1, 2, \dots, n\}$$

1.3 计算最优解

为了解决构造最优超平面的问题, 引入 Lagrange 函数^[8]:

$$L(w, b, a) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n a_i (y_i ((w \cdot x_i) + b) - 1)$$

式中, $a_i \geq 0, i=1, 2, \dots, n$ 为拉格朗日乘子, 为了求得 w 和 b 的最小值, 对 w 和 b 求偏导数, 并使得它们的偏导数为零, 即:

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^n y_i a_i x_i = 0$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^n y_i a_i = 0$$

对于已知训练样本 (x_i, y_i) , 求使得 $L(w, b, a)$ 取得最大值时的 a_i , 将二次规划问题转化为相应的对偶问题, 根据 Lagrange 对偶性原理, 将问题中的参数全部转变为仅以 Lagrange 乘子作为参数变量的目标式^[9-10]:

$$\max L(a) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i=1, j=1}^n a_i a_j y_i y_j x_i' x_j$$

$$\text{s.t. } \sum_{i=1}^n y_i a_i = 0, a_i \geq 0, i=1, 2, \dots, n$$

解得最优解 $a^* = (a_1^*, a_2^*, \dots, a_n^*)'$ 。最优法向量 w 和最优偏置量 b 分别为:

$$w^* = \sum_{i=1}^n a_i^* y_i x_i$$

$$b^* = y_i - \sum_{j=1}^n y_j a_j^* (x_i \cdot x_j)$$

根据最优值, 从而求得最优分类超平面 $(w^* \cdot x) + b^* = 0$, 则其对应的最优分类函数为:

$$f(x) = \text{sgn}(\sum_{i=1}^n a_i^* y_i x_i + b^*)$$

为实现样本空间从线性分划到非线性分划的过渡, 引入径向基(RBF)核函数, 将样本空间从低维映射到高维^[11-12]。径向基核函数表达式为:

$$K(x_i, x_j) = \exp[-\frac{\|x_i - x_j\|^2}{\sigma^2}]$$

式中, s 为径向基半径, 令 $g = \frac{1}{\sigma^2}$ 代入上式得

$$K(x_i, x_j) = \exp[-g(\|x_i - x_j\|^2)]$$

则最优分类函数为

$$f(x) = \text{sgn}(\sum_{i=1}^n a_i^* y_i K(x_i, x_j) + b^*)$$

2 基于优化 GridSearch 的 SVM 分类模型

2.1 GridSearch-SVM 原理

惩罚参数 C 和核函数参数 g 是影响 SVM 分类器性能的关键参数^[13], 其中 C 表示模型对误差的容忍度。 C 值太大容易导致过拟合的现象, 使得测试集的数据分类效果不佳; C 值太小容易导致欠拟合, 模型不能有效捕捉样本的数据特征, 泛化能力变差。 g 是选择 RBF 函数作为 kernel 后, 该函数自带的一个参数, 隐含地决定了数据映射到新的特征空间后的分布, g 越大, 支持向量越少, g 值越小, 支持向量越多。 g 值过大过小, 表明原始样本被映射至并不适用的高维空间, 无法建立较优的分类模型^[14-15]。

GridSearch 是用于 Libsvm 中的参数搜索方法^[16], 在 C 和 g 组成的二维参数矩阵中, 依次遍历网格内所有的点进行取值, 对于取定的 C 和 g 利用 $K-CV$ 方法得到在此组 C 和 g 下训练集验证分类准确率, 最终取使得训练集验证分类准确率最高的那组 C 和 g 作为最佳参数。使用 GridSearch 算法可以得到全局最优, 且 C 、 g 相互独立, 便于并行化进行^[17-18]。

基于 GridSearch-SVM 算法的服装版型分类识别, 具体操作是在 MATLAB 软件中利用 LIBSVM3.14 工具包中的 SVMcgForClass.m 函数, 使用 meshgrid 方法构建网格, 寻找最佳 C 和 g 参数, 实现用 GridSearch 优化 SVM 参数和服装版型识别, 算法如表 1 所示。

2.2 GridSearch-SVM 模型建立

在版型分类预测的过程中, 提取量体数据特征值, 并进行归一化处理, 引入 RBF 核函数, 将样本数据映射到高维空间, 再进行 SVM 模型训练、参数寻优以

及模型验证,整个版型的预测模型如图1所示。

表1 GridSearch-SVM 算法流程

步骤	流程
Step1	初始化网格大小及设置步距,定义初始 $C&g$ 参数值
Step2	循环遍历每一组参数值
Step3	分别对每一组参数值进行 SVM 模型训练,得到准确度
Step4	如果找到更优的 $C&g$ 参数,则代替之前的 best
Step5	将最好的 C 和 g 拼接,作为参数项,传入到 SVM 模型中
Step6	输出最优结果

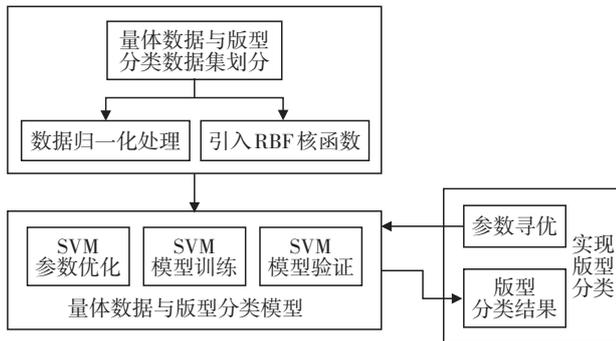


图1 GridSearch-SVM 版型预测模型

(1) 选定训练集与测试集,对数据集进行划分。

(2) 样本数据预处理:为了避免各个样本因子之间量级的差异,减少样本之间的相互影响,同时保证程序运行收敛加快,需要对样本因子进行归一化处理^[19-20],在 MATLAB 中用 mapminmax 函数实现其归一化。

$$y = \frac{(\max(y) - \min(y))}{\max(x) - \min(x)} + \min(y)$$

式中, x 、 $\min(x)$ 、 $\max(x)$ 分别是原始样本数据及其最小值、最大值; $\min(y)$ 、 $\max(y)$ 分别代表归一化后样本的最大值、最小值。

(3) 引入径向基核函数(RBF),通过调整 g 参数实现样本从低维空间到高维空间的映射,从而实现线性可分。

$$K(x_i, x_j) = \exp[-g(\|x_i - x_j\|^2)]$$

(4) 使用二度 GridSearch 算法进行 SVM 模型训练及 $C&g$ 参数寻优。

粗略搜索阶段 定义初始网格,设置大步距,获得局部最优参数区间;

精确搜索阶段 以最优参数组为搜索中心,设置小步距,不断扩大搜索范围,逐步跳出局部最优,获得全局最优解,从而实现分类效果最优。

3 试验结果与分析

(1) 试验数据来自于某大型服装企业平台,试验共采用 169 组量体数据与版型之间的对应数据,随机选取 131 个样本作为训练集,38 个样本作为测试集,见表 2。

表2 量体数据版型数据集(上衣部分示例)

编号	YC	XW	XB	ZY	JK	XC	y
1	71.0	100	86.6	98	43.8	56.5	24
2	78.4	114	102.0	112	48.0	61.0	47
3	71.4	100	86.6	98	43.8	56.0	20
4	78.4	114	102.0	112	48.0	61.0	41
5	62.0	101	88.0	107	41.6	59.0	46
6	82.5	126	114.6	124	52.0	65.0	105
7	79.3	116	102.6	114	28.6	62.0	107

(2) 首先设定所要搜索的 (C, g) 参数的初始网格搜索范围及初始步长,其中, $C&g$ 的初始网格搜索范围为 $[2^{-10}, 2^{10}]$, 初始步长为 4.5, 通过交叉验证方法获得局部最优参数组;

(3) 在其附近进行小范围的精确网格搜索,其 C 的网格搜索范围为 $[2^{-2}, 2^4]$, g 的网格搜索范围为 $[2^{-4}, 2^4]$, 其搜索步长为 0.05;

(4) 将最终得到的参数 (C, g) 重新传入到支持向量机的和函数中,建立基于二度网格搜索的支持向量机模型。

试验结果如图 2 所示。

从试验结果看出,将 GridSearch-SVM 算法运用于服装版型的预测分类研究中,有良好的分类效果,且将 GridSearch 网格搜索算法分步搜索,可大大缩短参数寻优时间,本次试验运行时间为 4.47 s,最佳 C 参数为 36.7, g 参数为 1.3,最终试验的版型分类准确率在 90% 以上,收敛情况好时可达 100%。

4 结语

为挖掘量体数据与服装版型的关系,构造出量体数据及服装版型之间的 SVM 数据模型,并通过 Matlab 进行仿真试验,使用 SVM 算法进行模型训练及模型验证,实现分类效果,并通过 GridSearch 二度网格搜索算法进行参数寻优,使准确率达到理想状态。量体数据及版型的预测研究实现了在服装定制推荐过程中根据量体数据进行版型推荐的过程,使得服装定制推荐的过程向前推进,逐步实现服装定制元素的推荐。针对量体数据及版型的对服装定制推荐有重要的理论与实践意义。

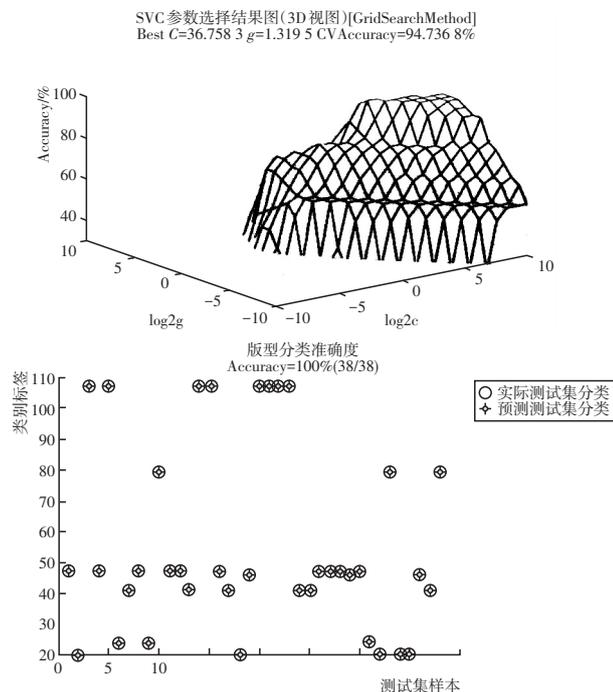


图2 试验结果

参考文献:

- [1] 胡耀雯,陈晓玲.非传统定制模式分析[J].纺织科技进展, 2019,(7):48-51.
- [2] 戴玉芳,杜岩冰,凌军,等.服装工业化定制中的信息交互[J].纺织高校基础科学学报,2019,32(1):30-36.
- [3] 石美红,艾磊,李维乾,等.基于约束满足的个性化西服定制推荐系统[J].西安工程大学学报,2015,29(3):312-319.
- [4] ZHU S, MOK P Y, KWOK Y L. Intelligent 3D accurate human body modeling for virtual try-on of clothing[C]// Web 3D 2012-17th International Conference on 3D Web Technology, 2012.
- [5] 冯娇.基于专家知识的个性化服装推荐搭配系统[D].上海:东华大学,2018.
- [6] 齐扬,朱欣娟.基于数据挖掘的服装推荐系统研究[J].西安工程大学学报,2010,24(4):438-443.
- [7] 王健峰,张磊,陈国兴,等.基于改进的网格搜索法的SVM参数优化[J].应用科技,2012,39(3):28-31.
- [8] 刘小生,章治邦.基于改进网格搜索法的SVM参数优化[J].江西理工大学学报,2019,40(1):5-9.
- [9] 顾凯成.改进网格搜索的支持向量机参数优化研究及应用[D].兰州:兰州理工大学,2016.
- [10] 李斌.支持向量机的模型与参数选择研究[D].广州:华南理工大学,2018.
- [11] SU C T, YANG C H. Feature selection for the SVM: An application to hypertension diagnosis[J]. Expert Systems with Application, 2008, 34(1): 754-763.
- [12] 张学工.关于统计学习理论与支持向量机[J].自动化学报, 2000, 26(1): 34-39.
- [13] CHAPELLE O, VAPNIK V, BOUSQUET O, et al. Choosing multiple parameters for support vector machines[J]. Machine Learning, 2002, 46(1-3):131-159.
- [14] 罗小燕,陈慧明,卢小江,等.基于网格搜索与交叉验证的SVM磨机负荷预测[J].中国测试,2017,43(1):132-135.
- [15] 李琼,董才林,陈增照,等.一种快速的SVM最优核参数选择方法[J].计算机工程与应用,2010,46(15):165-168.
- [16] LIU X L, JIA D X, LI H, et al. Research on Kernel parameter optimization of support vector machine in speaker recognition[J]. Science Technology and Engineering, 2010, 10(7): 1669-1673.
- [17] 温博文,董文瀚,解武杰,等.基于改进网格搜索算法的随机森林参数优化[J].计算机工程与应用,2018, 54(10): 154-157.
- [18] 刘梅.网格自适应直接搜索算法的研究[D].西安:西安电子科技大学,2018.
- [19] 汤荣志,段会川,孙海涛.SVM训练数据归一化研究[J].山东师范大学学报(自然科学版),2016,31(4):60-65.
- [20] 王剑.关于网络数据归一化处理探讨[J].信息系统工程, 2015,(12):60.

Research on Classification and Prediction of Garment Pattern Based on GridSearch-SVM Algorithm

ZHENG Wen-jing

(School of Computer Science, Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Clothing customization recommendation played an important role in the process of clothing customization. In order to realize the process of customization recommendation, the clothing pattern was predicted according to body-measure data. According to the combination relationship between the body-measure data and the pattern data of a clothing enterprise, SVM classification algorithm was used to train the classification model, and the clothing pattern was predicted according to the body-measure data. A two-dimensional grid search was proposed. The (C, g) parameters of support vector machine was optimized, to achieve the optimal classification effect and shorten the operation time.

Key words: garment customization; body-measure data; garment pattern; SVM; GridSearch