

Individualized Psychology Predict Model Based on Support Vector Machine

XIONG Zeng-Gang, ZHANG Xue-Min, CHEN Jian-Xin

School of Computer, Xiaogan University, Xiaogan, 432000, China
e-mail gavinwill2008@gmail.com

Abstract: Taking the guide system as background, an individualized psychology predict model is constructed by using SVM (support vector machine) algorithm and do the standardized proposal to sample data. This model can predict the unknown-type clothes and get the clothes which users like, then realize the aim of providing the shopping guide service for users. By comparing SVM algorithm with quantification theory I algorithm, genetic algorithm and BP neural network algorithm in the aspects of the predicting accuracy and users' satisfaction, this model is proved to be feasible and effective.

Keywords: SVM; individualized service; psychology predict; guide system; artificial psychology

支持向量机的个性化心理预测模型

熊曾刚, 张学敏, 陈建新

孝感学院计算机与信息科学学院, 孝感, 中国, 432000
e-mail gavinwill2008@gmail.com

【摘要】本文以服装导购系统为背景, 采用支持向量机(SVM, support vector machine)算法构建了一个个性心理预测模型, 并对样本数据进行了标准化处理。该模型可对未知类型的服装样本进行预测, 得到用户所喜爱的服装, 从而为顾客提供个性化的导购服务。通过与数量化 I 类理论、遗传算法及 BP 神经网络算法在预测结果准确率及用户满意度方面的比较, 充分显示了该预测模型的优越性。文档定义了。

【关键词】支持向量机; 个性化服务; 心理预测; 服装导购; 人工心理;

1 引言

“人性化、个性化”是当今时代大潮的呼唤, 个性化的导购系统是发展的必然趋势。目前, 国内外对导购系统的研究大多集中在系统网络功能的实现和导购的网络自动化技术研究上, 特别是信息搜索技术的实现, 并没有真正做到“个性化、人性化”, 即对用户的个性化偏好指导不够, 消费需求的差异化在导购信息中未能引起足够的重视与体现。针对这种情况, 本文旨在搭建一个个性化心理预测模型, 该模型可探寻不同用户对商品喜好的个性倾向, 为用户筛选出符合其心理的商品, 以真正实现商品导购的个性化。

由于人类心理是个复杂的大系统, 心理活动普遍存在模糊性、混沌性, 具有非线性特性。支持向量机采用结构风险最小化原则, 解决了小样本、非线性、高维数、局部极小点等实际问题, 具有很强的泛化能

力。本文选取了 8 个典型的形容词作为服装的评价指, 提取了服装的特征值, 并采用 SVM 算法建立了个性化心理预测模型, 最后通过实验证明了该模型的优越性。

2 支持向量机的基本原理

给定两类训练样本集为

$$s = \{x_i, y_i\}, i = 1, 2, \dots, n$$
$$x_i \in R^N, y_i \in \{+1, -1\}$$
(1)

公式 (1) 中, n 为训练样本个数; N 为每个训练样本向量的维数; y 为类别标号 (其值等于 +1 为一类, -1 为另一类)。

SVM 算法的目标是寻找一个最优超平面将两类样本分开, 即

$$w^T x_i + b \geq 1, y_i = +1$$
$$w^T x_i + b \leq -1, y_i = -1$$
(2)

公式 (2) 中, w 为权值向量, $w \in R^N$; b 为分类阈值。

资助信息: 孝感学院自然科学基金项目资助(Z2009017)。

若对于所有训练样本公式(2)都成立,则认为两类样本线性可分。在优化超平面学习过程中,SVM 算法极大化两类样本间分离距离 ρ ,这里为 $\rho=2/\|w\|$ 。因此,对于线性可分情况,可通过求解以下约束优化问题得到最优超平面:

$$\text{minimize } J(w) = \frac{1}{2} w^T w \quad (3)$$

使之满足于

$$y_i(w^T x_i + b) \geq 1, i=1,2,\dots,n \quad (4)$$

对于线性不可分情况,样本将不满足公式(2)的分离条件,则引入松弛变量(非负标量) $\xi_i(i=1,2,\dots,n)$,公式(4)改写为

$$y_i(w^T x_i + b) \geq 1 - \xi_i, i=1,2,\dots,n \quad (5)$$

对于 $0 \leq \xi_i < 1$,数据点落在两类分离区域且在正确决策面一侧;若 $\xi_i > 1$,数据点落在错误决策面一侧。寻找线性不可分情况下的最优超平面可通过求解以下约束优化问题得到:

$$\text{minimize } J(w, \xi) = \frac{1}{2} w^T w + C \sum_{i=1}^n \xi_i \quad (6)$$

使之满足于

$$y_i(w^T x_i + b) \geq 1 - \xi_i, i=1,2,\dots,n \quad (7)$$

$$\xi_i \geq 0, i=1,2,\dots,n \quad (8)$$

公式(6)中常数 $C > 0$ (用户定义),控制着对错分样本惩罚的程度,这是一个凸二次优化问题,能够保证找到的极值解就是全局最优解。可利用 Lagrange 函数使原问题转化为较简单的对偶问题,即在 $0 \leq \alpha_i \leq C(i=1,2,\dots,n)$ 和 $\sum_{i=1}^n \alpha_i y_i$ 约束条件下求

解下列函数的最大值:

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \quad (9)$$

根据Karush-Kuhn-Tucker条件,优化系数 α_i 必须满足

$$\alpha_i [y_i(w^T x_i + b) - 1 + \xi_i] = 0, i=1,2,\dots,n \quad (10)$$

非零的 α_i 对应于使公式(7)等号成立的样本为支持向量,只有为支持向量的样本才能决定最终的分类结果。按公式(9)求出优化系数 α_i 后,对于给定的测试样本 x ,SVM 分类器的分类决策函数的一般形式为

$$f(x) = \text{sign}(\sum_{sv} \alpha_i y_i x_i^T x + b) \quad (11)$$

式中,sign()为符号函数,根据分类决策函数 $f(x)$ 的正负即可判定 x 所属的类别。

应用 SVM 进行模式分类的基本思路可概括为首先将输入向量映射到一个特征空间,然后在特征空间中寻找优化的线性分界线,即构建一个可分离两类的

超平面,能将两类正确分开。SVM 的训练过程总是去寻找一个全局最优解。

3 建立服装选购的个性心理预测模型

3.1 建立评价指标

形容词创造的目的就是为了方便人们表述对某种事物的心理感受。同样的,要获取用户对服装商品的感受,也需要通过这个桥梁,让用户依照词语所指的形容方向,对服装进行评价、打分。为了方便情绪的描述,研究中将形容服装的词语按相反的意思配对,如“鲜艳的-灰暗的”。这样,便可将形容词对作为评价指标,来量化用户对服装的感受。系统中采用了文献[1]中的因子分析法,选取了8个贡献度最大的形容词对作为服装的评价指标,如表1所示。

表 1. 贡献度最大的形容词对

NO.	形容词对	NO.	形容词对
1	鲜艳的一灰暗的	5	华贵的一质朴的
2	稳重的一活泼的	6	性感的一保守的
3	凌乱的一单调的	7	典雅的一豪放的
4	正式的一随意的	8	个性的一大众的

3.2 提取服装属性特征值

众所周知,服装设计的三要素是:色彩、款式、面料。服装色彩与款式、面料的结合,能产生千变万化的服装风格。

服装各部分色彩的提取采用 PHOTOSHOP 测量^[2],然后采用文献[2]中提供的颜色计算公式计算颜色取值。

服装的款式造型可分为外轮廓造型和内部结构。现代服装的外部轮廓通常归纳为 A、H、X、Y 四个基本型。服装的内部结构主要是结构线、领袖等设计。结构线可归纳为直线、弧线和曲线三种。衣领主要有立领、平领、驳领和领口领和无领五种类型,而无领又分为圆领、一字领和 V 字领。袖子按照长短分为长袖、中袖、短袖、无袖及吊带五种类型。另外,还有一个影响服装款式造型因素是衣长。而根据季节的不同,服装的特点也随之产生变化。在秋冬装中,比如风衣,有短款、长款等等,我们把它分为以下几种:腰围线以上、腰围处、腰围和臀围中间、臀围处、臀围线以下。而在夏装中,上衣衣长的变化不是很大,我们主要用衣长来表示下装的长度。裙装以膝盖为界,可简单分为短裙、及膝裙和长裙,裤装分为短裤、中裤和长裤。

服装面料可分为五种：柔软型面料，挺爽型面料，光泽型面料，厚重型面料，透明型面料。

根据以上服装款式的介绍，我们提取关于款式的服装特征并将其量化。外部轮廓有四个类型 A、X、H、Y，分别编码为 1-4；结构线有直线、弧线和曲线三种类型，分别编码为 1, 2, 3；衣领有立领、平领、驳领、领口领、圆领、一字领和 V 字领七种类型，同理编码为 1-7；袖子有长袖、中袖、短袖、无袖及吊带等五种，编码为 1-5；秋冬装中，衣长有腰围线以上、腰围处、腰围和臀围中间、臀围处、臀围线以下等五种，分别编码为 1-5；夏装中，衣长简单分为短裙、及膝裙、长裙、短裤、中裤和长裤六种，分别编码为 1-6；同理，面料也依次编码为 1-5。表 2 为样本服装其中 4 套的特征值。

表 2. 样本服装的特征值表

服装 编号	外部 轮廓	结构线	领子	袖子	衣长	面料	上装 颜色	下装 颜色
1	1	2	2	1	2	4	38.1	38.1
2	3	1	5	1	5	1	144.4	154
3	4	3	7	2	3	5	121.6	0.2
4	2	2	2	2	3	3	10	11.8

3.3 样本数据预处理

首先，创建一个包含 100 套各类服装的数据库，并将每套服装的属性特征值录入数据库中。以 t_{ij} 表示第 i 套服装第 j 个特征值的取值，然后将此特征值按公式(14)进行归一化处理，这对 SVM 分类器是至关重要的。

$$x_{ij} = \frac{t_{ij} - t_{ij\min}}{t_{ij\max} - t_{ij\min}} \quad (14)$$

其中， x_{ij} 为第 i 套服装归一化后的特征值， $t_{ij\min}$ 为第 i 套服装第 j 个特征值的最小取值， $t_{ij\max}$ 为第 i 套服装第 j 个特征值的最大取值。

3.4 基于 SVM 算法的建模过程

根据 SVM 基本原理，建立服装选购的个性化心理预测模型的过程如下：

1) 确定系统的输入输出参数。将本文所提取的八个服装属性特征值构成一组一维变量作为系统的输入向量 X ，并对 X 作归一化处理；系统的输出即服装的评价指标，即八个形容词对。

2) 构建支持向量机的学习样本。抽选一定数量的服装图片，将其特征值作为样本输入参数。以服装图片为对象作评价实验，取得评价价值 Y ， Y 即为样本输

出参数。以形容词对“稳重的一活泼的”为例，如该服装给人以稳重感，则 Y 取 1，否则 Y 取-1。

3) 选择合适的内积核函数。SVM 中不同的内积核函数将形成不同的算法，目前研究最多的核函数主要有线性核函数、多项式函数、径向基函数、Sigmoid 函数 4 种。本文经过实验比较选择了径向基函数：

$$K(x, x_i) = \exp\left\{-\frac{|x - x_i|^2}{\sigma^2}\right\} \quad (15)$$

4) 确定核函数的参数 C 和 σ^2 。本文采用 LIBSVM 软件进行实验，以形容词对“鲜艳的-灰暗的”为例， σ^2 、 C 等参数采用交叉验证法，最后确定为 $\sigma^2=0.5$ ， $C=2$ ，其余参数取默认值。

5) 输入预测样本进行预测，取得预测值，并与用户的真实评价价值进行比较，计算预测正确率。

4 实验结果与分析

本文共有 15 位实验者参与实验，做了两个测试实验。

实验 1：为了验证该模型的准确性，本文选取“鲜艳的-灰暗的”形容词对，实验者依次对全部 100 个样本进行该形容词对的评价；然后抽取前 30 个样本作为训练样本，其余 70 个为测试样本；利用 SVM 算法对前 30 个训练样本进行训练后，得到该形容词对相应的 SVM 模型，并对测试样本进行预测，取得预测值。另外，本文还利用了数量化 I 类理论、遗传算法及神经网络算法作了对比实验。各算法预测结果正确率如表 3 所示。其中，正确率 P 的计算公式如公式(16)所示。

$$P = \frac{(S - E)}{S} \times 100\% \quad (16)$$

其中， E 为测试样本中分类错误的样本数， S 为测试样本总数，在这里 $S=70$ 。

表 3 各算法的预测结果正确率

算法	正确率
数量化 I 类理论	72.44%
遗传算法	80.35%
神经网络算法	83.14%
支持向量机	87.29%

实验 2：对八对形容词对的预测结果做用户满意度调查。从 100 个样本中抽取 30 个作为训练样本，利用 SVM 模型对剩余 70 个样本进行预测，并让实验者对预测结果进行满意度评价。最后的调查结果如表 4 所示。

表 4. SVM 模型预测结果与用户满意度调查结果

形容词	属于左词的服装件数(即 y=1 的形容词)	属于右词的服装件数(即 y=-1 的形容词)	用户满意度
鲜艳的-灰暗的	56	24	82.4%
稳重的-活泼的	15	65	92.5%
凌乱的-单调的	21	59	68.1%
正式的-随意的	34	46	74.5%
华贵的-质朴的	36	44	76.8%
性感的-保守的	19	61	81.6%
典雅的-豪放的	48	32	70.2%
个性的-大众的	37	43	83.3%

注：以鲜艳的-灰暗的为例，表中的左词即鲜艳的，右词即灰暗的，表中数据为预测属于该形容词类别的服装件数。

从表 3 中，可以看出 SVM 算法预测得到的结果的正确率为 87.29%，要比其它几种算法得到的结果的正确率高很多。说明在样本数较少的情况下，SVM 的泛化能力很好，这也充分体现了 SVM 算法适用于小样本、非线性、高维数情况的优点。

从表 4 可以看出，SVM 模型在某些形容词上的预测效果较好，在其它一些形容词上的预测效果则一般。其主要原因是，人类心理是个复杂的大系统，心理活动普遍存在模糊性、混沌性，很难一下子把握其非线性的特征。另外，服装本身也是一个复杂的对象，其特征值的提取也需要进一步的精确。只有在以前的基

础上不断改进，才能建立更加准确的心理模型，这也正是我们下一步所要做工作。

5 结论

本文选取了 8 个典型的形容词作为服装的评价指标，提取了服装的属性特征值，并采用了基于统计学理论的 SVM 算法建立了服装选购中的人性化心理预测模型。该模型可预测用户对服装的心理喜好，并取得良好效果。本文还将该模型的预测结果与数量化 I 类理论、遗传算法及神经网络算法的预测结果进行了比较，结果证明了该模型的优越性。

References (参考文献)

[1] Zhiliang Wang, Yanling Zhao, Chunhui Hao, et al. An Expert System of Commodity Choose Applied with Artificial Psychology[J]. Journal of University of Science and Technology Beijing. 2001, 23(4): 376-378.
王志良, 赵彦玲, 郝春辉等. 采用人工心理理论的商品选购专家系统[J]. 北京科技大学学报, 2001, 23(4): 376-378.

[2] Yun Ma, Zhiliang Wang. Personalized Space For Individualized Fashion-Shopping With HSV Color Model. The First International Conference on Affective Computing & Intelligent Interaction. Beijing, 2003, 110-116.

[3] Guozheng Li, Meng Wang, Huajun Zeng. An Introduction to Support Vector Machines[M]. Beijing: Publishing House of Electronics Industry, 2004, 50-150.
李国正, 王猛, 曾华军. 支持向量机导论[M]. 北京: 电子工业出版社, 2004, 50.