

# SVM 算法的兵王问题解决

何杭薇

中央民族大学信息工程学院 中国北京 100044

**摘要:** SVM, 即支持向量机, 是一种主要用于解决二分类问题的相关模型。SVM 可以分为两大类: 线性类和非线性类。其主要思想是在空间中找到一个超平面, 这个超平面能够将所有数据样本划分开, 而且使得数据集中的所有数据到这个超平面的距离最短。兵王问题存在国际象棋中。在国际象棋当中有一种残局现象: 剩余三子, 分别是黑方的王, 白方的王和兵, 那么无论这三子在棋盘的布局如何, 只有两种结果, 白方胜利和逼和。这就是一个非常典型的二分类问题。本文就是利用 SVM 算法来解决兵王问题。

**关键词:** SVM 算法; 兵王问题; 解决

**Abstract:** SVM, also known as support vector machine, is a binary model. Modified, it can also be used to classify multiple categories of problems. Support vector machines can be divided into two categories: linear kernel and nonlinear. The main idea is to find a hyperplane in the space which is more enough to divide all data samples and make the shortest distance from all data in the set to this hyperplane. The Soldier Problem exists in chess. In chess, there is a endgame phenomenon: the three remaining pieces, the black king, the white king and the pawn, so regardless of the layout of the three pieces on the board, there are only two outcomes, the white victory and close down. This is a very typical dichotomy problem. This paper is to use SVM algorithm to solve the problem.

**Keywords:** SVM algorithm; Soldier Problem; solution

## 一、SVM 算法原理

SVM 学习的基本想法是寻找一个超平面, 这个超平面能够将数据集训练集准确地划分出来并且保证其距离 (几何距离) 最大。如图 1 所示: 图中的  $wx+b=0$  就是我们要找的超平面<sup>[1]</sup>。

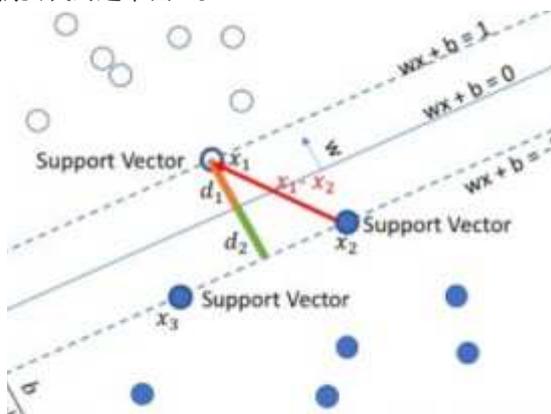


图 1 SVM 算法中的分离超平面示意图

在兵王问题中, 由 SVM 算法原理可得, 当使用线

性支持向量机, 并以高斯核 RBF 作为核函数时, 我们只需调节正则化参数 C 的范围和核函数参数  $\gamma$ 。

在这里我们选择使用网格搜索法来确定相关参数, 以此来保证在调整参数过程中的相关精准度。在使用 SVM 的过程中, 需要先将原始数据转化为 SVM 算法软件或包所能识别的数据格式, 在 IntelliJ IDEA 2020.3 环境中, 需将原始信息全部转化为数字, 并将数据标准化, 然后使用 RBF, 利用交叉验证网格搜索寻找最优参数 (C,  $\gamma$ ), 其中交叉验证防止过拟合, 网格搜索在指定范围内寻找最优参数, 最后使用最优参数来训练模型, 得到最优模型。

## 二、设计思路

### (一) 问题描述

国际象棋规则: 8\*8 棋盘, 黑、白兵各有 8 个。

兵: 第一次向前可以走一格或两格, 以后只能向前走一格, 不能后退; 吃对方子的时候可以斜前方走, 并落在对方位置。

王: 黑白各一个, 王被将死即告负, 走法是横着、

斜着都可以，但是每次只能走一格。

兵王问题：国际象棋的残局中，黑方只剩下一个王，白方剩一个兵和一个王。结局只有两种：白方将死黑方获胜，或者和棋。这是一个二分类问题。（国际象棋图示如图2）



图2 国际象棋棋盘

## （二）初始化参数设定

先将数据归一化，让其防止样本中不同特征数值大小相差较大影响分类器性能。接下来将各种变量以及含义放置如表1<sup>[2]</sup>：

表1 变量及其含义

变量	含义
iter	迭代次数
nu	选择的核函数类型的参数
obj	SVM 文件转换为的二次规划求解得到的最小值
rho	判决函数的偏置项 b
nSV	标准支持向量个数 ( $0 < a_i < c$ )
nBSV	边界上的支持向量个数 ( $a_i = c$ )
Total nSV	支持向量总个数 (Total nSV = nSV)
Accuracy	准确率

## （三）模型训练样本确定

训练样本为已有通用数据集，数据集内容为多条与“a,1,c,1,c,2,draw”相同结构的记录了下棋位置和行棋结果的数据条，draw表示逼和，英文数字 zero, one, two等表示经过多少步后白方获胜。

选择采用 Matlab 作为软件平台，代码编程实现 SVM 算法解决兵王问题。首先是样本预处理，数据归一化，划分训练集和测试集，然后设置 SVM 相关参数，开始进行相关校验，最后进行预测，获取预测结果，与正确标记对比，计算其准确率。

## 三、程序实现

### （一）调试配置过程

1. 将 libsvm 解压到 matlab\toolbox\ 目录下。
2. 打开命令行窗口运行“mex -setup”。
3. 通过 matlab 打开目录下的 make.m 文件，将其中的所有 CFLAGS 替换为 COMPFLAGS。
4. 直接运行文件 testSVMChessLibSVM.m（代码分析见后文），就可以出来相关结果。

### （二）程序分析

#### 1. 训练样本和测试样本划分：

总样本数为 28056，其中正样本为 2796，负样本为 25260 条。取其中训练集占 20%，测试集占 80%。正样本为和棋。

#### 2. 数据处理

将字符型数据数字化，设置字母为数字，并将和棋设置为 -1，1 为死：

```

1- clear all;
2- % load the data;
3- fid = fopen('chessp.DAT');
4- c = fread(fid, 3);
5-
6- vec = zeros(6,1);
7- sapp = [];
8- yapp = [];
9- while ~feof(fid)
10- string = [];
11- c = fread(fid,1);
12- flag = flag+1;
13- while c=='\n'
14- string = [string, c];
15- c=fread(fid,1);
16- end;
17- fread(fid,1);
18- if length(string)>10
19- vec(1) = string(1) - 96;
20- vec(2) = string(3) - 48;
21- vec(3) = string(5) - 96;
22- c = fread(fid,1);

```

图3：数据处理代码截图

将样本标准化，特征间的单位，也就是尺度可能不同，在对结果的影响过程当中，尺度大的特征会对结果起决定性作用，而尺度小的特征的作用可能可以被忽略不计，为了消除相关影响，于是对数据进行归一化处理。我们在训练样本上，求出每个维度的均值和方差，在训练和测试样本上同时归一化，代码如图3。

### （三）选择核函数和调参数

这里核函数选择高斯核，svm 的参数有两个，一个是 C，一个是 gamma，目前已知 5000 个训练样本，还不知道 C、gamma 参数，最后希望能够通过训练样本训练出 SVM 模型。

在获取到最优参数后建立 SVM 模型并得到精度，代码如图4：

```

118
119 %Train the SVM model by the optimal C and gamma.
120 cmd=['-t 2 -c ', num2str(maxC), ' -g ', num2str(maxGamma)];
121 model = svmtrain(yTraining, xTraining, cmd);
122 save model.mat model;
123

```

图4：核函数调用代码截图

#### (四) 结果分析

总样本数为 28056, 其中正样本为 2796, 负样本为 25260 条。取其中训练集占 20%, 测试集占 80%。通常训练的样本不能拿来测试, 因为模型本来就是通过训练样本进行创建的, 我们的目的是用训练得到的模型对其他样本进行分类, 考察的是模型的学习能力, 类似于学习过程中的知识迁移。

##### 1. 定量分析

本次实验选择高斯核作为核函数, 迭代次数共为 329430 次, 边界支持向量共 80 个, 支持向量总个数为 208 个, 正确率为 99.4015%, 结果较好, 如图 5。

```
optimization finished, #iter = 329430
nu = 0.024624
obj = -374832.583919, rho = 61.554409
nSV = 208, nBSV = 80
Total nSV = 208
Accuracy = 99.4015% (22918/23056) (classification)
```

图 5 定量分析结果图

##### 2. 定性分析

可以看见图 6 中最后曲线无限接近于 1, 显示结果较好, 模型准确率较好, 基本上能完成分类目标。

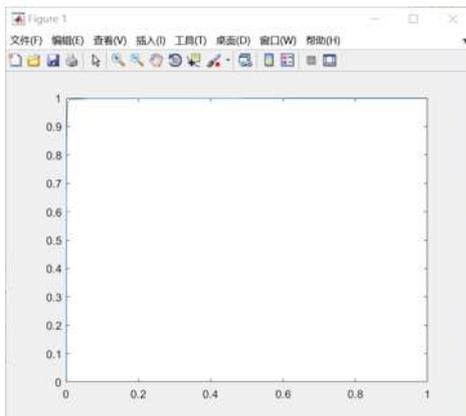


图 6 模型准确率分析图

#### 四、结束语

兵王问题是一个非常经典的问题, 本文通过 SVM 算法, 运用 MATLAB 对其进行了相关解决, 代码性能完成得不错。使用 SVM 进行了实际的二分类问题的求解的具体方法和程序的调试方式, 更显得其在二分类问题中的优越性和实用性。

#### 参考文献

- [1] 刘小洋, 刘万平, 刘超等. 基于 SVM 理论的专业学位硕士研究生教学质量评价研究 [J]. 科技资讯, 2019.
- [2] 刘振岩, 王勇, 陈立平等. 基于 SVM 的农业智能决策 Web 服务的研究与实现 [J]. 计算机技术与发展, 2010.

作者简介: 何杭薇 (2001—), 女, 大学本科, 现就读于中央民族大学信息工程学院 2018 级数据科学与大数据技术专业, 研究方向: 大数据分析。