

基于 SVM 的农业智能决策 Web 服务的研究与实现

刘振岩¹, 王 勇¹, 陈立平², 马俊杰¹, 陈天恩²

(1. 北京理工大学 软件学院, 北京 100081;

2. 国家农业信息化工程技术研究中心, 北京 100097)

摘 要:为了使农业用户能够轻松获得远程的智能决策服务, 研究支持向量机(SVM)在农业智能决策领域的应用, 并使用 LIBSVM 这个通用的支持向量机软件包, 在 J2EE 平台中实现基于 SVM 的智能决策 Web 服务, 用 Web 服务技术构建开放、松耦合的数字农业生产智能决策系统。采用径向基函数作为支持向量机模型的核函数, 径向基函数类型的 SVM, 实际上就是一个分类器, 可以用于农业领域中的分类问题。模型的构造过程就是通过对样本数据的学习, 自动确定支持向量的个数和支持向量的值, 无须人工参与, 从而使模型的构造更容易、更客观、更高效。

关键词:支持向量机; 径向基函数; 智能决策; Web 服务; 数字农业

中图分类号: TP181

文献标识码: A

文章编号: 1673-629X(2010)06-0213-04

Research and Implementation of Intelligence Decision Web Services Based on SVM for Digital Agriculture

LIU Zhen-yan¹, WANG Yong¹, CHEN Li-ping², MA Jun-jie¹, CHEN Tian-en²

(1. School of Software, Beijing Institute of Technology, Beijing 100081, China;

2. National Engineering Research Center for Information Technology in Agriculture, Beijing 100097, China)

Abstract: To easily access to remote intelligent decision services for agricultural users, support vector machine(SVM) is applied to intelligence decision in the agricultural field, and with the general LIBSVM support vector machine package, intelligent decision Web service based on SVM is developed in the J2EE platform. Through Web service the open and loose intelligence decision system for digital agriculture can be built. Radial basis function is the kernel function of support vector machine model, and in fact the SVM based on RBF is a classifier which can be used for classification in the agricultural field. Building this model is to get automatically the amount and value of support vector through learning samples without artificial intervention, so it becomes more easily, more impersonally and more efficiently.

Key words: support vector machine; radial basis function; intelligence decision; Web service; digital agriculture

0 引 言

近年来, 国内外众多科研工作者对农业智能决策技术进行了大量的研究, 研制了以专家系统为代表的各类农业智能决策系统^[1], 并在农业生产过程中得到了广泛的应用。然而, 随着农业信息化步伐的加快, 传统农业智能决策系统在决策方法、实现技术、服务模式等方面已难以满足需求。因此, 迫切需要深入研究新的智能决策方法, 并采用先进的智能决策提供模式, 建立随需应变的农业智能决策系统, 以满足当前和未来现代农业数字化生产决策的需求。

支持向量机(SVM, Support Vector Machine)作为

智能决策领域中一个很重要的研究分支正得到越来越多的重视, 在农业决策领域, 也已经有一些支持向量机的成功应用案例^[2,3]。但是这些应用研究虽然是结合实际需求, 但是其侧重点还是算法的理论研究, 离实际为农业生产提供随需应变的智能决策还有一段距离。

文中将针对数字农业条件下的主要农作过程中的重要决策问题, 使用 LIBSVM 这个通用的支持向量机软件包, 在 J2EE 平台中实现基于 SVM 的智能决策 Web 服务。即将用径向基函数(RBF, Radial Basis Function)类型的 SVM 实现的智能决策系统以 Web 服务的形式提供给用户, 用户不用购买整套的智能决策系统软硬件, 只需将自己的请求和数据发送给相应的远程服务就可以获得决策结果, 轻松享受按需决策的服务, 同时, Web 服务提供模式的采用也使得系统更加开放、更加易于维护和更新, 对于推进农业信息化的进程有着积极的意义。

收稿日期: 2009-09-22; 修回日期: 2009-12-10

基金项目: 国家高技术研究发展计划(863 计划)(2007AA10Z234, 2006AA10A306)

作者简介: 刘振岩(1973-), 女, 讲师, 研究方向为人工智能。

1 SVM & LIBSVM

支持向量机是 Vapnik^[4]在统计学习理论的基础上发展出来的一种全新的机器学习方法^[5]。近来,由于其优越的学习能力,在国内外学术界日益受到广泛重视。SVM 主要思想可以概括为两点:首先它通过使用非线性映射算法将低维输入空间的样本映射到高维属性空间,从而使得在属性空间采用线性算法对样本的非线性特性进行分析成为可能;其次,它通过使用结构化风险最小化的理论在属性空间构建最优化分割超平面,使得学习器得到全局最优化,并在整个样本空间的期望风险以某个概率满足一定上界。另外,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,这些特点都是包括神经网络在内的其它算法所不能及的。因此,无论是在理论基础还是应用前景上,SVM 在很多方面都具有其它学习方法难于比拟的优越性。

支持向量机的非线性变换是通过定义适当的内积核函数实现的,对应于选择核函数的不同,有以下几种支持向量机:

(1)采用多项式核函数,即 $K(x, x_i) = [(x \cdot x_i) + 1]^q$,得到的支持向量机是一个 q 阶多项式分类器。

(2)采用径向基函数,即 $K(x, x_i) = \exp\{-|x - x_i|^2/\sigma\}$,得到的支持向量机是一种径向基函数分类器。

(3)采用 Sigmoid 函数,即 $K(x, x_i) = \tanh(v(x \cdot x_i) + c)$,得到的支持向量机是一个两层感知器,此时隐层神经元的数目及其权值都是由算法自动确定的。

本研究基于台湾大学林智仁博士等开发设计的 LIBSVM 实现 SVM 算法,LIBSVM 的一个操作简单、易于使用、快速有效的通用支持向量机软件包,可以解决分类问题(包括 C-SVC、n-SVC)、回归问题(包括 e-SVR、n-SVR)以及分布估计(one-class-SVM)等问题,提供了线性、多项式、径向基和 S 形函数四种常用的核函数供选择,可以有效地解决多类问题、交叉验证选择参数、对不平衡样本加权、多类问题的概率估计等。

文中以解决农业生产领域中的分类问题为例,核函数选择径向基函数,这里径向基函数的数量及其中心都是由算法自动确定的,分别是支持向量的个数和支持向量的值,而传统径向基函数网络对这些参数的确定则依赖于经验知识^[6]。

基于 LIBSVM 构造 RBF 类型的支持向量机模型的一般步骤如下:

1)按照 LIBSVM 软件包所要求的格式准备数据

集;

- 2)对数据进行必要的预处理;
- 3)选用径向基函数作为核函数;
- 4)利用交叉验证选择最佳参数;
- 5)利用最佳参数对整个训练集进行训练获取支持向量机模型。

首先,按照 LIBSVM 所要求的格式准备数据,训练数据和测试数据的格式如下:

[label] [index1]:[value1] [index2]:[value2]...

其中[label]是训练数据集的目标值,对于分类,它是标识某类的整数(支持多个类);对于回归,是任意实数。[index]是以 1 开始的整数,表示特征的序号;[value]为实数,也就是特征值或自变量。当特征值为 0 时,特征序号与特征值 value 都可以同时省略,即 index 可以是不连续的自然数。

之后,使用 LIBSVM 的 svmtrain 实现对训练数据集的训练,获得 SVM 模型。

用法:svmtrain [options] training_set_file [model_file]

其中,options(操作参数)用于设置 SVM 类型、核函数类型,核函数中的相关参数等。training_set_file 是要进行训练的数据集;model_file 是训练结束后产生的模型文件,该参数如果不设置将采用默认的文件名,也可以设置成自己惯用的文件名。

使用实例:svmtrain train.scale train.model

训练数据文件是 train.scale,将模型保存于文件 train.model,并在 dos 窗口中输出如下结果:

```
optimization finished, # iter = 1756
nu = 0.464223
obj = -551.002342, rho = -0.337784
nSV = 604, nBSV = 557
Total nSV = 604
```

其中,#iter 为迭代次数,nu 设置 n-SVC、one-class-SVM 与 n-SVR 中参数 ν ,默认值 0.5,obj 为 SVM 文件转换为的二次规划求解得到的最小值,rho 为判决函数的常数项 b ,nSV 为支持向量个数,nBSV 为边界上的支持向量个数,Total nSV 为支持向量总个数。

训练后的模型文件 train.model,用记事本等文本浏览器打开可以看到其内容如下(其后“%”后内容为所加注释):

svm_type c_svc % 训练所采用的 svm 类型,此处为 C-SVC

kernel_type rbf % 训练采用的核函数类型,此处为 RBF 核

gamma 0.047619 % 即核函数中的 g
nr_class 2 % 分类时的类别数,此处为两分类问题
total_sv 604 % 总共的支持向量个数
rho -0.337784 % 决策函数中的常数项 b
label 0 1 % 类别标签
nr_sv 314 290 % 各类别标签对应的支持向量数
SV % 以下为支持向量
1 1:-0.963808 2:0.906788 ... 19:-0.197706
20:-0.928853 21:-1
1 1:-0.885128 2:0.768219 ... 19:-0.452573
20:-0.980591 21:-1
... ..
1 1:-0.847359 2:0.485921 ... 19:-0.541457
20:-0.989077 21:-1

% 对于分类问题,上面的支持向量的各列含义与训练数据集相同;对于回归问题,略有不同,与训练数据中的标签 label(即 y 值)所对应的位置在模型文件的支持向量中现在存放的是 Lagrange 系数 a 值。

在获得了 SVM 模型之后,就可以使用 LIBSVM 的 svmpredict 对测试数据进行分类或预测了。本研究将基于 LIBSVM 所构造的 RBF 类型的支持向量机模型封装成 Web 服务,以智能决策 Web 服务的形式提供给用户使用。

2 智能决策 Web 服务的实现

Web 服务是可以使用标准 Internet 协议进行访问的可编程的应用程序逻辑。它是基于 Web 标准来开发分布式应用程序的一种新型的分布式组件技术。使用 Web 服务可以实现动态的、开放的、松耦合的分布式应用程序^[7]。Web 服务的体系结构是面向服务的体系结构的一种典型实现,它有三个组成部分,分别是:Web 服务提供者,Web 服务代理和 Web 服务消费者,如图 1 所示。

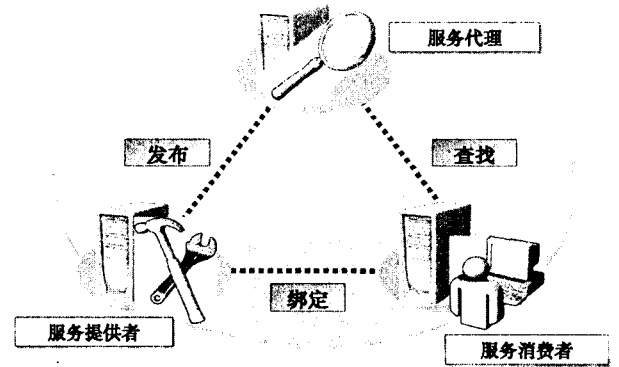


图 1 Web 服务的体系结构

Web 服务提供者向 Web 服务代理发布服务,其发布的信息包括 Web 服务的接口定义、Web 服务提供者所处的位置以及其他相关的支持信息或文档;当 Web 服务请求者需要某个 Web 服务时,首先向 Web 服务代理查找所需的 Web 服务,得到如何调用该 Web 服务的说明信息;然后根据这些信息去调用 Web 服务提供者提供的 Web 服务。当 Web 服务请求者从 Web 服务代理得到调用所需 Web 服务的说明信息之后,或者 Web 服务请求者本来就知道所需 Web 服务的说明信息,通信是在 Web 服务请求者和提供者之间直接进行,而无须经过 Web 服务代理。

服务请求者调用 Web 服务的过程是,首先根据获得的 Web 服务说明(WSDL)信息,生成一个代理类,该代理类可以根据服务说明中的精确定义与 Web 服务进行通信,使用代理可以封装远程调用的复杂细节,其他类可以像调用本地类一样通过代理类使用远程功能;然后创建一个客户端应用程序,该应用程序调用代理类的方法,从而完成 Web 服务的调用^[8]。

本研究基于 J2EE 平台实现相应的智能决策 Web 服务,即使用 AXIS 引擎将基于 SVM 的智能决策系统封装成 Web 服务并发布,发布方法是:将实现了 RBF 类型的 SVM 智能决策方法的类文件拷贝到 Web 服务器的 AXIS_HOME\WEB-INF\classes\目录下,然后使用 WSDD 文件来发布 Web 服务。编写 deploy.wsdd 文件,并将其存放在 AXIS /web-inf 文件夹下,deploy.wsdd 文件代码如下:

```
<deployment name="svmID" xmlns="http://xml.apache.org/axis/wsdd/"
xmlns:java="http://xml.apache.org/axis/wsdd/providers/java">
  <service name="urn:svmID" provider="java:RPC">
    <parameter name="className" value="SVMID"/>
    <parameter name="allowedMethods" value="*/>
  </service>
</deployment>
```

其中,<deployment name="svmID"...> 中的 svmID 是 Web 服务部署的名称,urn 为发布成的服务名称,是服务标识,name="className" value="SVMID" 中的 value 为即将发布成服务的程序所在 java 文件名,name="allowedMethods" value="*" 中的 value 为 java 程序中要作为 Web 服务发布的方法名,而 * 代表将所有的 public 方法都发布成服务。

在 deploy.wsdd 所在目录下使用命令行工具执行命令:

```
java org.apache.axis.client.AdminClient deploy.wsdd,
```

如果显示:

```
Processing file deploy.wsdd
```

```
<Admin>Done processing</Admin>
```

则表示发布服务成功。

调用 Web 服务则采用静态 stubs 方式。stubs 是为了隐藏远端调用而在系统里建立的一个代理,使用 stubs 可以封装远程调用的复杂细节,其他类可以像调用本地类一样通过 stubs 使用远程功能。Axis 提供的 WSDL2Java 工具可以通过 Web Service 的 WSDL 文件,自动生成 Java stubs 类。对已部署的名为 svmID 的 Web 服务,使用 WSDL2Java 工具会在 web-inf 目录下,生成一个 svmID 目录,其中包括 4 个 java 类文件:

```
SVMID_PortType.java
```

```
SVMID Service.java
```

```
SVMID ServiceLocator.java
```

```
SVMID SoapBindingStub.java
```

之后在编写 JSP 客户端时就可以直接调用这些 stubs 类,从而完成相应的 Web 服务调用。

3 结束语

实践证明,支持向量机用于农业生产决策领域是可行的,并且,本研究将径向基函数类型的 SVM 封装成 Web 服务提供给用户,用户只需将自己的请求和数据发送给相应的远程服务就可以获得决策结果,轻松享受按需决策服务。Web 服务的相关标准都是 W3C 的开放协议,具有与平台和操作系统无关的特性,采用 Web 服务技术构建的开放、松耦合的数字农业生产智能决策系统在数据共享、软件重用与降低系统集成成本等方面具有明显优势。

(上接第 212 页)

3 结束语

由于多媒体技术的发展,人们对多媒体的制作要求越来越高,图像序列合成视频作为其中一个方面,其实用化、普及化具有很强的现实意义。本系统基于时间线管理多媒体,特技合成更加方便快捷;随时预览编辑效果的功能,增加了系统的灵活性;保存格式的多样性,增加了编辑的视频的应用范围;同时本系统支持微软 DirectX 的第三方插件使用,使软件更易于扩展。

参考文献:

- [1] 胡春华,曹元大,张磊. 基于 DirectShow 的视频流媒体存储系统的设计与实现[J]. 计算机工程与设计, 2003(11):31-33.

本研究中的智能决策方法——径向基函数类型的 SVM,实际上就是一个分类器,可以用于农业生产中的分类问题。模型的构造过程就是通过样本数据的学习,自动确定支持向量的个数和支持向量的值,无须人工参与,从而使模型的构造更容易、更客观、更高效。它克服了传统 RBF 网络或多层感知器需要根据经验确定参数,局部最优等缺点,是一种较理想的非线性计算工具。但是,通过实验农业领域的不同决策问题发现,训练样本数据的选取对于模型的准确性影响很大,其中训练样本数、数据的维数和数据的相关性等等都是很重要的影响因素,在实际应用中对于数据预处理环节还需要更深入的研究。

参考文献:

- [1] 武向良,高聚林,赵于东,等. 农业专家系统研究进展及发展方向[J]. 农机化研究,2008(1):235-238.
- [2] 陈月华,胡晓光,张长利. 基于机器视觉的小麦害虫分割算法研究[J]. 农业工程学报,2007,23(12):187-191.
- [3] 陈冬冬,杨春. 支持向量回归机在农业供应链预测中的应用[J]. 四川农业大学学报,2008,26(3):290-292.
- [4] Vapnik V N. The Nature of Statistical Learning Theory[M]. New York:Springer,1995.
- [5] 杨斌,路游. 基于统计学习理论的支持向量机的分类方法[J]. 计算机技术与发展,2006,16(11):56-58.
- [6] 韩力群. 人工神经网络教程[M]. 北京:北京邮电大学出版社,2006.
- [7] 王晓东,姜浩. Web Service 同传统分布式技术的比较分析[J]. 计算机技术与发展,2008,18(3):125-127.
- [8] 微软公司. XML Web Service 开发[M]. 北京:高等教育出版社,2004.

- [2] 张勇,罗静. 基于 DirectShow 的多媒体文件音视频的重新压缩[J]. 现代电视技术,2005(5):87-91.
- [3] 陈志峰,田裕鹏,王珊珊. MPEG-2 音视频编辑软件的实现方案[J]. 现代电子技术,2009(10):100-103.
- [4] 胡海峰,陈喜,张文渊,等. DirectShow 非线性音频、视频编辑应用的实现[J]. 微计算机应用,2004,25(1):58-63.
- [5] 韩云,陈祖爵,郑尚志. MPEG-4 编码技术应用及 FPGA 实现[J]. 计算机技术与发展,2007,17(10):219-222.
- [6] 严权锋. 基于 MPEG-4 的综合抗误码方法[J]. 计算机技术与发展,2008,18(4):169-173.
- [7] Wiegand T, Sullivan G J, Bjntegaard G, et al. Overview of the H.264/AVC Video Coding Standard[J]. IEEE Trans, Circuits Syst, Video Technol., 2003,1(7):560-578.
- [8] 陆其明. DirectShow 开发指南[M]. 北京:清华大学出版社,2003.
- [9] Microsoft Corporation. DirectX 9.0 Programmer's Reference [CP/DK]. 2002.