# 基于多级SVM/KNN混合模型的电磁干扰信号分离

**作者：****段柯宇**

**摘要**

在电磁兼容（EMC）领域中，电磁干扰（EMI）信号的分类一直是热点问题，但是由于电磁干扰信号的高维性和强非线性，单一的传统分类聚类算法已经无法胜任电磁干扰信号的分类工作。支持向量机（SVM）作为发展成熟的分类聚类算法，因其优秀的泛化能力而成为最常用的、效果最好的分类器之一。作为改进，本文拟将SVM与K最近邻（KNN）算法结合，并建立多级SVM/KNN混合模型来解决EMI信号分类问题。

**关键词**：电磁干扰，分类，SVM，KNN

**Abstract**

In the field of electromagnetic compatibility, the classification of electromagnetic interference (EMI) signals has always been a hot issue, but due to the high dimensionality and strong nonlinearity of EMI signals, a single traditional classification and clustering algorithm has been unable to compete with EMI signal classification. As a well-developed classification clustering algorithm, support vector machine (SVM) has become one of the most commonly used and best-performing classifiers because of its excellent generalization ability. As an improvement, this paper proposes to combine SVM and K Nearest Neighbors (KNN) in parallel and establish a multi-level SVM/KNN hybrid model to solve the EMI signal classification problem.

**Keywords**: EMI, classification, SVM，KNN,

**1 引言**

随着科技的发展，从智能家居到航空航天，从智能医疗设备到智能城市交通，各个领域电子设备的数量愈来愈多，并向着高频率、宽频带、高集成度、高可靠性、高精度和高灵敏度[1]方向发展，而与此同时，大量电子设备带来的电磁干扰问题也逐渐成为不可忽略的问题之一。例如，在航空航天领域，各个系统工作频带宽错综复杂、接受设备灵敏度高等因素均影响了航空器的设计和制造[2]；在医疗领域，现代医疗设备的电磁信号在一定程度上使敏感设备受到干扰，从而影响医生的判断[3]。据研究[4]，强电磁场会对人体健康带来一定危害，而相对较弱的电磁干扰会对设备或系统造成性能和安全的威胁。

本文尝试解决电磁干扰信号的分类问题。由于电子设备和电子系统的高度复杂性，电磁干扰信号呈现出高维、非线性等特性，常用的分类聚类算法，包括K-means, KNN，support vector machine（SVM）等，均可应用于电磁干扰信号的分类，但是各自有其利弊。

K-means算法简单且应用广泛，但是对初始化的聚类中心点敏感，且不能处理非凸形簇和不同密度、不同尺寸的簇，而高维、非线性的电磁干扰信号势必会严重影响K-means算法的效果。

KNN算法作为懒惰学习的代表算法，学习开销为0，易于执行，原理简单且应用广泛，但由于其计算复杂度较高，不适于高维且包含大量数据的分类问题。因此，单一的KNN算法在电磁干扰信号分类中很难取得理想效果。

SVM作为最为高效的算法之一，拥有优秀的泛化能力，可以很好地完成对数据的结构化描述，对数据规模和数据结构要求较低，可以较好的满足电磁干扰信号高维和非线性的特征。但是，对于复杂的分类问题，SVM存在时间复杂度较高[5]，且在分类超平面附近的分类误差较大[6]的问题。

为了提高SVM在电磁干扰信号分类问题中的分类精度，结合KNN原理简单、应用方便、适用于低数据量、非监督的特点，本文提出使用SVM-KNN并行结合的多级混合模型[7-9]来解决电磁干扰信号分类问题。

**2 算法原理**

**2.1 SVM**

由Vapnik[10]提出的支持向量机算法是一种高效的二值分类算法，其核心思想为将在当前特征空间中不可分的数据模型通过核技法映射到更高维的空间中，依据几何间隔最大化的准则得到最优分类的超平面。

为了解释SVM算法在电磁干扰信号分类问题中的应用，我们需要了解几个概念，首先假设数据集线性可分，并做如下规定：

1. 类别
2. 决策函数
3. 目标函数

**2.2.1 函数间隔与几何间隔**

点到由决定的超平面的函数间隔为：

规定，超平面与整个训练集合的函数间隔为

为了解决成比例改变，函数间隔也会成比例改变的问题，对函数间隔引入限制条件,得到几何间隔：

**2.2.2 最优间隔分类器**

最优间隔分类器，目的是寻求最优分类的超平面，使得超平面对数据集的几何间隔最大化，表示如下：

等价转化为：

通过对原始问题的对偶化，得到满足KKT条件的、与原问题等价的优化问题：

**2.2.3 核技法**

我们提到，SVM的核心思想是将低维空间的数据集投射到高维空间中，从而解决非线性可分的问题，对于最优问题（6）来说，令表示关于的高维向量，用得到高维空间中的优化问题：

但是，不可避免的是，经过映射之后的向量维数会很高，在很大程度上导致向量内积的运算复杂度过高，为了解决这一问题，我们引入核函数：

核函数的使用大大减小了向量内积的计算量，且核函数的选择对于模型的训练影响很大，在本文SVM的模型训练中将使用高斯核函数，它可以将向量映射到无限维，在非线性可分问题中表现优秀。

**2.2.4 软间隔分类器**

我们将所有样本都必须划分正确的分类方法称为硬间隔，相对的，允许有分类错误的样本出现的分类方法就称为软间隔分类器。软间隔分类器的提出从根本上解决了线性不可分的问题。根据软间隔分类器的定义，将分类的原始问题变为：

目标函数分为两部分，w部分与惩罚部分，即允许存在几何间隔比1小的点，但是这些点要收到相应的惩罚，C为惩罚因子，在本模型构建中，C值设为1。

**2.2 KNN**

KNN学习是一种监督学习算法，在本次实验分类问题中的算法原理如下：

给定测试样本，寻找与某点A欧氏距离最近的K个点，称为点A的K个紧邻，对点A的类别判断采用投票法，即选择这K个点中出现最多的类别作为点A的类别。

值得一提的是，KNN并没有显示的学习过程，它的学习过程只是将所有的样本数据储存起来，其训练时间开销几乎为零，因此很适合作为二次修正的算法，应用与SVM中。

在本次实验中，单级SVM/KNN模型均为二分类问题，选取K值为奇数可以保证每次得到符合算法要求的结果。

**2.3 多级KNN/SVM混合模型**

**2.3.1 多值SVM/KNN分类器**

通常的SVM多值分类问题大致分为两种：

1. 将多个分类问题参数的求解合并在一个最优化问题中，通过求解最优化问题而直接实现多值分类，如[11]。
2. 通过构建多个SVM分类器，并将它们结合起来实现多值分类，传统方法有1-v-r与1-v-1等

本文拟采用第二种方法实现多值SVM/KNN分类器，具体流程如下：

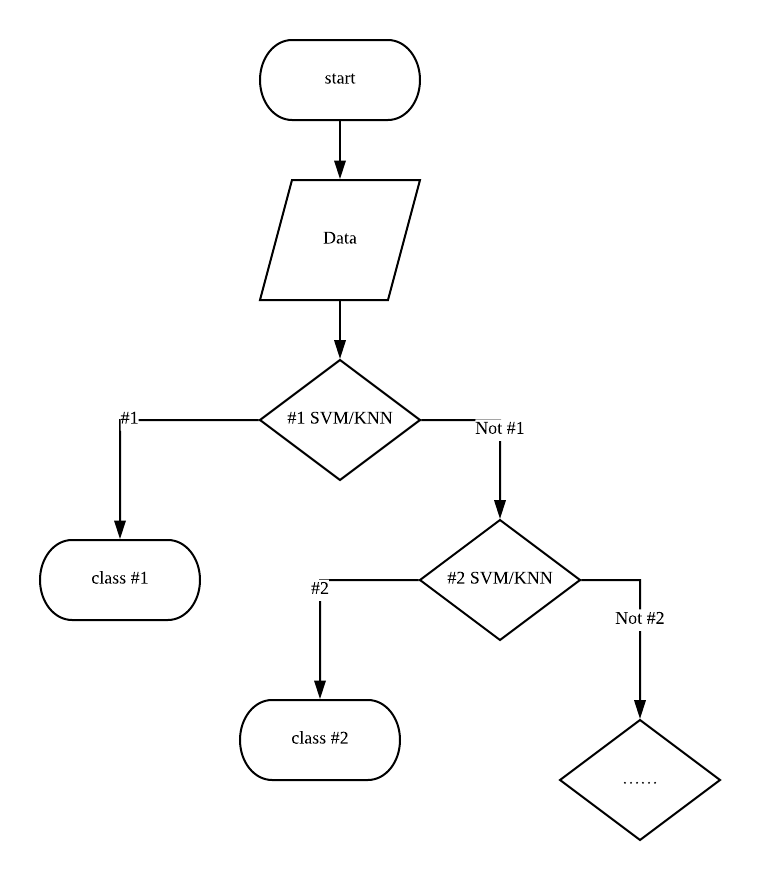


图1 多级SVM/KNN分类器流程图

对于输入的数据集，在第一层分类器中，根据公式（1），将所有非#1类数据的y值设为-1，#1类数据y值设为1，通过SVM混合模型得到关于类1的超平面分类器，将非#1类数据作为下一级分类器的数据输入，继续如上的操作，知道将所有的电磁干扰信号分离开。注意这里的分类器层级数为训练样本中的样本种类数。

**2.3.2 SVM/KNN混合模型**

我们在引言中提到，SVM的分类误差主要集中在分类超平面附近，为了解决这一问题，我们引入KNN来对SVM超平面附近的数据进行检查与二次分类，从而实现提高SVM分类精度的目的。

对于训练样本，使用传统SVM的训练方法得到关于各个分类样本的超平面，并得到各级分类模型。

对于测试样本，首先使用SVM对样本进行二分类，得到测试结果，然后将与超平面的几何距离小于阈值的点进行KNN分类，设置K值为9，设定如果超过2/3的临近点与SVM的分类结果不同，则将该点的类别更改为与其大部分临近点相同的类别。

**4 实验分析**

**4.1 实验数据介绍**

**4.2 实验性能评价指标**

**4.3 实验结果**

**5 结语**

本文通过结合SVM算法和KNN算法建立多级SVM/KNN混合模型，用来解决电磁干扰信号多分类问题。SVM/KNN混合模型在一定程度上解决了SVM在分类超平面误差较大的问题。此算法做到结构化分类电磁干扰信号，和基础神经网络相比，在一定程度上降低了训练时间，且提高了预测精度和泛化性。

在目前电磁兼容的理论研究中，关于将机器学习应用于电磁兼容方向的研究甚少。在人工智能蓬勃发展的今天，笔者希望通过更多的机器学习的应用推动电磁兼容的发展。

**感谢**

**参考文献**

[1]李岐新. 研究电磁兼容技术的意义与展望[J]. 机电一体化, 1997(2):12-14.

[2]苏东林,雷军,王冰切.系统电磁兼容技术综述与展望[J].宇航计测技术,2007(S1):34-38.

[3]房媛媛.电磁干扰对医疗电子设备影响的研究[J].生命科学仪器,2013,11(06):23-25.

[4]张林昌.发展我国的电磁兼容事业[J].电工技术学报,2005(02):23-28.

[5]Vlachosa A.Active learning with support vector ma-chines[D].Scotland:University of Edinburgh，2004:12-14.

[6]Chin K K.Support vector machines applied to spech patern clasification[D].Cambridge:Cambridge Uni-versity，1998.

[7]李胜东,吕学强,施水才,石俊涛.KNN和SVM并行结合的算法[J].华中科技大学学报(自然科学版),2013,41(S2):113-116.

[8]赵涓涓,马瑞良,张小龙.基于决策树和改进SVM混合模型的语音情感识别[J].北京理工大学学报,2017,37(04):386-390+395.

[9]PENG NanBo,ZHANG YanXia,ZHAO YongHeng.A SVM-kNN method for quasar-star classification[J].Science China(Physics,Mechanics & Astronomy),2013,56(06):1227-1234.

[10]Vapnik V. The Nature of Statistical Learning Theory. New York: Springer, 1995

[11] TianLong Gu, Parisa A Bahri. A survey of petri net applications in batch processes[J]. Computers in Industry, 2002;47:99~111