**PSO-LSSVM算法**

**1.支持向量机理论**

早在60年代，Vapnik等人开始研究统计学习理论，并和 Chervonenkis共同提出了VC（Vapnik-Chervonenkis Dimension）维理论，用于衡量函数集的学习性能，并在此基础上建立了结构风险最小化（Structural Risk Minimization），即SRM 准则。结构风险最小化是针对经验风险最小化（Empirical Risk Minimization，ERM）准则提出的。在SRM准则提出之前，人们一直是基于 ERM 准则实现基于数据驱动的建模，研究中发现将训练误差最小化作为优化问题的目标，很多情况下并不能获得满意的效果，甚至在解决某些问题时，会导致模型的预测能力快速下降，发生过学习。实际应用中风险既与经验风险相关，还与置信范围有关，SRM 综合考虑了经验风险和置信范围，因而在用于机器学习时具有更好的效果。

在1995 年，Vapnik等人基于SRM 提出了支持向量机（Support Vector Machine，SVM）算法，最初用于二分类问题，由于其较好地实现了SRM 思想，具有较好的学习能力，被广泛应用于多分类和回归问题。在求解非线性复杂问题时，SVM 通过将样本空间映射至高维特征空间，从而将非线性问题转化为特征空间中的线性问题，并引入核函数的相关定理，而不用求解具体映射问题，从而避免了“维数灾难”问题。分类、回归等问题最终可以描述为一个带有不等式约束条件的二次规划问题，当样本数量较大使得问题的求解复杂，导致 SVM 算法出现训练时间长、计算速度慢、精度低等问题。

针对这一问题，Suykens在 SVM中引入最小二乘线性系统，最早提出了最小二乘支持向量机（Least square support vector machine, LSSVM），算法中损失函数选择最小二乘线性系统，将不等式约束替换为等式约束，从而将求解问题转化为线性方程组的求解，并转变成简单的矩阵逆运算，提高了计算速度。

**2.最小二乘支持向量机**

最小二乘支持向量机(LSSVM)算法的基本原理如下：

给定N个训练样本，其中为n维训练样本输入，为训练样本输出。

LSSVM算法的数学模型可描述为：



式中，表示结构风险，为权值矢量，为惩罚参数，为偏置量，为和空间映射函数，为回归误差。

引入拉格朗日函数求解优化函数，得到：



式中，为拉格朗日乘子。

根据KKT条件，对上式求偏导可得：



可将求解的优化问题转化为求解线性方程组，即：



式中，是单位矩阵。

根据Mercer条件，核函数可表示为：



通过联立，可得LSSVM回归函数为：



式中，表示核函数。

**3.核函数与模型参数选择**

3.1核函数

核函数的引入主要用于解决非线性问题，避免求解非线性映射的具体形式，既提高了非线性处理能力，同时在不影响样本高维空间推广能力前提下，保持高维空间中的内在线性关系。不同的核函数对应于不同的内积，即 LSSVM算法对样本的相似性和相似程度的评估标准不同。常用的核函数主要有以下几种：多项式核函数、径向基（RBF）核函数和神经网络核函数等。

3.2参数选择的常用方法

LSSVM参数选择是其在解决实际问题时的关键技术，主要包括两类参数的选择，一类是惩罚参数，另一类是和参数，如RBF（径向基）核函数中的核参数等。用于对超出设定误差的样本进行控制，越大，经验风险越小，但模型的泛化能力则相对较差；是RBF核函数的宽度，用于控制函数的径向作用范围，当确定时，过小，则算法性能差，若超过一定范围，则算法性能趋于稳定。模型的计算精度及其推广能力主要取决于参数的选择是否恰当，所以在应用LSSVM算法解决实际问题时，首先要解决的便是参数的选择是否恰当，所以在应用LSSVM算法解决实际问题时，首先要解决的便是参数选择的问题。由于RBF核函数的参数只有，复杂度相对较低，研究表明在缺少先验知识时，RBF核函数相对较其他核函数具有更好的效果。本文选取RBF核函数作为LSSVM算法的核函数，需要优化的参数是和，且在参数选择时需同时考虑乘法参数和和参数。

目前对于参数的选取没有统一的方法，比较常用的方法包括基于网络搜索的交叉验证法、实验验证法、经验法等，其中最常用的方式包括基于网格的交叉验证法。根据其是否依赖其他算法完成参数的优化选择，LSSVM参数选择方法主要分为两类：

（1）基于网格搜索的交叉验证法

基于网格搜索的交叉验证法是结合网格搜索选择参数和交叉验证的方法，首先对两个参数确定的搜索平面进行的网格划分，得到对模型参数组合；

然后，对每对参数组合进行 k 重交叉验证，将样本数据分成k 个子集，每次随机选择k-1个子集作为训练样本，对LSSVM 模型进行训练，将剩余的一个子集作为验证样本并计算误差。不断重复上述过程，最终得到最小的 k 重交叉验证误差，若满足误差计算要求，则将其作为模型参数，反之则对当前参数组合进行更细的网格平面划分，继续计算误差，直到满足结束条件。实践表明，这种参数选择方法非常耗时，需要对每一对参数进行训练和验证，计算量较大。

（2）优化算法

近年来，随着学者对优化算法的不断研究，尤其是现代优化算法的发展，如遗传算法、模拟退火、蚁群算法、粒子群算法等，越来越多的研究者将优化算法应用于 LSSVM 等参数的选择中，并在实际应用中取得了比较好的结果。

粒子群算法作为一种并行优化算法，相对于遗传算法等其它现代优化算法，其原理更简单、易实现、效率高、搜索速度快等特点，被广泛应用于函数优化、神经网络等模型参数的优化、数据挖掘等领域。因此，采用粒子群算法对LSSVM 的参数进行选择和优化。

**4.粒子群算法的改进**

粒子群算法(Particle Swarm Optimization，PSO)是Kennedy和Eberhart在1995年提出的一种全局优化进化算法，通过模拟鸟群觅食行为而实现基于群体协作的随机搜索。其基本思想是利用生物群体模型和信息共享机制，实现群体的进化过程。粒子群的每个个体通过粒子速度和粒子位置来描述个体特征，速度决定了粒子在解空间中的变化方向和大小，粒子位置向量是优化问题的一个可能解，可将其对应的目标函数值作为该粒子的适应度值，从而评价粒子的优劣，实现群体的进化过程

LSSVM参数和对预测精度有较大的影响，通常采用参数空间穷尽搜索法对LSSVM的参数进行优化，其缺点是很难确定一个合理的参数范围，因此，采用 PSO 优化 LSSVM 的参数，其算法步骤如下

1）初始化PSO 的各种参数：群体规模、学习因子、迭代的最大次数、粒子的初始位置和速度等。

2）分别用每个粒子向量所对应的LSSVM对学习样本进行预测，得到各粒子当前位置值的预测误差，并将其作为各粒子的适应度值，再将各粒子的当前适应度值与该粒子自身的最优适应度值进行比较，如果更优，则将粒子当前的位置作为该粒子的最优位置。

3）将各粒子的自身最优位置适应度值与群体最优位置的适应度值比较，如果更优，则将该粒子的最优位置作为群体的最优位置。

4）计算惯性权值，更新粒子的速度和位置。

5）检查是否满足寻优结束条件（达到预先设定的最大迭代次数或预设精度），若满足则结束寻优，求出最优解；否则转至步骤2），继续新一轮搜索。