从自我研究方向出发-船舶的异常行为，对于异常的定义、探测方法、应用进行简要梳理。

**#摘要**

异常探测是数据挖掘领域的一个重要研究内容，旨在从海量数据中挖掘不符合普适性规律、表现出“与众不同”特性的数据或模式，其在金融欺诈、公共卫生、极端气候事件发现、交通拥堵判别、环境污染监测等领域具有重 要应用价值。按照传统的分类方式，异常探测方法可以分为三类：基于传统统计学的探测方法；基于集成检测方法以及现比较应用多的基于深度学习的检测方法。 为了更好地满足应用需求，发展性能更高、适应性更强的异常探测方法，从数据维度角度对于异常进行分类可以划分为典型的传统／空间／时空异常探测方法。

**#一、异常的定义**

2001年，Portnoy：“异常检测方法对正常数据建立模型，然后尝试检测观测数据中与正常模型的偏差” 2008年，Roy:"与正常相悖的即为异常" 2009年，Chandola：“异常定义为不符合预期正常行为的模式”

总结： 现有研究针对不同领域对于异常的定义各有细化，比如在海上交通领域对于传播行为的异常，定义为“船舶偏离其固定航线或违背历史行进规律的行为”。 但是现有研究对于异常探测比较偏向黑盒子，大部分还是认为不符合模型训练的目标具有异常行为。

**#二、探测方法及其优缺点**

接下来以船舶异常行为探测为方向，对其进行方法梳理

2.1 基于传统统计学的异常探测方法

基于统计学的异常检测中，轨迹正常模型的形式是一种轨迹点的概率模型。 其检测原理是判断输入数据是否符合已有的概率模型，认为正常的船舶轨迹数据在模型中则输出较高概率数值，如果该输入数据在该模型的概率值较低，则将其识别为异常。 其中比较典型的算法有高斯混合模型(Gaussian Mixture Model, GMM)、核密度估计(Kernel Density Estimation, KDE)以及贝叶斯网络(Bayesian Network ,BN)。 采用统计学的方法检测船舶的异常行为，其特征模型较好地对于船舶特征包括目标瞬时运动状态特征进行融合探测，但有效依赖于对给定的船舶轨迹数据所作的统计模型假设是否成立。 其优点是统计学方法具有坚实的理论基础，当存在充分的船舶轨迹数据和所用的检测类型知识时，检测结果非常有效。 缺点是计算开销依赖于所建立的模型，在使用简单的参数模型如高斯模型时，拟合参数所需要的时间是线性的，但是，使用复杂的模型时，确定最佳参数值通常需要多次迭代，比较耗时。