

# Some Materials on Streaming Prediction in Non-Stationary Environments

Cheng Hang

Machine Learning

Streaming Prediction

Online Learning

## 基本出发点

在很多现实场景中，数据往往是以数据流的形式给出的。同时，很多现实场景中，数据的生成过程中往往有一些非稳态的因素，如环境发生变化等等。

在非稳态环境下的流式数据学习中，我们需要考虑以下的问题：

- 数据分布发生变化；
- 特征空间发生变化；
- 类别空间发生变化；
- 评价指标发生变化
- .....

我们现在重点关心第一个：数据分布发生变化，常称为concept drift或者distribution change

## 参考文献及软件包

- (2014-CSUR) A Survey on Concept Drift Adaptation  
这篇文章是一篇非常好的关于survey文章
- (2017-CSUR) A Survey on Ensemble Learning for Data Stream Classification  
由于ensemble方法非常灵活，可以动态调整分类器个数、形态，所以常用于解决数据流中分布发生变化问题。这篇文章重点关注这类方法。
- (2017-MLJ) Adaptive random forests for evolving data stream classification  
一篇最近的利用随机森林来处理数据流分布变化的文章
- (2011-JMLR) MOA - Massive Online Analysis  
MOA是这个领域很经典的一个软件包，由weka团队开发的，是基于weka实现的。你需要研究明白这个软件的使用。

参考文献下载链接：[\[zip\]](#)