**时间序列相似性度量指标**

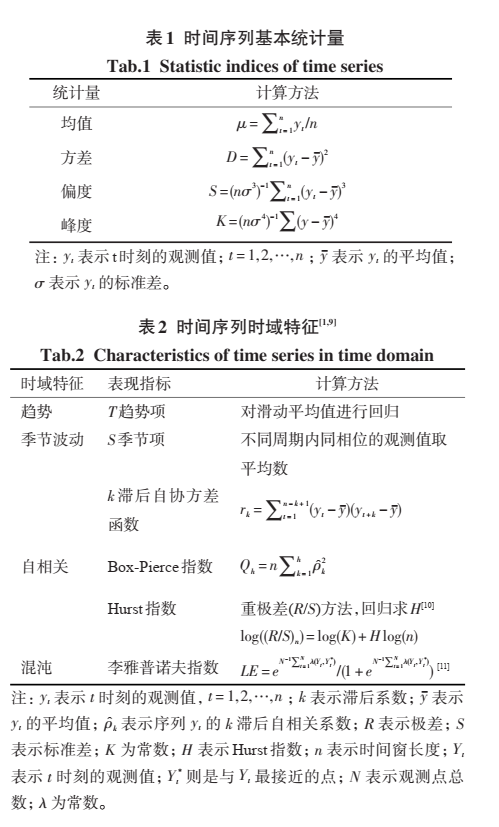
**按照时间轴比较可以分为两类**：锁步度量和弹性度量。锁步度量是时间序列进行 “一对一” 的比较；弹性度量允许时间序列进行 “一对多” 的比较。很明显，弹性度量在判别时间轴有缩放时更有效。

传统的DTW距离是基于欧氏距离的弹性度量，是基于动态规划从两个可以不等长的时间序列的距离矩阵出发（可以是任意的距离比如欧式比如闵可夫斯基），求两者之间最短距离，时间复杂度是O（n2）。有很成熟的包可以调用。优点：支持非等长序列、支持有断点序列，缺点：不符合三角不等式、对噪音敏感。

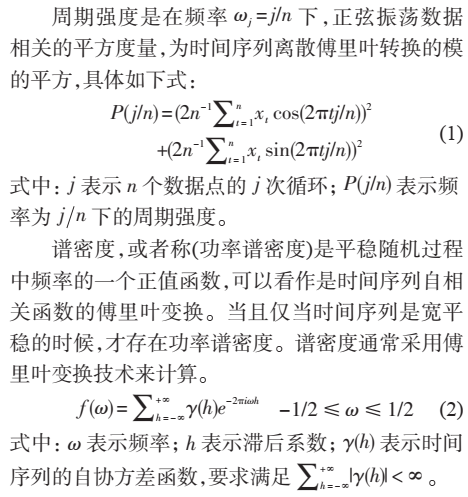
**按照比较方法可以分为**：基于欧式距离比较；基于编辑距离比较；基于特征比较；基于模型比较。其中前两种类似，后两种类似（降维思想，常用于聚类，一并介绍了）。

编辑距离比如LCSS距离，当两个时间序列在大部分时间段都有相似形态, 仅在小范围内有扭曲突变或断点时（噪声）, 欧氏距离将不再适用, 这时应采用 LCSS 距离度量，不过处理振幅平移以及时间轴伸缩等形变问题时的效果不理想。如果一个序列 S 分别是两个或多个已知序列中符合条件的最长子序列, 则 S 就是已知序列的最长公共子序列 (LCSS)，其长度就是两序列之间的距离。LCSS 用两字符串最大公共字串的长度与最长字符串长度的比值进行相似性度量。

时间序列特征可以分为形态特征，结构特征和模型特征。其中形态特征指可以直观看出的特征，如全局时间序列的上升下降，如局部的极值点突变点；结构特征是需要对原始数据进行处理才能得出的非直观特征，具体又分为：基本统计特征如均值、方差、峰度、偏度，时域特征如趋势、季节性、自相关，如右下图所示。

频域特征如周期强度、谱密度（这俩也不是很懂，涉及傅里叶变换）具体计算公式如下图所示。

模型特征一般表现为不同的参数特征，包括：高斯过程模型、ARMA(自回归滑动平均模型)以及 ARIMA 模型(差分自回归移动平均模型)、马尔科夫链模型、隐马尔科夫模型等

寻找超前滞后关系

**寻找超前滞后关系**

赋予两个时间序列不同的滞后阶数，然后通过时间序列相似度度量给出是否相似，从滞后阶数的大小，推断是A影响B还是B影响。

更进一步，其实可以通过VAR（向量自回归模型）,ECM（误差校正模型）,VECM（向量误差校正模型）来确定两个时间序列中任意一个与其各自的滞后项之间的回归方程，不知道对于超前滞后关系有没有用。

**多维时间序列处理**

如果每个时间点，是多维数据，可以根据每个分量计算相似度指标，然后按照分量的特征重要性加权或者等权处理。

在DTW方法中的基础距离矩阵无论采取的是欧式距离或者编辑距离，都是有多维形式的，也可以采用余弦相似度作为两个横截面之间的相似度度量。

补：

时间序列的形态特征主要指时间序列的形状变化特征，包括全局特征和局部特征。全局特征描述了时间序列的起伏变化，如上升、下降等；局部特征则表现为时间序列局部时间点上的异常观测值，如不连续点，极值点、突变点、转折点等。在时间序列最开始的研究中，人们通常是先将时间序列画出来，然后直观地通过观察来研究时间序列的起伏变化或异常点。这类反映时间序列整体变化或局部异常，可以直观看出的特征，称为时间序列的形态特征。基于形态特征的时间序列聚类，可以发现具有相同形状的时间序列簇，寻求时间序列的起伏变化规律。时间序列形态特征可以在一定程度上表现时间序列的特性，通常适用于描述短时间序列。当序列较长时，起伏变化往往比较复杂，难以用简单的“上升，下降”描述。虽然可以采用分段描述的方法，但这割裂了时间序列的整体性，不能很好地反映时间序列的全局特征；异常点特征主要描述时间序列上的某些特殊点的特征，同样难以反映其全局特征。