

时间序列分析作业1

502023370016 黄彦骁

October 19, 2023

1 DataSet

我选取的数据集为illness/national_illness.csv国家疾病数据集，总共包含七个特征，针对该数据集采样200个时间戳绘图，其中前五个特征如图1所示,可以看出这五个特征都具有相似的季节性。

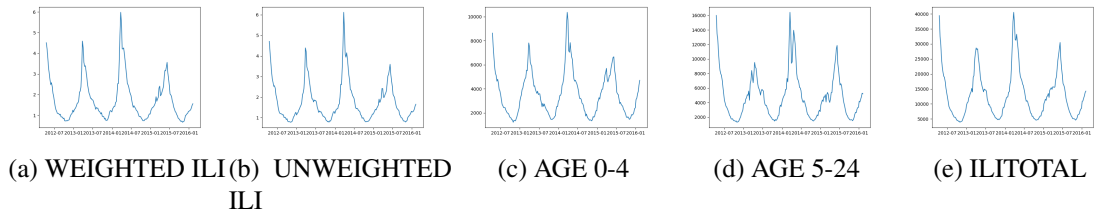


Figure 1: 疾病数据集前五个特征

而其余的两种特征如图2所示，季节性的周期与前五个特征不同，但同样具有比较明显的季节性。

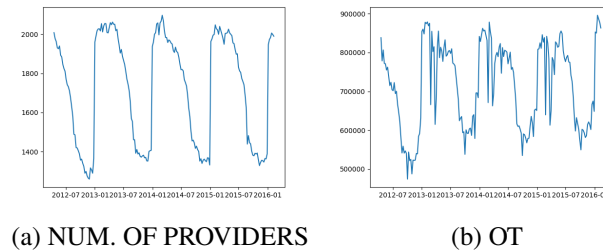


Figure 2: 疾病数据集后两个特征

2 Transform

Transformation的数学公式如下所示:

$$Normalization : y'_t = \frac{y_t - y_{min}}{y_{max} - y_{min}}$$

$$Standardlization : y'_t = \frac{y_t - \mu}{\sigma}$$

$$MeanNormalization : y'_t = \frac{y_t - \mu}{y_{max} - y_{min}}$$

$$Box - Cox : y_t^\lambda = \begin{cases} \frac{(sign(y_t)|y_t|^\lambda - 1)}{\lambda} & \lambda \neq 0 \\ \log(y_t) & \lambda = 0 \end{cases}$$

其中 λ 默认值设为0.7。

3 Metrics

Metrics的数学公式如下所示:

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i|$$

$$MAPE = \frac{100}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{|y_i|}$$

$$MASE = \frac{1}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{\frac{1}{T-m} \sum_{j=m+1}^T |y_j - y_{j-m}|}$$

$$sMAPE = \frac{200}{T} \sum_{i=1}^T \frac{|y_i - \hat{y}_i|}{|y_i| + |\hat{y}_i|}$$

4 Model

在线性回归模型中，将传入的训练数据进行相似的滑动窗口处理，采样到长度为seq_len+pred_len的数据，切分该数据作为特征值 \mathbf{X} 和目标值 \mathbf{Y} 进行训练，通过公式:

$$\hat{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

计算得到线性回归的权重值，进而进行预测。

指数平滑模型采用Holt两参数指数平滑，模型具有两个参数 a, b ，利用传入的test_X进行参数更新，参数初始化为：

$$\hat{a}_0 = X_0, \hat{b}_0 = X_1 - X_0$$

参数更新公式为：

$$\hat{a}_t = \alpha X_t + (1 - \alpha)(\hat{a}_{t-1} + \hat{b}_t) \quad \hat{b}_t = \beta(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta)\hat{b}_{t-1}$$

模型在不同数据规范方式下的表现如表1所示：

Dataset	Model	Transform	MSE	MAE	MAPE	SMAPE	MASE
illness.csv	LR	None	41873290900	157843	14.20	14.55	0.67
		Normalization	36804787168	141201	12.94	12.91	0.60
		Standardlization	36982933248	142314	13.04	13.03	0.60
		MeanNormalization	36887273471	142008	13.02	13.00	0.60
		Box-Cox	42293278157	158684	14.29	14.67	0.67
	ES	None	289615114805	389240	34.80	35.81	1.67
		Normalization	289615114805	389240	34.80	35.81	1.67
		Standardlization	289615114805	389240	34.80	35.81	1.67
		MeanNormalization	289615114805	389240	34.80	35.81	1.67
		Box-Cox	266024252291	387878	35.00	35.36	1.73

Table 1: 模型表现

5 TsfKNN

实现了manhattan距离和chebyshev距离，TsfKNN在不同距离下的表现如下表2所示：

Dataset	Model	Distance	MSE	MAE	MAPE	SMAPE	MASE
illness.csv	TsfKNN	euclidean	231178157011	426077	34.59	43.44	1.96
		manhattan	234921841396	426752	34.63	43.52	1.97
		chebyshev	250262723868	452147	37.13	47.06	2.08
ETTh1.csv	TsfKNN	euclidean	18.51	3.41	13673	55.41	3.12
		manhattan	21.72	3.62	15843	56.02	3.32
		chebyshev	16.93	3.28	13481	54.02	2.96

Table 2: TsfKNN表现