**数据预处理**

将数据集划分为训练集和测试集，

为了得到准确的预测结果和清晰的展示，商品销售数据按照时间序列，将商店和商品类别分类整理。采用特殊值填充法对测试集进行缺失值处理，将所有的销售额用“0”值填充。为了保持训练集和测试集的同步性，让训练集和测试集中的列名和顺序保持一致。

**特征工程**

将日期按照one-hot进行日、周、月编码处理，以便确保我们的网络模型可以跟好的额处理时间序列数据，对季节性的的数据更加敏感。预测方法选择在线单步预测，因此数据的输入采用追加方式将数据在训练的过程中不断地输入到模型中来进行模型调节。由于神经网络的数据输入会造成一定的抖动，所以采用滑动窗口算法来消除这种抖动，窗口大小=∆N，步长=n，时间间隔=∆T。

**数据归一化**

使用min-max标准化(Min-max normalization)方法对数据集中的销售额进项归一化处理。

最后将数据集中的数据日期、商店、商品和销售额转换为矩阵向量，输入到模型中。

|  |
| --- |
| 算法：数据预处理 |
| Input：训练集和测试集 |
| Begin：  For: sales  For: 商店和商品归类  For: 商店归类  For: 商品归类  按照日、周、月进行One-hot编码  将日、周、月的销售额连接为矩阵  单步预测：  For: sales:  将每天的数据追加到每一行  数据归一化：  min-max标准化(Min-max normalization)方法将销售额归一化  End |

|  |
| --- |
| 算法：滑动窗口算法 |
| Input： |
| Begin：  窗口大小=∆N=一周，步长=n=1天，时间间隔=∆T=1天，窗口中的第一天=θ   1. 选取一周（∆N）数据输入到模型中 2. 训练模型 3. When ∆T=1天   ∆N=∆N-θ+n（窗口滑动）  继续第一步  End |



