

商品期货涨跌幅预测问题

1 摘要

本文针对商品期货 30 分钟涨跌幅预测问题，提出了一种基于 LSTM 的时序预测模型。通过对 1 分钟级行情数据进行滑动窗口特征工程，提取了包含价格动量、波动率、成交量异动等 36 维特征。采用分层时间序列分割方法构建训练集与测试集，使用贝叶斯优化进行超参数调优。实验表明，在螺纹钢主力合约数据上，模型取得 $MAE=0.45\%$ 、 $R^2=0.72$ 的预测效果。进一步分析揭示了市场微观结构特征对短期价格预测的有效性，同时指出高频数据噪声和突发事件响应的局限性。本文为程序化交易策略提供了可靠的预测基准。

2 问题重述

2.1 问题背景

商品期货（如螺纹钢、铁矿石、焦炭、焦煤等）是金融市场中的重要交易品种，其价格波动受到多种因素的影响，包括供需关系、宏观经济政策、国际市场变化等。若能利用历史数据预测商品期货未来的涨跌幅，则可帮助投资者更好地进行交易决策。

2.2 问题描述

现有数据集为 1 分钟级数据，包括时间戳、开盘价、最高价、最低价、收盘价、成交量、持仓量等。请基于该数据集建立数学模型，预测商品期货未来 30 分钟的涨跌幅。涨跌幅定义为 $\text{涨跌幅} = \frac{P_{t+30} - P_t}{P_t} * 100\%$ 其中 P_t 是当前时刻的价格， P_{t+30} 是 30 分钟后的价格。要求从 1 分钟级数据中提取出可能影响 30 分钟涨跌幅的特征，选择合适的机器学习模型对未来 30 分钟的涨跌幅进行预测。解释模型的选择理由，并使用适当的评价指标评估模型的性能，讨论模型的局限性及可能的改进方向。

3 数据预处理与特征提取

去掉无效数据，归一化，缺省处理，异常值处理，提取特征。

3.1 去掉无效数据

去掉开盘价、最高价、最低价的数据。

3.2 归一化

....

3.3 提取特征

提取可能影响 30 分钟涨跌幅的特征.....。

4 模型选择

选择居于 RNN 的时间序列预测模型 LSTM

4.1 选择模型的理由

。。。。

4.2 模型的具体实现

。。。。

5 模型训练与验证

训练集: 测试集: 交叉验证集 =4:3:3

6 模型预测效果分析与改进方向

7 源码与文档