# MSBD Project 2: M5 Forecasting Accuracy and Uncertainty

Shihan BU, Yuxin YANG and Yilin Li {sbuaa, yyanget, ylime}@connect.ust.hk

## 1. Introduction

In this project, we are asked to predict item sales for two 28-day time periods from both accuracy and uncertainty standpoints. The main difference between these 2 standpoints lies in the loss formulation -- for accuracy it's WRMSSE and for uncertainty it's WSPL. We experimented with different versions of RNN, seq2seq and transformer models obtained a final WRMSSE of 0.67845 and WSPL of 0.19283.

## 2. Data

The dataset consist of unit daily sales record of 3049 items in 10 stores of 3 states in US in a 1941-day period, the price data of the products sold per store and date, and auxiliary calendar data which contains information on events and holidays. The dataset is hierarchical which means that it can be aggregated on different levels. The details of aggregation levels can be viewed from below.

| Level | Aggregation Level | Number of Series |
|---|---|---|
| 1 | Unit sale of all products, aggregated for all stores/states | 1 |
| 2 | Unit sale of all products, aggregated for each State | 3 |
| 3 | Unit sale of all products, aggregated for each store | 10 |
| 4 | Unit sale of all products, aggregated for each category | 3 |
| 5 | Unit sale of all products, aggregated for each department | 7 |
| 6 | Unit sale of all products, aggregated for each State and category | 9 |
| 7 | Unit sale of all products, aggregated for each State and department | 21 |
| 8 | Unit sale of all products, aggregated for each store and category | 30 |
| 9 | Unit sale of all products, aggregated for each store and department | 70 |
| 10 | Unit sale of product x, aggregated for all stores/states | 3049 |
| 11 | Unit sale of product x, aggregated for each | 9147 |
| 12 | Unit sale of product x, aggregated for each | 30490 |
| | Total | 42840 |

## 3. Evaluation Metric

The point forecast submission are being evaluated using the Root Mean Squared Scaled Error. And the uncertainty distributions are being evaluated using the Weighted Scaled Pinball Loss.

$$\boldsymbol{RMSSE} = \sqrt{\frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t-\widehat{Y}_t)^2}{\frac{1}{n-1}\sum_{t=2}^{n}(Y_t-Y_{t-1})^2}}, \quad \boldsymbol{WRMSSE} = \sum_{i=1}^{42,840} w_i * RMSSE$$

$$\mathbf{SPL(u)} = \frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t - Q_t(u))u\mathbf{1}\{Q_t(u) \le Y_t\} + (Q_t(u) - Y_t)(1-u)\mathbf{1}\{Q_t(u) > Y_t\}}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|}$$

$$\boldsymbol{WSPL} = \sum_{i=1}^{42,840} w_i * \sum_{j=1}^{9} SPL(u_j)$$

## 4. EDA

**Missing Value**

No missing values in our sales training data. But a lot of zero values among all the time series.

**Subjective/Objective Analysis**

We draw scatterplots and bar plots to see the relation between the features with target. We also do the objective analysis to draw the heatmap of and choose top5.

**Time Series Analysis**

We draw some plots to analysis the time series of both individual and aggregated time series about sales. We also dig from the time series about shift and lag, change of sell price, different Rolling-Window sizes.

Findings from EDA which guide us to do the feature engineering and setting model parameters.

## 5. Models

3 types of models –LSTM, Seq2Seq and Transformer were used to fit our data. In order to find the model with the best fit, we experimented with different versions of the seq2seq including a dilated seq2seq model as well as a seq2seq with attention on hidden layer model.

## 6. Experiments, Results and Analysis

For accuracy task, we applied 3 models: LSTM, seq2seq with attention on hidden layers and dilated seq2seq model. For uncertainty task, we adopted 3 models: LSTM, transformer, and seq2seq with attention. We set sliding window size equals to 28*13 in training process for prediction preparation. We also use 3-fold validation and early stopping for preventing model overfitting.

We compared multiple models and the results are shown in the table. The lower the leaderboard score, the better the performance. The best performance is achieved by dilated seq2seq model for accuracy task. And seq2seq with attention model achieves the best score for uncertainty task. In the accuracy task, we performed 2 epochs for both seq2seq with attention and dilated seq2seq model. In the Uncertainty task, due to resource and time limitation, we performed 1 epoch for transformer model and 2 epochs for seq2seq with attention model.

| | Model Architecture | Private Leaderboard Score | Private Leaderboard Rank |
|---|---|---|---|
| Accuracy | LSTM | 0.77957 | 1729/5558 |
| | seq2seq_w_attn_on_hid | 0.69061 | 482/5558 |
| | **dilated_seq2seq** | **0.67845** | **467/5558(top 8%)** |
| Uncertainty | LSTM | 0.60455 | 776/976 |
| | Transformer | 0.24819 | 524/976 |
| | **seq2seq_w_attn_on_hid** | **0.19283** | **131/976(top 13%)** |

And seq2seq with attention model achieves the best score for uncertainty task. Performance is 0.67845 and 0.19283 respectively for these two tasks.

Training performance and efficiency trade-off can be considered. However, considering its highest accuracy and the depth that the model gets into the dataset, we still recommend better performed models as the optimal model for daily sales forecasting or even time-series forecasting.

## 6. Conclusion and Future Work

For this Walmart sales prediction, we achieve 0.678 (top 8%) weighted RMSSE with dilated seq2seq model in M5 forecasting accuracy task and achieve 0.193 (top 13%) weighted SPL in uncertainty task for 28-day predictions. In the future work, more epochs can be performed to see whether other models can achieve better performance. Having more features, more precise models can be built and as accompaniments to ensemble for better aggregated results. Hyperparameter tuning can also be performed further.

## 7. References

Makridakis, Spyros, Evangelos Spiliotis, and Vassilios Assimakopoulos. "M5 accuracy competition: Results, findings, and conclusions." *International journal of forecasting (2022).*

Chang, S., Zhang, Y., Han, W., Yu, M., Guo, X., Tan, W., Cui, X., Witbrock, M., Hasegawa-Johnson, M.A., & Huang, T.S. (2017). *Dilated Recurrent Neural Networks. NIPS.*

## 8. Contribution

**Accuracy:** Yuxin YANG, Yilin LI, **Uncertainty:** Shihan BU

All group members have been fully participated in coding and poster writing.