M5 Forecasting: Accuracy and Uncertainty

Shihan BU |sbuaa@connect.ust.hk

Yuxin YANG |yyanget@connect.ust.hk

Yilin Li |ylime@connect.ust.hk

Introduction

- Given: sales data provided by Walmart
- Goal: predictions 28 days into the future

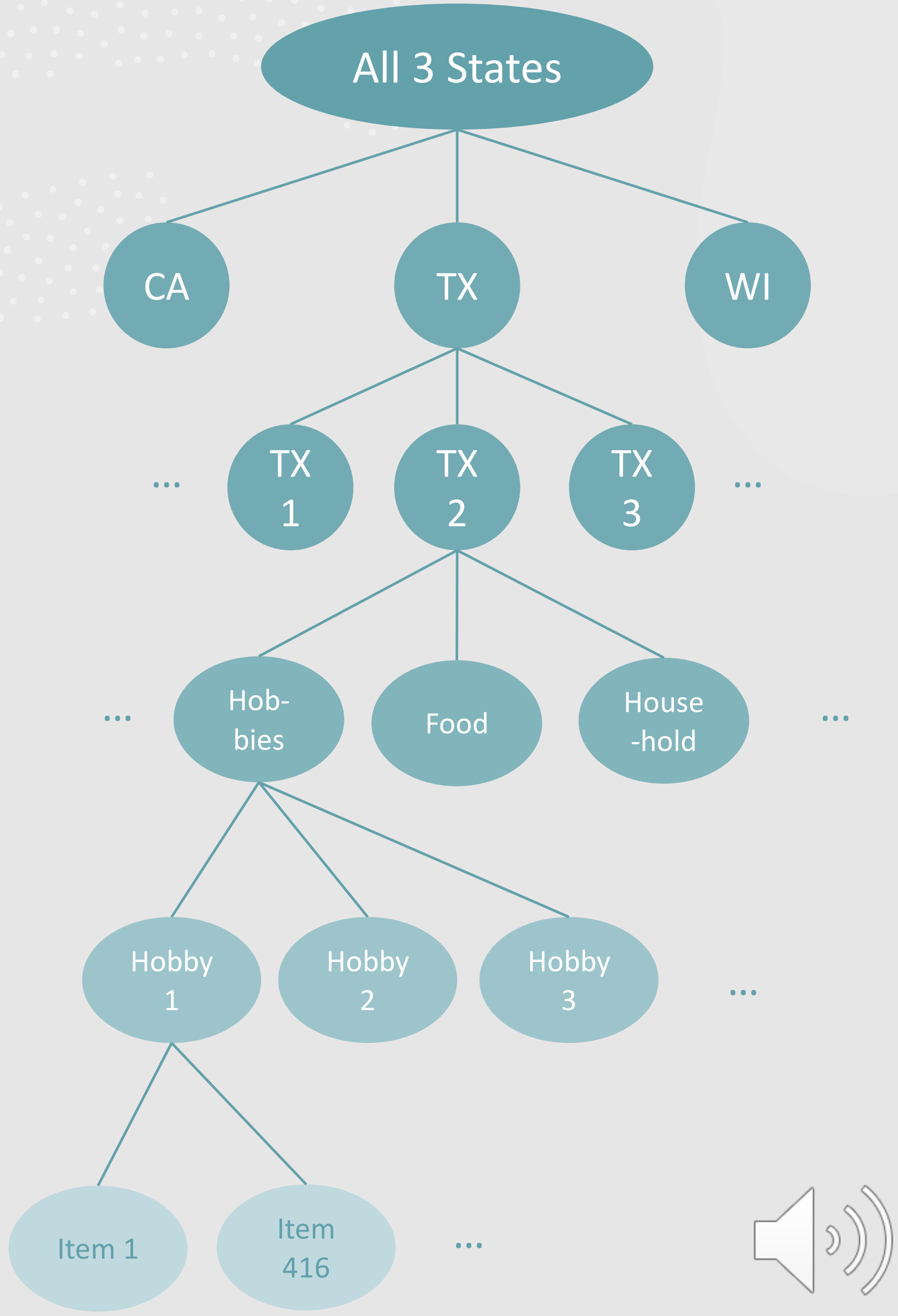
- Accuracy
  - Point forecast (PFs)
- Uncertainty
  - Prediction Intervals (PIs)
    - Median
    - 50%
    - 67%
    - 95%
    - 99%

Data

- **sales_train_evaluation.csv**
  Includes sales [d_1 - d_1941] (labels used for the Public leaderboard)
- **sales_train_validation.csv**
  Contains the historical daily unit sales data per product and store [d_1 - d_1913]
- **sell_prices.csv**
  Contains information about the price of the products sold per store and date.
- **calendar.csv**
  Contains information about the dates on which the products are sold.

Data

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



Evaluation Metrics

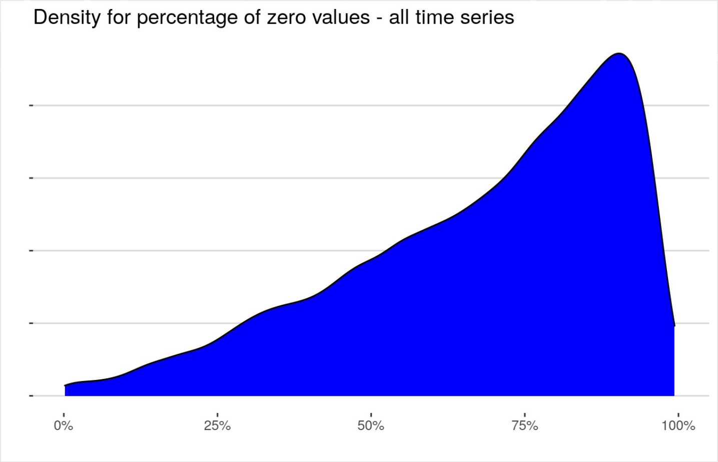
- Accuracy
  - $\boldsymbol{RMSSE} = \sqrt{\frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t - \widehat{Y_t})^2}{\frac{1}{n-1}\sum_{t=2}^{n}(Y_t - Y_{t-1})^2}}$
  - $\boldsymbol{WRMSSE} = \sum_{i=1}^{42,840} w_i * RMSSE$
- Uncertainty
  - $\mathbf{SPL(u)} =$ $\frac{1}{h}\frac{\sum_{t=n+1}^{n+h}(Y_t - Q_t(u))u\mathbf{1}\{Q_t(u) \leq Y_t\} + (Q_t(u) - Y_t)(1-u)\mathbf{1}\{Q_t(u) > Y_t\}}{\frac{1}{n-1}\sum_{t=2}^{n}|Y_t - Y_{t-1}|}$
  - $\boldsymbol{WSPL} = \sum_{i=1}^{42,840} w_i * \frac{1}{9}\sum_{j=1}^{9} SPL(u_j)$

Exploratory Data Analysis

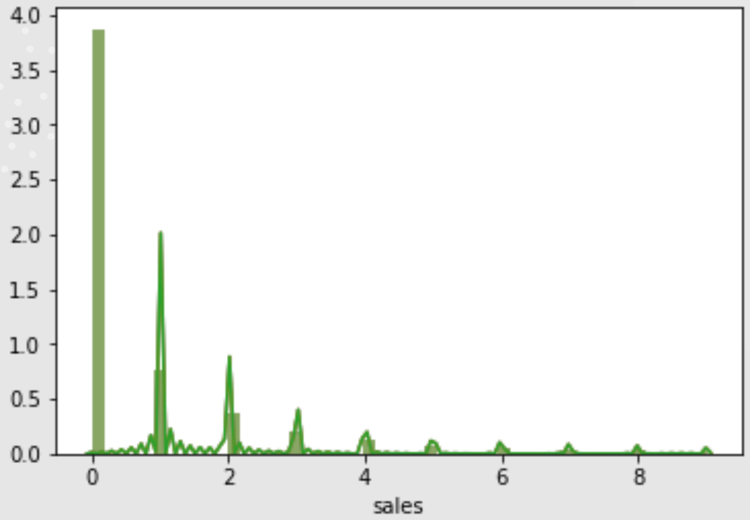
- Data structure arrangement
- Missing Value
- Subjective analysis
- From the target – sales analysis
- Relation between top 5 and target
  - Numerical
  - Categorical
- Dig from time-series
  - Shift and lag
  - Change of sell-price
  - Different Rolling-Window Size

Exploratory Data Analysis

Missing Value



From the Target – Sales Analysis

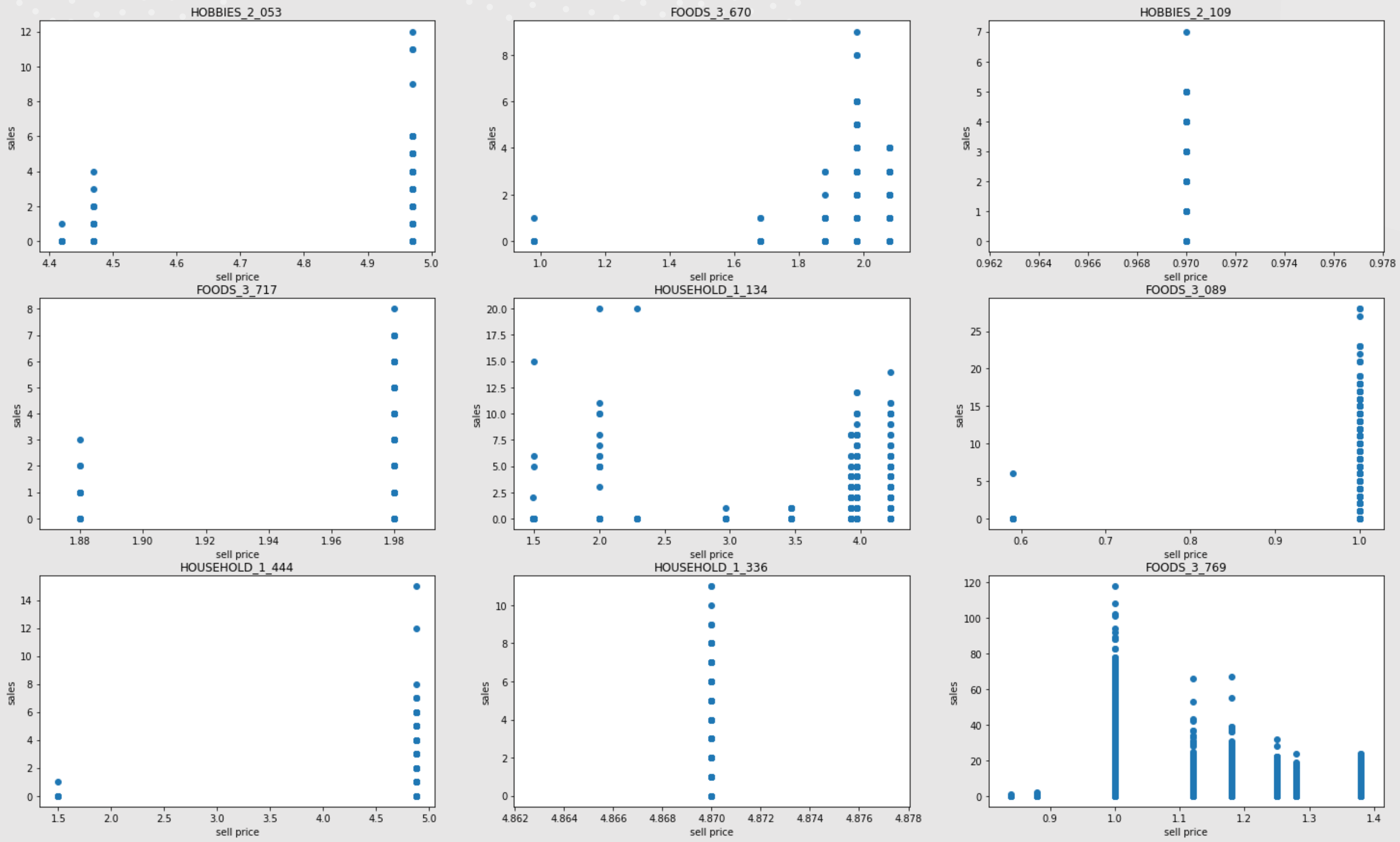


Subjective analysis

name	desc	type	segment	expectation
item_id	id for item	categorical	product	high
weekday	day of week	categorical	context	low
year	year	categorical	context	low
dept_id	id for item dept	categorical	product	middle
cat_id	id for item category	categorical	product	middle
sell_price	sell price for item	numerical	product	middle
store_id	id for store	categorical	store	middle
state_id	id for state	categorical	store	middle
wday	day of week in number	categorical	context	middle
month	month	categorical	context	middle

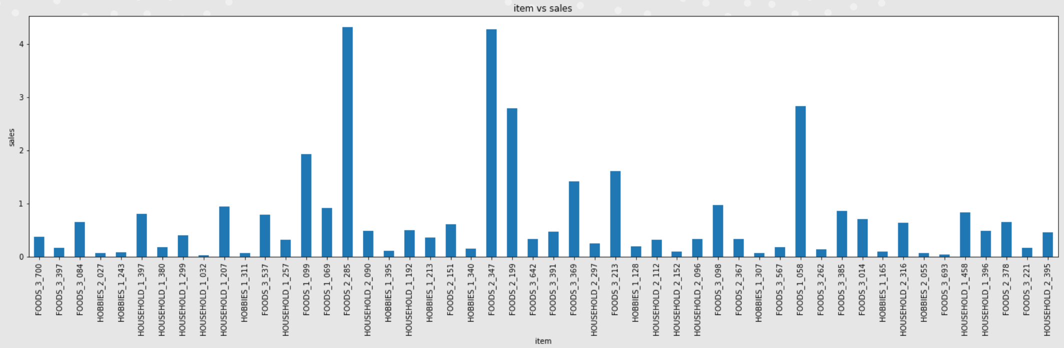
Exploratory Data Analysis

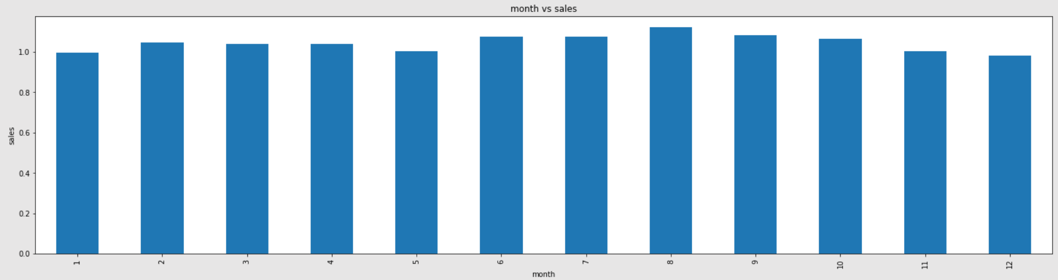
Relation between Top 5 and Target - Numerical

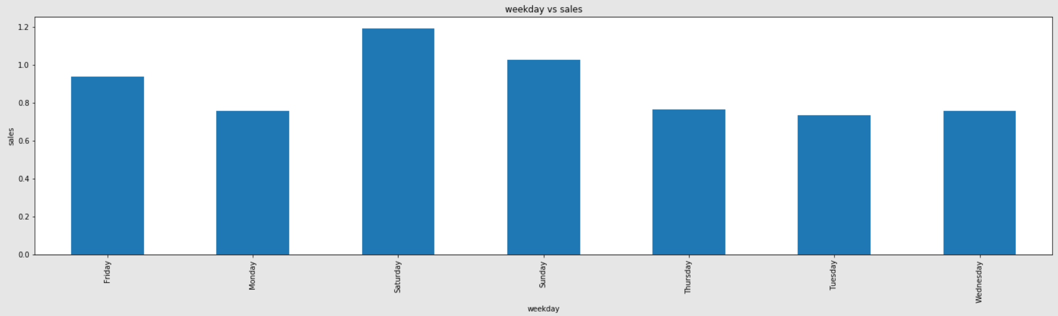


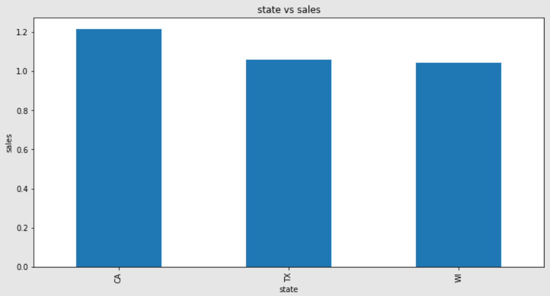
Exploratory Data Analysis

Relation between Top 5 and Target - Categorical



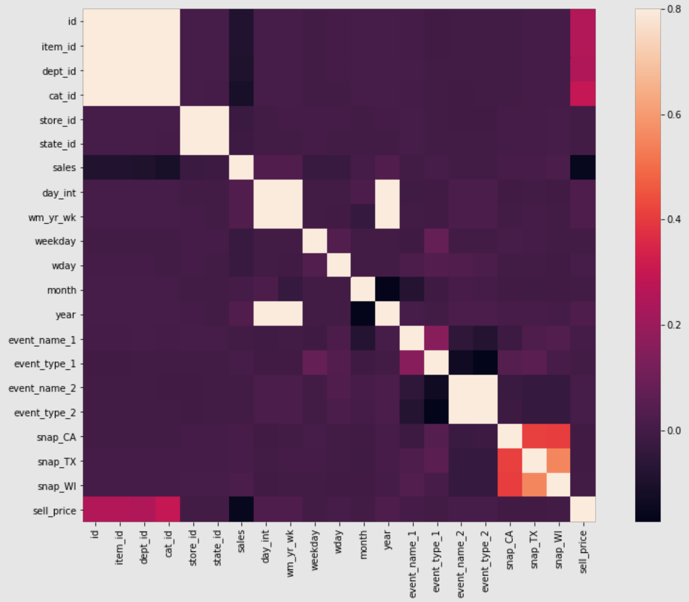




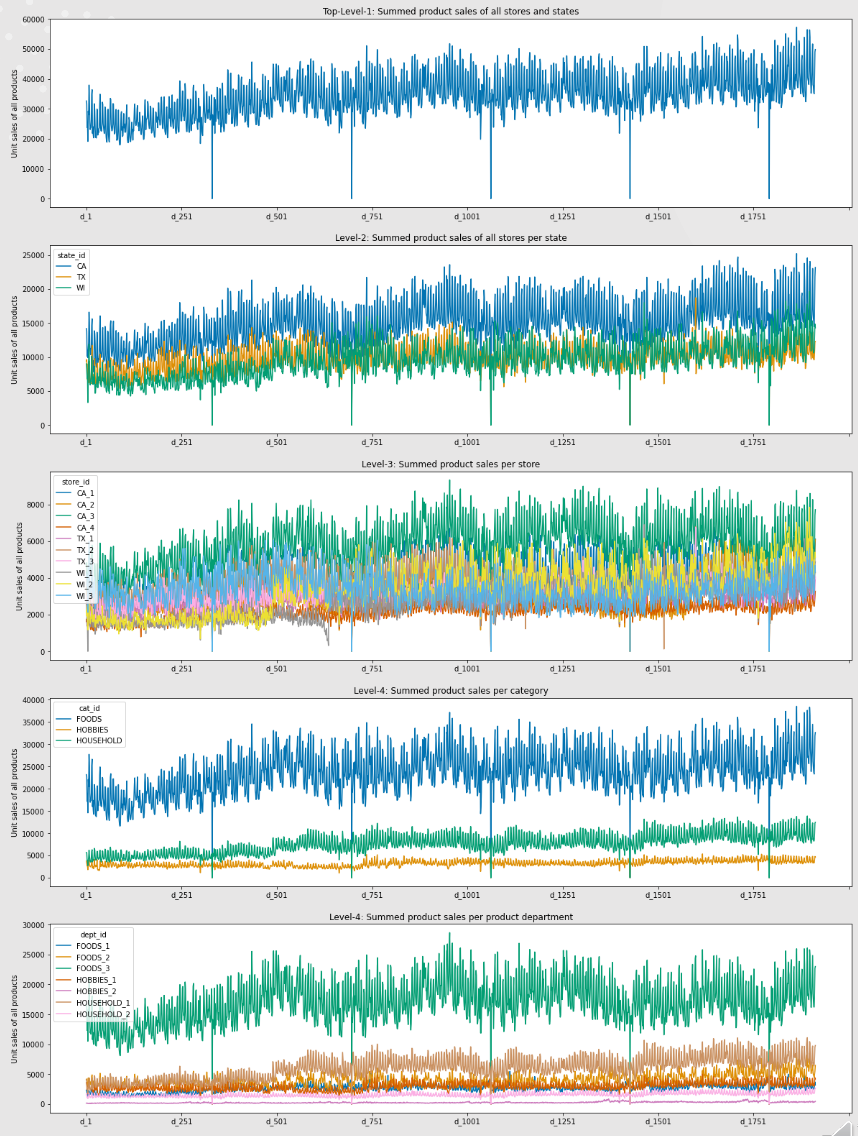


Exploratory Data Analysis

Objective Analysis

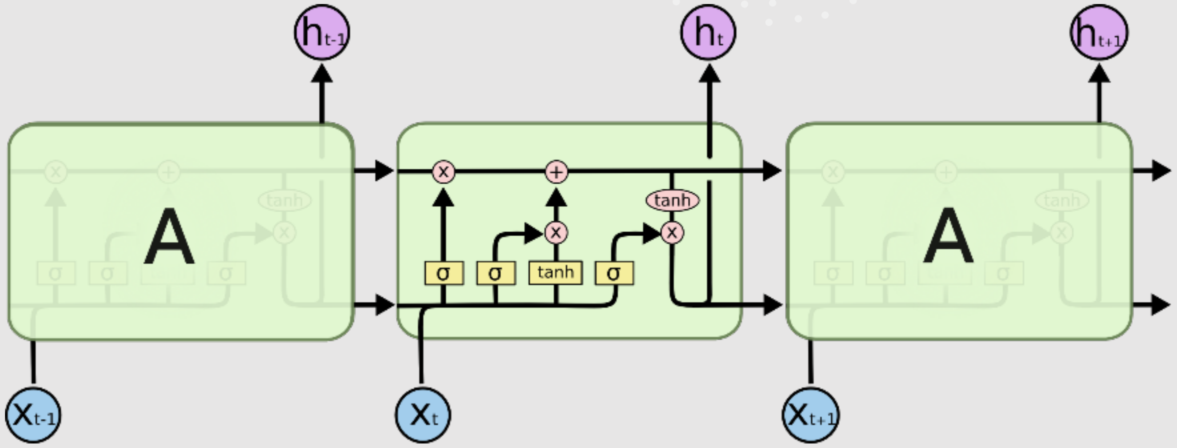


Dig from time-series



Models

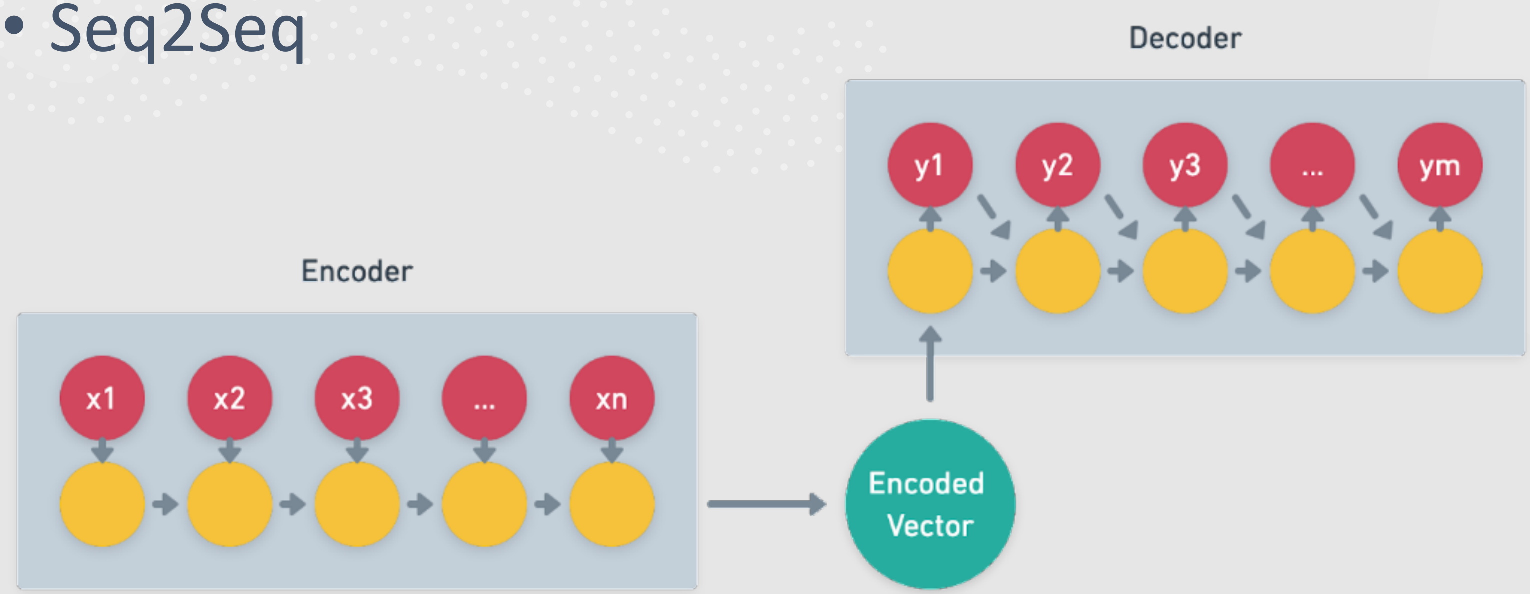
- LSTM (Long Term Short Memory)



- $\begin{pmatrix} i \\ f \\ o \\ g \end{pmatrix} = \begin{pmatrix} \sigma \\ \sigma \\ \sigma \\ tanh \end{pmatrix} W \begin{pmatrix} h_{t-1} \\ x_t \end{pmatrix}$
- $c_t = f \odot c_{t-1} + i \odot g$
- $h_t = o \odot \tanh(c_t)$

Models

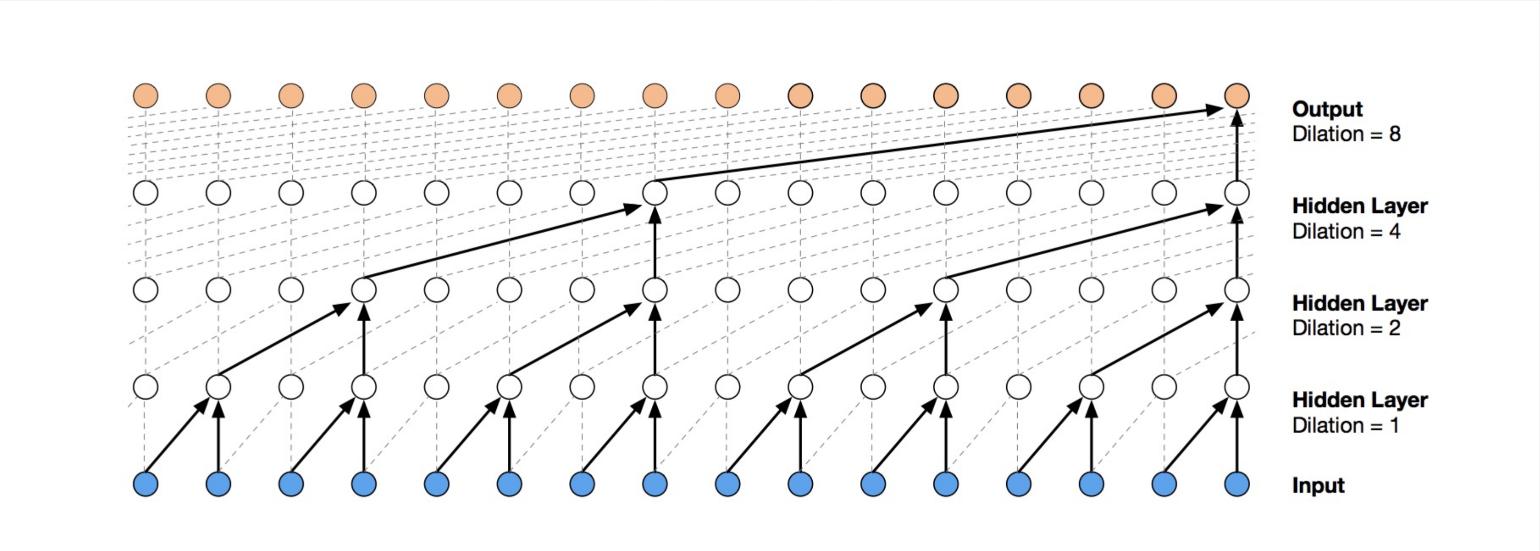
- Seq2Seq



- A type of Encoder-Decoder model using RNN
- Can be used for machine translation, machine interaction or time series
- Input and output vectors need not be fixed

Models

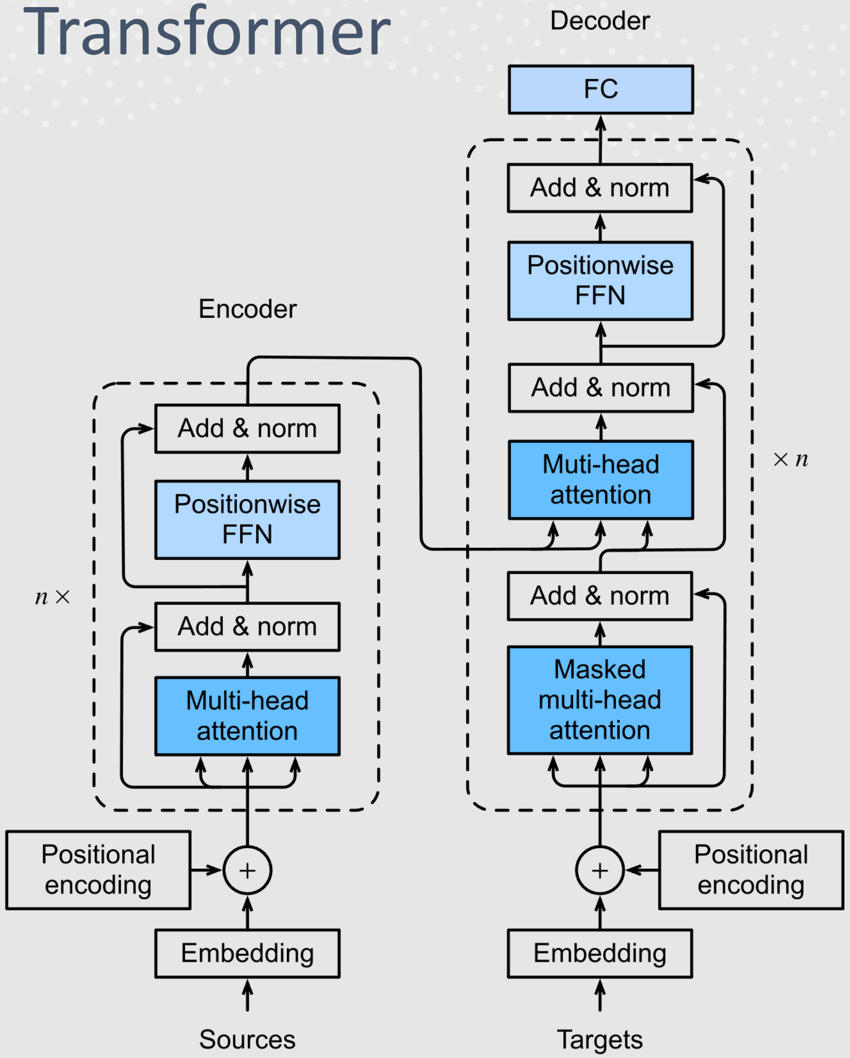
- Dilated-Seq2Seq



- Solve Challenges of RNN
  - Complex dependencies
  - Vanishing and exploding gradients
  - Efficient parallization
- Reduce the number of parameters needed and enhance efficiency

Models

- Transformer



- Attention mechanism
- Benefits:
  - Avoid recursion
  - Allow parallel computation
  - Reduce the drop in performances

Experiments

Parameter Settings

	Model Architecture	Parameter settings
Accuracy	LSTM	model = Sequential() Three layers, each layer L with setting layer_L_units=40 model.add(LSTM(units = layer_L_units, return_sequences = True, input_shape = (X_train.shape[1], X_train.shape[2]))) model.add(Dropout(0.2))
	dilated_seq2seq	(sliding window size: 28*13) rnn_num_hidden = 128, rnn_num_layers = 2, bidirectional = True, enc_rnn_dropout = 0.2 num_epochs = 2 batch_size = 160 learning_rate = 0.0003
	seq2seq_w_attn_on_hid	

Experiments

Parameter Settings

	Model Architecture	Parameter settings
Uncertainty	LSTM	Same as accuracy, Epochs= 32 Batch_size=32 optimizer = 'adam'
	Transformer	(sliding window size: 28*13) rnn_num_hidden = 2, rnn_num_layers = 1, bidirectional = False, enc_rnn_dropout = 0.2 num_epochs = 1 batch_size = 128 learning_rate = 0.001
	seq2seq_w_attn_on_hid	(sliding window size: 28*13) rnn_num_hidden = 128, rnn_num_layers = 2, bidirectional = True, enc_rnn_dropout = 0.2 num_epochs = 2 batch_size = 160 learning_rate = 0.0003

Result and Analysis

	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%)**

Trade-off between performance & training time!!!

Conclusion

✓For Accuracy task

- LSTM, seq2seq with attention, dilated seq2seq
- output “weighted RMSSE” and 28-days predictions
- dilated_seq2seq model has the best performance (0.678, top8%)

✓For Uncertainty task

- LSTM, transformer, and seq2seq with attention
- output “weighted SPL” and predictions for each product
- seq2seq_with_attention has the best score (0.193, top13%)

Future Work

Recommendation for future studies

- Running more epochs for further comparison
- More data on the products and customer behaviors should be collected and analyzed
- Focus more on Feature Engineering part (more features can be added into our model based on more domain knowledge)
- Hyperparameter tuning can be performed for higher performance of prediction
- Our models can be further embedded and stacked together

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.

Thank you!