NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

Informer & Applications

AI6103 Deep Learning and Applications Presentation

Group Project in Deep Learning

Presenter :
Wang Yu Tian (G2204884F)
Group Member:
Zhao YuHan (G2204624C)
Wang Qi Yue (G2204708F)
Koo Chia Wei (G2202810D)
Eugene Ho (G2204889B)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

CONTENTS

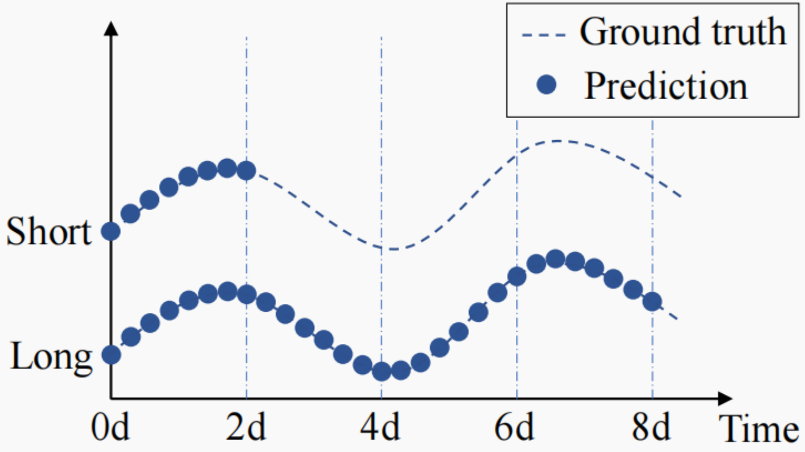
PART 01

Introduction

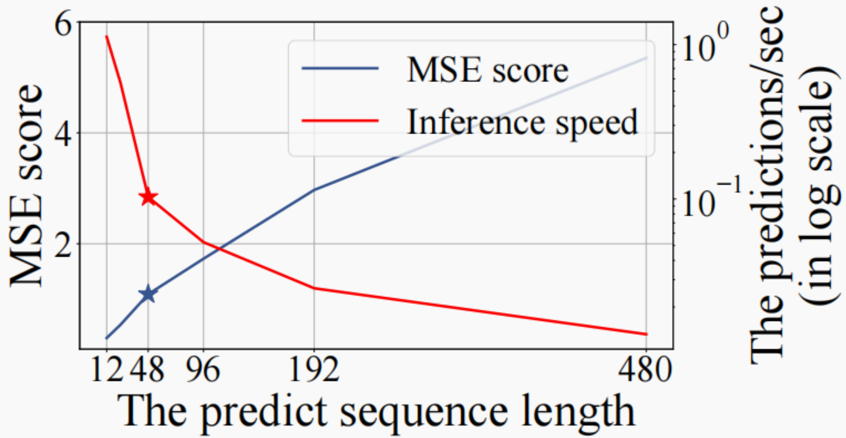
Transformer → Informer

Problems happended in LSTF

LSTF: Long series time-series forecasting



(a) Sequence Forecasting.



(b) Run LSTM on sequences.

PART 02

Model Informer

Summary of Model

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Honor Time

Win best paper in 2021 AAAI

ProbSparse self-attention

replace the canonical self-attention.
achieves $O(L\log L)$ time complexity
achieves $O(L\log L)$ memory usage on dependency alignments.

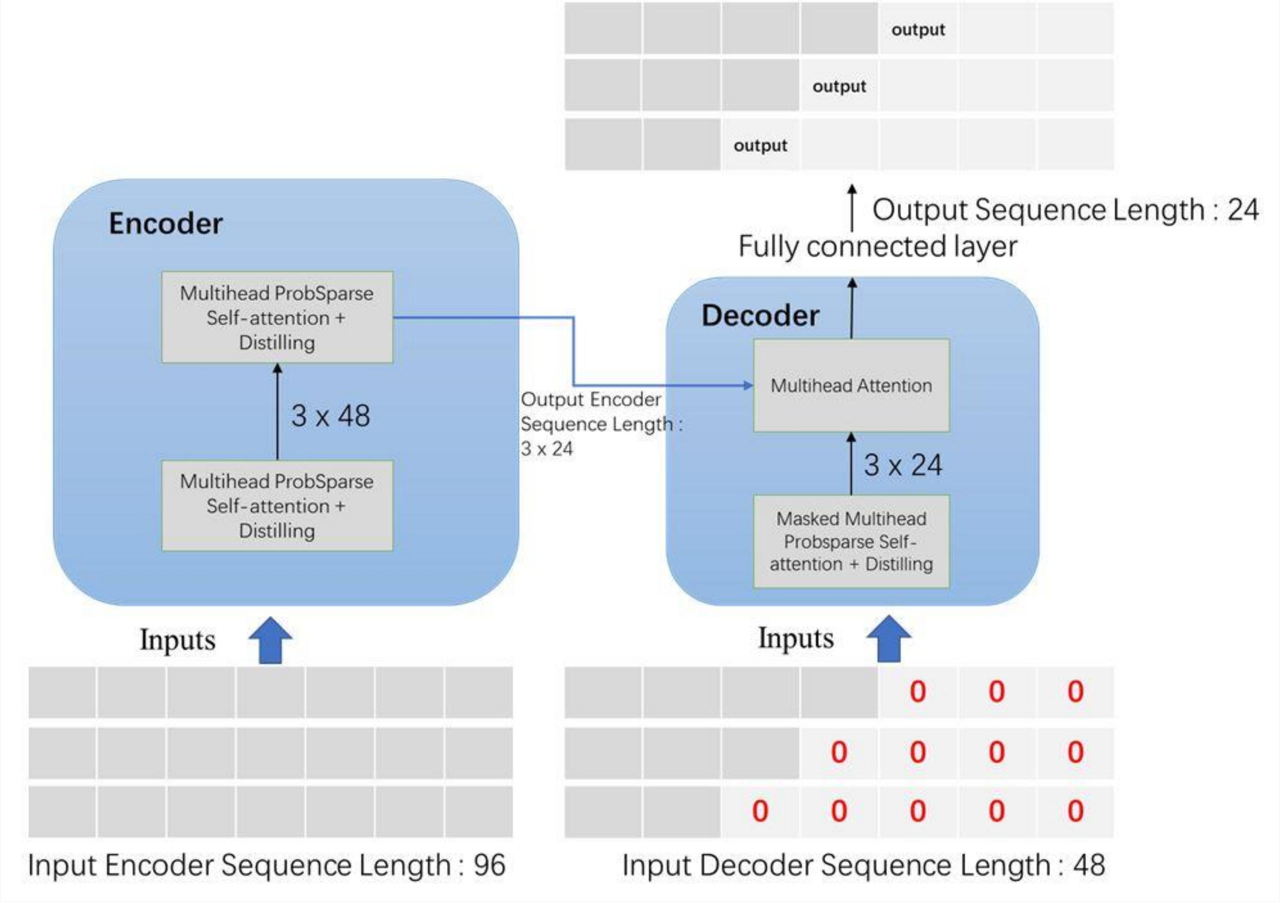
self-attention distilling

privilege dominating attention scores in J-stacking layers
sharply reduce space complexity to $O((2-e)L\log L)$,

generative style decoder

acquire long sequence output with only one forward step needed,
avoiding cumulative error spreading during the inference phase.

2.1 Informer Structure

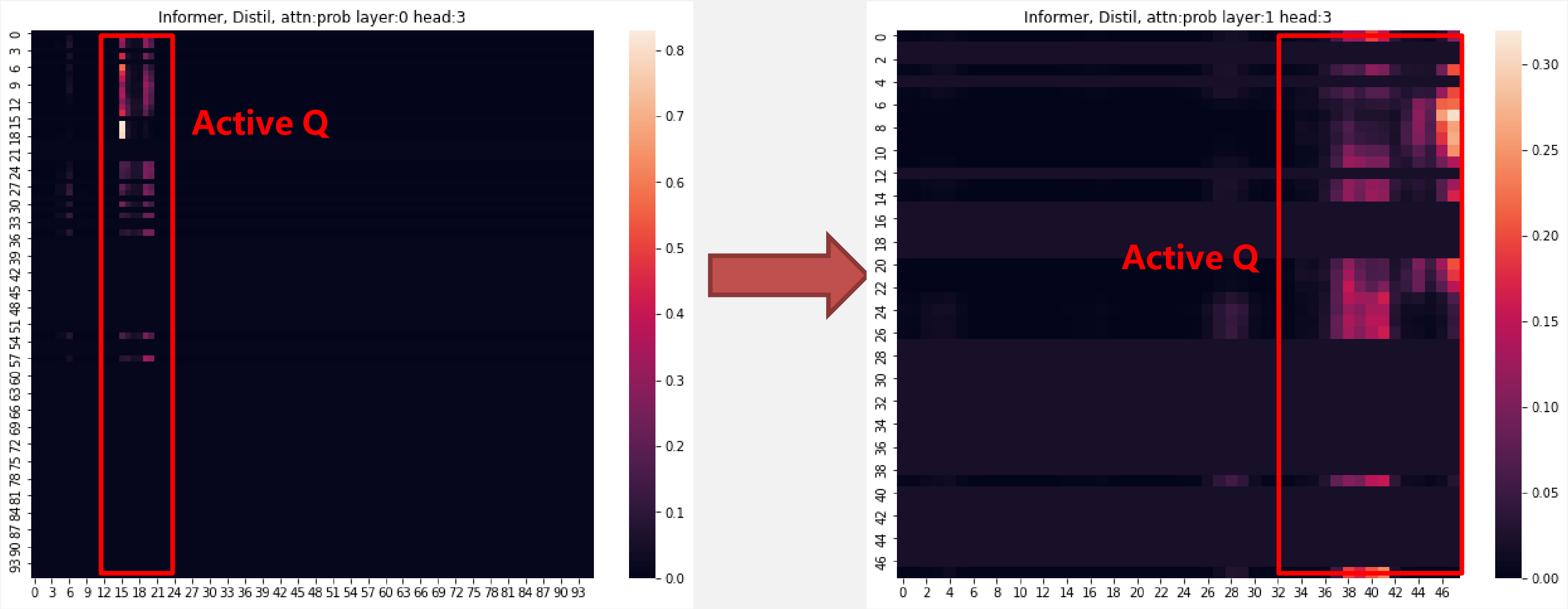


2.2 Probsparse Self-attention

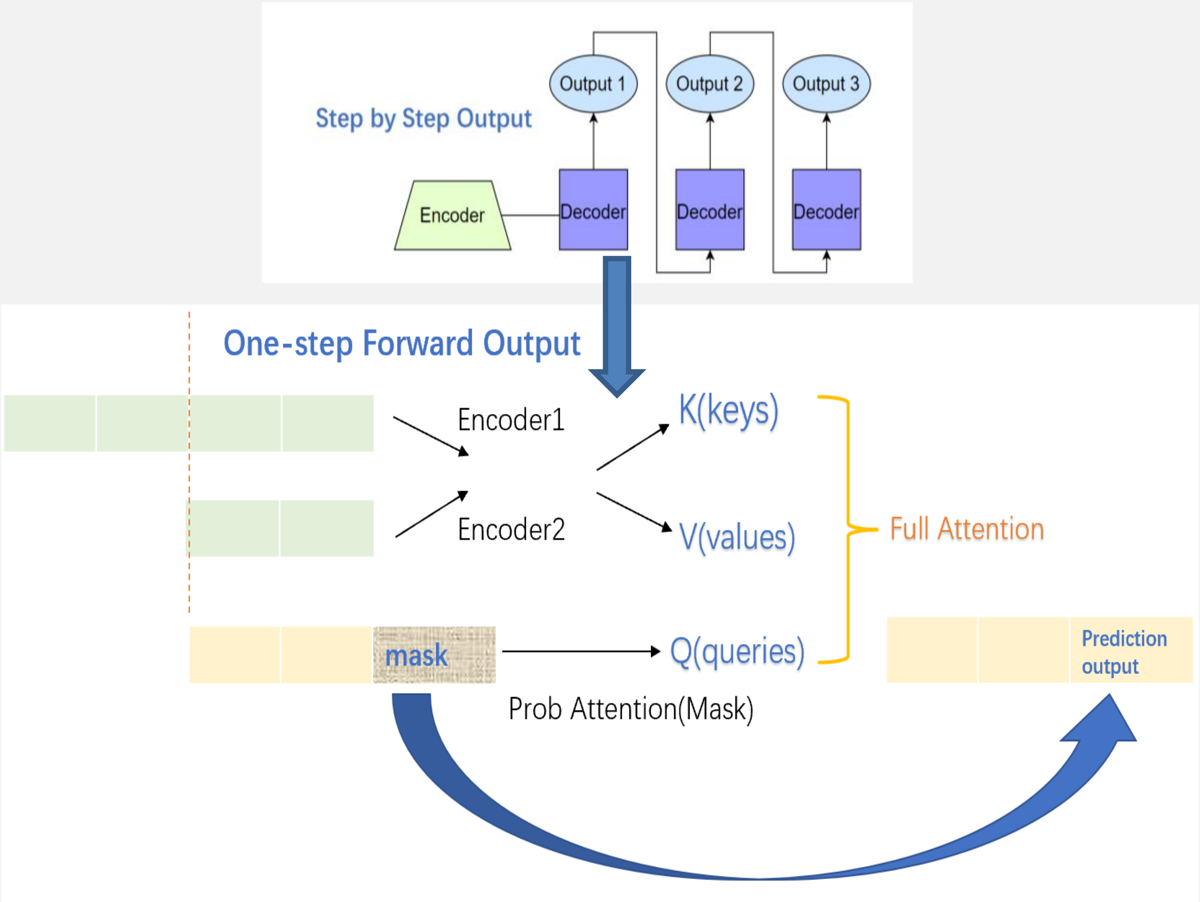
$$A(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d}}\right)V$$

$$A(q_i,K,V) = \sum_j \frac{k(q_i,K_j)}{\sum_l k(q_i,K_l)} V_j = E_p(k_j|q_i)[V_j]$$

2.2 Probsparse Self-attention



2.3 Stacked Decoder



PART 03

Experiments

3 datasets

3.1 WTH Experiment

Models	Informer					Informerstack		
Optimizers	Adam			Adammax		Adam	Adammax	
Loss Function	MSE	L1 loss	HuberL1	MSE	HuberL1	MSE	MSE	HuberL1
Metric	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE	MSE MAE
WTH	0.311 0.366	0.326 0.344	**0.320** **0.356**	0.305 0.360	**0.303** **0.346**	0.319 0.363	0.326 0.373	0.317 0.356

3.2 Global Temparature Experiment

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- The dataset was obtained from **data.world** and generated by **Berkeley Earth**
- Global land and ocean temperature using monthly data from **1850 to 2015**
  - Remove data prior to 1850 due to empty entries
- **Training set (60% - 1195), validation set (20% - 398), testing set (20% - 398)**
- Forecast for a **time horizon ~20 months**
- **Multivariate** Forcasting
- The dataset includes **8 variables**:
  1. LandAverageTemperature
  2. LandAverageTemperatureUncertainty-95% Interval
  3. LandMaxTemperature
  4. LandMaxTemperatureUncertainty-95% Interval
  5. LandMinTemperature
  6. LandMinTemperatureUncertainty-95% Interval
  7. **LandAndOceanAverageTemperature (Target)**
  8. LandAndOceanAverageTemperatureUncertainty

BERKELEY EARTH

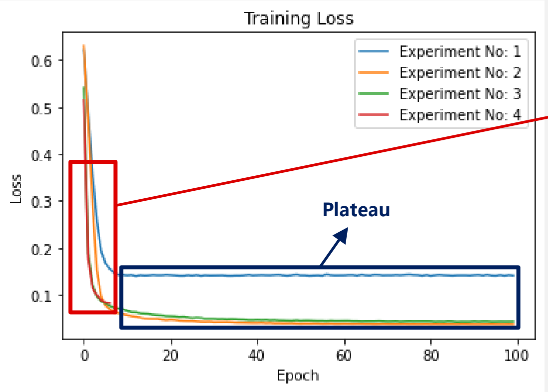
data.world

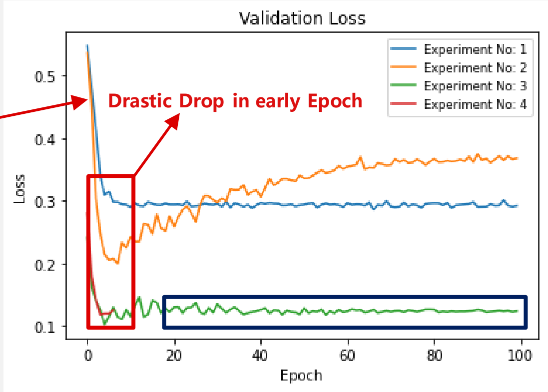
3.2 Global Temparature Experiment

Experiment No.	Epoch	Learning Rate Decay per Epoch	Sequence, Label, Prediction Length	Test Error, MSE	Test Error, MAE
1	100	0.5 (50%)	[96,48,24]	0.401	0.493
2	100	0.05 (5%)	[96,48,24]	0.302	0.391
3	100	0.05 (5%)	[48,24,12]	0.141	0.288
4	**10 (Early Stoppage, Patience = 3)**	**0.05 (5%)**	**[48,24,12]**	**0.165**	**0.313**

Shorter > Longer

Comparable Performance



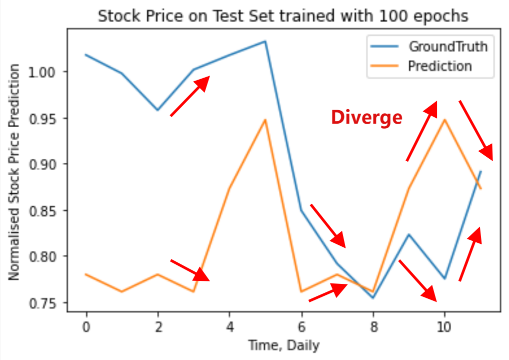


- **Lowering learning rate decay** for **longer epochs** result in an increase in validation loss, indicating **overfitting** possibility
- Experiment 2's Validation Loss < Experiment 3, **benefits to reduce learning rate decay**
- **Shorter** sequence, label and prediction **outperform longer** ones
- **Training and validating loss decrease** drastically at early **epoch < 15**
- Validation loss may plateau or increase during **higher epochs**, indicating **diminishing returns** on model performance.
- **Early Stoppage with Patience = 3** can achieve **similar performance** as model trained in **100 epochs**

3.3 Stock Experiment

Experiment No.	Epoch	Learning Rate Decay per Epoch	Sequence, Label, Prediction Length	Test Error, MSE	Test Error, MAE
1	5 **(Early Stoppage, Patience = 3)**	0.05 (5%)	[96,48,24]	0.168	0.339
2	10 **(Early Stoppage, Patience = 3)**	0.05 (5%)	[48,24,12]	0.0934	0.237
3	13 **(Early Stoppage, Patience = 3)**	0.05 (5%)	[24,12,6]	0.0379	0.152
4	**100**	**0.05 (5 %)**	**[24,12,6]**	**0.0371**	**0.151**

Comparable Performance



- Dataset from Singapore Exchange Limited Stock Exchange **(S68.SI)** from **Nov 2000 to Dec 2017**
- The dataset includes **4 variables**:
  1. Market Open Price
  2. Market Close Price **(Target)**
  3. Low Price (Per Day)
  4. High Price (Per Day)
- No financial crisis indicator, remove all datapoint prior to Nov 2000 and after Dec 2017
- Performance from **early stoppage** are **comparable** to model that is trained for **100 epochs**
- Model trained with 4 features is **unable** to **forecast and predict accurately**
- **Trends** generated by the model are **inconsistent** with the **ground truth**
- Require **additional financial indicators** (Price-to-earning Ratio, Debt-to-earning ratio and etc.)

PART 04

Conclusion
&
Prospect

Conclusion and Prospect

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3 conclusions：

1. Different datasets：
   different optimization methods and loss functions

2. high epochs :
   reducing the learning rate → increase the risk of higher validation loss
   【performs well in predicting regular time series trends】

3. poorly in predicting complex time series trends
   【achieve relatively good results in a short number of epochs】

Overall, we achieved the expected performance of the informer model.

Prospect: Our outlook is to construct algorithms that can accurately predict both short-term and long-term time series trends in a short amount of time.

Thanks for your listening