



南京大學  
NANJING UNIVERSITY

RINC



# 基于前馈神经网络的时间序列预测问题研究

- 答辩人：郝鸿延 MF1833023
- 导师：申富饶 教授  
宋方敏 教授



# 目录

## CONTENTS

### 1 研究背景

### 2 研究内容

- 前馈序列网络 - FSN
- 多尺度 DTW 损失函数 - MS-DTWI

### 3 研究成果

- 商品管理和销量预测系统
- 硕士研究生期间成果总结

### 4 工作总结



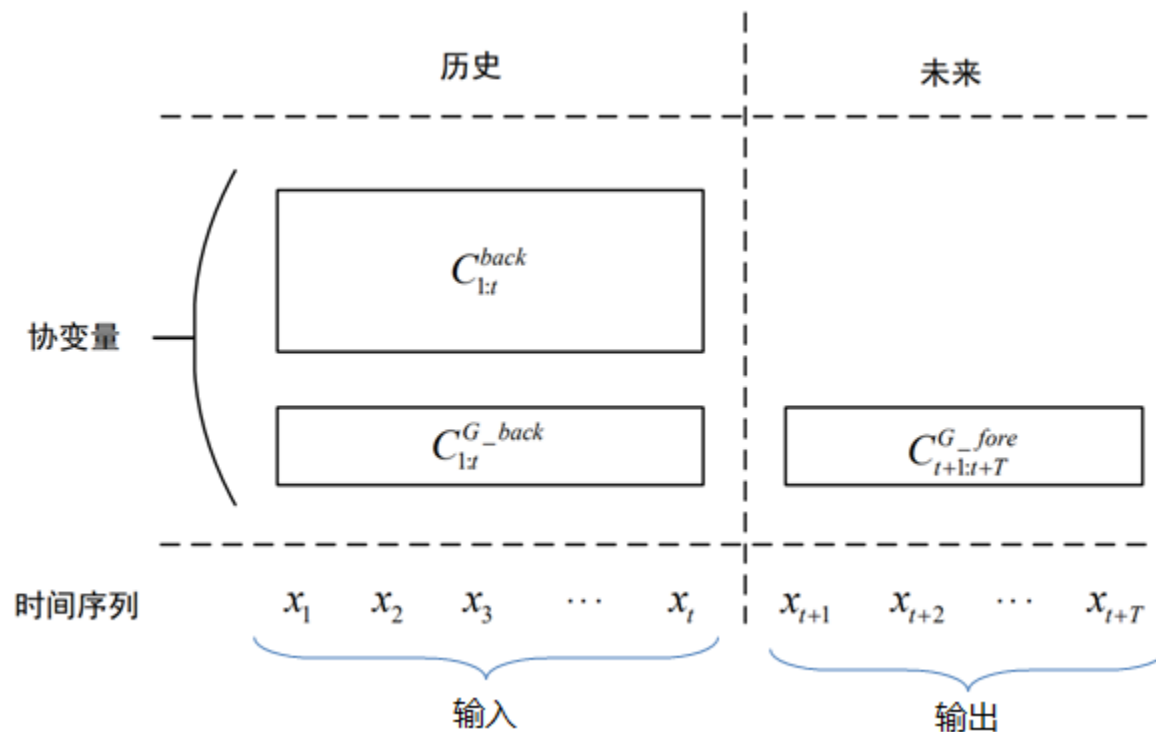
# 第一部分

## Research Background 研究背景



- 时间序列：一组按照时间发生先后顺序排列而成的数据点序列
- 时间序列预测任务：用**历史序列值**和**协变量**，预测**多个未来序列值**

$$\hat{X}_{t+1:t+T} = f\left(X_{1:t}, C_{1:t}^{G\_back}, C_{t+1:t+T}^{G\_fore}, C_{1:t}^{back}\right)$$





# 研究背景

## ■ 应用场景



销量预测



量化交易

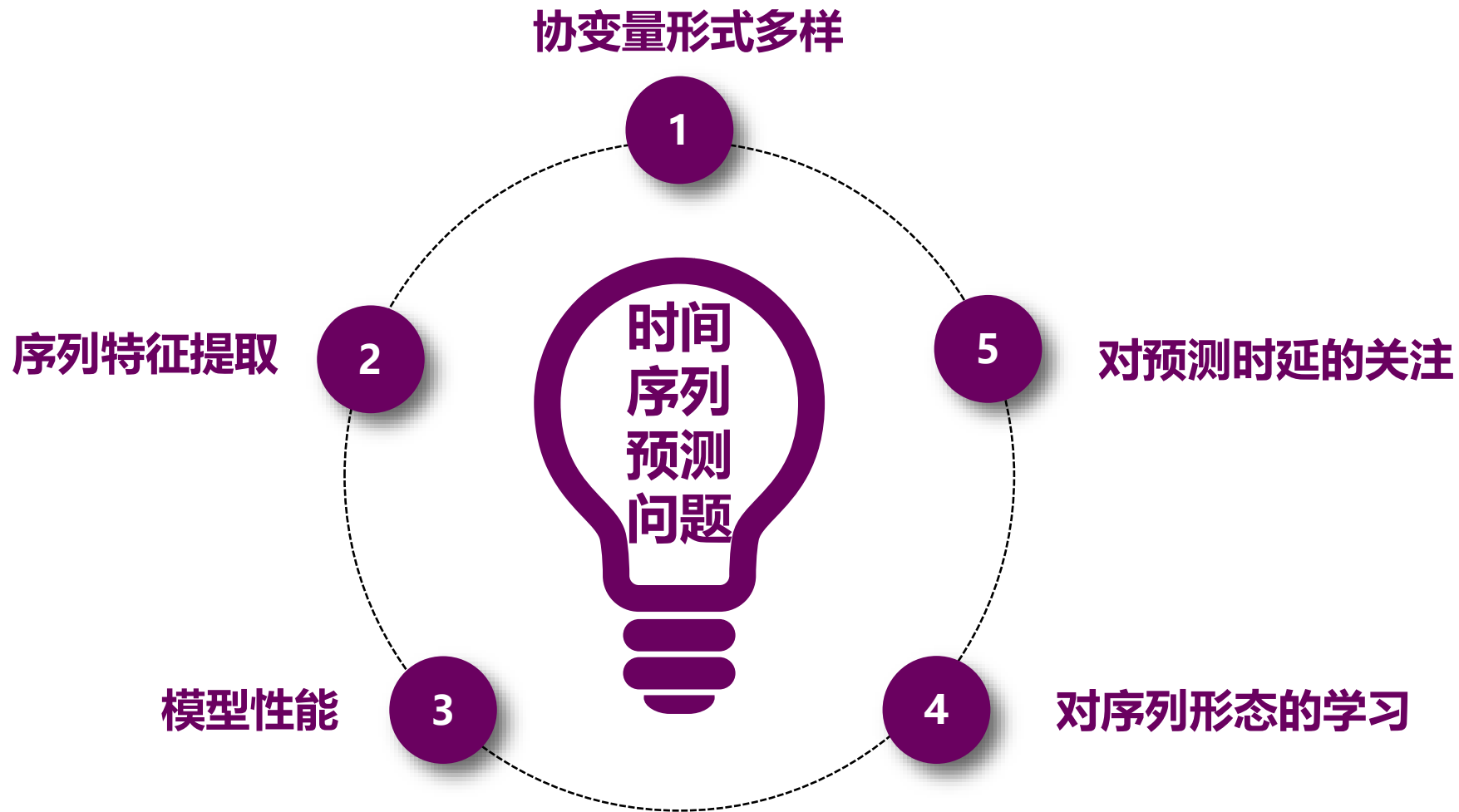


能源消耗预测

时间序列预测无处不在



## 困难与挑战





## 第二部分

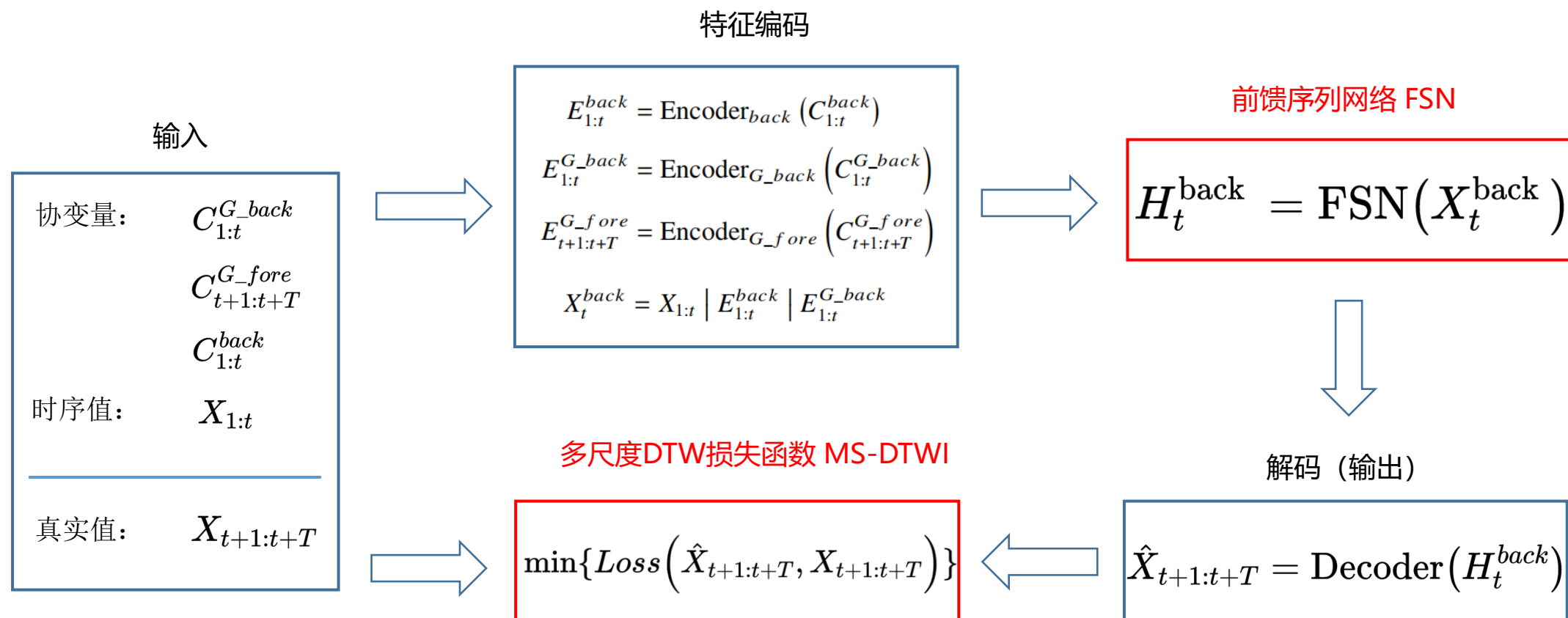
## Research Contents

### 研究内容

- 前馈序列网络 - FSN
- 多尺度 DTW 损失函数 - MS-DTWI



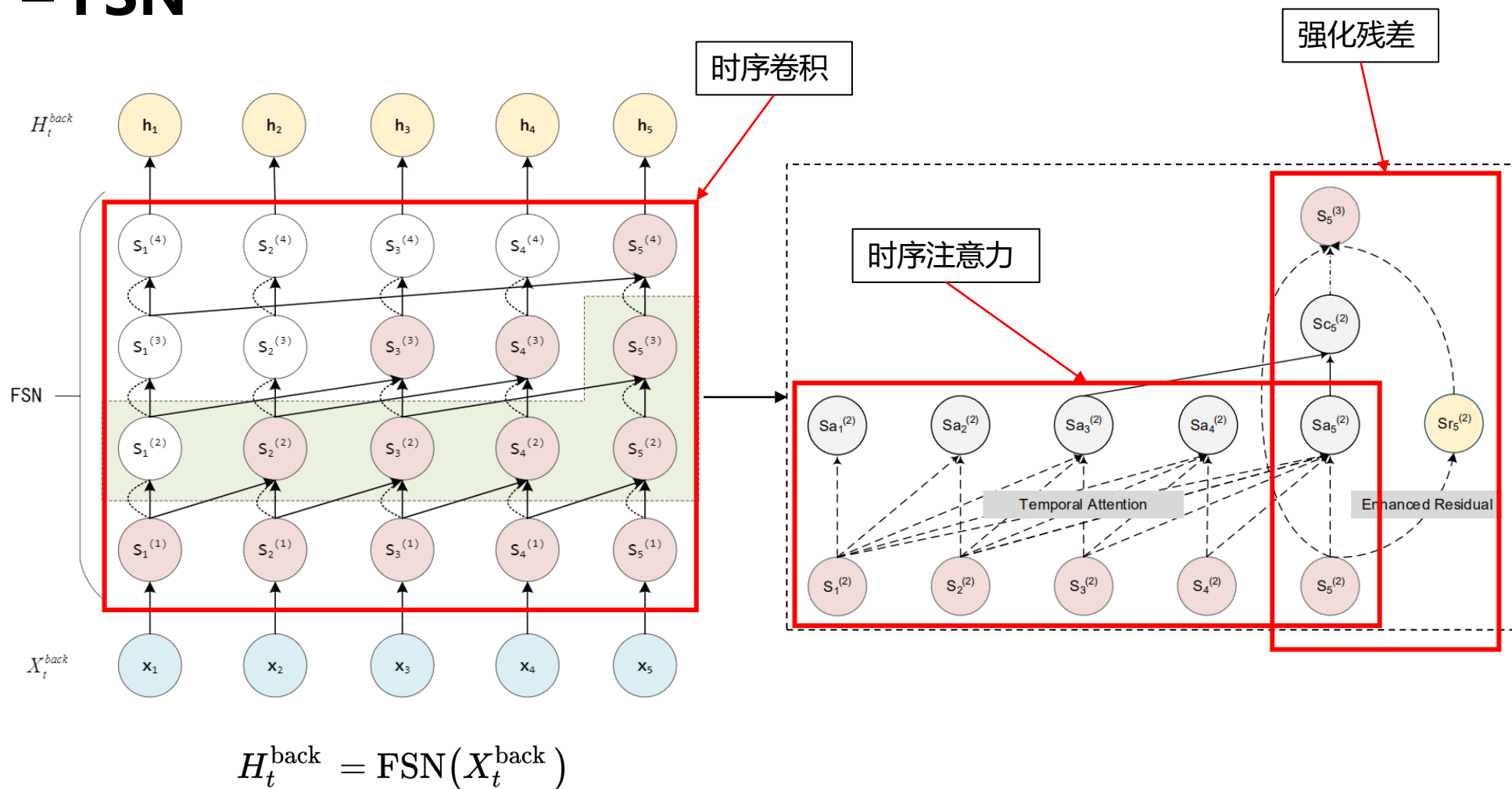
## 基于神经网络方法的时间序列预测流程





## 前馈序列网络 – FSN

- 时序卷积
- 时序注意力
- 强化残差



## 时序卷积

- 卷积网络避免递归计算
- 扩展卷积增大感受野
- 因果卷积建模序列相对位置关系

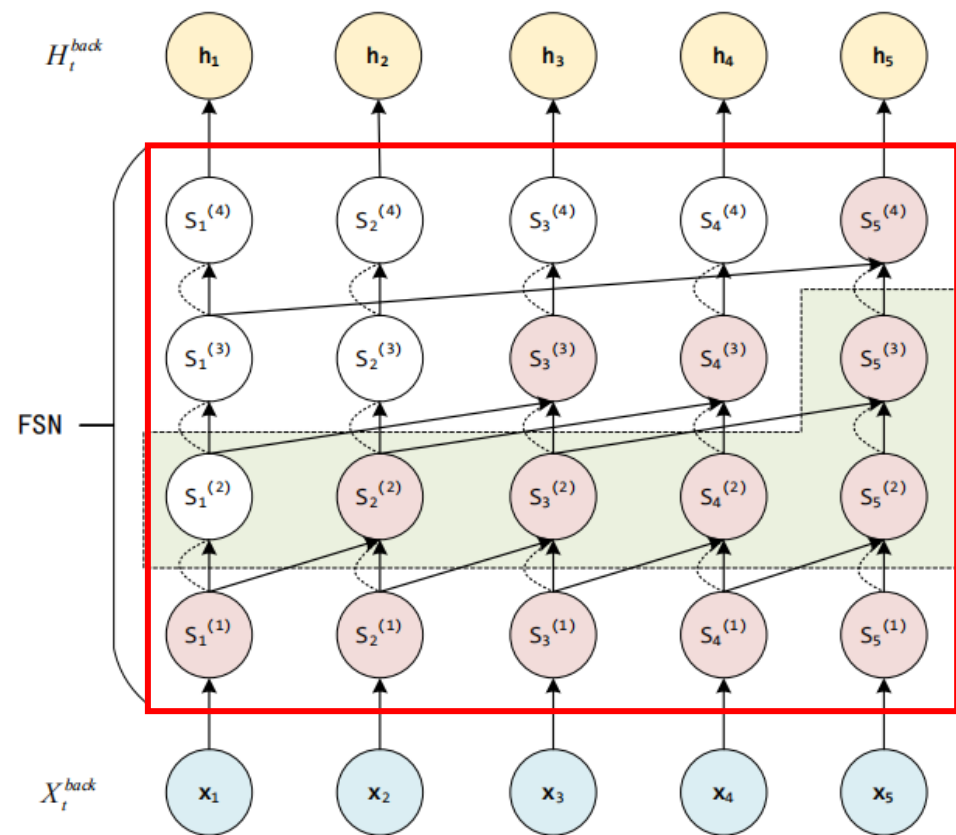
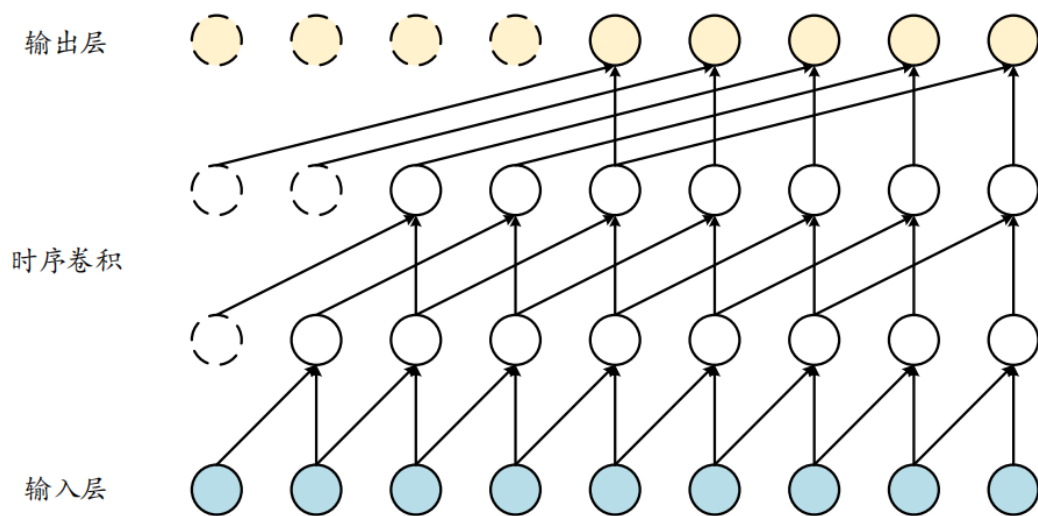


图 3-6: 前馈序列网络结构

## 时序注意力

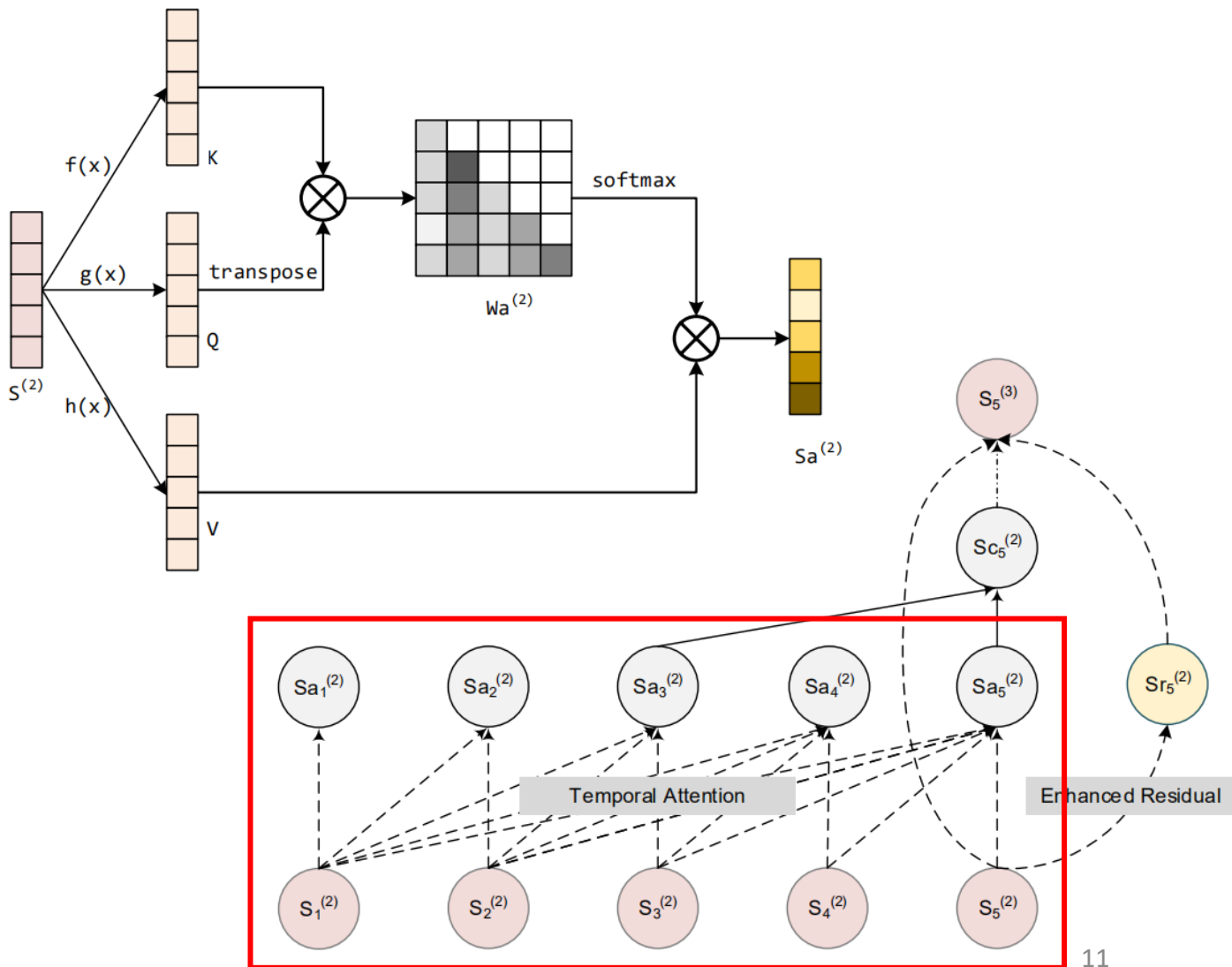
- 为防止序列信息泄露，取权重矩阵的下三角部分：

$$Wl_{i,j} = \begin{cases} (W)_{i,j}, & \text{if } i \geq j \\ 0, & \text{if } i < j, \end{cases}$$

$$Wa_{i,j} = \frac{e^{-\frac{Wl_{i,j}}{\sqrt{p}}}}{\sum_{j=1}^t e^{-\frac{Wl_{i,j}}{\sqrt{p}}}} \text{ for } j = 1, 2, \dots, t$$

- 权重矩阵与序列特征相乘得到包含历史序列点信息的新特征：

$$Sa_i = \sum_{j=1}^t Wa_{i,j} \cdot V_j$$



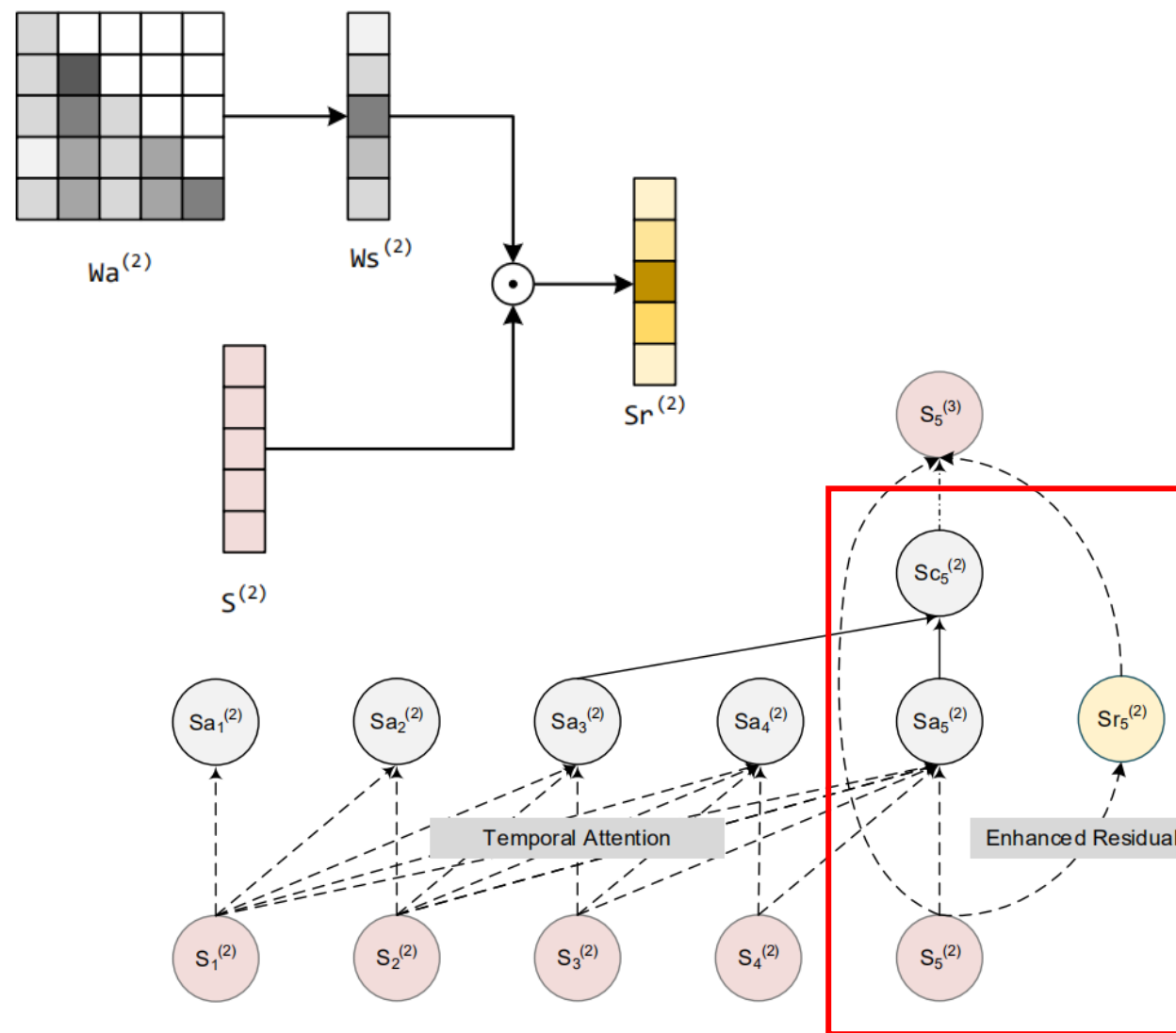
## 强化残差

- 利用时序注意力层的下三角权重矩阵，在第二个维度上求和得到权重向量：

$$w_i = \sum_{j=1}^t W a_{i,j}$$

- 将权重向量和序列特征相乘：

$$Sr_i = S_i \times w_i$$





# 研究内容

## 实验验证

表 3-2: FSN 网络和其他方法对比结果

Share Bike					TSA2016				
Model	MAE	ND	NRMSE	SMAPE	Model	MAE	ND	NRMSE	SMAPE
SVR	0.140	0.725	0.974	0.855	SVR	0.111	0.212	0.266	0.216
LSTM	0.065	0.542	0.901	0.997	LSTM	0.099	0.206	0.261	0.194
TCN	0.066	0.596	0.837	0.743	TCN	0.083	0.172	0.211	0.171
FSN	<b>0.050</b>	<b>0.449</b>	<b>0.677</b>	<b>0.738</b>	FSN	<b>0.049</b>	<b>0.103</b>	<b>0.132</b>	<b>0.102</b>
Traffic Bits					PRSA				
Model	MAE	ND	NRMSE	SMAPE	Model	MAE	ND	NRMSE	SMAPE
SVR	0.211	0.494	0.671	0.513	SVR	0.070	0.828	1.054	0.821
LSTM	<b>0.048</b>	<b>0.116</b>	<b>0.157</b>	<b>0.152</b>	LSTM	0.052	0.551	0.778	0.634
TCN	0.100	0.242	0.288	0.359	TCN	0.050	0.529	0.780	0.745
FSN	0.063	0.152	0.190	0.199	FSN	<b>0.021</b>	<b>0.223</b>	<b>0.331</b>	<b>0.327</b>

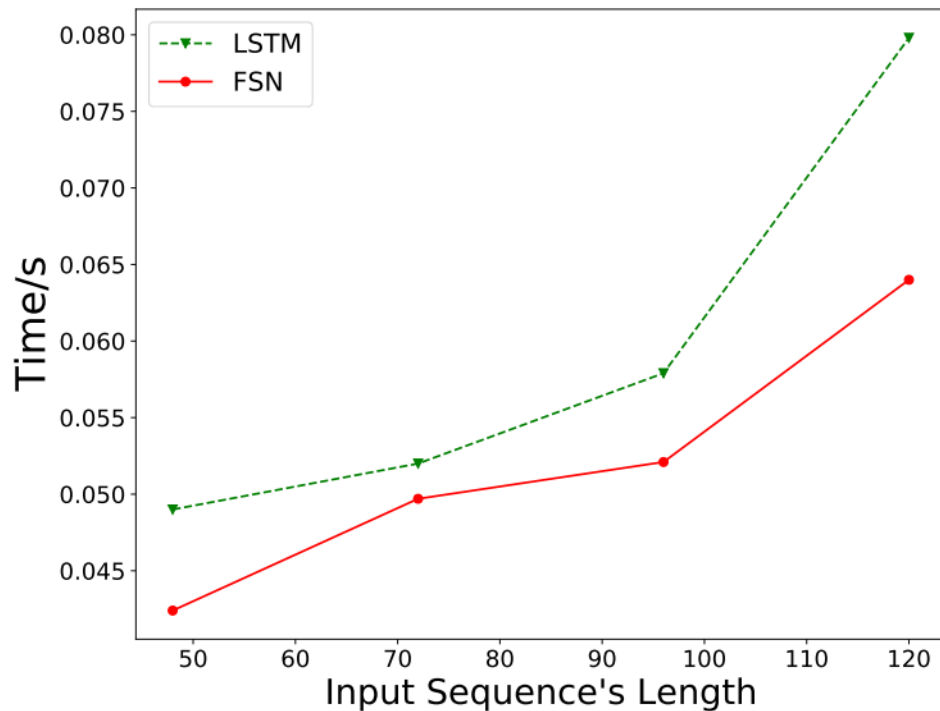


图 3-13: 不同输入长度模型耗时比较



# 研究内容

## 实验验证

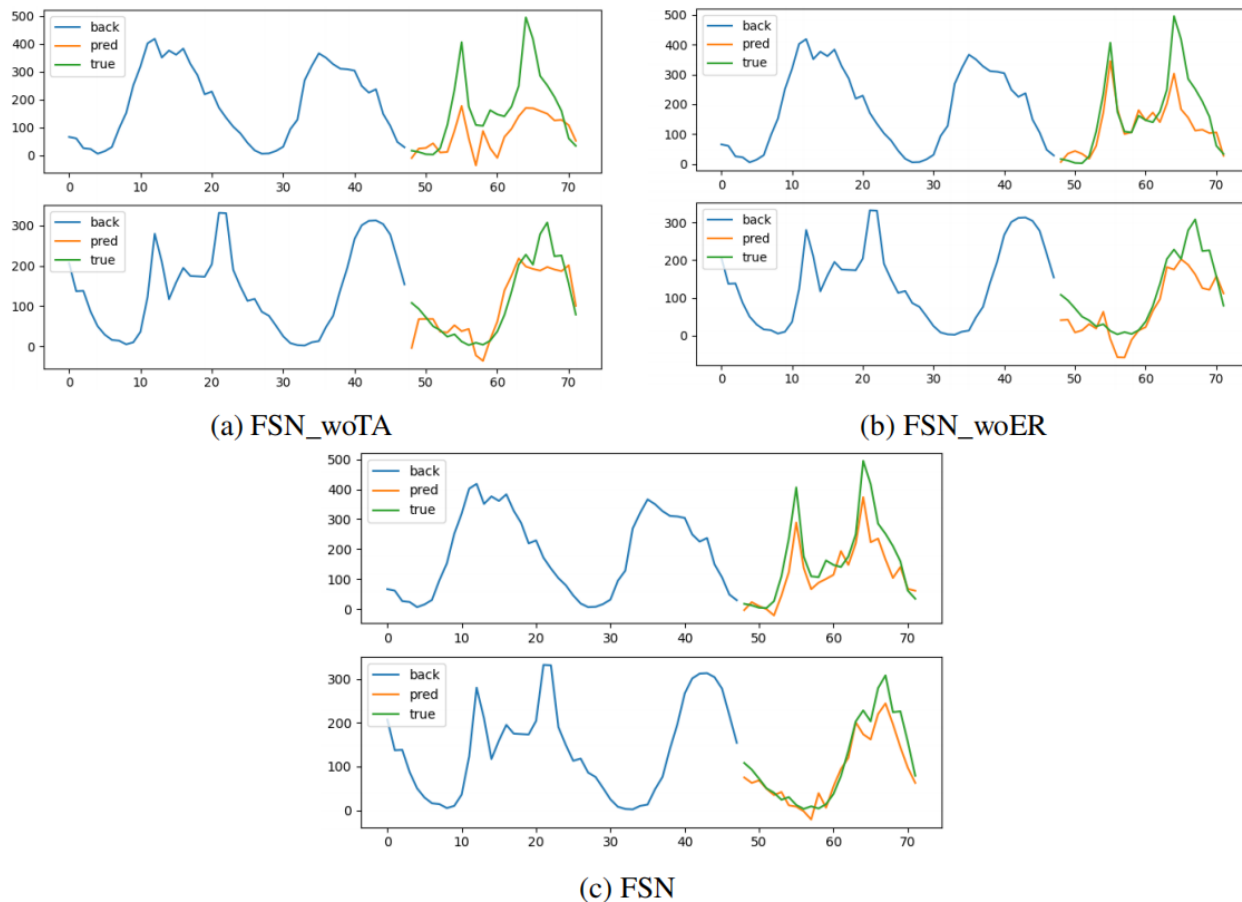


图 3-12: 消融实验结果序列图

表 3-3: 消融实验结果

Models	ND	NRMSE	MAE
FSN_woTA	0.5959	0.8373	0.0660
FSN_woER	0.5586	0.7783	0.0618
FSN	0.4485	0.6765	0.0496

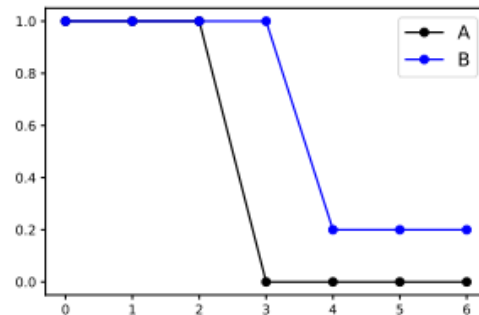


## 多尺度 DTW 损失函数 - MS-DTWI

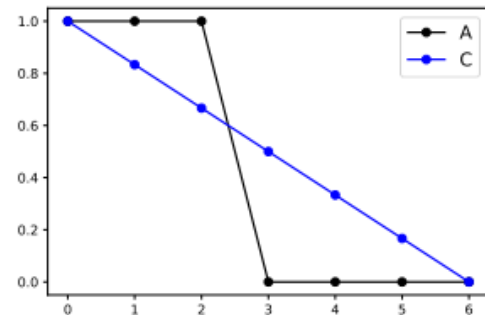
$$\text{MSE}(A,B) = 0.23 > \text{MSE}(A,C) = 0.21$$

$$\text{DTW}(A,B) = 1.40 < \text{DTW}(A,C) = 1.83$$

MSE 未反映形态相似度，但 DTW 可以



(a) A 和 B

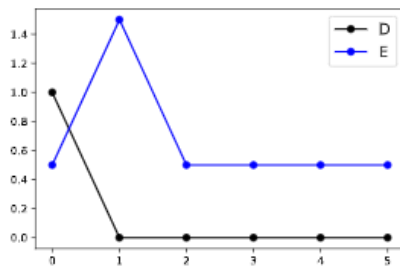


(b) A 和 C

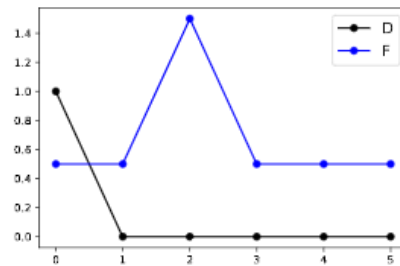
$$\text{MSE}(D,E) = \text{MSE}(D,F) = \text{MSE}(D,G) = 0.67$$

$$\text{DTW}(D,E) = \text{DTW}(D,F) = \text{DTW}(D,G) = 5.5$$

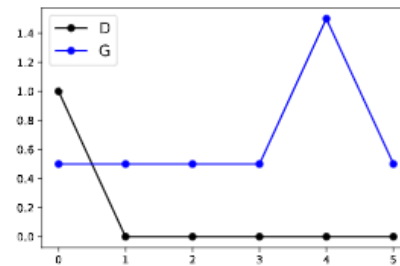
MSE 和 DTW 都无法反映序列的延时情况



(a) D 和 E



(b) D 和 F



(c) D 和 G

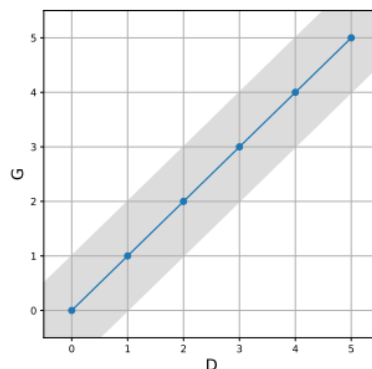


## MS-DTW

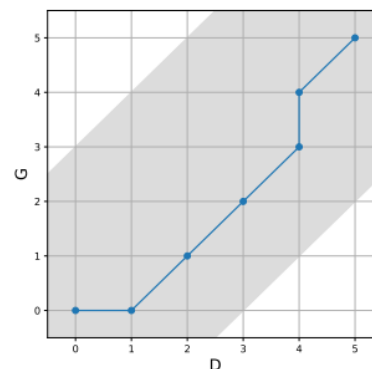
( Multi-Scale Dynamic Time Warping )

- DTW 本身能够表征序列形态相似度;
- DTW 的规整窗口表示序列点最大对应距离, 多个窗口大小可以表征不同程度的延时性:

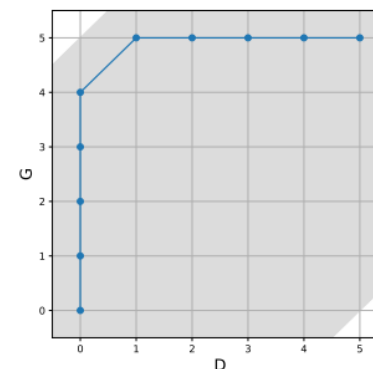
$$MS-DTW^n = \frac{1}{n} \sum_{i=1}^n DTW^{u_i}, u_i \in \mathcal{U}$$



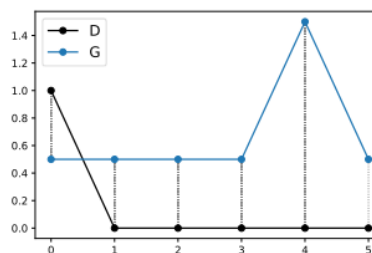
(a)  $w = 1$



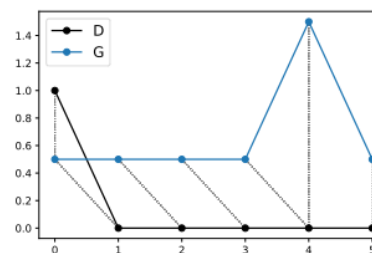
(b)  $w = 3$



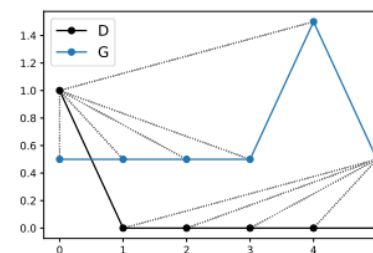
(c)  $w = 5$



(d)  $w = 1$



(e)  $w = 3$



(f)  $w = 5$





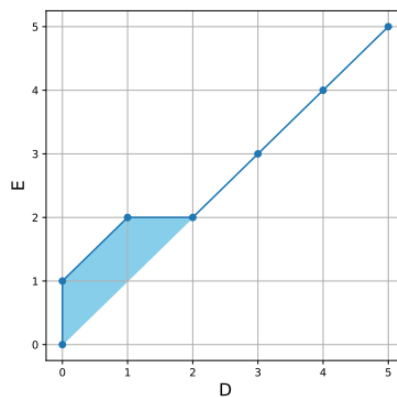
## TDI

( Temporal Distortion Index )

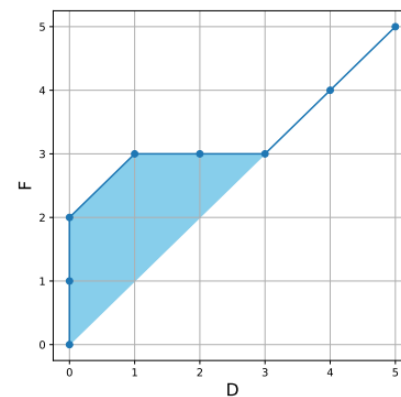
- TDI 指标为序列点距离矩阵中**最短路**径与对角线包围的区域占路径空间的比例：

$$S_l = \int_{i_l}^{i_{l+1}} \left( x - \frac{(x - i_l)(j_{l+1} - j_l)}{(i_{l+1} - i_l)} + j_l \right) dx$$
$$TDI = \frac{2 \sum_{l=1}^{k-1} |S_l|}{N^2}$$

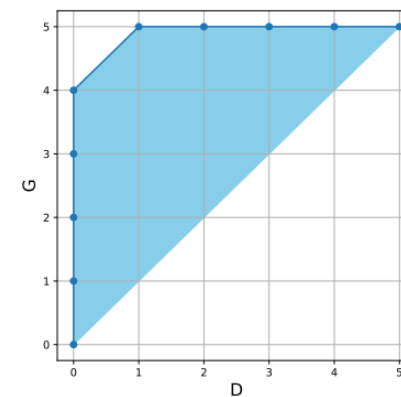
- 此指标也可以表征两条序列的延时性大小



(a) D 和 E



(b) D 和 F

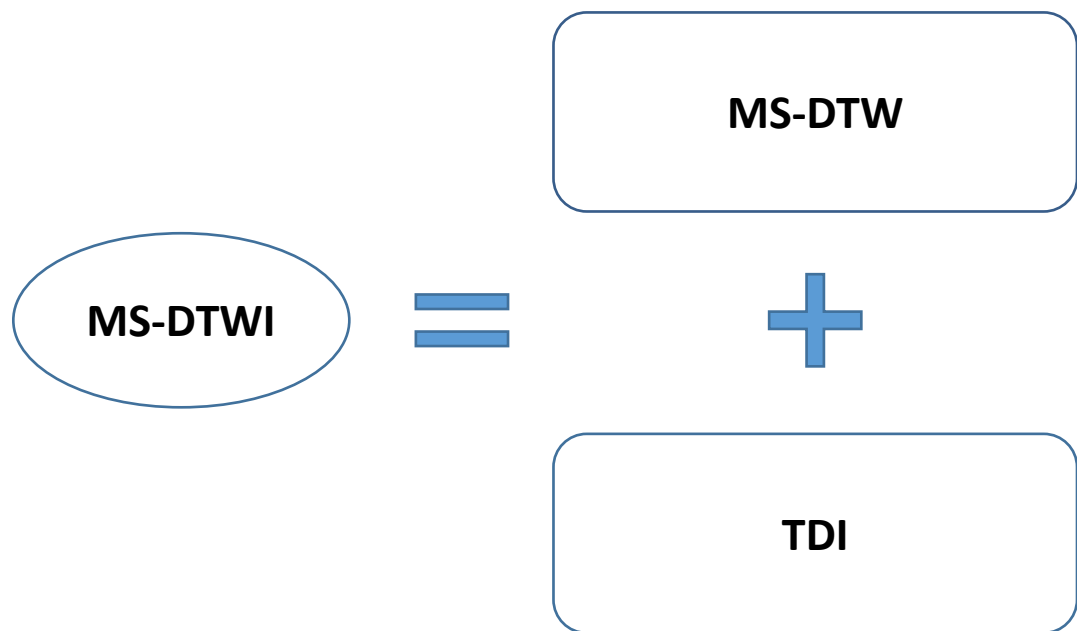


(c) D 和 G



## 多尺度 DTW 损失函数 - MS-DTWI

$$\text{MS-DTWI} = \alpha \cdot \text{MS-DTW} + (1 - \alpha) \cdot \text{TDI}$$



度量指标	D 与 E	D 与 F	D 与 G
MSE	0.67	0.67	0.67
DTW	5.50	5.50	5.50
TDI	0.10	0.37	1.67
MS-DTW(3)	0.99	1.08	1.25
MS-DTW(6)	1.03	1.10	1.25

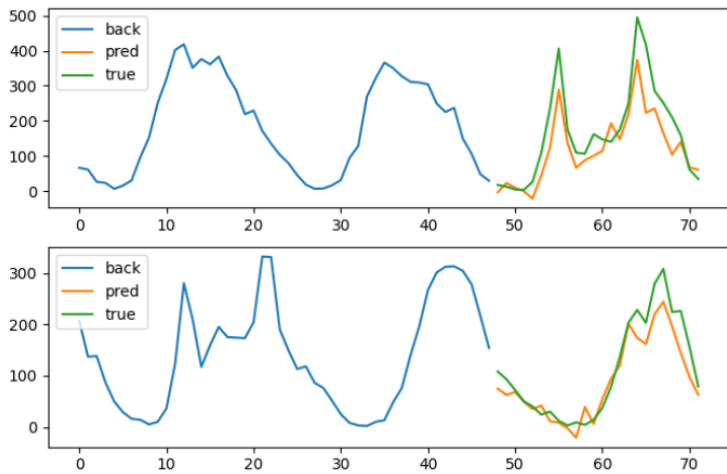


## 实验验证

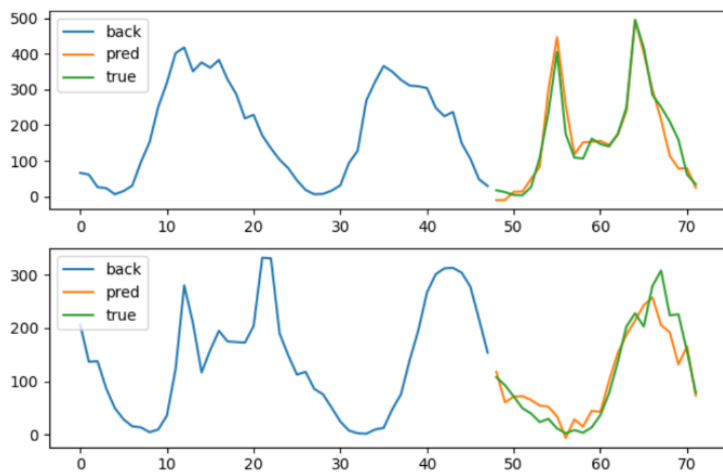
### 对比实验

表 4-2: 不同损失函数实验对比结果

损失函数	验证集			测试集			
	MAE	ND	NRMSE	MAE	ND	NRMSE	MS-DTW
MSELoss	0.063	0.334	0.473	0.051	0.461	0.632	0.060
HuberLoss	0.060	0.315	0.440	0.055	0.496	0.701	0.078
DILATE	0.066	0.350	0.500	0.058	0.523	0.772	0.086
MS-DTWI	<b>0.055</b>	<b>0.291</b>	<b>0.412</b>	<b>0.042</b>	<b>0.379</b>	<b>0.544</b>	<b>0.044</b>



(a) MSELoss



(b) MS-DTWI



## 实验验证

敏感性分析：n 和  $\alpha$  对 MS-DTWI 的影响

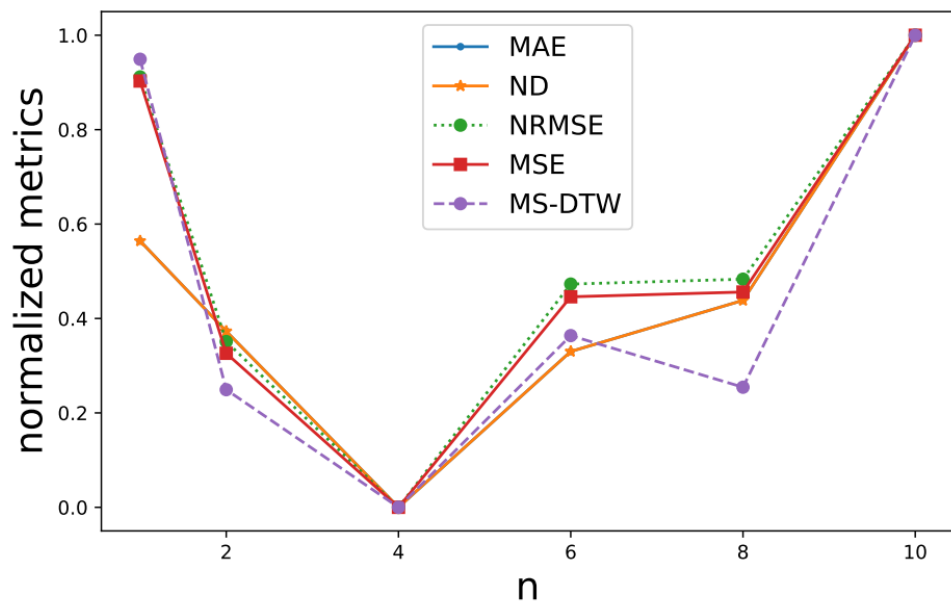


图 4-10: 不同尺度系数 n 时效果对比

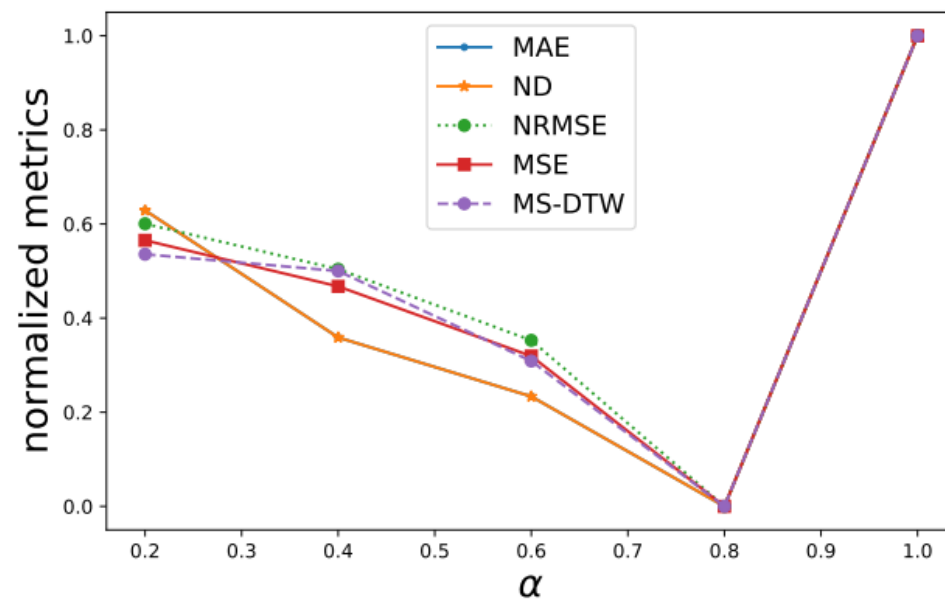


图 4-11: 不同  $\alpha$  时效果对比



## 实验验证

敏感性分析:  $\gamma$  对 MS-DTWI 的影响

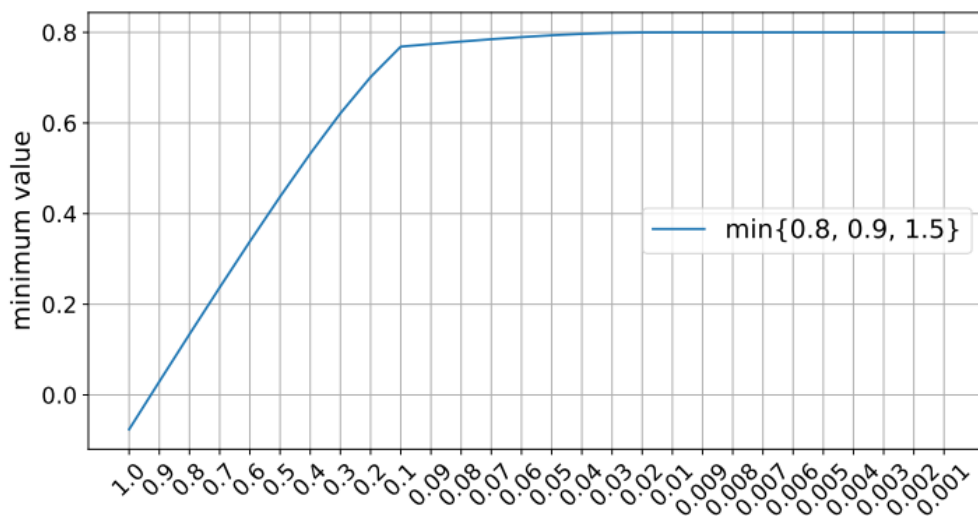
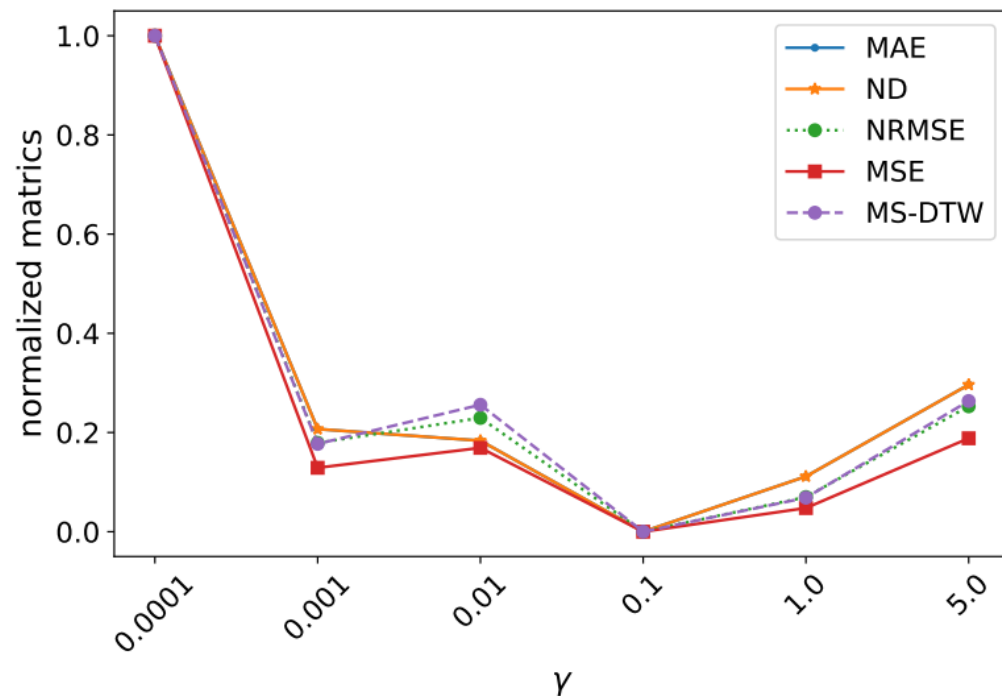


图 4-8: 不同  $\gamma$  时平滑最小化函数计算结果

平滑最小化函数:

$$\min^\gamma \{a_1, \dots, a_n\} := \begin{cases} \min_{i \leq n} a_i, & \gamma = 0 \\ -\gamma \log \sum_{i=1}^n e^{-a_i/\gamma}, & \gamma > 0 \end{cases}$$





## 第三部分

## Achievements

### 研究成果

- 商品管理和销量预测系统
- 研究生期间成果总结



## 商品管理和销量预测系统

- 基于 B/S 的系统结构
- 商品管理
  - 库存管理
  - 商店管理
  - 交易管理
- 分析和预测
  - 关联分析
  - 销量预测

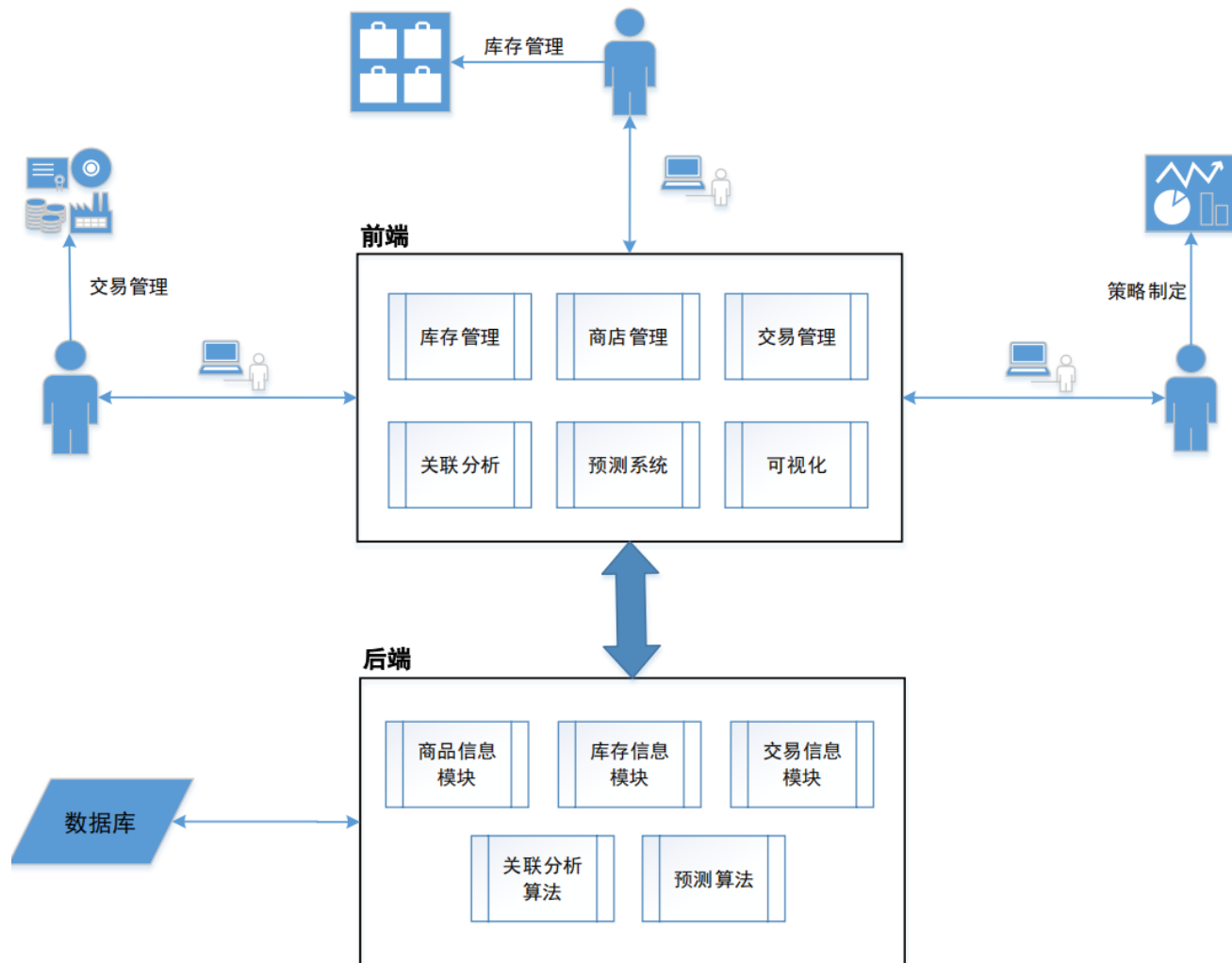


图 5-1: 整体系统架构图



实际界面展示

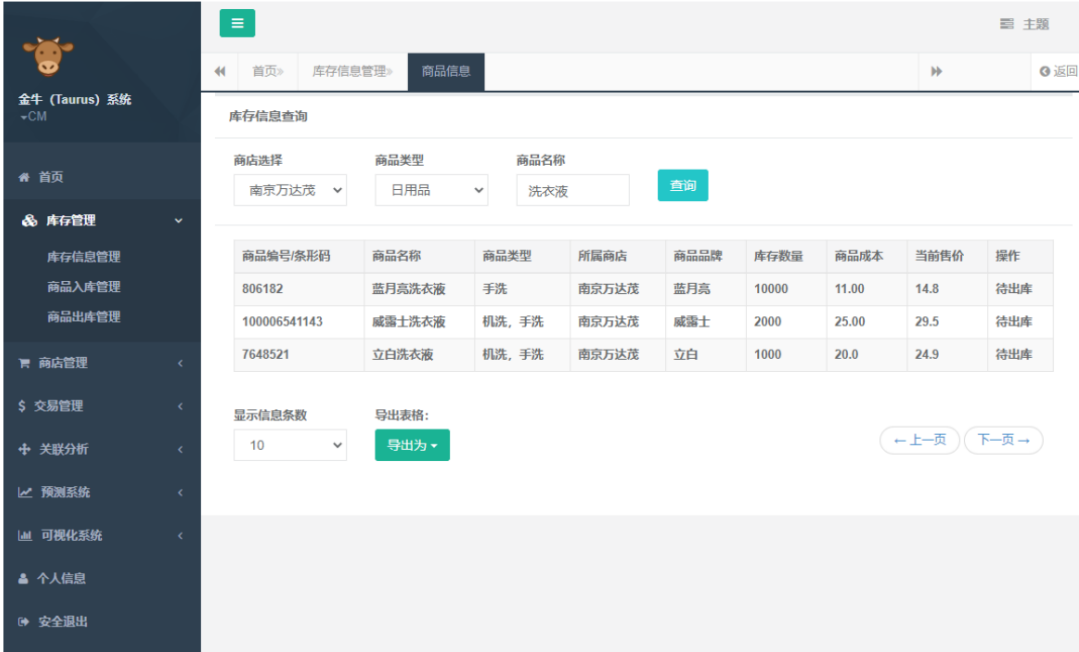


图 5-3: 库存管理

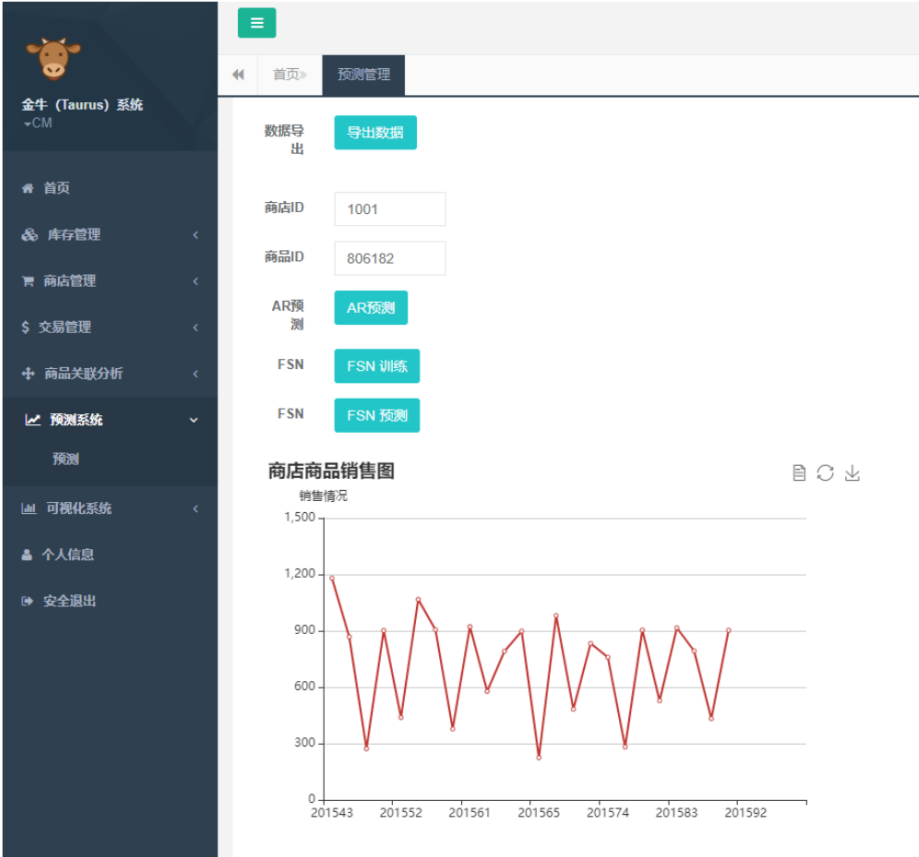


图 5-5: 销量预测





## 学术论文

- **Hongyan Hao**, Yan Wang, Jian Zhao, Furao Shen, “Temporal Convolutional Attentionbased Network For Sequence Modeling” in arXiv preprint arXiv:2002.12530, 2020.
- Yuanjie Yan, **Hongyan Hao**, Baile Xu, Jian Zhao, Furao Shen, “Image clustering via deep embedded dimensionality reduction and probability-based triplet loss” in IEEE Transactions on Image Processing 2020, 29: 5652-5661. (CCF, A)
- Siqiao Xue, Xiaoming Shi, **Hongyan Hao**, et al, “A Graph Regularized Point Process Model For Event Propagation Sequence” in 2021 International Joint Conference on Neural Networks (IJCNN, CCF C)

## 专利

- 申富饶, 郝鸿延, 张旭. ‘一种基于多生物特征的身份验证系统’ (201910933448)

## 项目

- 声纹验证与识别
- RINC 多生物特征身份验证系统



## 第三部分

## Summary 工作总结



### 前馈序列网络

- **时序卷积**建模序列相对位置属性，增大感受野
- **时序注意力**增加输入序列点特征间差异化程度
- **强化残差**防止梯度消失，为序列点特征加权

### 多尺度DTW损失函数

- **多尺度 DTW** 结合不同大小规整窗口的 DTW
- **TDI** 利用最短路径占路径空间面积比表征延时大小
- **MS-DTWI** 为两者加权结合后的损失函数

### 商品管理和销量预测系统

- **商品管理**包含库存管理、商店管理和交易管理
- **销量预测**包含商品关联分析和销售量预测



南京大學  
NANJING UNIVERSITY



谢谢！

