CondTSF: One-line Plugin of Dataset Condensation for Time Series Forecasting

Jianrong Ding*, Zhanyu Liu*, Guanjie Zheng†, Haiming Jin, Linghe Kong

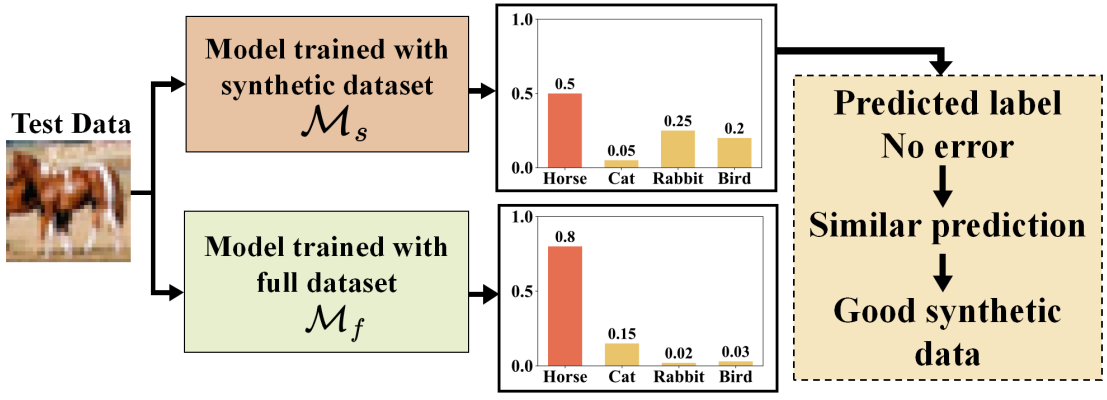
上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Background

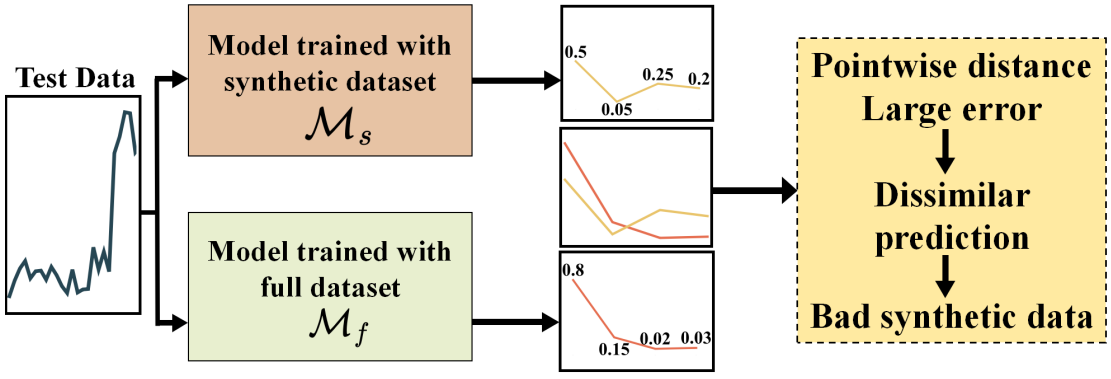
- Dataset Condensation
  - Use full dataset to train a model M_f, use a small synthetic dataset to train a model M_s.
  - Dataset condensation aims to generate a small synthetic dataset so that M_f and M_s have comparable performance on an unseen test dataset.

Background

- Motivation: Evaluation of classification tasks are distinct from time series forecasting.



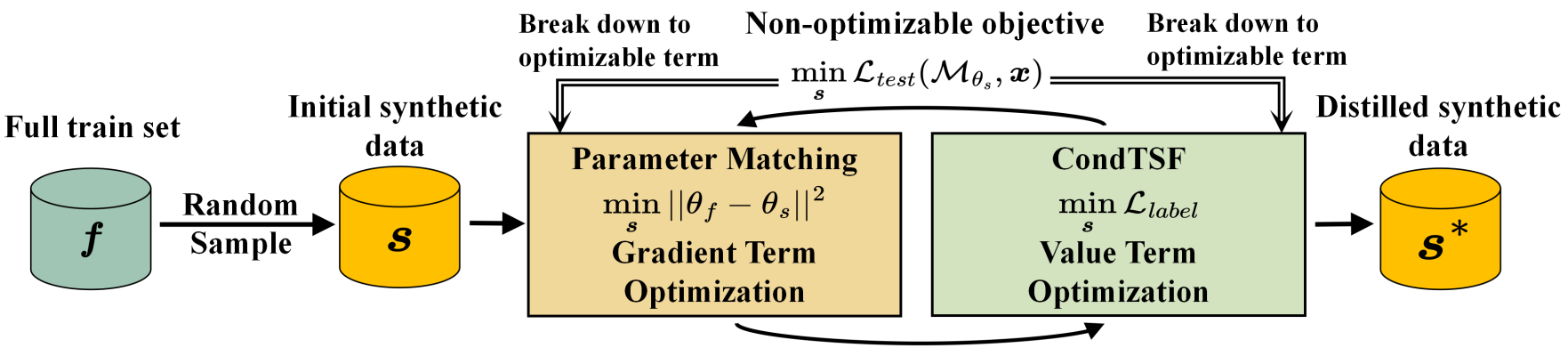
(a) Evaluation of synthetic data for image classification



(b) Evaluation of synthetic data for time series forecasting

Method Overview

- We propose a lightweight plugin module called **CondTSF** that can improve the performance of previous methods on time series forecasting tasks.



Decomposition of Optimization Objective

- Original optimization objective is to minimize the test error of the model M_s.

$$\min \mathcal{L}_{test}(\mathcal{M}_\theta, \boldsymbol{x}) \triangleq \sum_t ||\mathcal{M}_\theta(\boldsymbol{x}_{t:t+m}) - \boldsymbol{x}_{t+m:t+m+n}||^2$$

- Use Taylor Expansion to decompose the optimization objective.

$$\mathcal{L}_{test}(\mathcal{M}_{\theta_{s,test}}, \boldsymbol{x}) \leq \sum_t ||\boldsymbol{\epsilon}||^2 + \underbrace{||\mathcal{M}_{\theta_{s,test}}(\boldsymbol{s}_{t':t'+m}) - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m})||^2}_{\textbf{Value Term}}$$

$$+ \underbrace{||(\nabla \mathcal{M}_{\theta_{s,test}}(\boldsymbol{s}_{t':t'+m}) - \nabla \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m}))^\top (\boldsymbol{x}_{t:t+m} - \boldsymbol{s}_{t':t'+m})||^2}_{\textbf{Gradient Term}}$$

Value Term Optimization

- We reformulate the value term into an optimizable upper bound.

$$
\begin{aligned}
& ||\mathcal{M}_{\theta_{s,test}}(\boldsymbol{s}_{t':t'+m}) - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m})||^2 \\
= \; & ||\mathcal{M}_{\theta_{s,test}}(\boldsymbol{s}_{t':t'+m}) - \boldsymbol{s}_{t'+m:t'+m+n} \\
& \qquad + \boldsymbol{s}_{t'+m:t'+m+n} - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m})||^2 \\
\leq \; & \sum_{t'} ||\mathcal{M}_{\theta_{s,test}}(\boldsymbol{s}_{t':t'+m}) - \boldsymbol{s}_{t'+m:t'+m+n}||^2 \\
& \qquad + \sum_{t'} ||\boldsymbol{s}_{t'+m:t'+m+n} - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m})||^2 \\
\leq \; & 2 \cdot \sum_{t'} ||\mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}_{t':t'+m}) - \boldsymbol{s}_{t'+m:t'+m+n}||^2
\end{aligned}
$$

CondTSF

- We use a simple additive method to optimize the value term. In each iteration, with $0 < \beta < 1$ we have

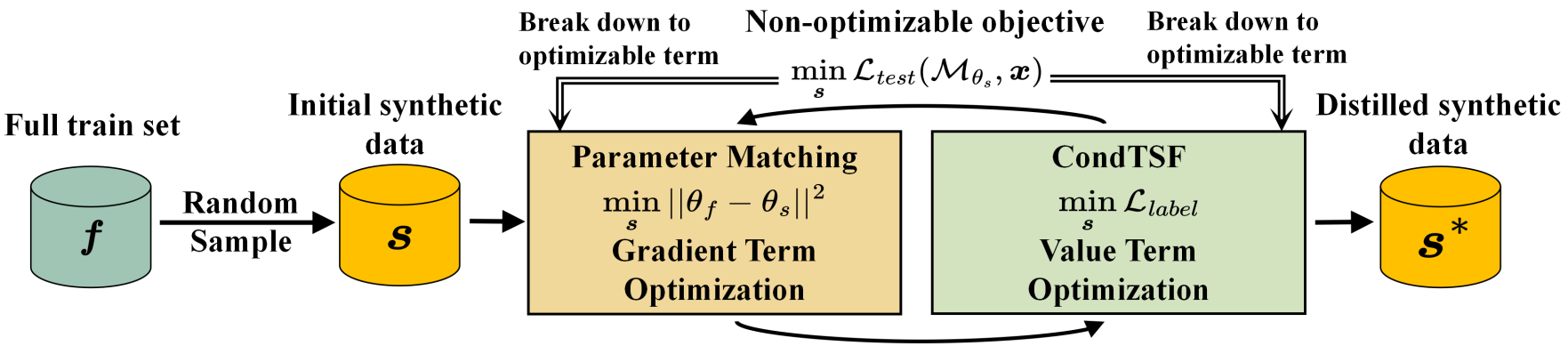
$$\boldsymbol{s}^{(i+1)}_{t'+m:t'+m+n} = (1-\beta) \cdot \boldsymbol{s}^{(i)}_{t'+m:t'+m+n} + \beta \cdot \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}^{(i)}_{t':t'+m})$$

- It can exponentially decrease the value term we want to optimize.

$$\begin{aligned}
\mathcal{L}^{(i+1)}_{label} &= \sum_{t'} ||\boldsymbol{s}^{(i+1)}_{t'+m:t'+m+n} - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}^{(i+1)}_{t':t'+m})||^2 \\
&= (1-\beta)^2 \sum_{t'} ||\boldsymbol{s}^{(i)}_{t'+m:t'+m+n} - \mathcal{M}_{\theta_{f,test}}(\boldsymbol{s}^{(i)}_{t':t'+m})||^2 \\
&= (1-\beta)^2 \mathcal{L}^{(i)}_{label}
\end{aligned}$$

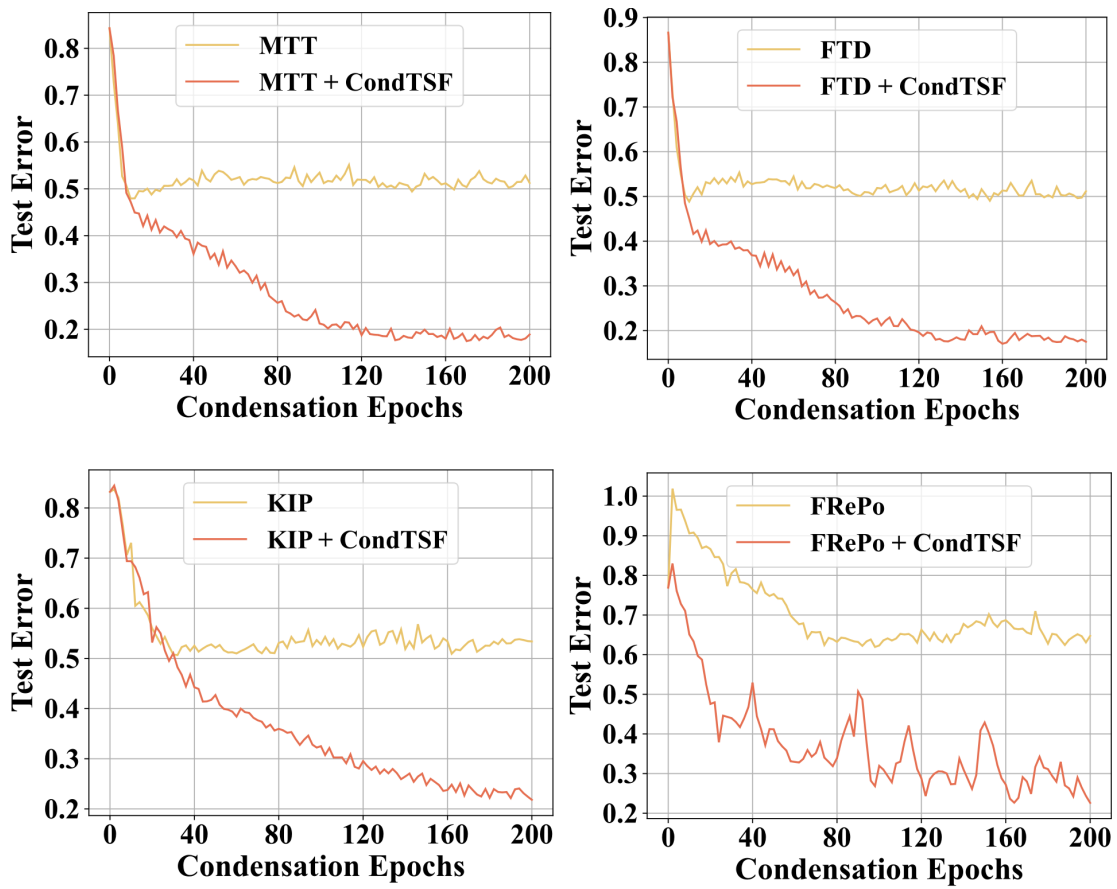
CondTSF

- CondTSF is used every few updates of previous methods.



Result

- By plugging CondTSF into previous dataset condensation methods, a huge improvement can be observed.



Thank you!

Presenter: Jianrong Ding
rafaelding@sjtu.edu.cn