Towards Diverse Perspective Learning with Selection over Multiple Temporal Poolings

Jihyeon Seong [1 *] Jungmin Kim [1 *] Jaesik Choi [1 2]

*: Equal Contribution [1]KAIST [2]INEEJI

Problem & Motivation

- Global pooling in CNN disregards temporal information in time series classification.
- Existing temporal poolings exhibit data dependency based on segmentation type.

To address these issues,

- Selection ensemble learning is applied to the pooling level.

Data-dependency on Temporal Poolings

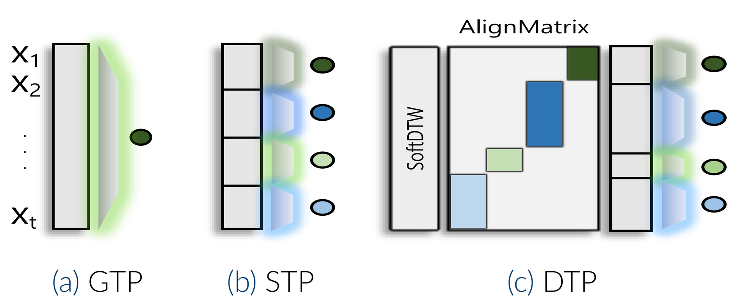

Figure 1. Different Perspectives of Temporal Poolings.

- Global View: Without segmentation, GTP best performs on time series data with dominant features.
- Uniform Local View: With segmenting in equal lengths, STP best performs on the recursive patterns.
- Dynamic Local View: With segmenting in dynamic lengths based on DTW metric, DTP best performs on complex and noisy patterns.

Multiple Choice Learning

MCL generates M solutions $\hat{Y}_i = (\hat{y}_i^1, ..., \hat{y}_i^M)$, and learns a mapping $g : \mathcal{X} \rightarrow \mathcal{Y}^M$ that minimizes oracle loss $min_m \ell(y_i, \hat{y}_i^m)$. The ensemble mapping function g consists of multiple predictors, $g(x) = \{f_1(x), f_2(x), ..., f_M(x)\}$.

1. **Ambiguous evidence** refers to situations with insufficient information to make a definitive prediction.
   - Presenting a small set of reasonable possibilities can alleviate the over-confidence problem of deep learning rather than striving for a single accurate answer.
2. **Bias towards the model** indicates the model's tendency to learn a mode-seeking behavior to reduce the expected loss across the entire dataset.
   - When only a single prediction exists, the model learns to minimize the average error.
3. **MCL** generates multiple predictions, allowing classifiers to cover the lower-density regions of the solution space without sacrificing performance on the high-density regions.
   - With multiple predictions from multiple classifiers, MCL solves 'ambiguous evidence' and 'bias toward the mode' problems.
4. **SoM-TP**, motivated by MCL, achieves selection ensemble learning at the pooling level.
   - To achieve MCL in the pooling level, SoM-TP dynamically selects optimal pooling by attention mechanism and achieves effective optimization through perspective loss.
   - In a single classifier, SoM-TP reflects all temporal poolings with $\mathcal{O}(N)$ optimization cost, while MCL has $\mathcal{O}(N^2)$.

Contributions

- Found data dependency in temporal poolings arising from distinct perspectives.
- Propose SoM-TP, leveraging diverse perspectives through a selection ensemble.
- Employ an attention mechanism to enable non-iterative ensemble learning.
- Define DPLN and perspective loss as a regularizer for effective optimization.

Selection over Multiple Temporal Poolings

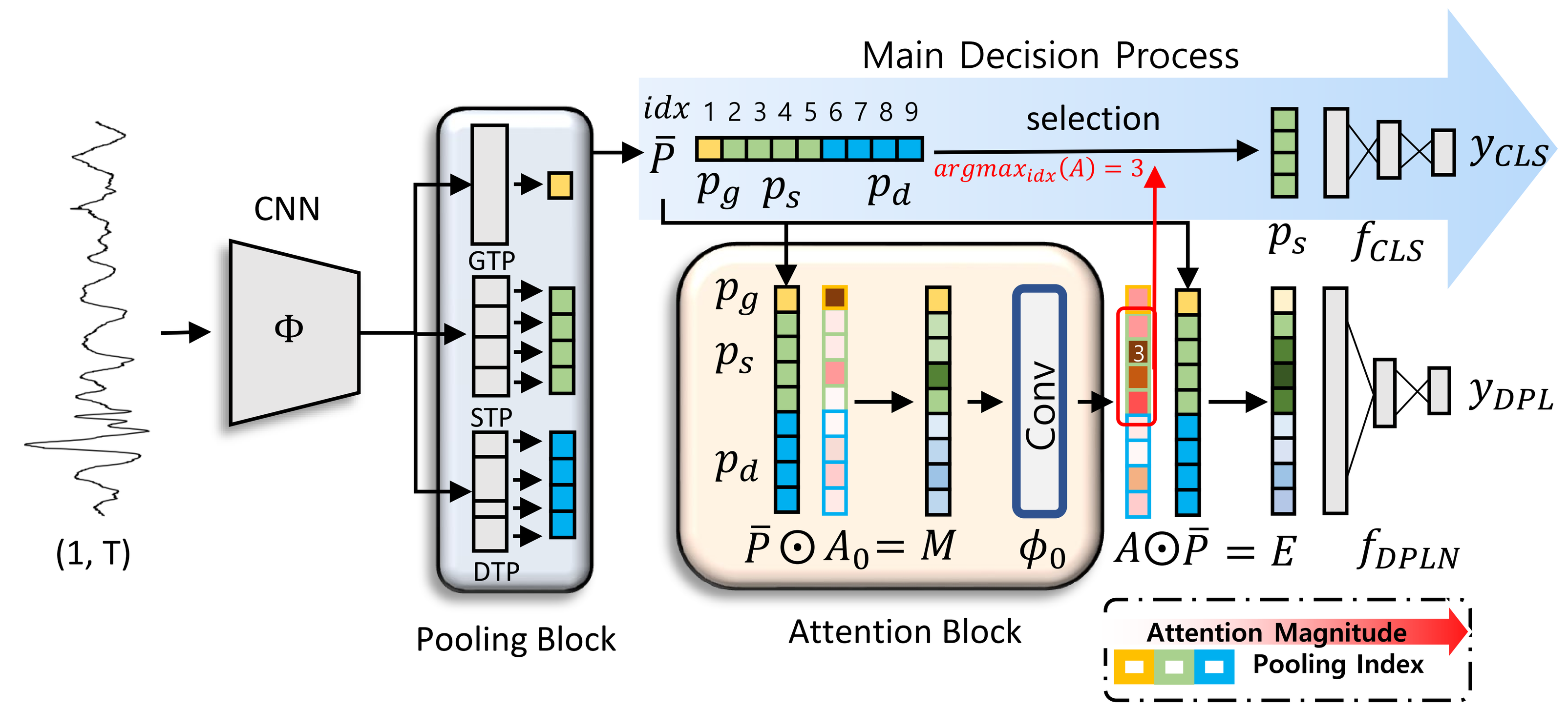

Figure 2. SoM-TP Architecture

- From heterogeneous temporal poolings, all pooling outputs are generated.
- Attention block automatically selects optimal pooling at batch level.
- Diverse Perspective Learning Network (DPLN) considers all pooling perspectives for effective selection ensemble learning.

DPL Attention

- Weight vector A_0 cumulatively learns the overall characteristics of dataset.
- A convolutional layer ϕ_0 captures batch-level characteristics.

DPLN & Perspective Loss

- DPLN is a regularization network to accelerate diverse perspective learning from all poolings through perspective loss.
- Perspective loss is a cost function to maximize utilization of DPLN, linking the two networks by KL term.

$$KL(y_{CLS}, y_{DPL}) = y_{DPL} \cdot log\frac{y_{DPL}}{y_{CLS}},$$

$$\mathcal{L}_{DPLN} = -\frac{1}{t}\sum_{n=1}^{t} logP(y = y_n|\mathbf{X}_n),$$

$$\mathcal{L}_{perspective} = KL(y_{CLS}, y_{DPL}) + \mathcal{L}_{DPLN}.$$

Cost Function & Optimization

$$\mathcal{L}_{CLS} = -\frac{1}{t}\sum_{n=1}^{t} logP(y = y_n|\mathbf{X}_n),$$

$$\mathcal{L}_{cost} = \mathcal{L}_{CLS} + \lambda \cdot \mathcal{L}_{perspective},$$

$$\mathcal{L}_{attn} = -y_{CLS} \cdot y_{DPL},$$

$$\mathbf{A_0} \leftarrow \mathbf{A_0} - \eta \cdot \partial\mathcal{L}_{attn}/\partial\mathbf{A_0},$$

$$\mathcal{W}_\Phi \leftarrow \mathcal{W}_\Phi - \eta \cdot \partial\mathcal{L}_{cost}/\partial\mathcal{W}_\Phi,$$

$$\mathbf{W} \leftarrow \mathbf{W} - \eta \cdot \partial\mathcal{L}_{cost}/\partial\mathbf{W}.$$

Diverse Perspective Learning Results

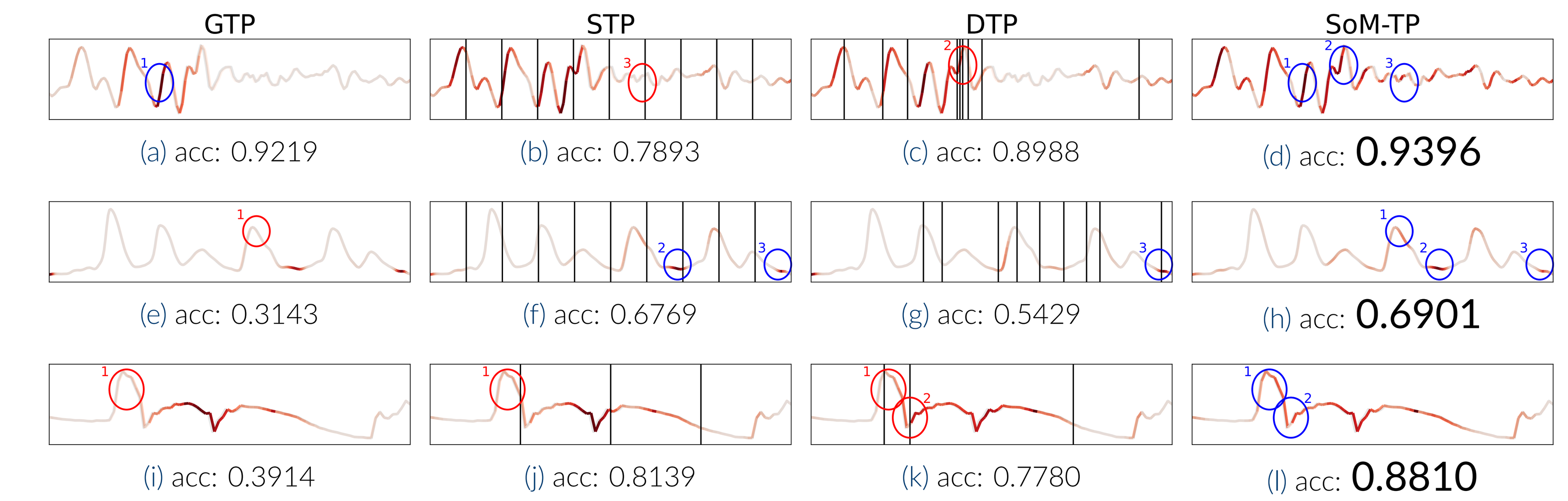

Figure 3. Layer-wise Relevance Propagation (LRP): Fixed vs Diverse Perspective Learning

- Utilizing the input attribution method, LRP, the pooling perspective can be analyzed.
- The blue circle indicates well-captured points, while the red circle shows the lost points that should be captured for enhanced performance.
- Due to the unique characteristics of time series data, temporal poolings exhibit drawbacks with single perspectives based on segmentation type.
- SoM-TP addresses the limitations of a single perspective by incorporating diverse perspective learning while also capturing the benefits from individual perspectives.

Dynamic Selection Results

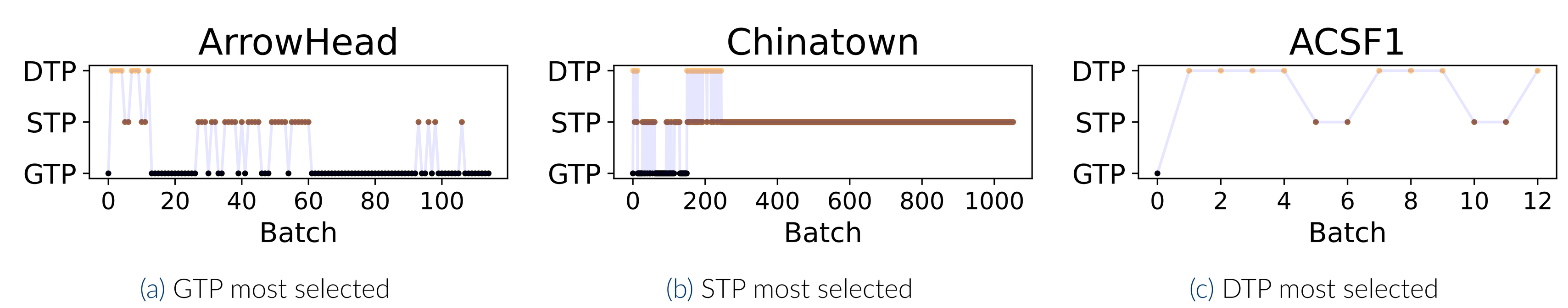

Figure 4. Dynamic Pooling Selection in SoM-TP.

Comparison Results with Advanced Methods

Methods	UCR Baseline wins	UCR SoM-TP (Ours) wins	UCR Tie	UCR Average Rank	UEA Baseline wins	UEA SoM-TP (Ours) wins	UEA Tie	UEA Average Rank
Vanilla-Transformer	10	**99**	4	5.2	3	**18**	1	4.4
TCN	28	**80**	5	4.0	4	**17**	1	4.3
TST	32	**78**	3.8	3.1	6	**15**	1	3.6
ConvTran	35	**71**	7	3.1	8	**13**	1	3.1
MLSTM-FCN	46	**60**	6	2.6	6	**12**	4	3.2
SoM-TP - MAX	-	-	-	**2.5**	-	-	-	**2.5**

Methods	UCR Acc	UCR F1-score	UCR ROC AUC	UCR PR AUC	UEA Acc	UEA F1-score	UEA ROC AUC	UEA PR AUC	Complexity
ROCKET	<u>0.7718</u>	<u>0.7478</u>	0.8899	0.7841	0.6785	<u>0.6592</u>	0.7926	0.6940	$\mathcal{O}(N^2)$
InceptionTime	0.7713	0.7455	<u>0.9056</u>	<u>0.8164</u>	0.6612	0.6360	0.7984	<u>0.7106</u>	
OS-CNN	0.7663	0.7324	0.9005	0.8139	<u>0.6808</u>	0.6547	**0.8118**	**0.7137**	
DSN	0.7488	0.7230	0.8838	0.7968	0.5648	0.5433	0.7575	0.6265	
SoM-TP - MAX	**0.7773**	**0.7489**	**0.9182**	**0.8261**	**0.6920**	**0.6621**	<u>0.8105</u>	0.7099	$\mathcal{O}(N)$

Table 1. Performance Comparison with Advanced Methods.

jihyeon.seong@kaist.ac.kr, aldirl7@kaist.ac.kr, jaesik.choi@kaist.ac.kr