Enhancing Time Series Classification: An In-Depth Analysis of Interval-Based Methods and Their Methodological Components

Student:
Mabrouka Salmi

Supervisor:
José María Luna Ariza

Cosupervisor:
Sebastián Ventura Soto

Master en ICIOT, Escuela Politecnica Superior de Cordoba
University of Cordoba

26 de Septiembre de 2024

Content

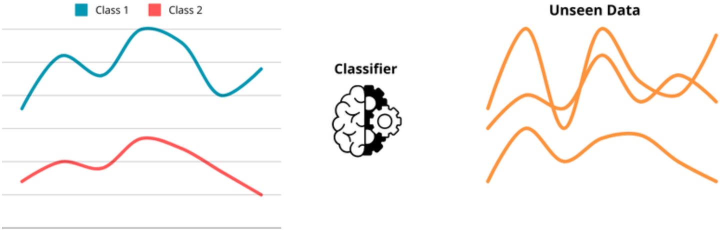
Time series classification and its approaches

Definition

A univariate time series **x** is an ordered real-valued vector with m data points (observations) $(x_1, x_2, ..., x_m)$ taken over fixed equal-spaced intervals

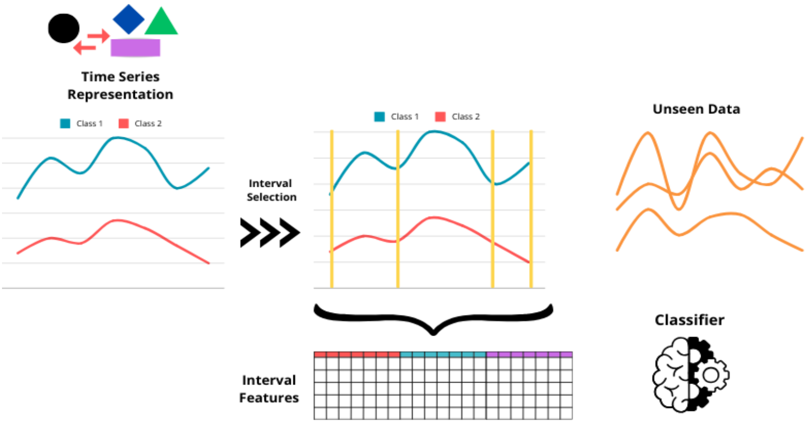
- an instance is a pair $(\mathbf{x}, y)$ that is the time series **x** and its label y which is a discrete value c in C $(C = \{c_1, c_2, ..., c_k\})$

Time series classification and its approaches

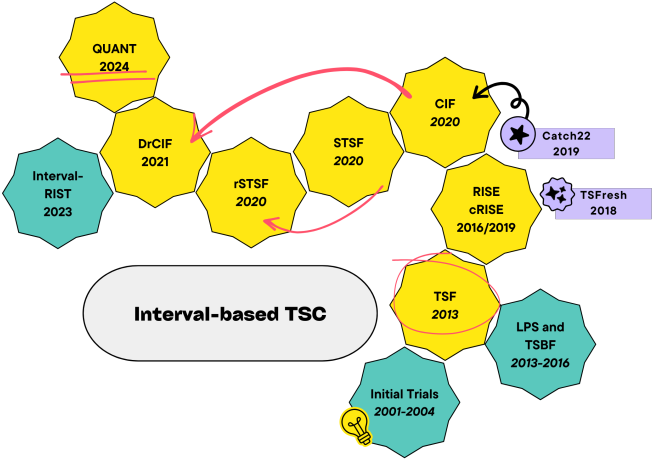


Interval-based | Feature-based | Distance-based | Deeplearning-based | ...

Interval-based time series classification



The progress of the approach



Challenges

- The enhancement in the performance is unexplained

Challenges

- The enhancement in the performance is unexplained
- Error-and-trial processes/ heuristics for the methodological components

Challenges

- The enhancement in the performance is unexplained
- Error-and-trial processes/ heuristics for the methodological components
- The lack of critical review of the methodological components of interval-based methods is evident

Challenges

- The enhancement in the performance is unexplained
- Error-and-trial processes/ heuristics for the methodological components
- The lack of critical review of the methodological components of interval-based methods is evident
- The impact of the fundamental elements of the method on performance is undiscovered

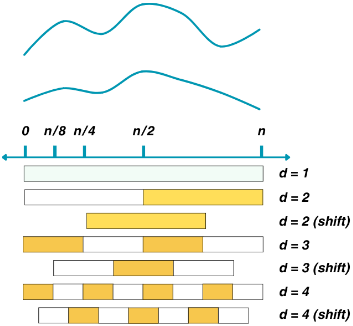
Objectives

How the constituent methodical elements of interval-based methods could affect the time series classification, focusing on:

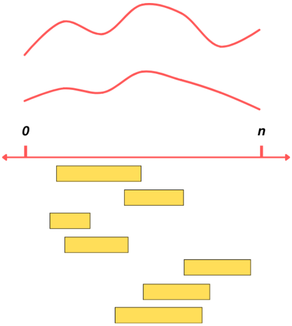
- Intervals (selection)
- Interval features (feature extraction from intervals)
- The time series representation (spectral, derivatives, etc.)

Content

Interval selection

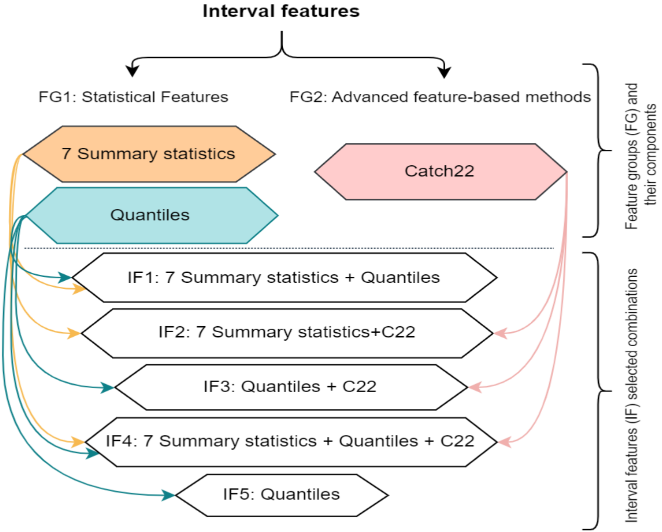


Fixed dyadic Intervals
with depth = 4 and shifted intervals

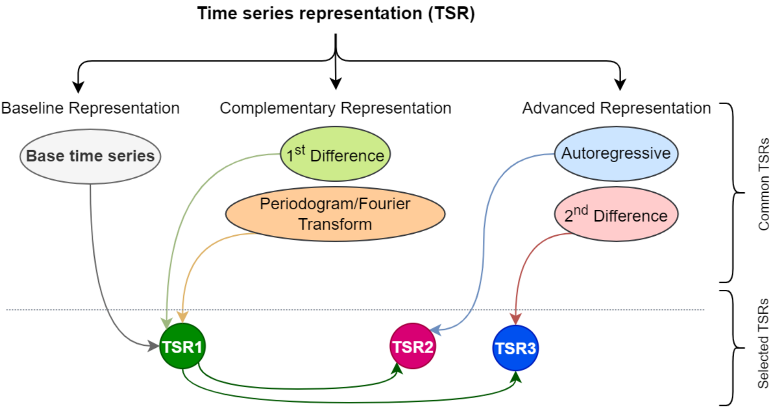


Random interval selection
with random position and length

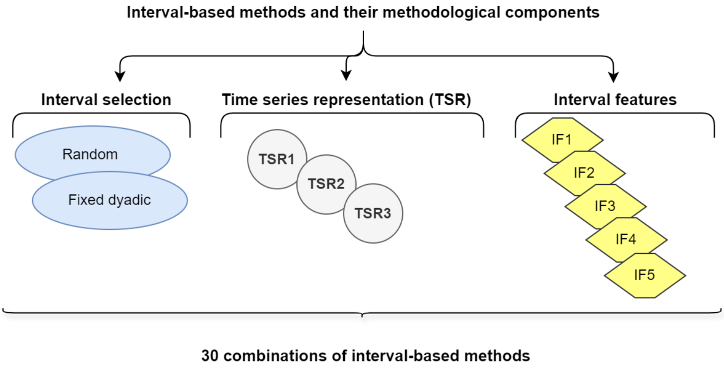
Interval features



Time series representation



Framework of the study



△ **Datasets:** The UCR_{112} Time Series Classification Archive.

- **Datasets:** The UCR_{112} Time Series Classification Archive.
- **Classification Algorithm:** The Extra Trees algorithm with default hyperparameters.

- **Datasets:** The UCR_{112} Time Series Classification Archive.
- **Classification Algorithm:** The Extra Trees algorithm with default hyperparameters.
- **Evaluation**

- **Datasets:** The UCR_{112} Time Series Classification Archive.
- **Classification Algorithm:** The Extra Trees algorithm with default hyperparameters.
- **Evaluation**
  - Training time and multi-metric evaluation (accuracy-related)

- **Datasets:** The UCR_{112} Time Series Classification Archive.
- **Classification Algorithm:** The Extra Trees algorithm with default hyperparameters.
- **Evaluation**
  - Training time and multi-metric evaluation (accuracy-related)
  - Statistical tests and the critical difference plot.

- **Datasets:** The UCR_{112} Time Series Classification Archive.
- **Classification Algorithm:** The Extra Trees algorithm with default hyperparameters.
- **Evaluation**
  - Training time and multi-metric evaluation (accuracy-related)
  - Statistical tests and the critical difference plot.
  - Analysis of variables: study of the impact of each variable in isolation from the others.

Content

1. Comparison of all models
2. Comparison of the best models to the state-of-the-art (SOTA)

Comparison of all models: Overall difference with Friedman test

Table: Friedman Test Statistics and p-values for Different Metrics

Metric	**Friedman Stat**	**p-value**
Accuracy	692.53	3.59×10^{-127}
F1	672.44	5.57×10^{-123}
AUC	823.18	1.56×10^{-154}
Balanced Accuracy	630.88	2.50×10^{-114}
Training time	2420.63	0

- The difference is statistically significant in the performance and training time of all models

Comparison of all models: Accuracy-related metrics

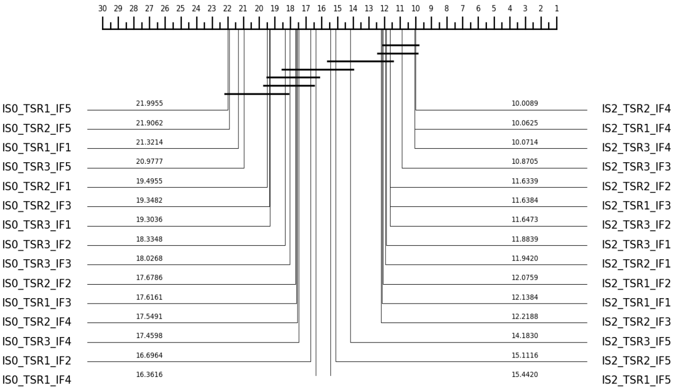


Figure: Critical difference diagram of model accuracy ranks on the UCR_{112} archive datasets

Comparison of all models: Accuracy-related metrics

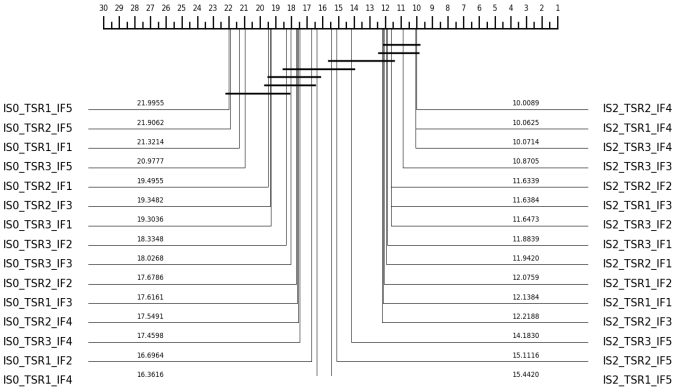


Figure: Critical difference diagram of model accuracy ranks on the UCR_{112} archive datasets

- The advance of the models using the fixed dyadic intervals method (IS2)

Comparison of all models: Accuracy-related metrics

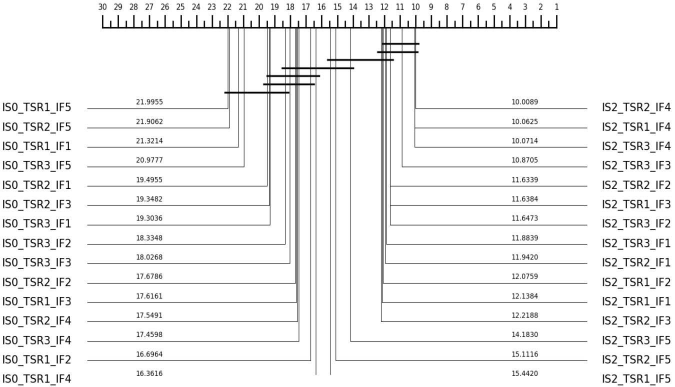


Figure: Critical difference diagram of model accuracy ranks on the UCR_{112} archive datasets

- The models with IF1, IF2, and IF3 usually perform well when combined with the fixed dyadic interval selection

Comparison of all models: Training time

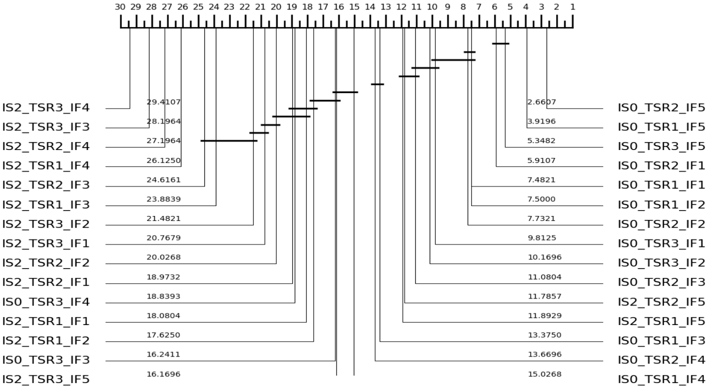


Figure: Critical difference diagram of model training time ranks on the UCR_{112} archive datasets

- The high training time of fixed dyadic intervals-based methods compared to the random intervals-based models.

Comparison of the best models to the SOTA: Overall difference with Friedman test

Table: Friedman statistics and p-values for different metrics

Metric	**Friedman stat**	**p-value**
Accuracy	14.480	0.025
Balanced accuracy	11.819	0.066
F1	13.151	0.041
AUC	8.441	0.208
Training time	531.478	1.387×10^{-111}

- The SOTA and the three best models were statistically different (5%) in terms of accuracy, F1, and training time

Comparison of the best models to the SOTA: Accuracy

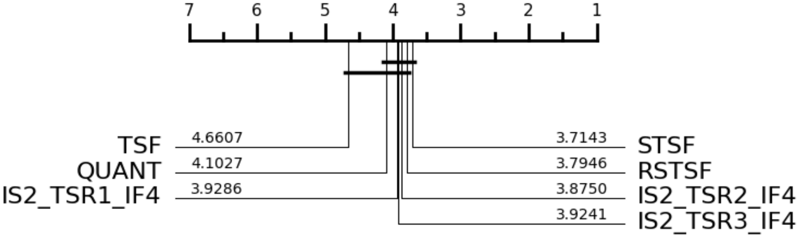


Figure: Critical difference diagram of SOTA models' accuracy ranks on the UCR_{112} archive datasets

- Comparable results, statistically, with the SOTA models

1. The impact of interval features
2. The impact of interval selection
3. The impact of time series representation

The impact of interval features: Overall difference with Friedman test

Table: Friedman Test Statistics and p-values for Different Metrics in the IF groups

Metric	Friedman stat	p-value
Accuracy	88.46	2.79×10^{-18}
Balanced accuracy	64.71	2.97×10^{-13}
AUC	108.55	1.48×10^{-22}
F1	86.09	8.89×10^{-18}
Training time	366.94	3.85×10^{-78}

- The models based on different interval features (IF1 to IF5) demonstrate a statistically significant difference.

The impact of interval features: Accuracy-related metrics

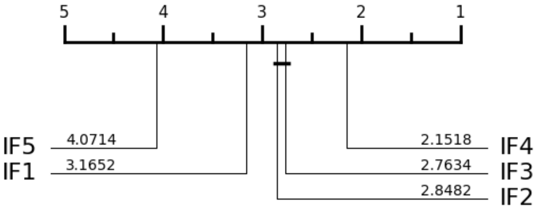


Figure: Critical difference diagram of averaged accuracy ranks by the IF variable on the UCR_{112} archive datasets

- The high performance of models based on IF4 (catch22, 7 summary stats, quantiles)
- Models based on IF3 (quantiles and catch22) and IF2 (7 summary statistics and catch22) show similar results statistically.

The impact of interval features: Training time

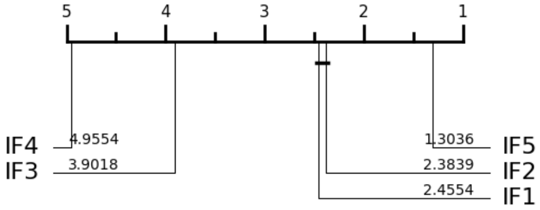


Figure: Critical difference diagram of averaged training time ranks by the IF variable on the UCR_{112} archive datasets

- The models based on IF5 (quantiles) are far more efficient in terms of training time

The impact of the interval selection: Wilcoxon test

Table: Wilcoxon Test Results for Different Metrics in the IS-groups

Metric	Wilcoxon Statistic	p-value
Averaged Accuracy	2792	0.280
Averaged F1	2769	0.251
Averaged AUC	2788	0.275
Averaged Balanced Accuracy	2799	0.289
Averaged Training time	2236	0.007

- Contrary to what has been observed in the performance comparison of all models!

On the main effect of IS/IF and their interaction effect: Friedman test

Table: Friedman Test Results for Averaged Metrics from the IS-IF groups

Metric	Friedman_stat	p-value
Averaged Accuracy	55.367	1.037×10^{-8}
Averaged F1	54.811	1.323×10^{-8}
Averaged AUC	62.994	3.537×10^{-10}
Averaged Balanced Accuracy	39.470	9.476×10^{-6}
Averaged Training time	269.010	9.582×10^{-53}

- The models based on ISk_IFj (with $j \in \{1, 2, 3, 4, 5\}$ and $k \in \{2, 0\}$) demonstrate a significant difference statistically in terms of training time and accuracy-related metrics

On the main effect of IS/IF and their interaction effect: Accuracy

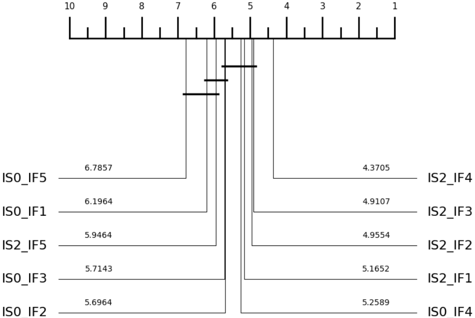


Figure: Critical difference diagram of averaged accuracy ranks by the IS-IF combination on the UCR_{112} archive datasets

- IF4 (quantiles, c22, and 7 summary stats) improves the random interval selection

The impact of time series representation: Friedman test

Table: Friedman Test Results for Different Metrics in the TSR groups

Metric	Friedman Statistic	p-value
Averaged Accuracy	0.125	0.939
Averaged Balanced Accuracy	0.696	0.706
Averaged AUC	0.071	0.965
Averaged F1	0.054	0.974
Averaged Training time	4.625	0.099

- The difference in time series representation doesn't statistically affect the performance and training time of the models

Large datasets: Accuracy

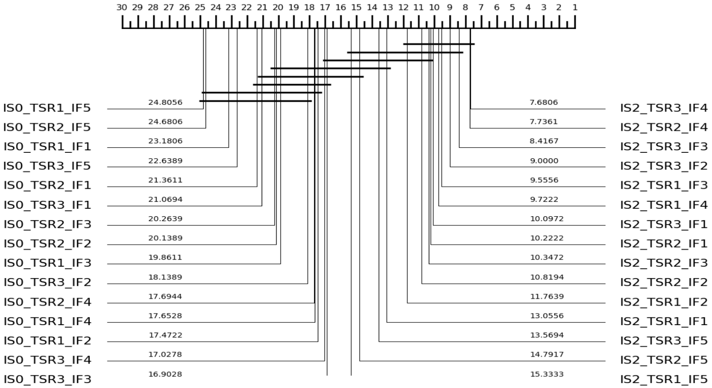


Figure: Critical difference diagram of model accuracy ranks on large datasets from the UCR_{112} archive

- The models IS2_IF3 (3^{rd}) compared to those IS2_IF1 (7^{th})

Small datasets: Accuracy

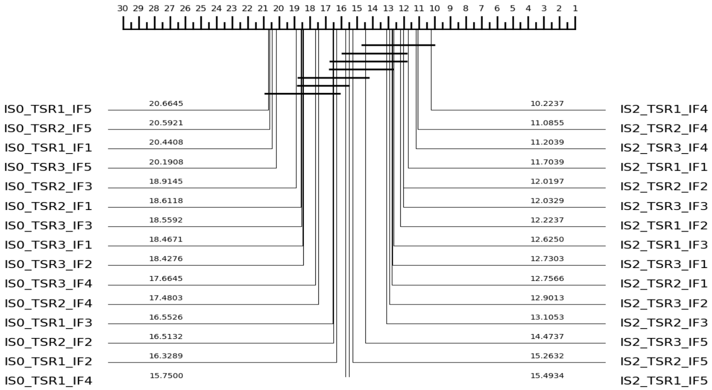


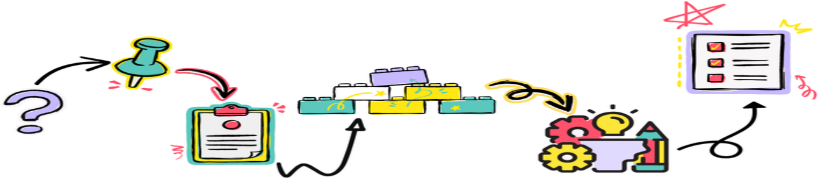
Figure: Critical difference diagram of model accuracy ranks on small datasets from the UCR_{112} archive

- Shift in their accuracy ranks

Content

Key findings

- The Interval feature effect
- Optimal feature combinations
- Performance of Interval Selection Methods
- Interaction effects of interval selection and feature
- Model sensitivity to dataset size



Limitations and future research directions

Limitations

- Exclusion of supervised interval selection
- TSFresh and advanced features integration in interval-based approach

Future research directions

- Interval selection approaches
- Other aspects in interval-based TSC
- Representative sample of component methods