Incremental and Adaptive Feature Exploration over Time Series Stream

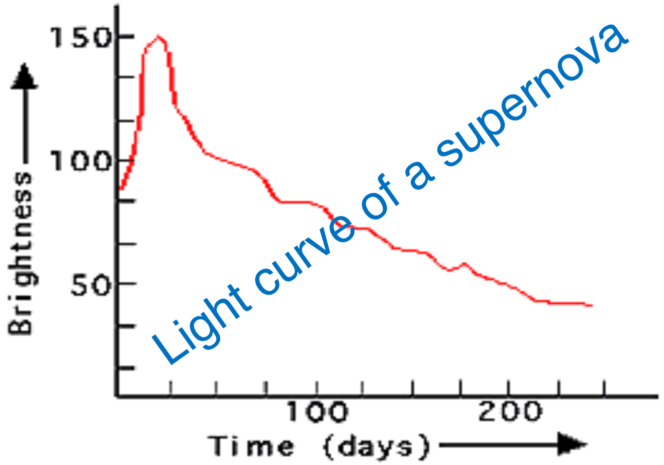
Jingwei ZUO, Karine ZEITOUNI, Yehia TAHER

IEEE Big Data 2019 - December 2019

Co-funded by
the European Union

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under Marie Sklodowska- Curie grant agreement No 777695
Disclaimer: This work reflects only the author's view and that the EU Agency is not responsible for any use that may be made of the information it contains.

1. Context & Background



Knowledge Discovery in Time Series (TS)

- Motif Matching
- (Frequent) Pattern Discovery
- Anomaly Detection
- Time Series Classification/Clustering, etc.

Knowledge Discovery in Data Streams (DS) & Challenges[1]

- Infinite Length ☞ *Memory Cost*
- Feature Evolution ☞ *Incrementality of learning model*
- Concept Drift ☞ *Adaptive adjustment of learning model*
- Concept Evolution ☞ *Emergence of new classes*

1. M. M. Masud, Q. Chen, J. Gao, L. Khan, J. Han, B. Thuraisingham, "Classification and Novel Class Detection of Data Streams in a Dynamic Feature Space", ECML-PKDD'10

1. Context & New Mining directions

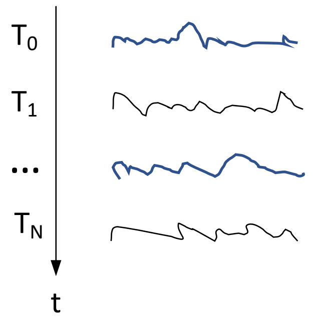
Time Series + Data Stream = ?

A combination which covers more practical scenarios !

1. Context & Definitions

Time Series Stream S_{TS}

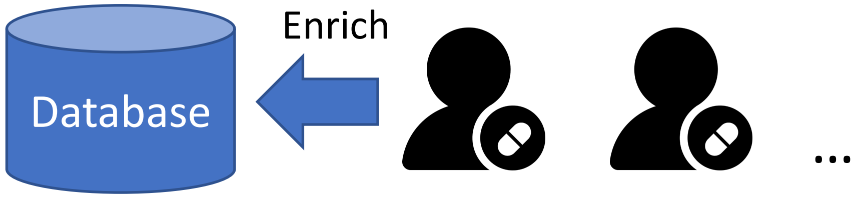
- A continuous input data stream where each instance is a Time Series: $S_{TS} = (T_1, T_2, \dots, T_N)$, notice that N increases with each new time-tick.



Our context

Use Cases:

- Medical domain
  Patient TS database is getting bigger and bigger
- Astronomy discovery
  New detection of the star light curves, update the features inside the Learning Model

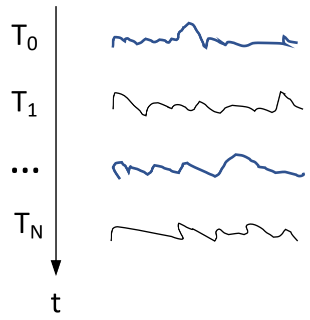


Health Care

1. Context & Definitions

Time Series Stream S_{TS}

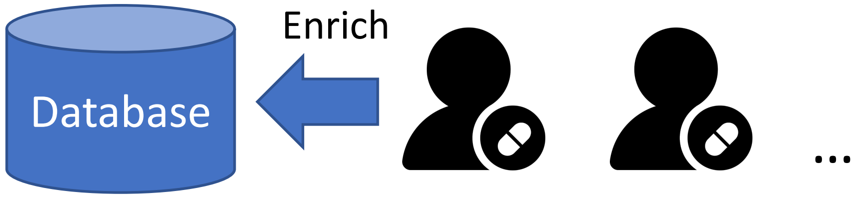
- A continuous input data stream where each instance is a Time Series: $S_{TS} = (T_1, T_2, \dots, T_N)$, notice that N increases with each new time-tick.



Our context

Use Cases:

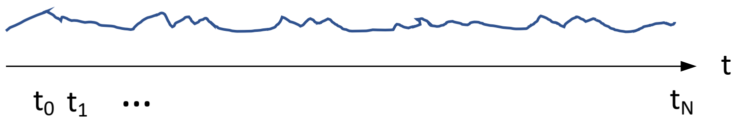
- Medical domain
  Patient TS database is getting bigger and bigger
- Astronomy discovery
  New detection of the star light curves, update the features inside the Learning Model

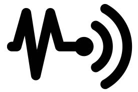


Health Care

Streaming Time Series S

- A continuous input data stream where each instance is a real-valued data: $S = (t_1, t_2, \dots, t_N)$





Smart City Sensor

Out of scope

1. Objectives of Time Series Stream Mining

Time Series Mining

- Real valued data with high temporal dependence
- Feature Representation is the essential part in the mining process

Classic Data Stream Mining

- Row or vector data with multiple attributes without assumption of temporal dependence

Interpretable, **Incremental**, **Adaptive** features in streaming context

1. Filling the Gap between TS & DS Mining

Time Series Feature Representations	
R1	Global features of entire series (1-NN)
R2	Summary statistics of intervals/sub-series
R3	Motif (frequent patterns)
R4	Shapelet [1] (shape-based features)
…	

Data Stream Challenges [2]	
C1	Infinite Length
C2	Feature Evolution
C3	Concept Drift
C4	Concept Evolution

R4 + {C1, C2, C3}

1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009
2. M. M. Masud, Q. Chen, J. Gao, L. Khan, J. Han, and B. Thuraisingham, "*Classification and Novel Class Detection of Data Streams in a Dynamic Feature Space*", ECML-PKDD'10

Why Shapelet[1]?

Definition

- A representative shape in Time Series which is capable of distinguishing one class from the others

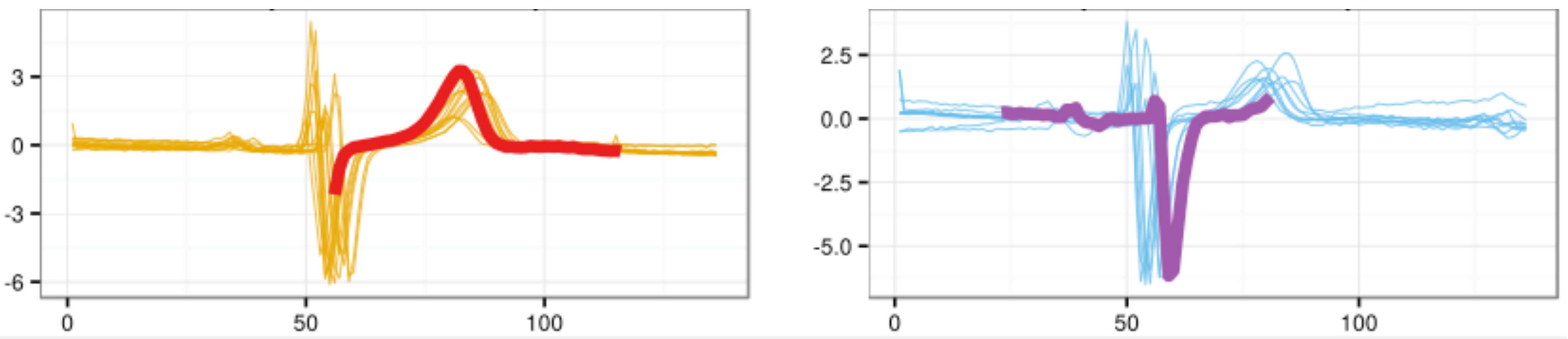
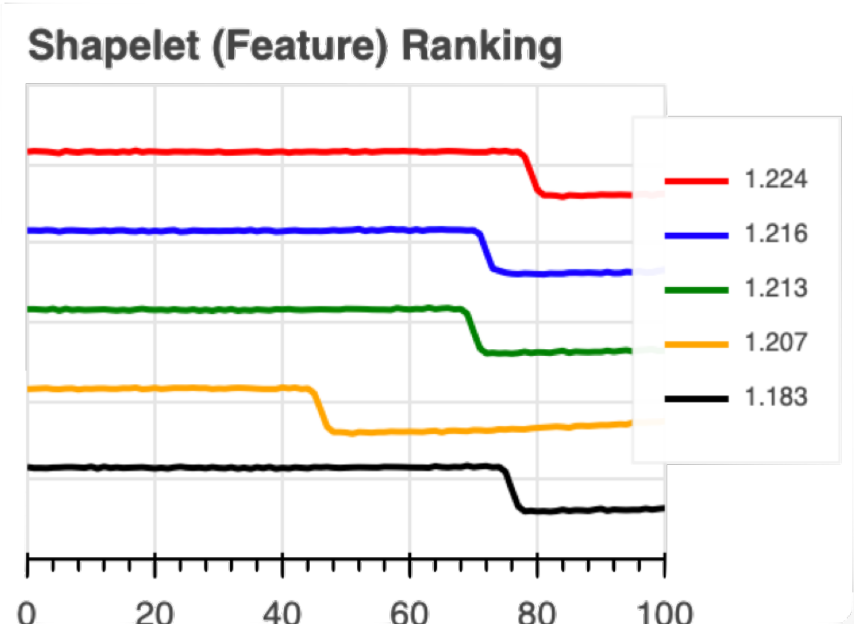


Figure 1[2]: Two classes from the "ECGFiveDays" dataset and the *best representative patterns (Shapelets)*

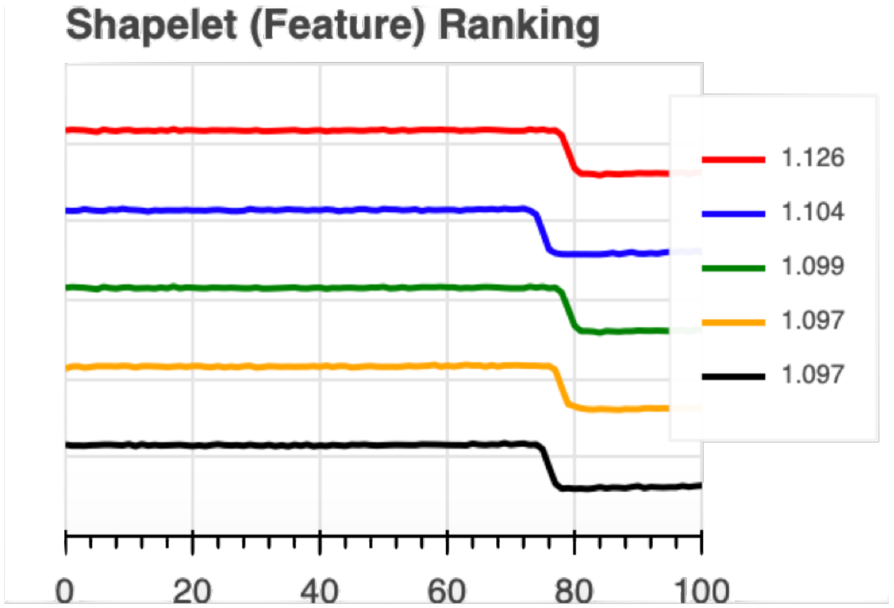
1. L. Ye and E. Keogh. "Time series shapelets: A New Primitive for Data Mining." In Proc. SIGKDD 2009
2. X. Wang et al. "RPM: Representative Pattern Mining for Efficient Time Series Classification." , In Proc. EDBT'16

Feature Evolution over Shapelets

Dataset *Trace*[1] (class 2)



Time stamp = 20

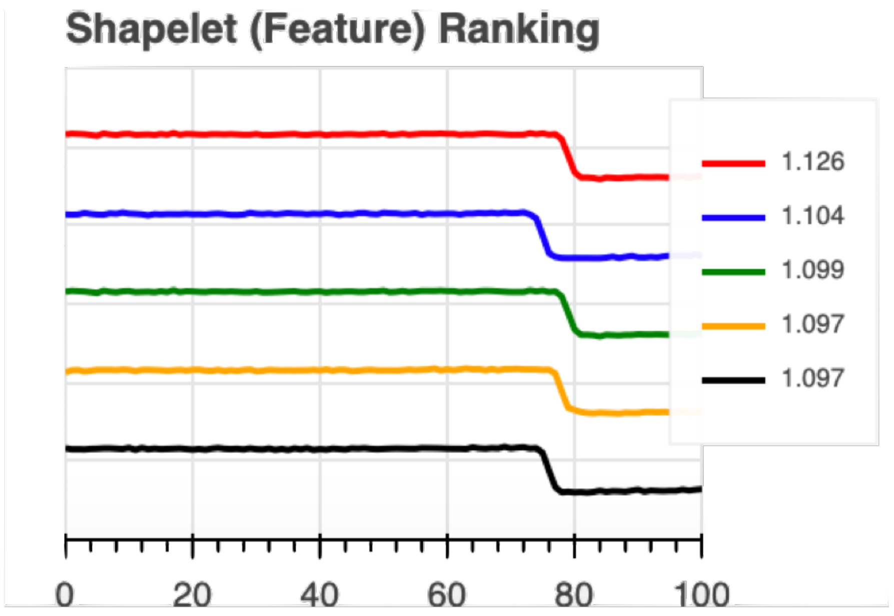


Time stamp = 100

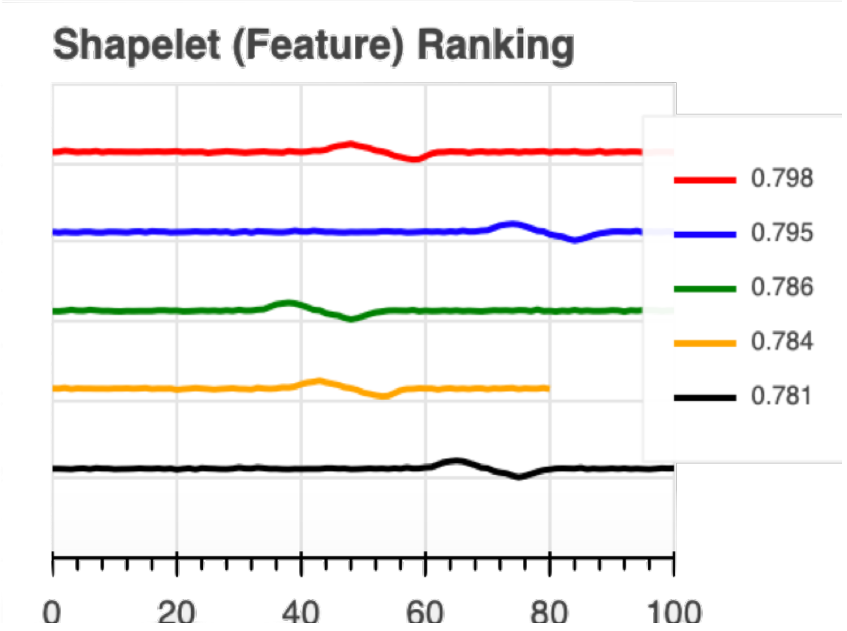
1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

Concept Drift over Shapelets

Dataset *Trace*[1] (class 2)



Time stamp = 100



Time stamp = 200

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

1. Problems to tackle

- **Low Scalability** and **Incrementality** of Shapelet approaches
- Classic **Shapelet Evaluation** is not suitable in streaming context
- **Concept Drift detection** in TS Stream model
- **Memory cost** of infinite TS instances

2. Preliminaries

Distance Profile & Matrix Profile[1]

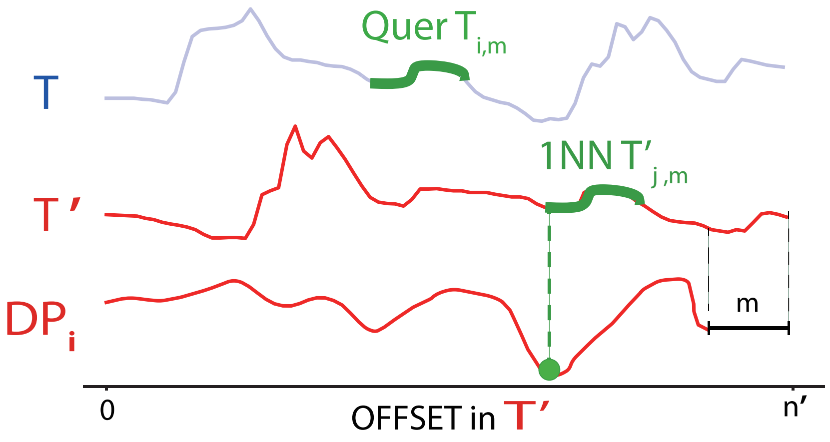


Figure 2.1: ***Distance Profile*** *between Query $T_{i,m}$ and target time series T', where n' is the length of T'. $DP_{i,j}$ can be considered as a meta TS annotating target T'*

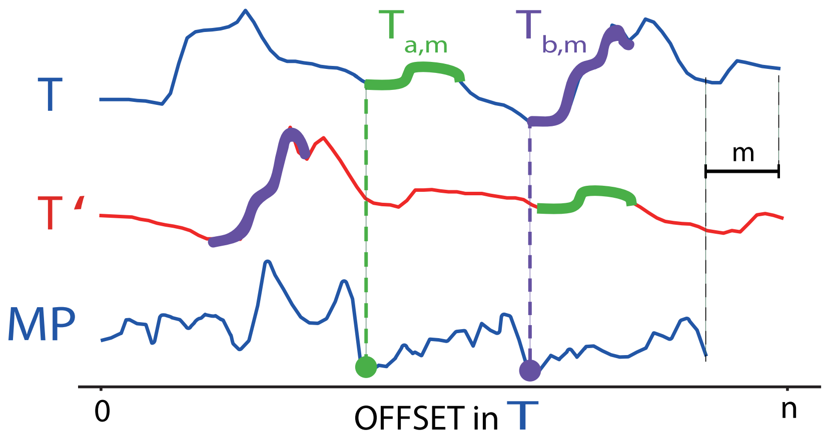


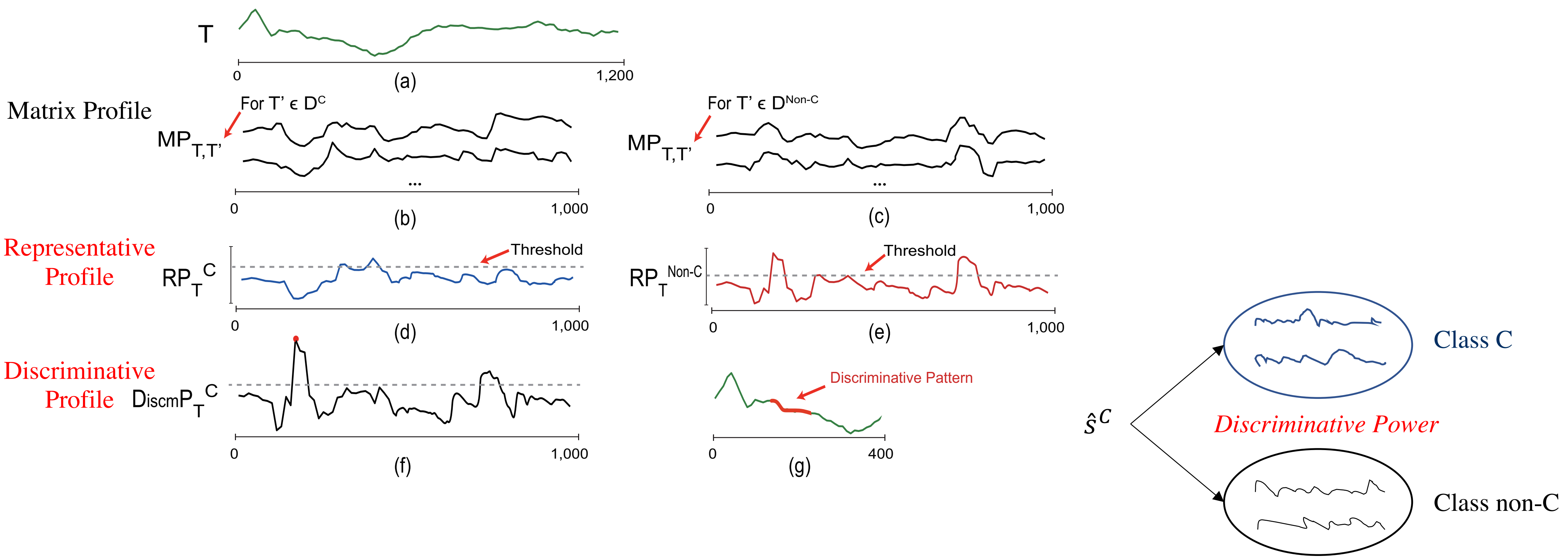
Figure 2.2: ***Matrix Profile*** *between Source T and Target T', where n is the length of T. Intuitively, MP_i shares the same offset as source T*

- ➢ Find the Nearest Neighbor of the Query
- ➢ Find the closest pairs between two TS

1. Chin-Chia Michael Yeh et al. "*Matrix Pro le I: All Pairs Similarity Joins for Time Series: A Unifying View That Includes Motifs, Discords and Shapelets*." In Proc. ICDM 2016

2. Preliminaries - Our previous work

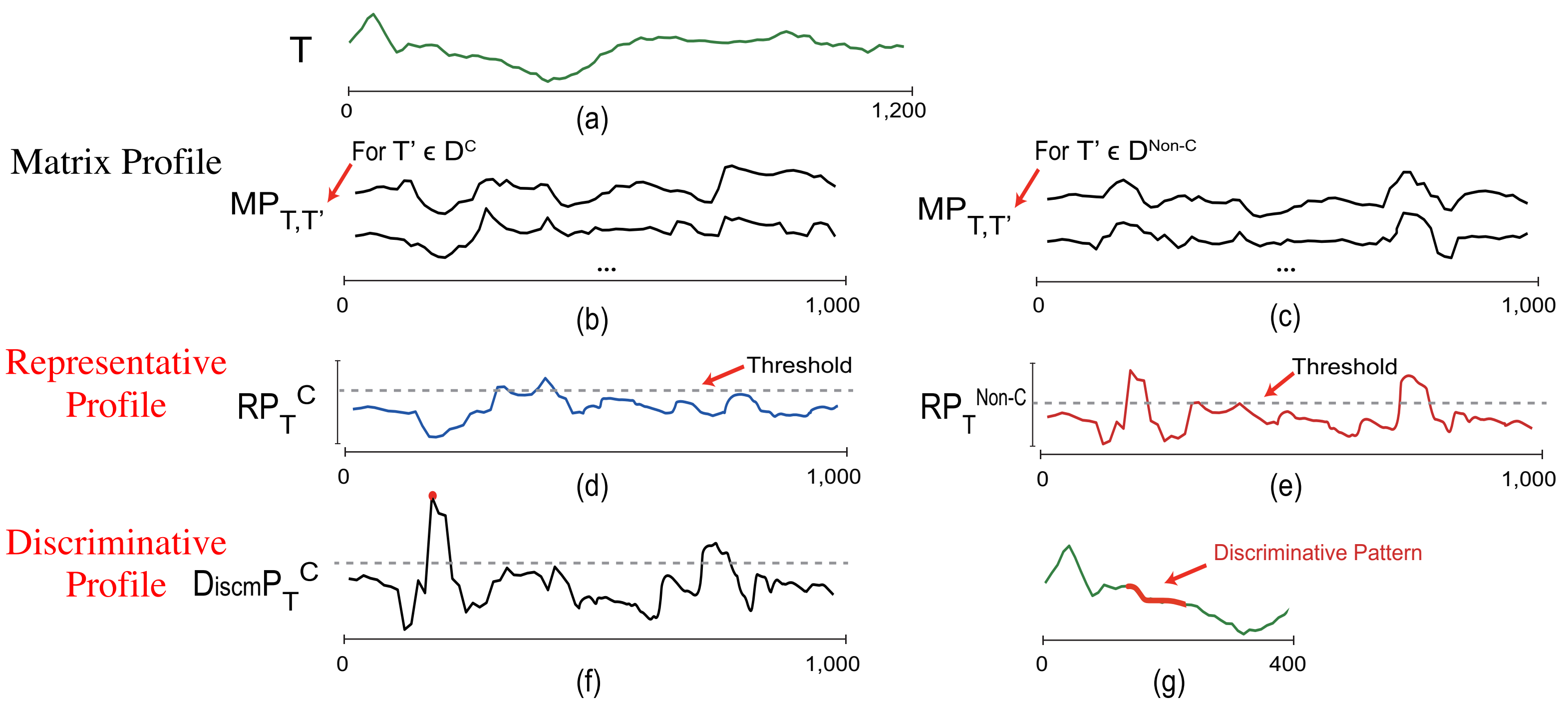
SMAP [1] (Shapelet Extraction on Matrix Profile)



1. J. Zuo, K. Zeitouni, and Y. Taher, "*Exploring interpretable features for large time series with SE4TeC*." In Proc. EDBT 2019,

2. Preliminaries - Our previous work

SMAP [1] (Shapelet Extraction on Matrix Profile)



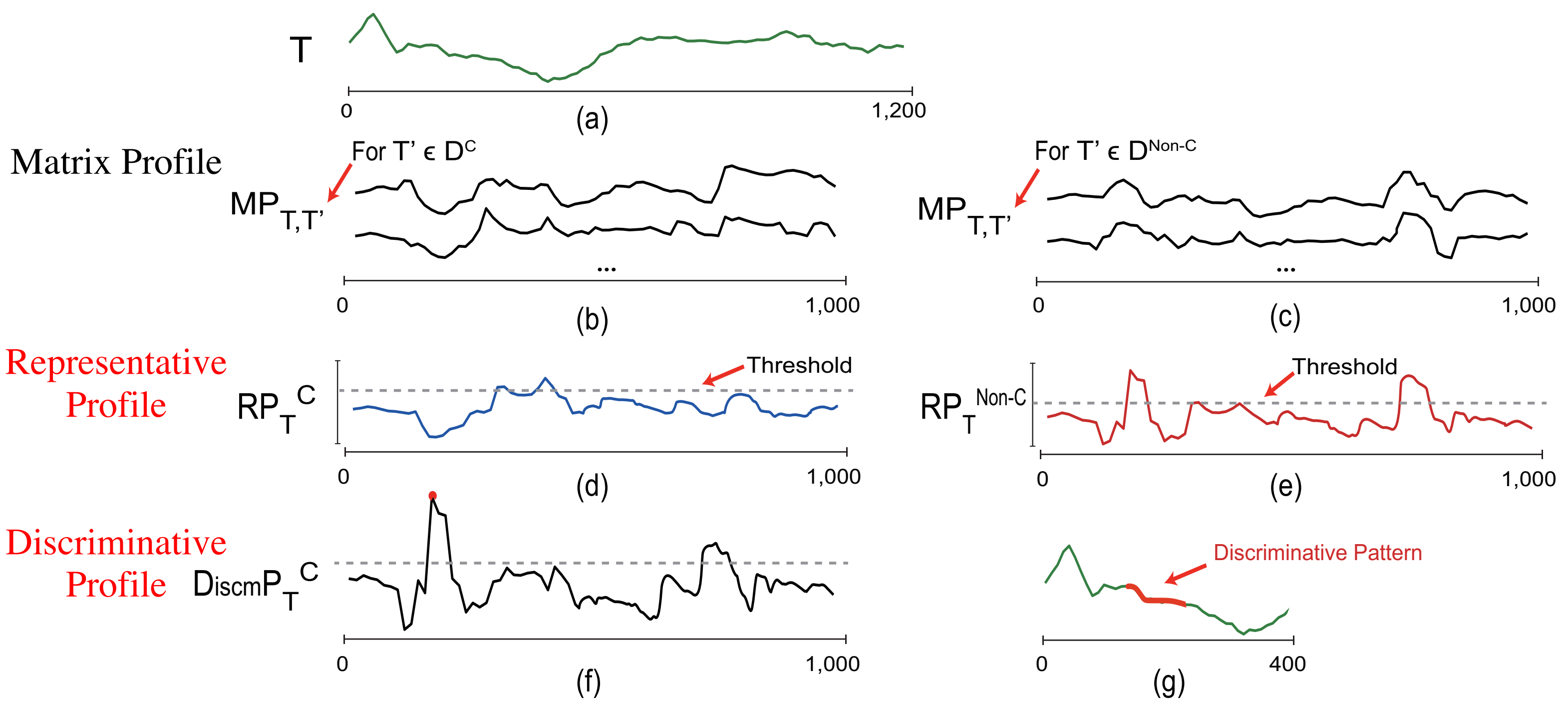
Cache Dataset in HDFS.

- MapPartition (*Set of <ID, T>*)
  - T. $dist_{Thresh} \leftarrow RepresentativeProfile(T, D^C)$
  - T. $DiscmP \leftarrow ComputeDiscriminativeProfile(T, D)$
  - **emit** *(ID, T)*

1. J. Zuo, K. Zeitouni, and Y. Taher, “*Exploring interpretable features for large time series with SE4TeC*.” In Proc. EDBT 2019

2. Preliminaries - Our previous work

SMAP[1] (Shapelet Extraction on Matrix Profile)



Cache Dataset in HDFS.

- MapPartition (*Set of <ID, T>*)

  T. $dist_{Thresh} \leftarrow RepresentativeProfile(T, D^C)$

  T. $DiscmP \leftarrow ComputeDiscriminativeProfile(T, D)$

  emit *(ID, T)*

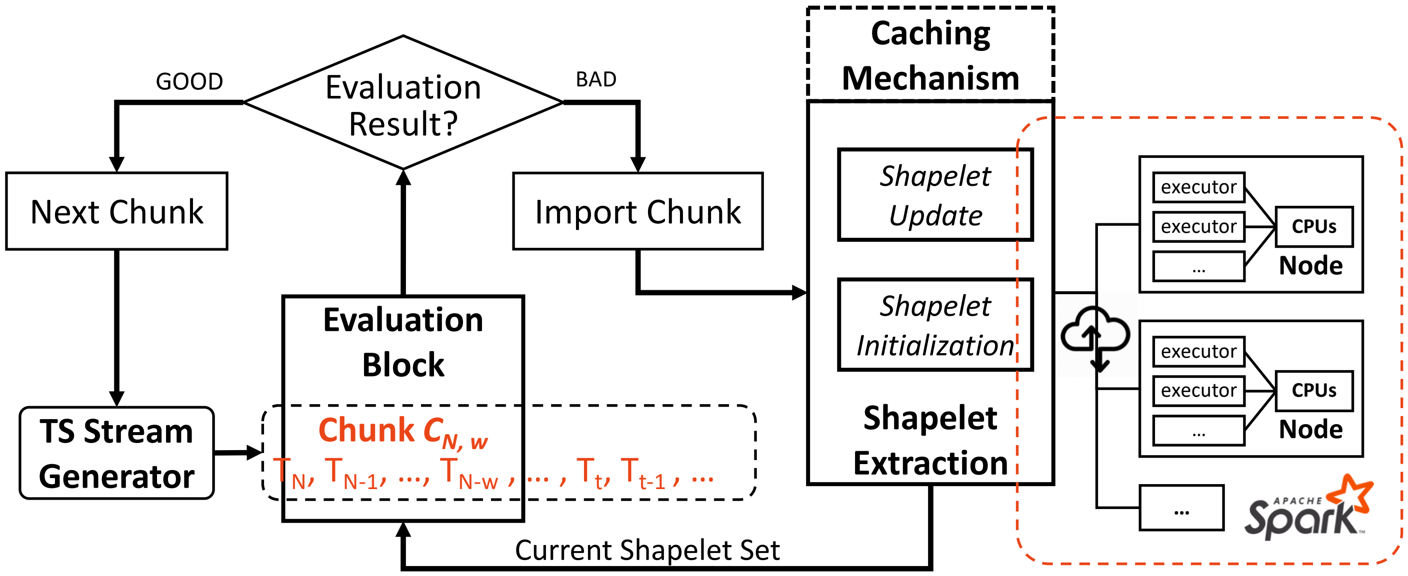
- MapAggregation (*class*, *(ID, T)*)

  $\hat{S} \leftarrow getTopK(aggregation(T.DiscmP))$

  return $\hat{S}$

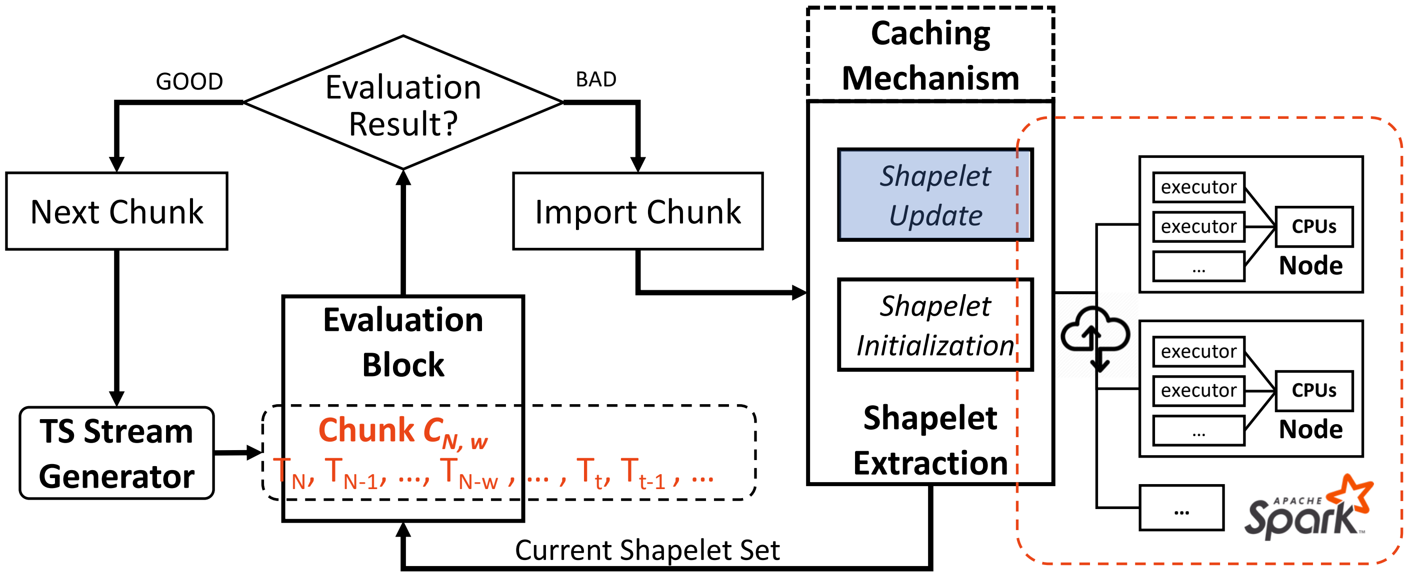
1. J. Zuo, K. Zeitouni, and Y. Taher, "*Exploring interpretable features for large time series with SE4TeC*." In Proc. EDBT 2019

3. Our proposal



Test-then-Train strategy

3. Our proposals - Incremental SMAP (ISMAP)



Cache TS Chunk / T_N in HDFS

1. MapPartition (*Set of (ID, T)*)

computeMP(T, T_N)

$T.dist_{Thresh} \leftarrow UpdateRepresentativeProfile(T, T_N)$

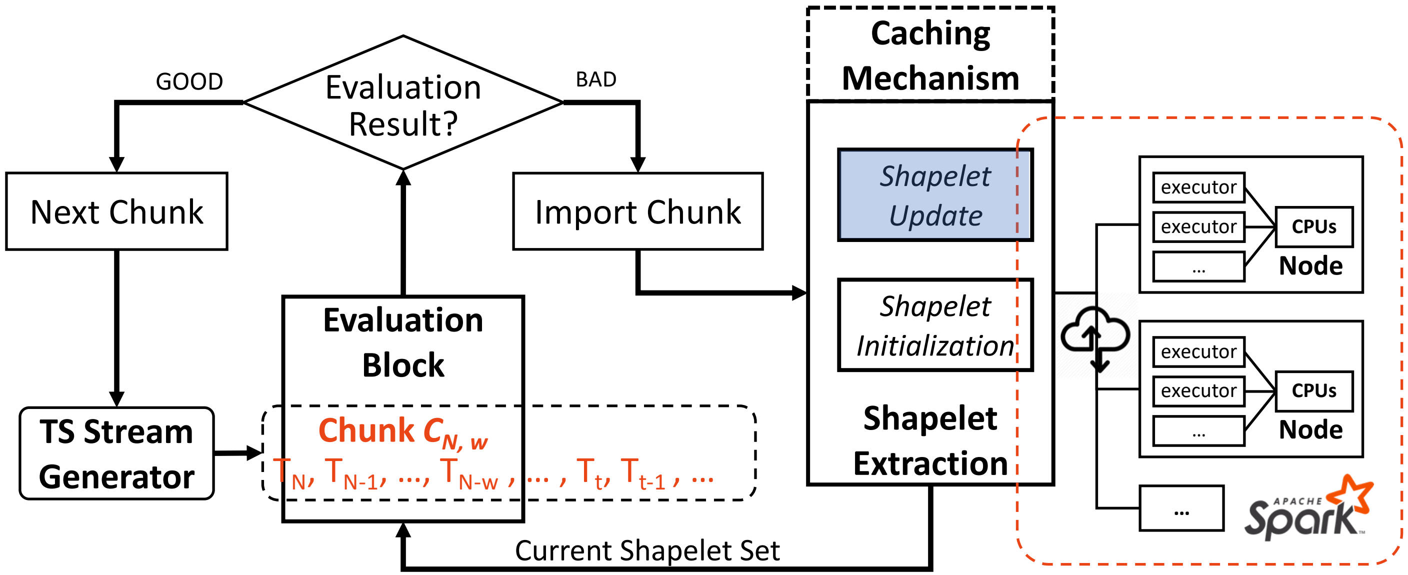
$T.DiscmP \leftarrow UpdateDiscriminativeProfile(T, T_N)$

$MP_{T_N} \leftarrow computeMP(T_N, T)$

emit (ID, T, MP_{T_N})

Update the discriminative power of existing Shapelets

3. Our proposals - Incremental SMAP (ISMAP)



Cache TS Chunk / T_N in HDFS

1. MapPartition (*Set of (ID, T)*)

   $T.\,dist_{Thresh} \leftarrow UpdateRepresentativeProfile(T, T_N)$

   $T.\,DiscmP \leftarrow UpdateDiscriminativeProfile(T, T_N)$

   $\boxed{MP_{T_N} \leftarrow computeMP(T_N, T)}$

   emit (ID, T, MP_{T_N})

2. MapAggregation $(*, (ID, T, MP_{T_N}))$

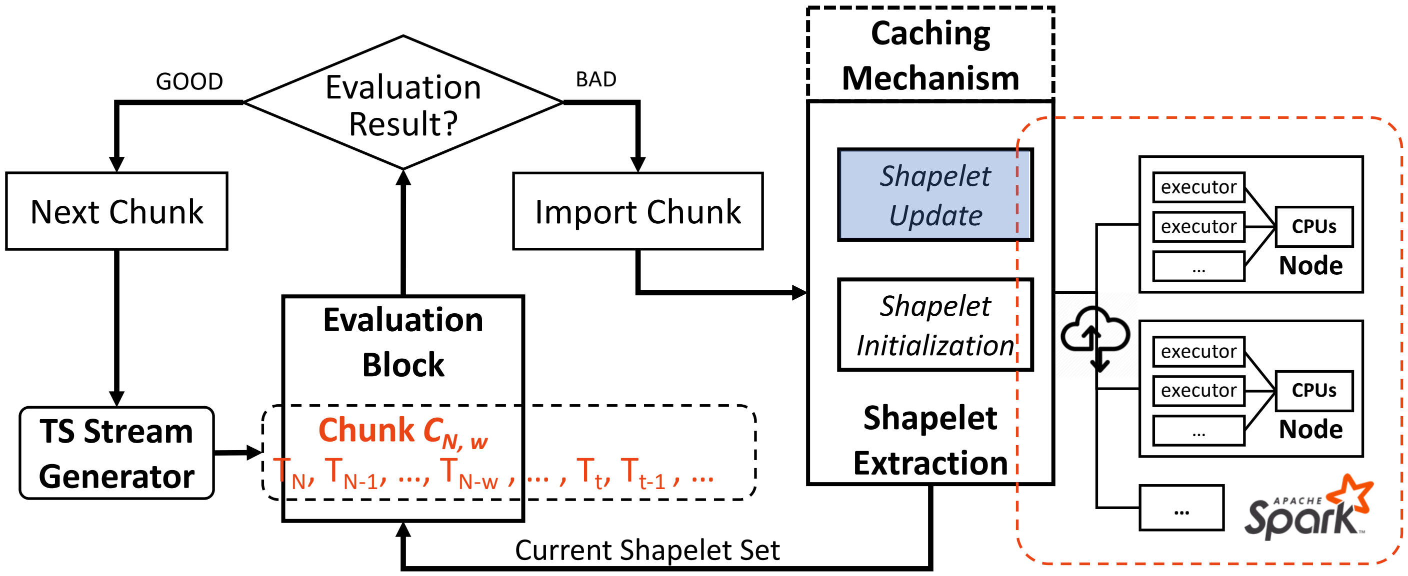
   $T_N.\,dist_{Thresh} \leftarrow RepresentativeProfile(agg(MP_{T_N}))$

   $T_N.\,DiscmP \leftarrow DiscriminativeProfile(agg(MP_{T_N}))$

   return(ID, T_N)

Introduce new candidate Shapelets, compute their discriminative power

3. Our proposals - Incremental SMAP (ISMAP)



Cache TS Chunk / T_N in HDFS

1. MapPartition (*Set of (ID, T)*)

   $T.\,dist_{Thresh} \leftarrow UpdateRepresentativeProfile(T, T_N)$

   $T.\,DiscmP \leftarrow UpdateDiscriminativeProfile(T, T_N)$

   $MP_{T_N} \leftarrow computeMP(T_N, T)$

   emit (ID, T, MP_{T_N})

2. MapAggregation $(*, (ID, T, MP_{T_N}))$

   $T_N.\,dist_{Thresh} \leftarrow RepresentativeProfile(agg(MP_{T_N}))$

   $T_N.\,DiscmP \leftarrow DiscriminativeProfile(agg(MP_{T_N}))$

   return (ID, T_N)

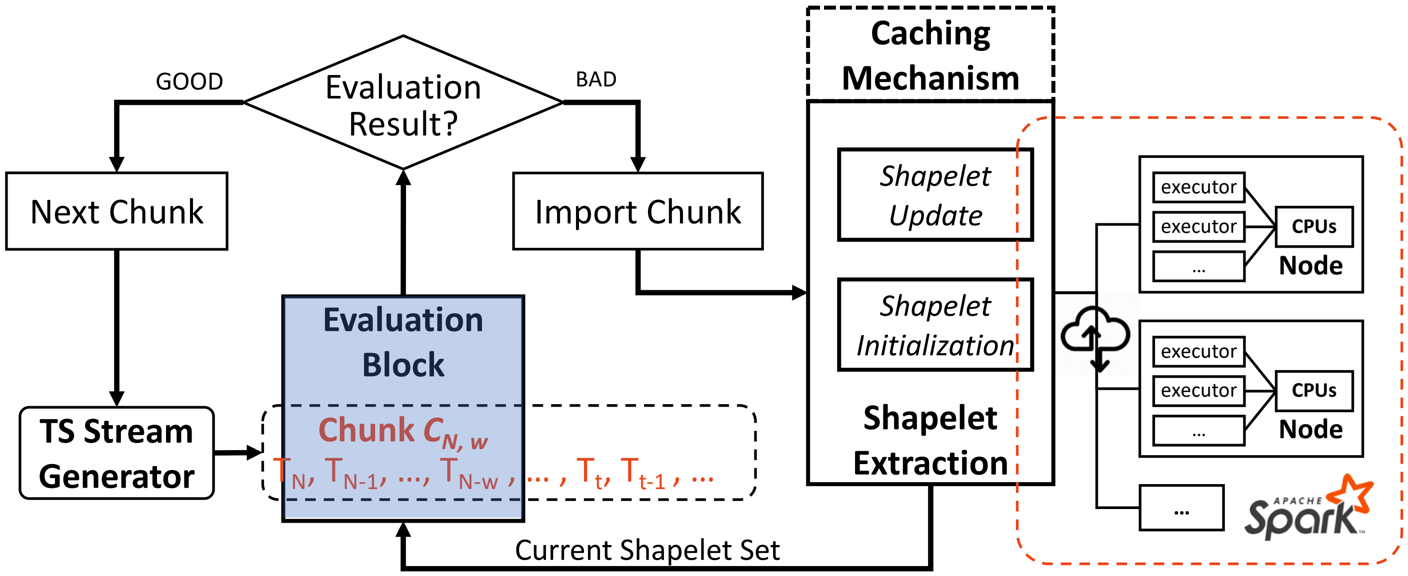
3. MapAggregation (*class*, *(ID, T)*)

   $\hat{S} \leftarrow getTopK(aggregation(T.\,DiscmP))$

   return $\hat{S}$

Update the Shapelet Set

3. Our proposals - Evaluation Block



Evaluation from two aspects

1. Shapelet Evaluation

   ⇢ *Only when loss > threshold, import TS into extraction process*

   ⇢ Select the most informative TS chunks

   Memory & Computation Saving

2. Concept Drift Detection

   ⇢ Distinguish from Shapelet loss

Evaluation Block (Shapelet Evaluation + Concept Drift Detection)

Shapelet Evaluation

- 0-1 Loss Function (classic methods)

$$L(Y, h(T)) = \begin{cases} 0, & Y = h(T) \\ 1, & Y \neq h(T) \end{cases}$$

where

$$h(T) = \begin{cases} C, & if\ dist(T, \hat{s}) \leq \hat{s}.dist_{Thresh} \\ nonC, & otherwise \end{cases}$$

- Sigmoid Loss Function (our proposal)

$$L(Y, h(T)) = \frac{1}{1 + e^{-(x-\sigma)}}, \quad \sigma = \hat{s}.dist_{Thresh}$$

$$x = min(dist(T^C, \hat{s})), \quad \hat{s} \in \hat{S}^C$$

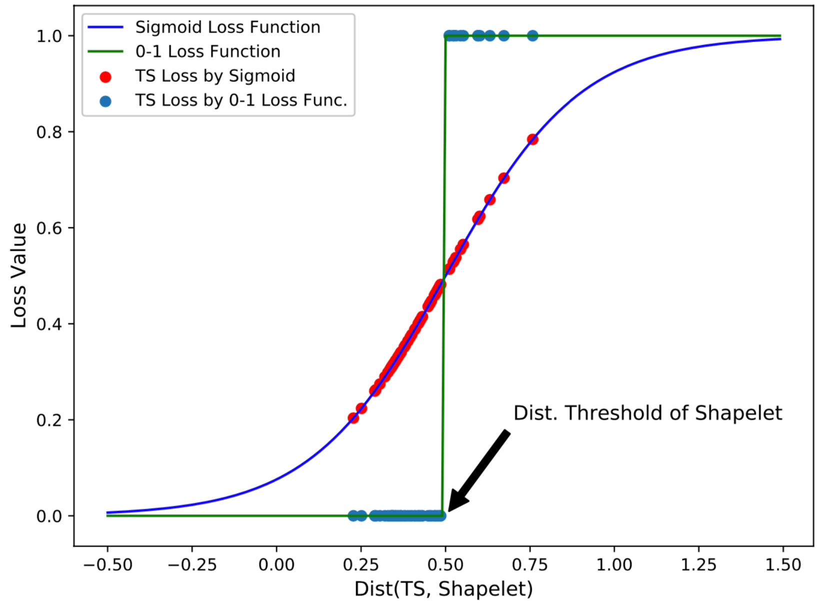


Figure 2: Shapelet Evaluation over newly input TS instances

Evaluation Block (Shapelet Evaluation + Concept Drift Detection)

Shapelet Evaluation

- 0-1 Loss Function (classic methods)

$$L(Y, h(T)) = \begin{cases} 0, & Y = h(T) \\ 1, & Y \neq h(T) \end{cases}$$

where

$$h(T) = \begin{cases} C, & if\ dist(T, \hat{s}) \leq \hat{s}.dist_{Thresh} \\ nonC, & otherwise \end{cases}$$

- Sigmoid Loss Function (our proposal)

$$L(Y, h(T)) = \frac{1}{1 + e^{-(x-\sigma)}}, \quad \sigma = \hat{s}.dist_{Thresh}$$

$$x = min(dist(T^C, \hat{s})), \quad \hat{s} \in \hat{S}^C$$

Loss Threshold

$\Delta = 0.3$

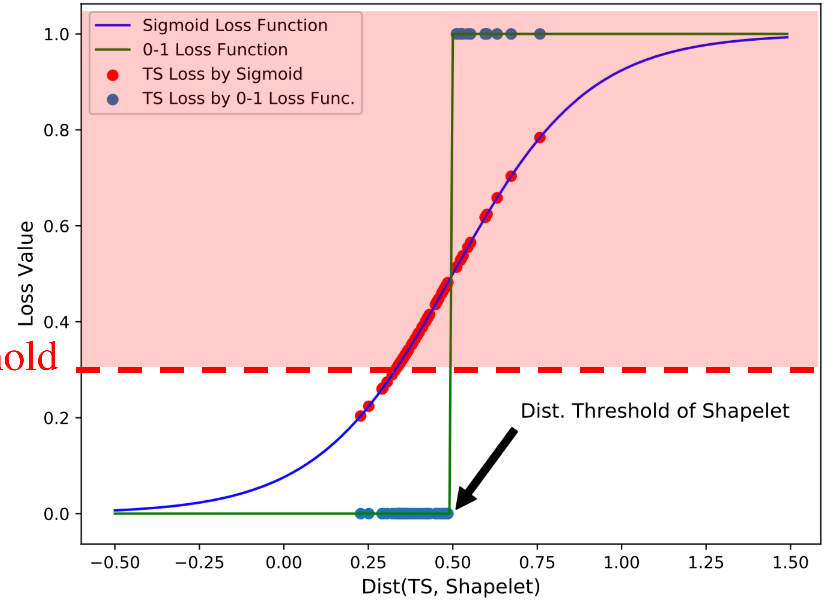


Figure 2: Shapelet Evaluation over newly input TS instances

A Loss Threshold Δ can be set to import incrementally the valuable instances.

Evaluation Block (Shapelet Evaluation + Concept Drift Detection)

Concept Drift detection

- Page-Hinkey (PH) Test[1]: a typical technique for change detection in signal processing.

$$L_C(N) = \frac{1}{w}\sum_{k=1}^{w} L(Y_{N-w+k}, h(T_{N-w+k}))$$

$$m_N = \sum_{t=0}^{N}(L_C(t) - L_{avg}(t) - \delta)$$

$$M_N = min(m_t, t = 1 \ldots N)$$

$$PH_N = m_N - M_N$$

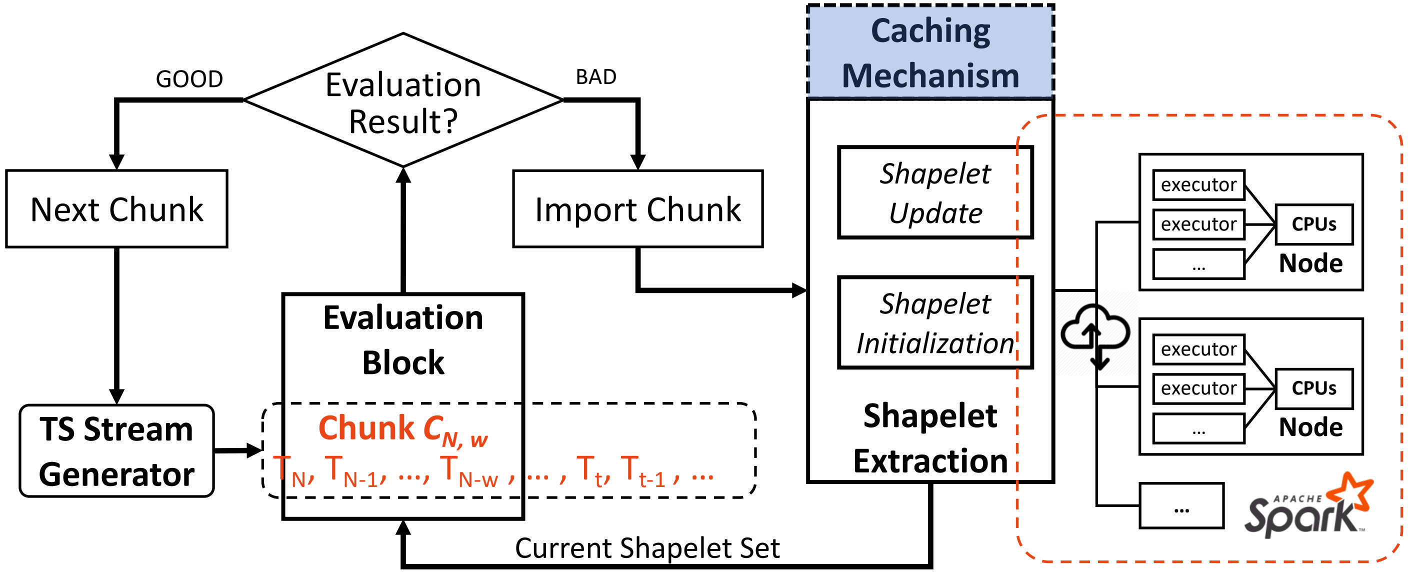
- $L_C(N)$: the average loss of newly input TS chunk
- $L_{avg}(t)$: the average loss of all historical TS chunk until t
- m_N: the cumulative difference between the chunk loss and average loss until the current time. δ: *Loss Tolerance*
- M_N: the minimal cumulative difference recorded

- λ: *PH threshold* to detect a Concept Drift
- $Concept\ Drift = \begin{cases} True, & PH_N \geq \lambda \\ False, & otherwise \end{cases}$

Loss -> Signal
Change point detection

1. J. Gama, I. Zliobait E, A. Bifet, M. Pechenizkiy, and A. Bouchachia, "A Survey on Concept Drift Adaptation,". ACM Comput. Surv. Article, vol. 1, 2013.

3. Our proposals - Elastic Caching Mechanism

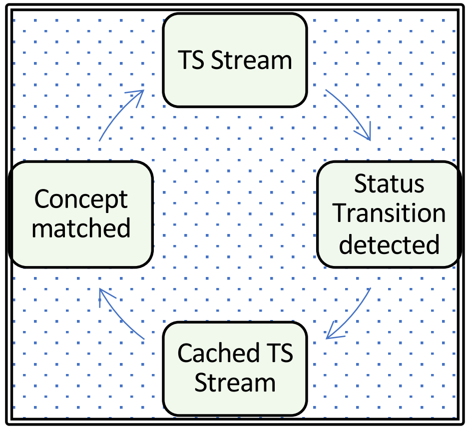


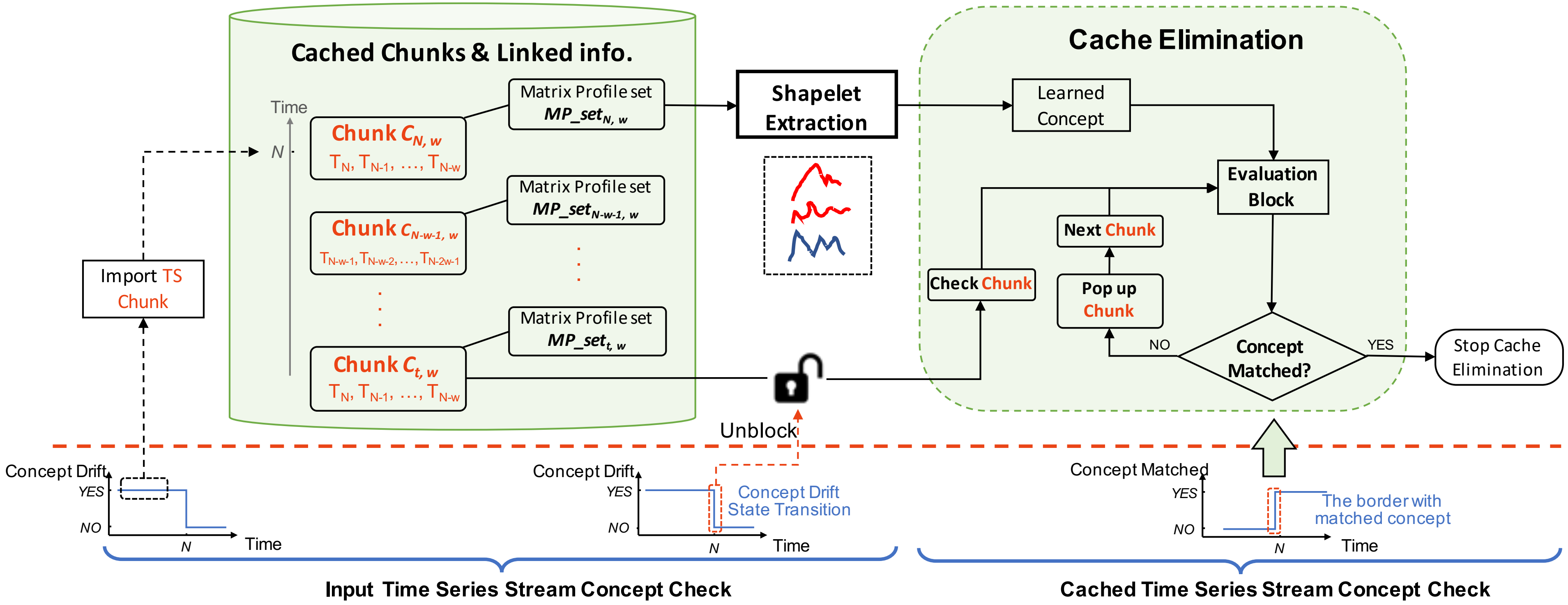
Dependence on cached data

Shapelet Extraction relies on a set of TS instance

- Current Learned concept
- Out-of-date concept

3. Our proposals - Elastic Caching Mechanism





Intuition: Fresh learned concept might be inapplicable for the old instances in the cache

-> Delete them from the cache

4. Experiments

Experimental Designs:

- Accuracy & Incrementality of ISMAP

Datasets:

- 14 datasets from UCR Archive[1, 2]

Baseline: Shapelet Tree classifiers

- Information Gain (IG)[3]
- Kruskall-Wallis (KW)[4]
- Mood's Median (MM)[4]

Evaluation:

- Incrementality: captured by **Compression Ratio** $Comp.Ratio = \frac{nbr.instance_{imported}}{nbr.instance_{training}}$
- Accuracy & Time

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/
2. A. Bagnall, J. Lines, A. Bostrom, J. Large, and E. Keogh, "The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances," Data Mining and Knowledge Discovery, vol. 31, no. 3, pp. 606–660, 2017
3. Lexiang Ye and Eamonn Keogh, "*Time series shapelets: A New Primitive for Data Mining*" In Proc. SIGKDD 2009
4. Jason Lines, and Anthony Bagnall, "*Alternative Quality Measures for Time Series Shapelets*", IDEAL 2012

4. Experiments

Experimental Designs:

- Concept Drift Detection & Adaptive Features

Datasets:

- Synthetic *Trace*[1] dataset:
  - Randomly put noise for Data Augmentation
  - 1000/1000 training/testing instances
  - Two drifts are inserted at time *333* and *667*
- Synthetic *ECG5000*[1] dataset:
  - 500/500 training/testing instances
  - Two drifts are inserted at time *167* and *233*

Evaluation:

- Drift detection
- Elastic caching mechanism
- Reliability of Adapted features

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/

4.1 Accuracy & Incrementality of ISMAP

Baseline: Shapelet Tree classifiers

- Information Gain (IG)[2]
- Kruskall-Wallis (KW)[3]
- Mood's Median (MM)[3]

Type	Name	Train/Test	Class	Length	IG	KW	MM	ISMAP(best)	Para. (Δ)	Comp. Ratio
Simulated	SyntheticControl	300/300	6	60	**0.9433**	0.9000	0.8133	0.7007	0.35	46.7%
Sensor	Trace	100/100	4	275	0.9800	0.9400	0.9200	**1**	0.5, 0.45	26.0%
	MoteStrain	20/1252	2	84	0.8251	0.8395	0.8395	**0.9169**	0.45	60.0%
	SonyAIBO.I	20/601	2	70	0.8453	0.7281	0.7521	**0.9151**	0.4	95.0%
	SonyAIBO.II	27/953	2	65	0.8457	-	-	**0.8583**	0.4	63.0%
	ItalyPower.	67/1029	2	24	0.8921	0.9096	0.8678	**0.9466**	0.45	25.4%
ECG	ECG5000	500/4500	5	140	0.7852	-	-	**0.9109**	0.4	9.4%
	ECGFiveDays	23/861	2	136	0.7747	0.8721	0.8432	**0.9826**	0.4	51.2%
	TwoLeadECG	23/1189	2	82	0.8507	0.7538	7657	**0.9337**	0.5	47.8%
Images	Symbols	25/995	6	398	0.7799	0.5568	0.5799	**0.8113**	0.35	96.0%
	Coffee	28/28	2	286	**0.9643**	0.8571	0.8671	0.9286	0.4	78.6%
	FaceFour	24/88	4	350	0.8409	0.4432	0.4205	**0.9886**	except 0.45	62.5%
	DiatomSize.	16/306	4	345	0.7222	0.6111	0.4608	**0.8758**	0.5	50.0%
Motion	GunPoint	50/150	2	150	0.8933	0.9400	0.9000	**0.9733**	0.45	42.0%

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/
2. Lexiang Ye and Eamonn Keogh, "*Time series shapelets: A New Primitive for Data Mining*" In Proc. SIGKDD 2009
3. Jason Lines, and Anthony Bagnall, "*Alternative Quality Measures for Time Series Shapelets*", IDEAL 2012

4.1 Accuracy & Incrementality of ISMAP

Baseline: Shapelet Tree classifiers

- Information Gain (IG)[2]
- Kruskall-Wallis (KW)[3]
- Mood's Median (MM)[3]

ISMAP can be concatenated with Shapelet Transform[4] methods for higher accuracy

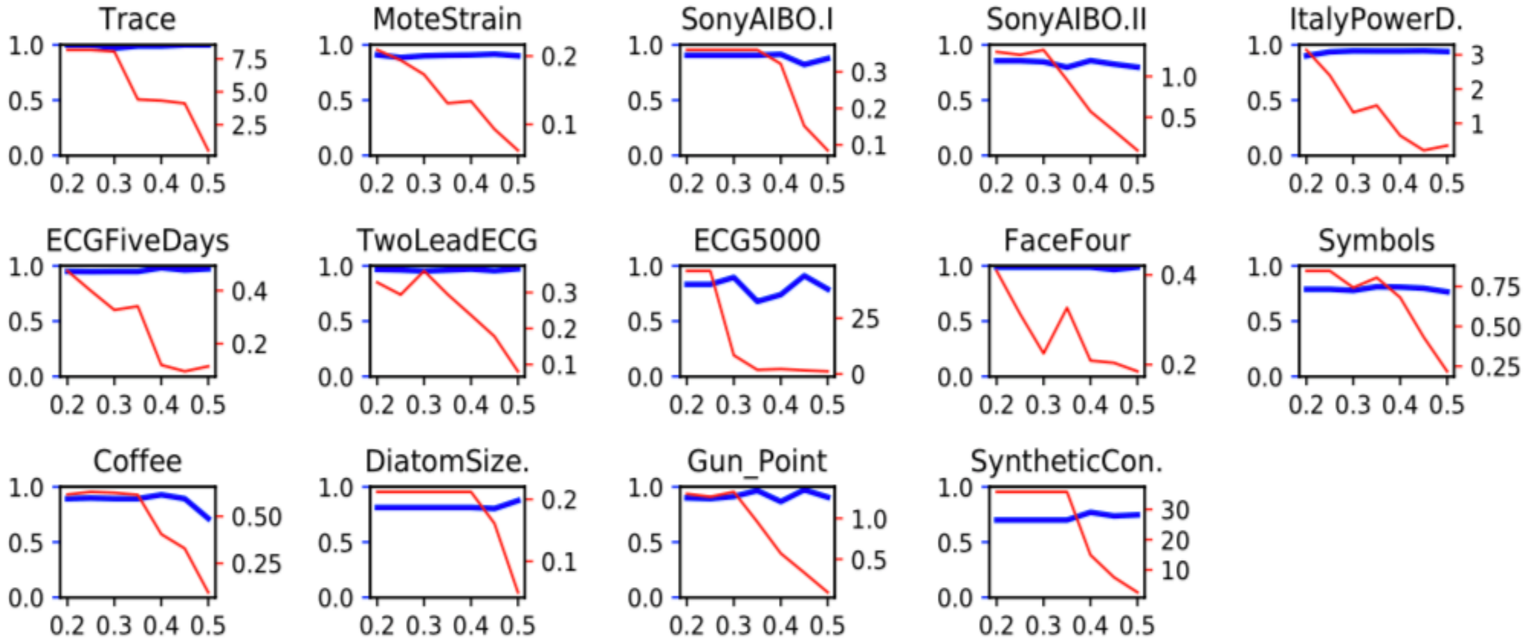
ISMAP can be integrated into TS ensemble classifiers, e.g., HIVE-COTE[5]

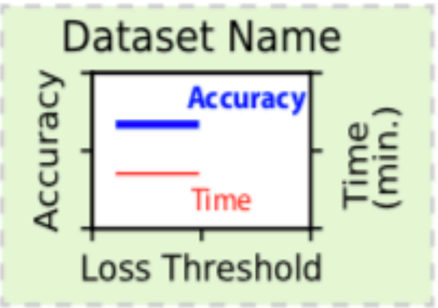
Type	Name	Train/Test	Class	Length	IG	KW	MM	ISMAP(best)	Para. (Δ)	Comp. Ratio
Simulated	SyntheticControl	300/300	6	60	**0.9433**	0.9000	0.8133	0.7007	0.35	46.7%
Sensor	Trace	100/100	4	275	0.9800	0.9400	0.9200	**1**	0.5, 0.45	26.0%
	MoteStrain	20/1252	2	84	0.8251	0.8395	0.8395	**0.9169**	0.45	60.0%
	SonyAIBO.I	20/601	2	70	0.8453	0.7281	0.7521	**0.9151**	0.4	95.0%
	SonyAIBO.II	27/953	2	65	0.8457	-	-	**0.8583**	0.4	63.0%
	ItalyPower.	67/1029	2	24	0.8921	0.9096	0.8678	**0.9466**	0.45	25.4%
ECG	ECG5000	500/4500	5	140	0.7852	-	-	**0.9109**	0.4	9.4%
	ECGFiveDays	23/861	2	136	0.7747	0.8721	0.8432	**0.9826**	0.4	51.2%
	TwoLeadECG	23/1189	2	82	0.8507	0.7538	7657	**0.9337**	0.5	47.8%
Images	Symbols	25/995	6	398	0.7799	0.5568	0.5799	**0.8113**	0.35	96.0%
	Coffee	28/28	2	286	**0.9643**	0.8571	0.8671	0.9286	0.4	78.6%
	FaceFour	24/88	4	350	0.8409	0.4432	0.4205	**0.9886**	except 0.45	62.5%
	DiatomSize.	16/306	4	345	0.7222	0.6111	0.4608	**0.8758**	0.5	50.0%
Motion	GunPoint	50/150	2	150	0.8933	0.9400	0.9000	**0.9733**	0.45	42.0%

1. UCR Archive: https://www.cs.ucr.edu/~eamonn/time_series_data_2018/
2. Lexiang Ye and Eamonn Keogh, "*Time series shapelets: A New Primitive for Data Mining*" In Proc. SIGKDD 2009
3. Jason Lines, and Anthony Bagnall, "*Alternative Quality Measures for Time Series Shapelets*", IDEAL 2012
4. J. Lines, L. M. Davis, J. Hills, and A. Bagnall, "A shapelet transform for time series classification," in Proc. SIGKDD 2012
5. J. Lines, S. Taylor, and A. Bagnall, "Hive-cote: The hierarchical vote collective of transformation-based ensembles for time series classification," IEEE ICDM 2016

4.1 Accuracy & Incrementality of ISMAP

- Trade-off between Accuracy and Loss Threshold Δ





In theory

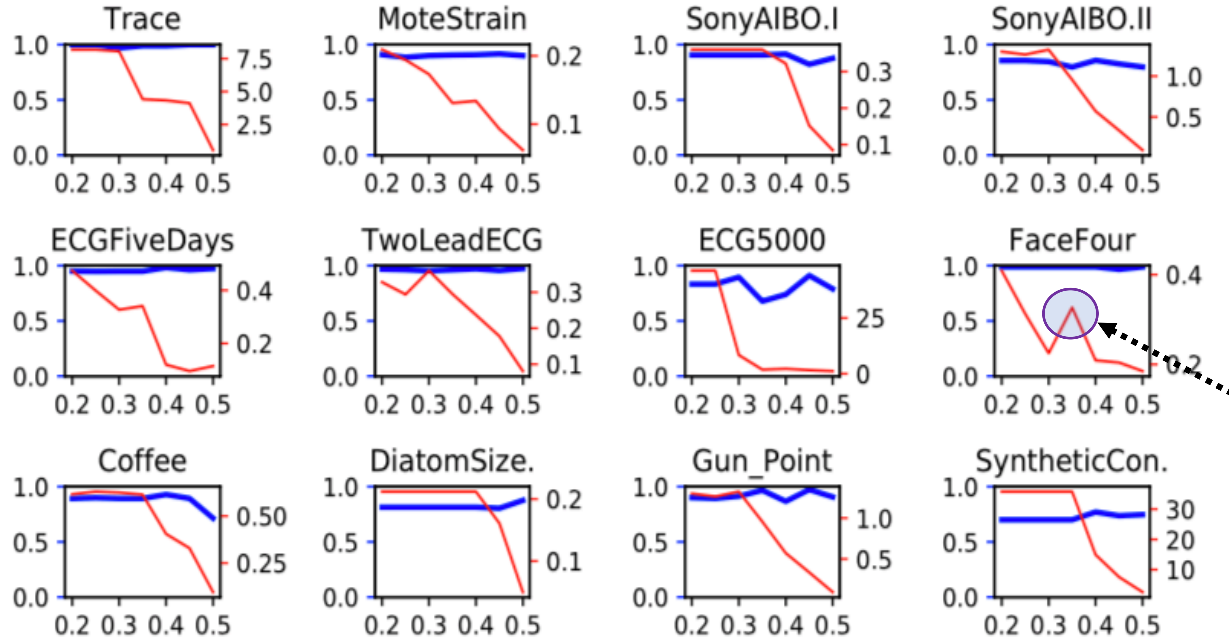
- Loss threshold ↗, the efficiency ↗, the accuracy ↘

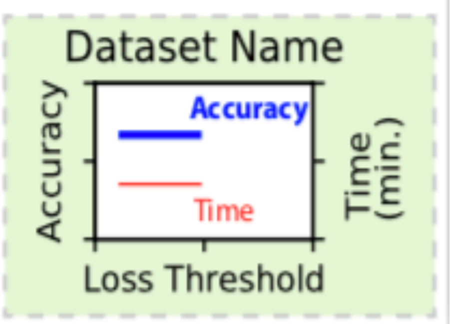
In practice

- The highest accuracy falls in the range $\Delta \in [0.35, 0.45]$.

4.1 Accuracy & Incrementality of ISMAP

- Trade-off between Accuracy and Loss Threshold Δ





In theory

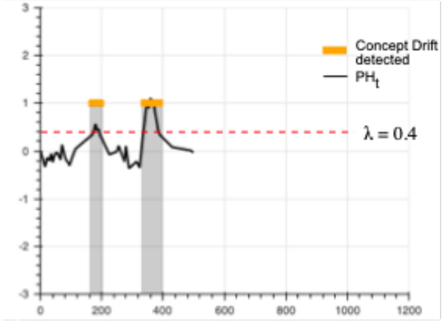
- Loss threshold ↗, the efficiency ↗, the accuracy ↘

In practice

- The highest accuracy falls in the range $\Delta \in [0.35, 0.45]$.
- Small uncertainty for the number of instances to be imported into the system

4.2 Concept Drift Detection & Adaptive Features

ECG5000



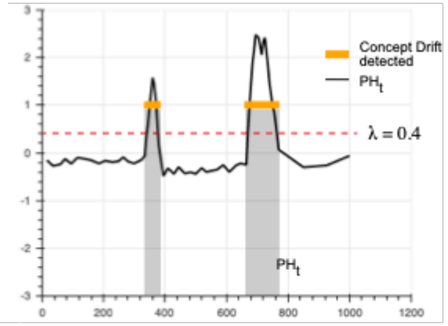
a) ECG5000: Tolerance $\delta = 0.10$, PH thresh. $\lambda = 0.4$

Drift 1 inserted	167	Drift 2 inserted	333
Drift 1 detected	170	Drift 2 detected	340
Adapted Concept 1	195	Adapted Concept 2	390

Adaptation period: **25**,
Accuracy: **0.9018**

Adaptation period: **50**
Accuracy: **0.8927**

Synthetic Trace
δ=0.15



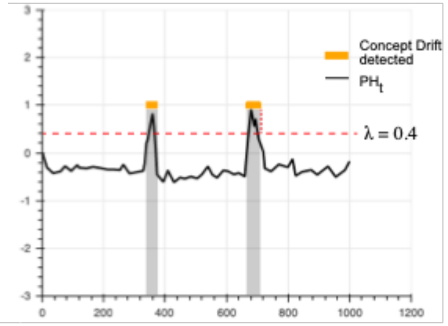
b) Aug. Trace: Tolerance $\delta = 0.15$, PH thresh. $\lambda = 0.4$

Drift 1 inserted	333	Drift 2 inserted	667
Drift 1 detected	345	Drift 2 detected	670
Adapted Concept 1	380	Adapted Concept 2	790

Adaptation period: **35**
Accuracy: **0.99**

Adaptation period: **120**
Accuracy: **0.98**

Synthetic Trace
δ=0.30



c) Aug. Trace: Tolerance $\delta = 0.30$, PH thresh. $\lambda = 0.4$

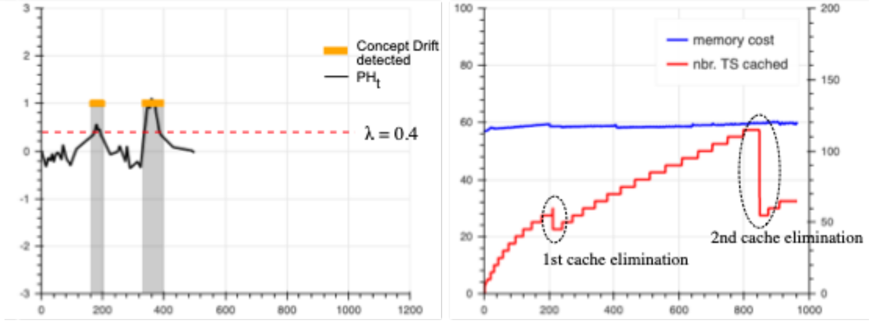
Drift 1 inserted	333	Drift 2 inserted	667
Drift 1 detected	350	Drift 2 detected	675
Adapted Concept 1	365	Adapted Concept 2	700

Adaptation period: **15**
Accuracy: **0.98**

Adaptation period: **25**
Accuracy: **0.97**

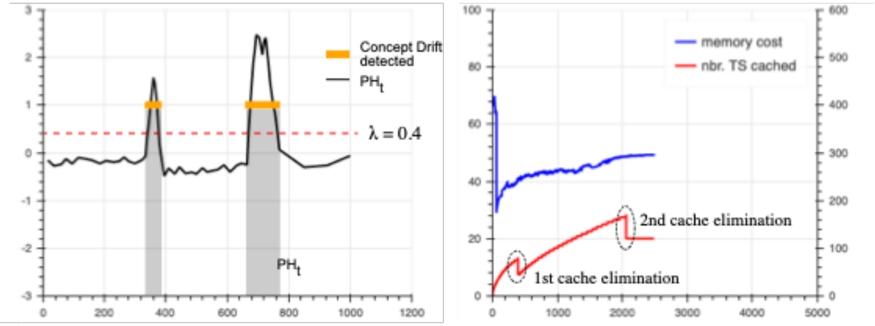
4.2 Concept Drift Detection & Adaptive Features[1]

ECG5000



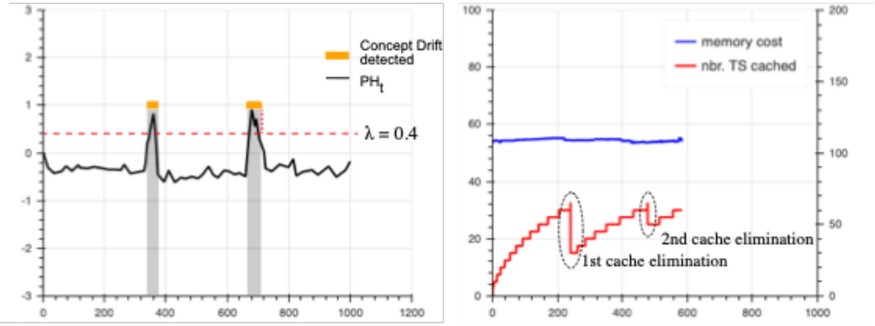
a) ECG5000: Tolerance δ = 0.10, PH thresh. λ = 0.4

Synthetic Trace
δ=0.15



b) Aug. Trace: Tolerance δ = 0.15, PH thresh. λ = 0.4

Synthetic Trace
δ=0.30



c) Aug. Trace: Tolerance δ = 0.30, PH thresh. λ = 0.4

Number of instances:

- 0 -> **65 -> 45**
- 45 -> **115 -> 65**
- 65 of 500 instances cached

δ=0.15

- 0 -> **70 -> 40**
- 40 -> **170 -> 120**
- 120 of 1000 instances cached

δ=0.30

- 0 -> **65 -> 30**
- 30 -> **65 -> 50**
- 50 of 1000 instances cached

1. J. Zuo, K. Zeitouni, and Y. Taher, "*ISETS: Incremental Shapelet Extraction from Time Series Stream*", demo paper in ECML-PKDD'19

5. Conclusion

- ✓ First attempt to explore incremental and adaptive features in Time Series Stream.
- ✓ We propose a novel Shapelet Evaluation approach which allows the transition from Time Series to Data Stream Mining.
- ✓ We propose an elastic caching mechanism which is capable of eliminating out-of-date concepts & data proactively in the Time Series Stream model.
- ✓ The system is applicable in the scenario where:
  - New TS instances enrich the learned concept
  - New TS instances bring Concept Drift
- ➢ **Future work:**
  - Extend to Streaming TS context
  - Focus on weak-labelled data

Project page in Github
(Demo video[1] included)

1. J. Zuo, K. Zeitouni, and Y. Taher, "*ISETS: Incremental Shapelet Extraction from Time Series Stream*", demo paper in ECML-PKDD'19

Questions?