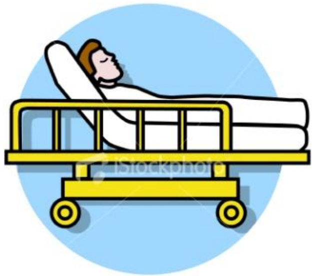
Advanced Classification for Streaming Time series and Data Streams

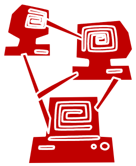
Presented by:	Nguyen Hai Long
Supervisor:	Assoc. Prof. Ng Wee Keong
Co-supervisor:	Dr. David Woon

Massive Sequential Data

Many applications can produce very big, massive data.



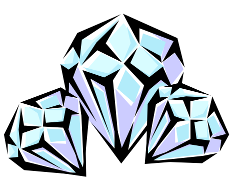
- Hospital data: A large hospital with thousands of patients can easily generate terabytes of physiological data per day.



- IP Network Traffic: Each router receives up to 1 billion packets per hour.

- Telecommunication data: There are 3 billion telephone calls in US daily.

→ We need to analyze this massive data to extract useful knowledge.



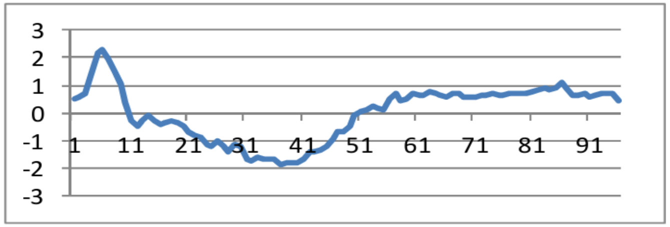
Sequential Data

The two most popular types of sequential data:

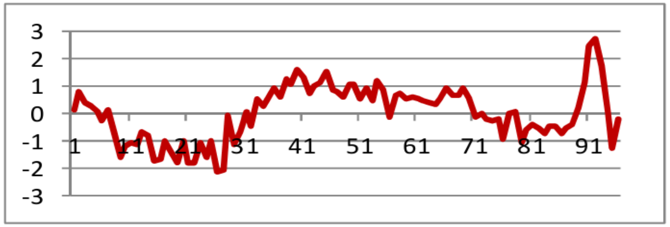
✓ Streaming time series: is a string of massive chronological observations, which can be found in biomedicine, sensor networks and stock markets.

✓ Data streams: are massive amount of continuous multivariate data, which are often generated by real-time surveillance industry, online transactions and scientific experiments.

They both share some unique characteristics: *huge volume, time-ordered and dynamically changing.*



Normal ECG



Abnormal ECG

Streaming time series (univariate domain)

Tid	SrcIP	Start time	Dest IP	Dest Port	Number of bytes	Attack
1	206.135.38.95	11:07:20	160.94.179.223	139	192	No
2	206.163.37.95	11:13:56	160.94.179.219	139	195	No
3	206.163.37.95	11:14:29	160.94.179.217	139	180	No
4	206.163.37.95	11:14:30	160.94.179.255	139	199	No
5	206.163.37.95	11:14:32	160.94.179.254	139	19	Yes
6	206.163.37.95	11:14:35	160.94.179.253	139	177	No
7	206.163.37.95	11:14:36	160.94.179.252	139	172	No
8	206.163.37.95	11:14:38	160.94.179.251	139	285	Yes
9	206.163.37.95	11:14:41	160.94.179.250	139	195	No
10	206.163.37.95	11:14:44	160.94.179.249	139	163	Yes

Data streams (multivariate domain)

Streaming Data Mining

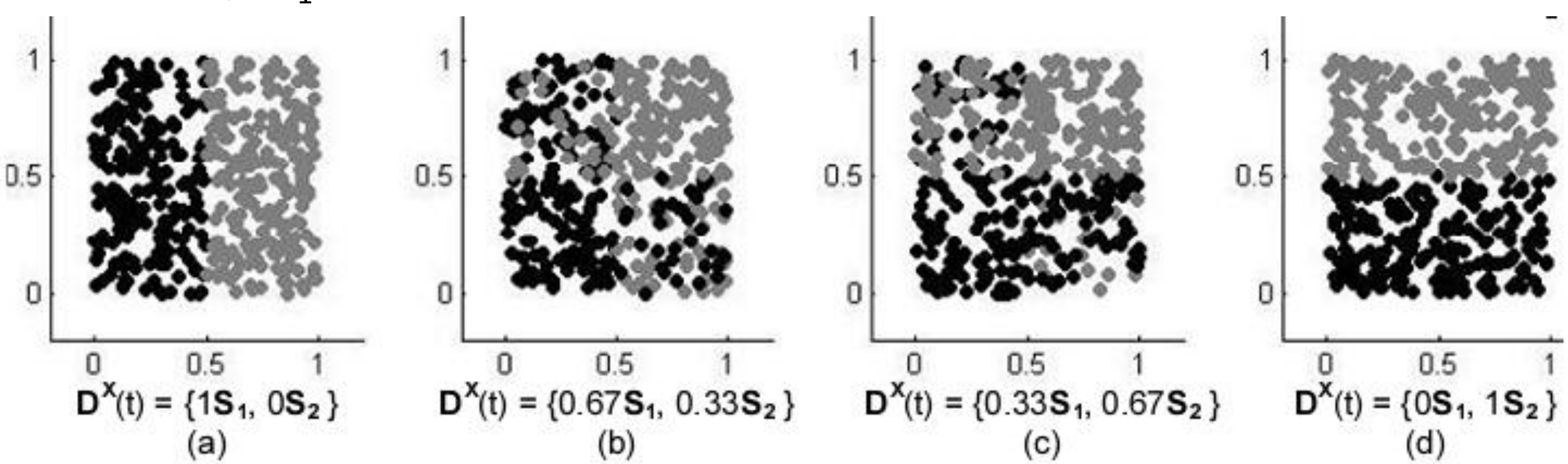
Constraints:

	Traditional Data Mining	Streaming Data Mining
Number of passes	multiple	single
Time	unlimited	real-time
Memory	unlimited	bounded
Number of concepts	one	multiple
Result	accurate	approximate

Concept Drift:

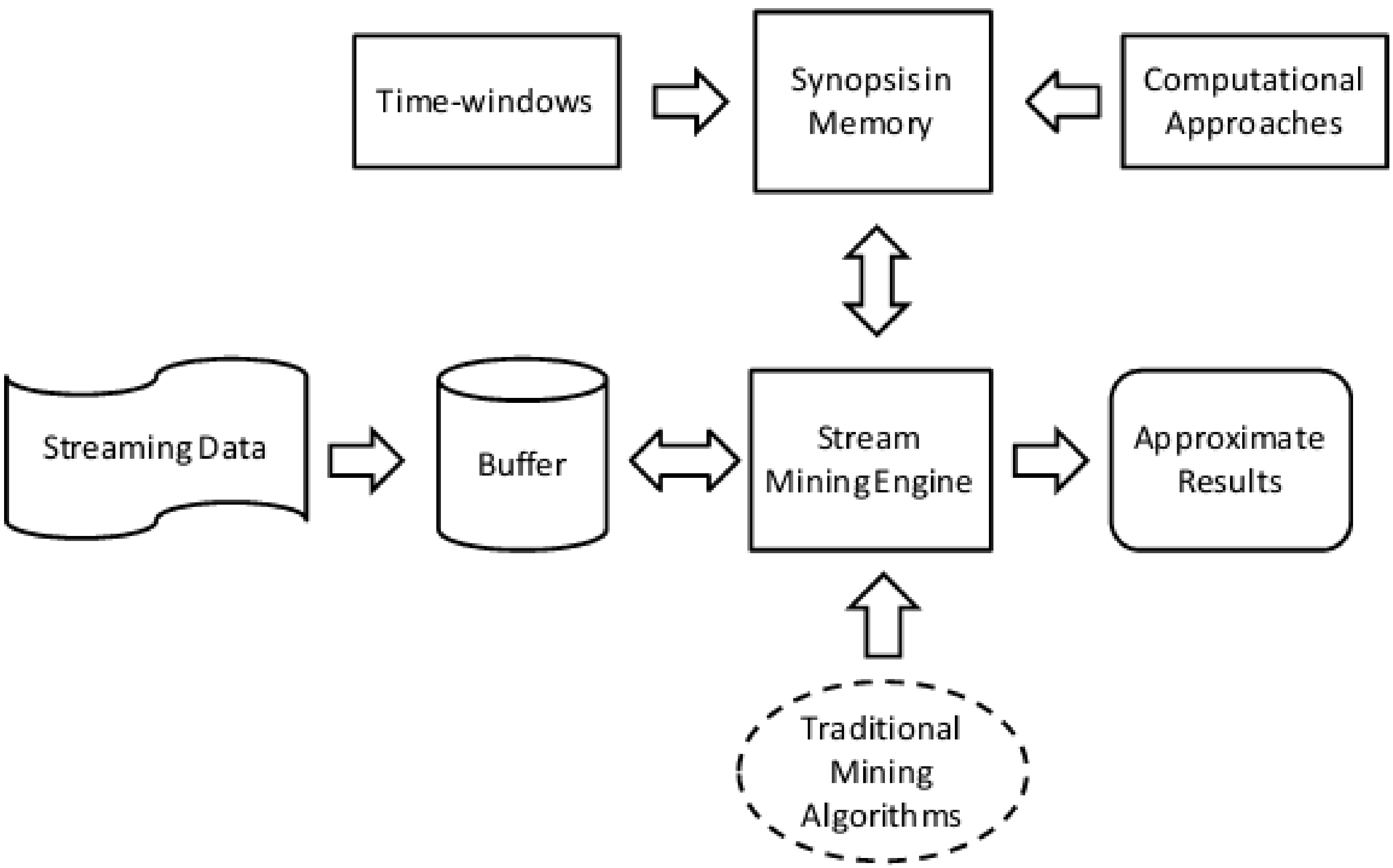
Concept S_1

Concept S_2

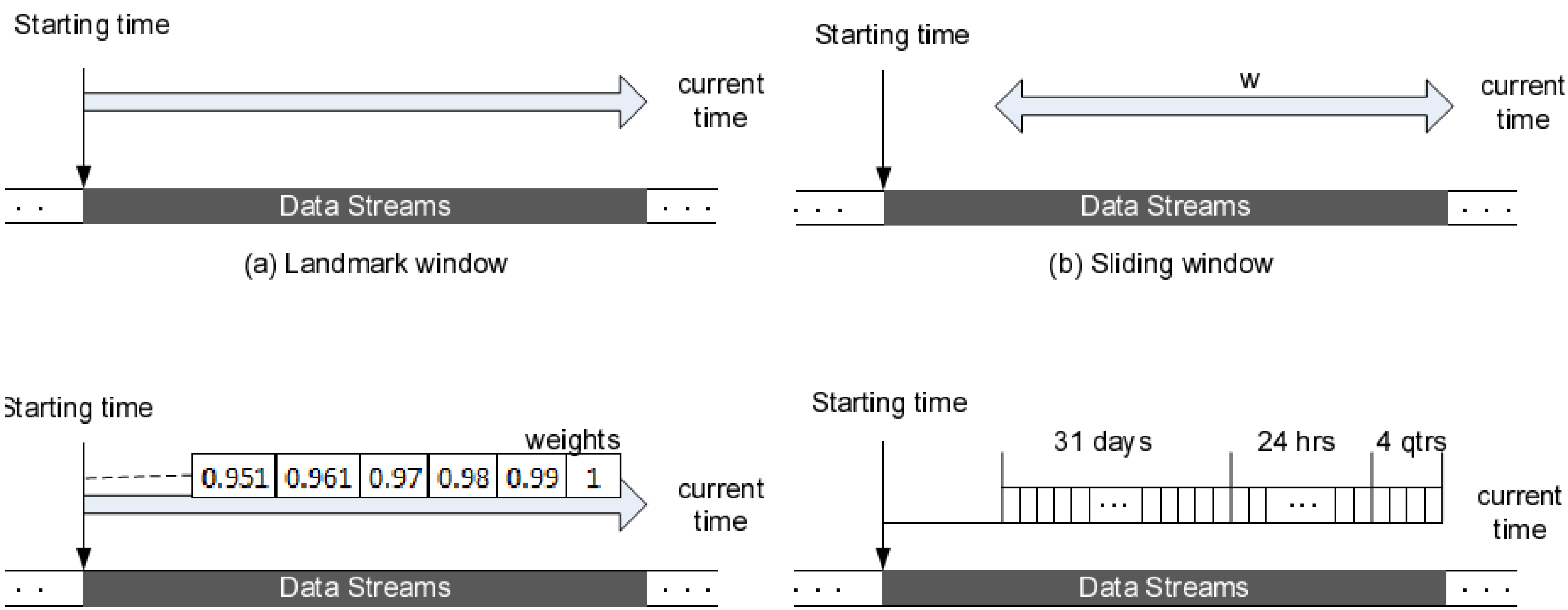


Example of a gradual change from a data source S1 to a data source S2.
Class y1 is depicted with grey dots, and class y2 with black dots

A General Mining Model



Time Windows



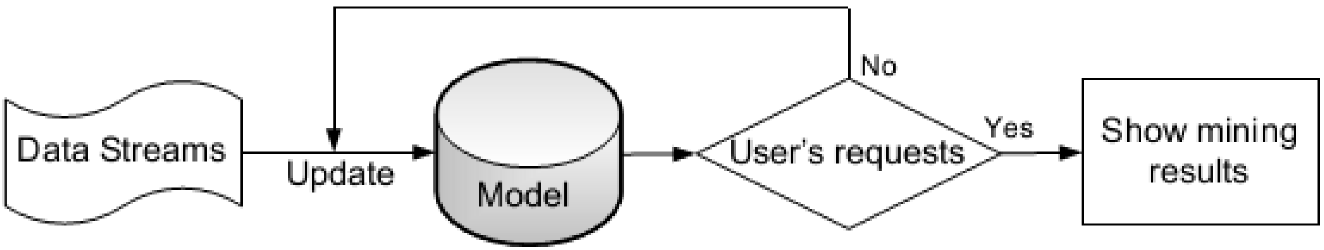
(a) Landmark window

(b) Sliding window

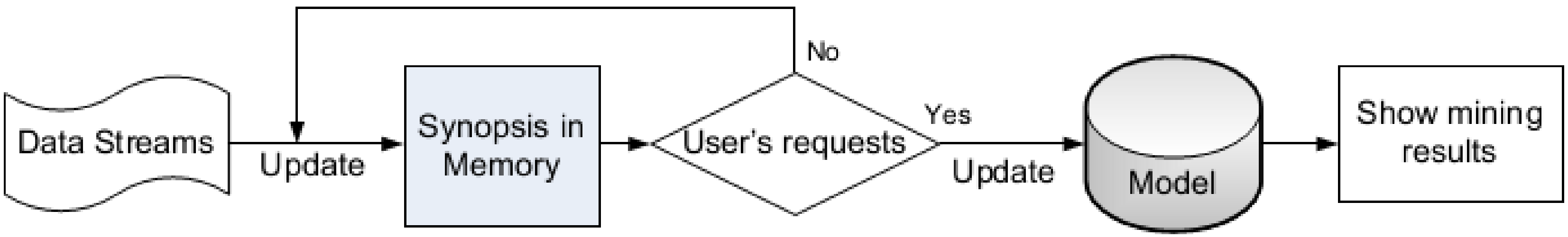
(c) Fading window (λ = 0.99)

(d) Tilted-time window

Computational Approaches

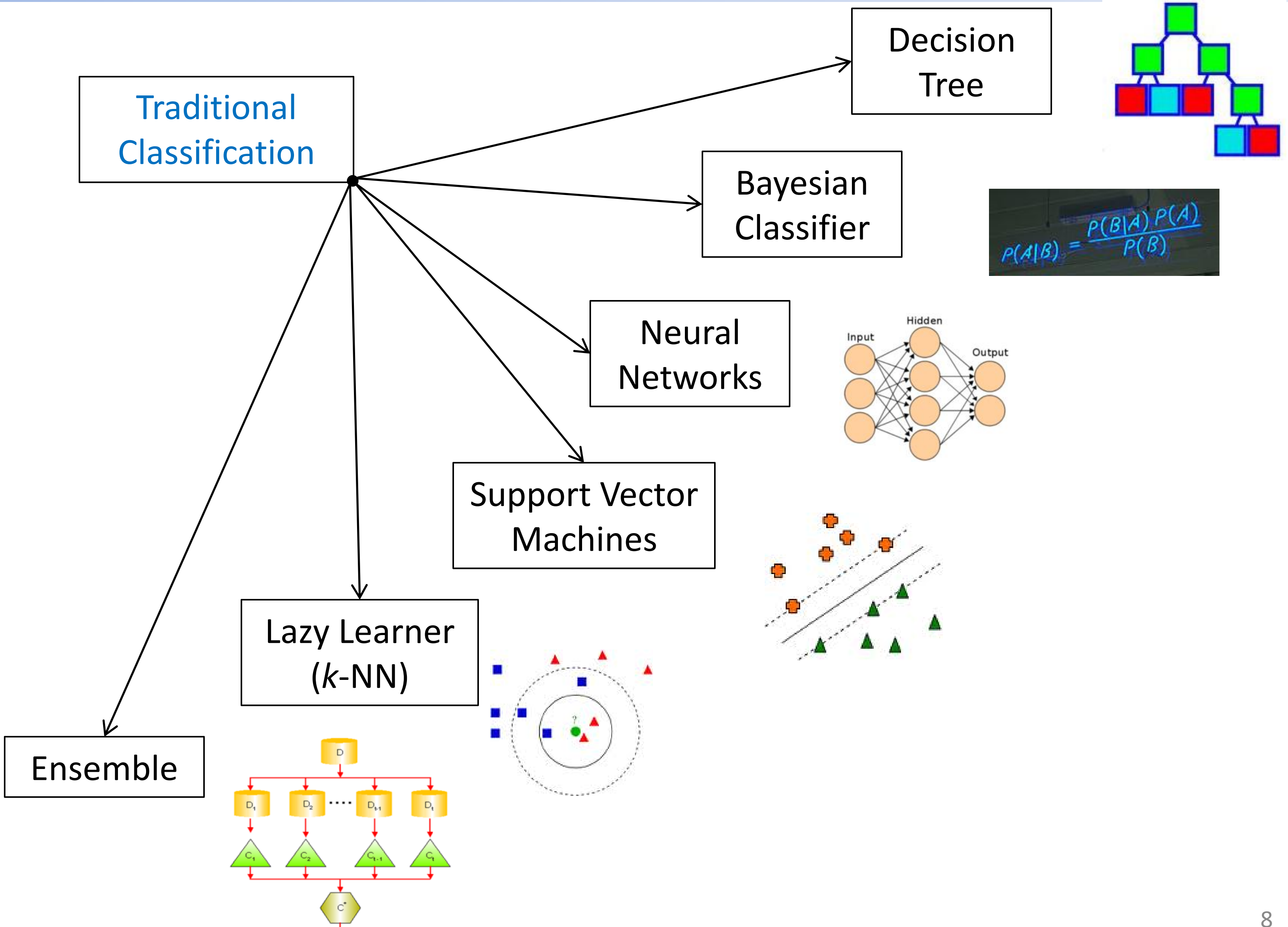


(a) Incremental Learning



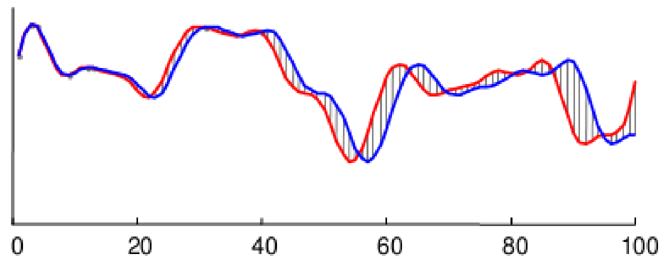
(b) Two-phase Learning

Traditional Classification

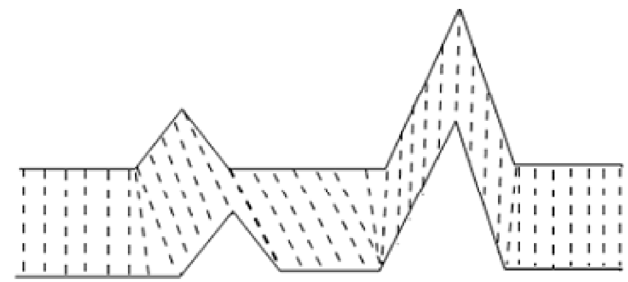


Time-series Classification

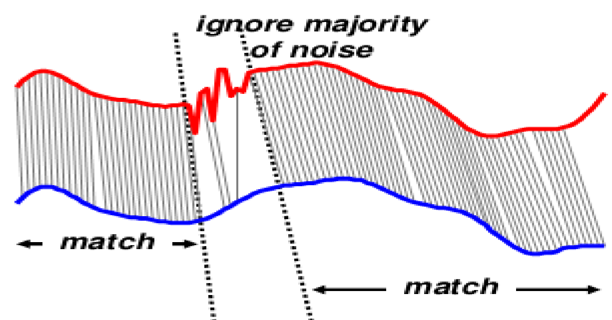
- Whole-series classification
  - Distance measures play an important role in whole-series classification
    - ✓ Euclidean distance



    - ✓ Dynamic Time Warping

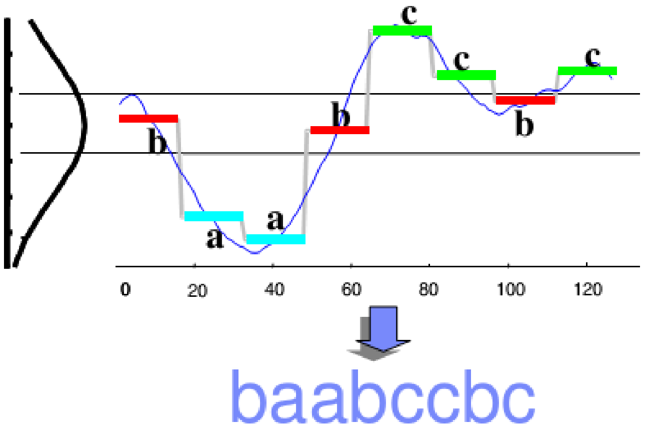


    - ✓ Longest Common Subsequence

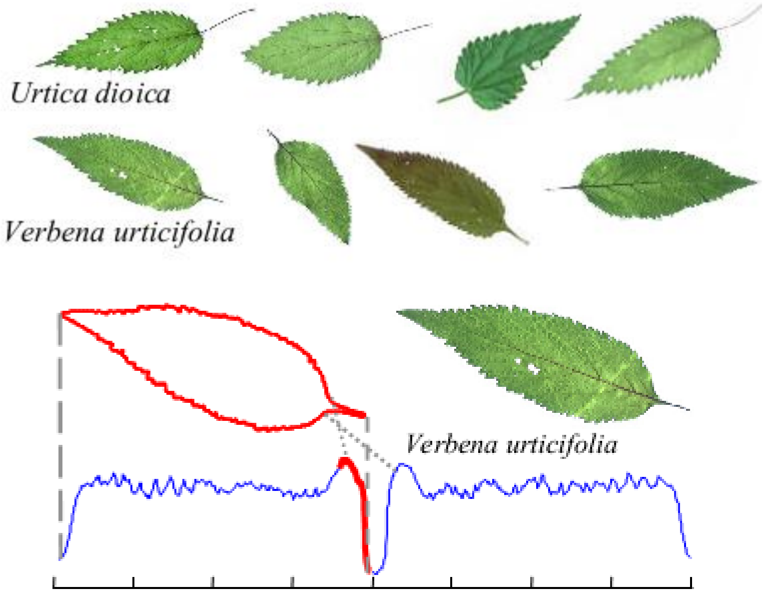


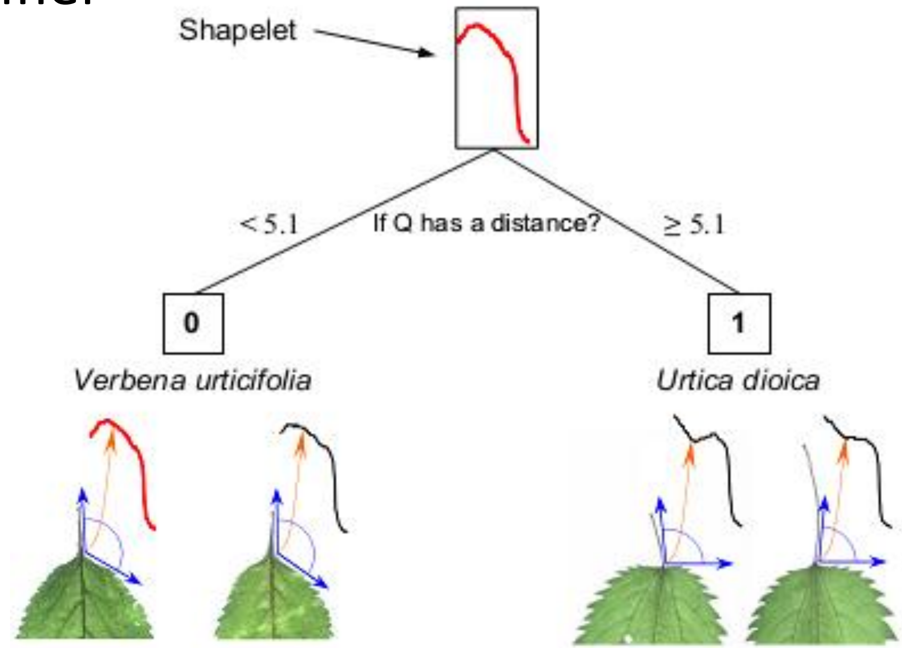
Time-series Classification

- Motif-based Classification
  - Motif Discovery:
    - Symbolic Aggregate approXimation (SAX)



- Example of Decision Tree classifier





Ye, L. and E. Keogh (2009). Time series shapelets: a new primitive for data mining, ACM.

Time-series Classification

Traditional classification

Decision Tree (DT)
Bayesian Classifier (BC)
Neural Network (NN)
SVM
Lazy Learner (kNN)
Ensemble (∑E)

Whole-series classification

DT	*TClass* [93]	Global features: mean, min, max, trend
BC	*mixed-state DBN* [95]	HMM + Bayesian network
NN	*PREMONNS* [94]	
SVM	*Zeus* [96]	Evolutionary computation for feature extraction
kNN	*SequentialSearch* [97] *FastAWARD* [7], *INSIGHT* [8] , *1NN-comparison* [6]	1NN with Euclidean distance, DTW, instance selection
∑E	*TEC* [100]	Data transformations

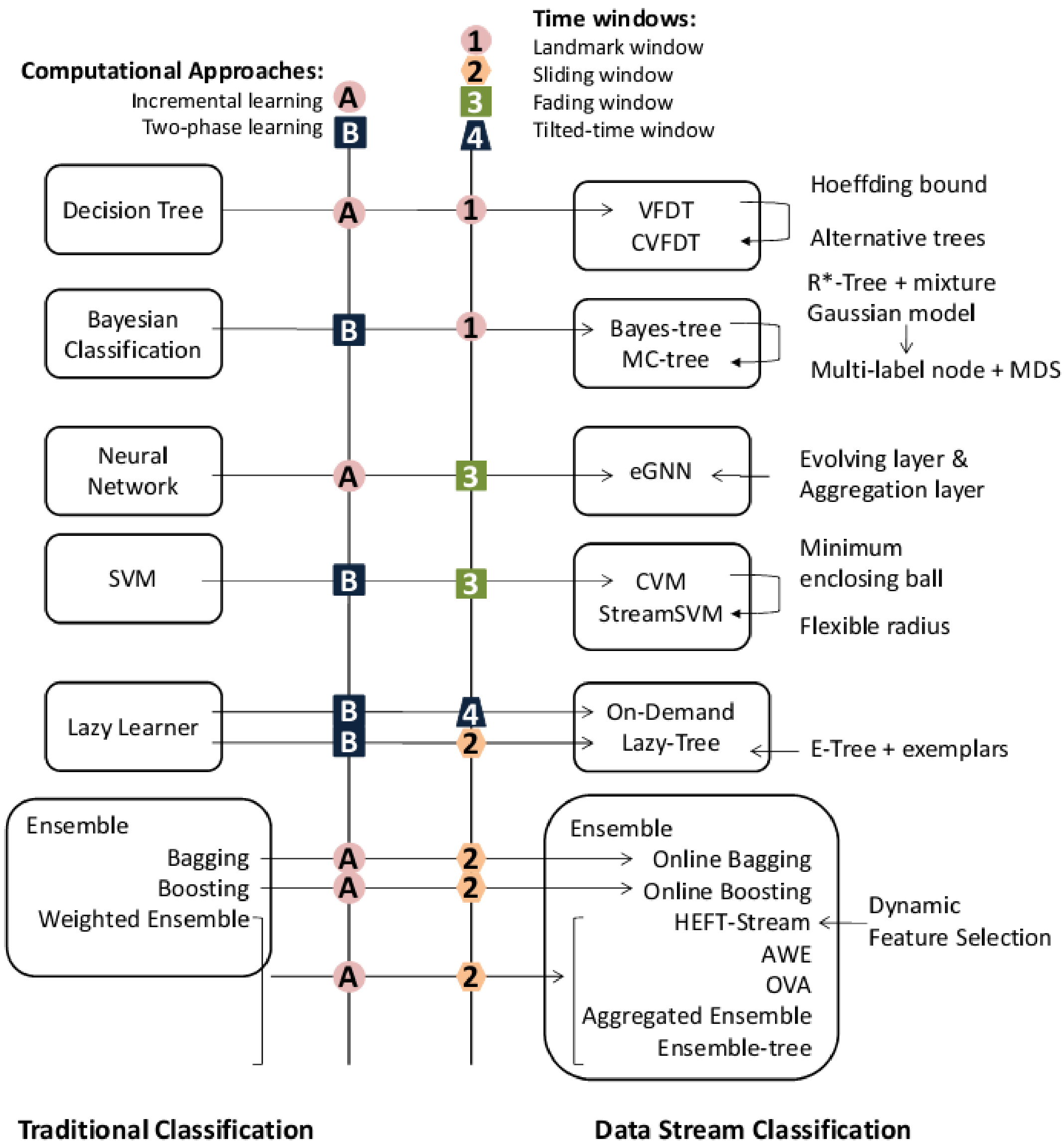
Motif discovery

SAX [3],
iSAX [44]

Motif-based classification

DT	*PCT* [117], *Shapelet* [2], *Logical-Shapelet* [9]
BC	*MCSC* [110], *BC+GSC-motifs* [120]
NN	
SVM	*MEX* [119], *SVM+GSC-motifs* [120]
kNN	
∑E	

Data Stream Classification



Algorithm	Bounded Memory	Single-pass	Real-time Response	Concept-drift Adaptation	Concept-drift Classification	High-dimensional Data
VFDT [10]	√	√	√			√
CVFDT [11]	√	√	√	√		√
Bayes tree [12]	√	√	√	√		
MC-tree [13]	√	√	√	√		
eGNN [83]	√	√		√		
CVM [14]	√	√		√		
StreamSVM [15]	√	√		√		
On-Demand [81]	√	√	√	√		
Lazy-Tree [16]	√	√	√	√		
Online Bagging & Boosting [17]	√	√	√			
AWE [18]	√	√	√	√		√
OVA [21]	√	√	√	√		√
Aggregated Ensemble [20]	√	√	√	√		
Ensemble-tree [22]	√	√	√	√		√

Capabilities of data stream classification algorithms.

Limitations

- Time series classification:
  - Most of the time series classifiers do not work in a streaming manner.
  - Many time series classifiers are based on finding motifs with predefined length.

Closed Motifs for Streaming Time Series Classification – **Chapter 3**

- Data stream classification:
  - Many data stream classifiers do not work well with high dimensional data.
  - Most data stream classifiers do not adapt well to different types of concept drifts.

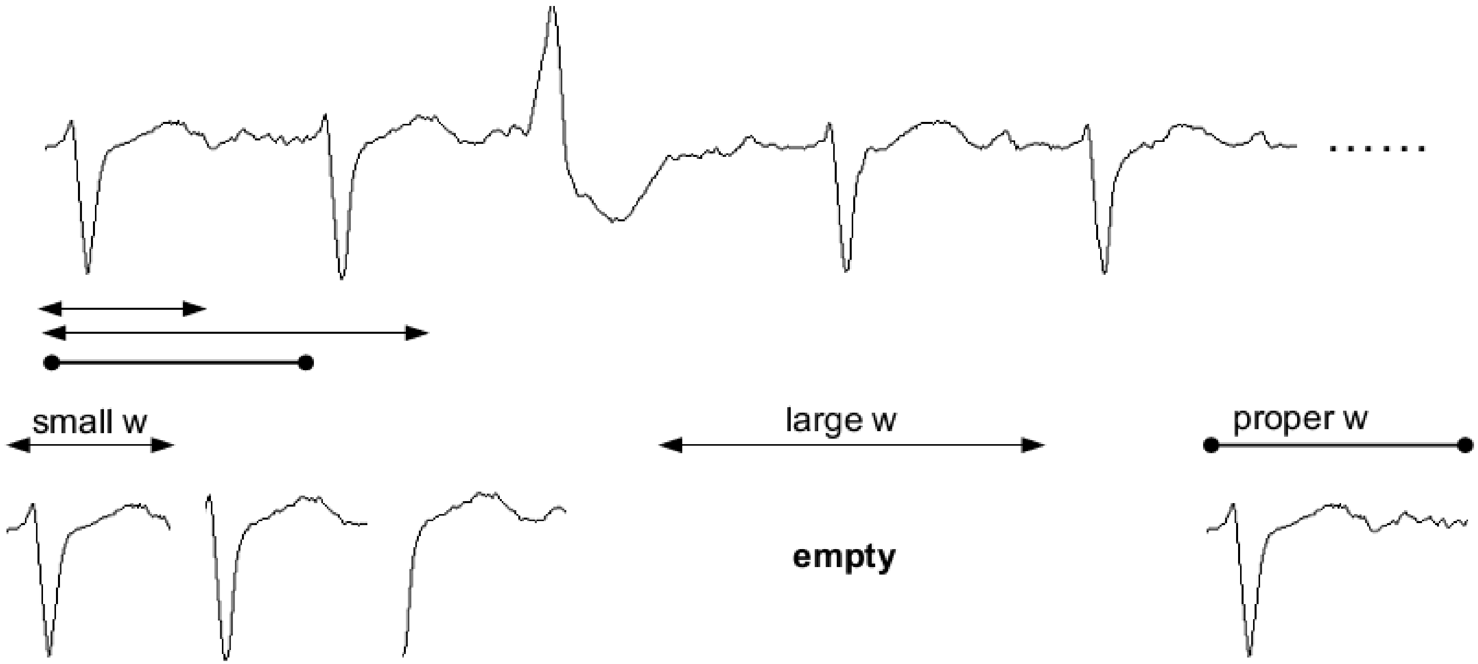
Ensemble Learning of Feature Drifts – **Chapter 4**

  - Most data stream algorithms focus on a single, stand-alone mining task.
  - Most data stream classifiers do not work well with very sparsely labeled datasets.

Concurrent Semi-supervised Learning – **Chapter 5**

Closed Motifs for Streaming Time Series Classification

- Closed motif is a frequent subsequence with no parent-sequence having the same number of occurrences.
  - Closed motifs can be considered as representatives for each class in time-series classification



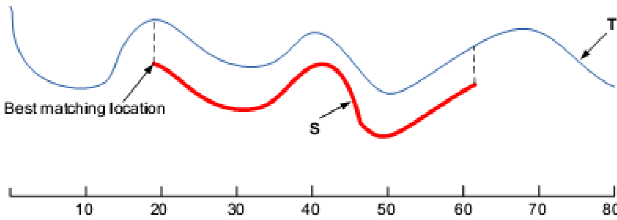
(a) small w with many redundant motifs

(b) large w with no motif found

(c) proper w with closed motif

Closed Motifs for Streaming Time Series Classification

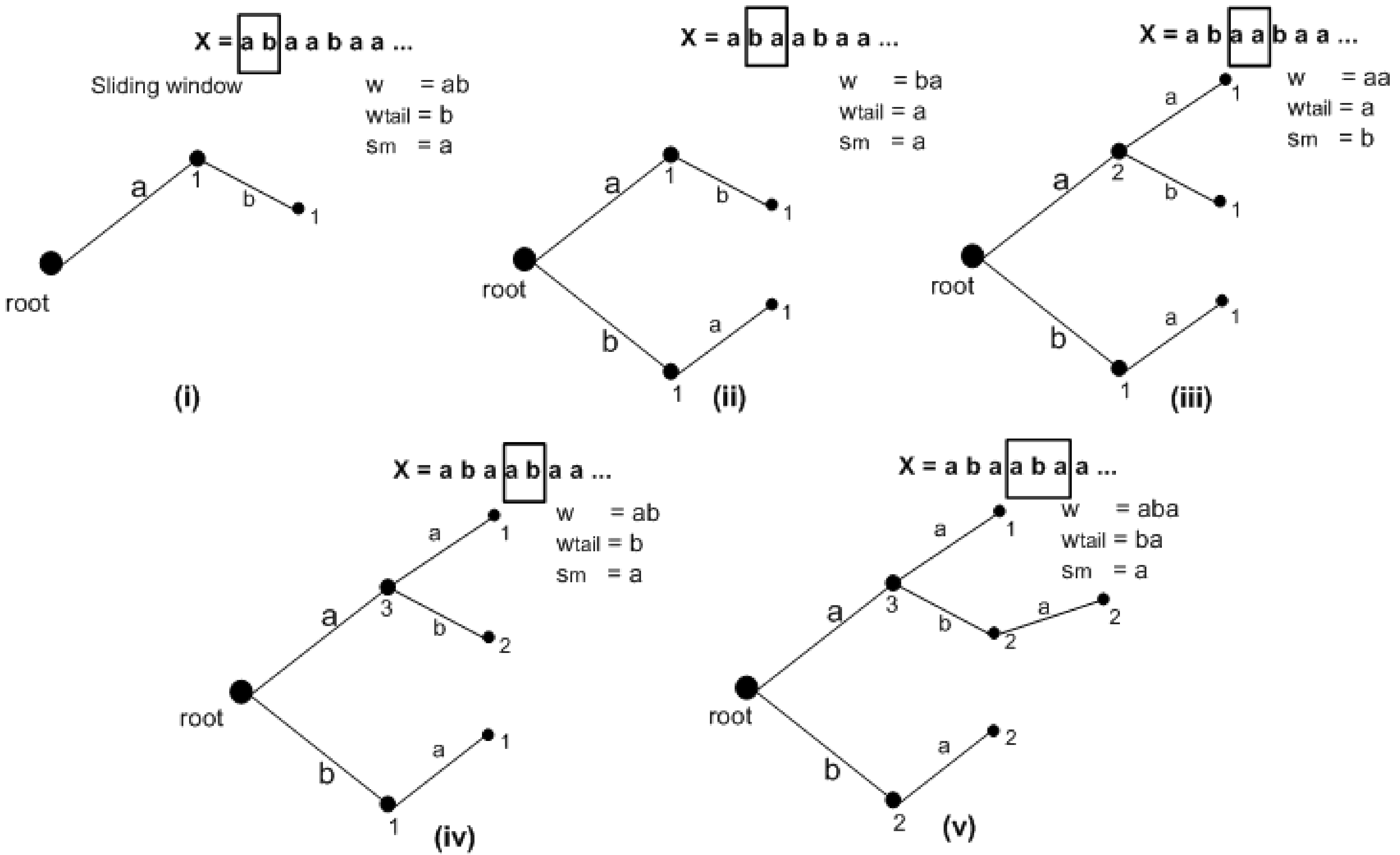
- Suffix Tree Construction
  - A variant of SAX with sliding window
  - Single pass + count updating with a probabilistic model
- Closed motifs discovery
  - Depth-first traverse the tree
- ARC-VIEW:
  - A visualization toolkit with color coding
  - Highlight closed motifs
- Closed Motifs for Classification
  - Rank closed motifs based on distinctive power (purity + coverage)
  - 1-NN classifier with subsequence distance



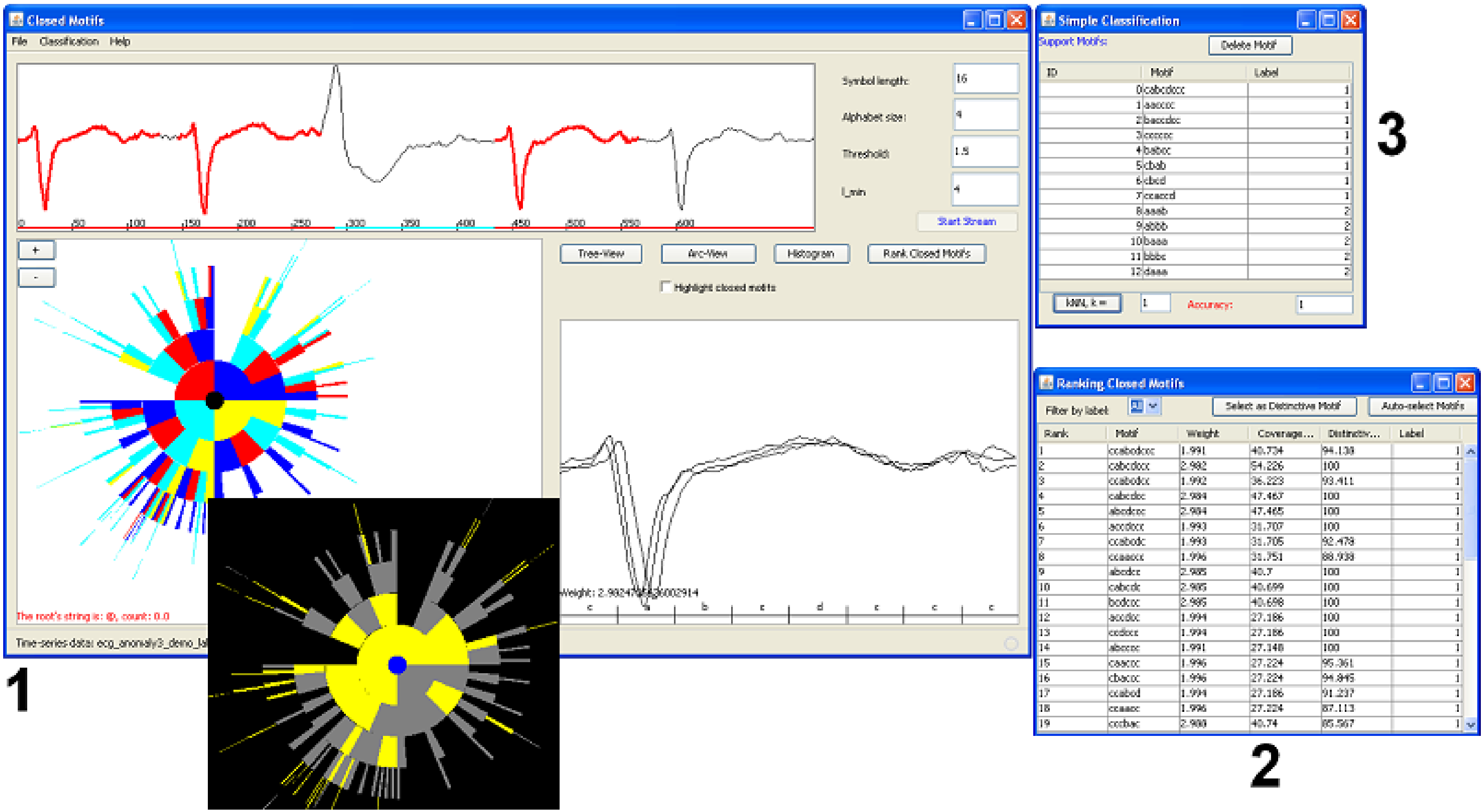
Suffix Tree Construction

Corollary: If a subsequence S is a motif, all subsequences of S are also motifs.

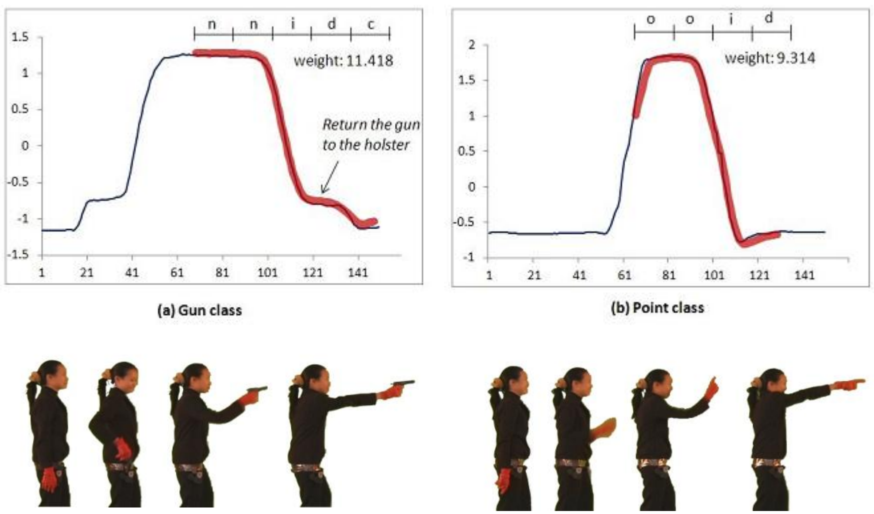
→ we only need to index a word $(a_1a_2 \ldots a_{k-1}\ a_k)$, if two of its subsequence $(a_1a_2 \ldots a_{k-1})$, $(a_2 \ldots a_{k-1}\ a_k)$ are motifs.



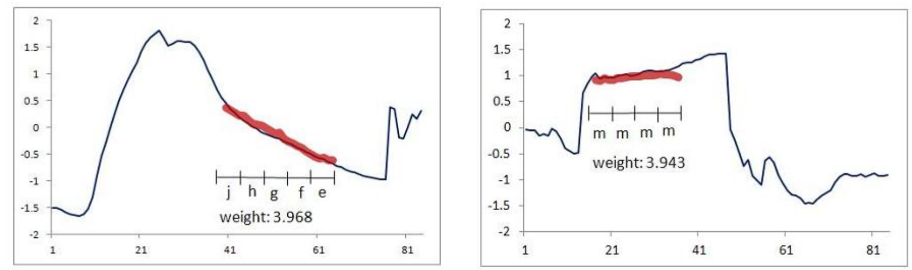
ARC-VIEW: a visualization toolkit



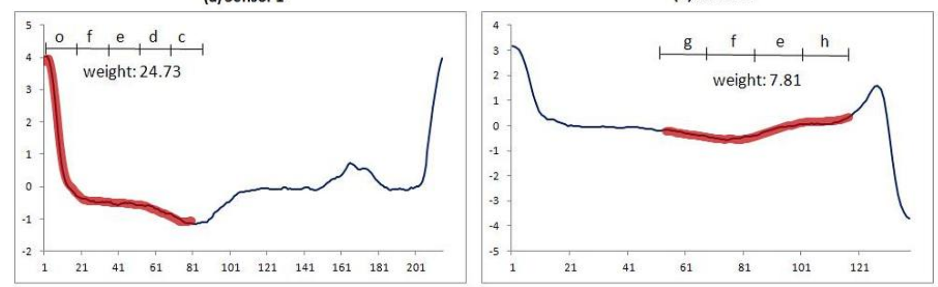
Experimental Results



(a) Gun class (b) Point class



(a) Sensor 1 (b) Sensor 2



(a) Normal ECG (b) Arrhythmia ECG

Dataset	#Classes	Size of training set	Size of testing set	Time series length	1-NN ED	Logical-Shapelet	1-NN closedMotif
Gun-Point	2	50	150	150	91.30%	93.30%	**94.00%**
MoteStrain	2	20	1252	84	87.90%	83.23%	**89.54%**
ArrhythmiaECG	2	100	113	285	85.50%	76.11%	**85.50%**
CBF	3	30	900	128	85.00%	88.56%	**99.67%**
TwoLeadECG	2	23	1139	82	74.70%	85.60%	**89.03%**
DiatomSizeReduction	4	16	306	345	93.5%5	75.16%	**96.41%**
FaceFour	4	24	88	350	78.40%	90.91%	**95.45%**
Trace	4	100	100	275	76.00%	**100.00%**	88.00%

Mueen, A., E. Keogh and N. Young (2011). Logical-shapelets: an expressive primitive for time series classification, 17th ACM SIGKDD, 2011.

Limitations

- Time series classification:
  - Most of the time series classifiers do not work in a streaming manner.
  - Many time series classifiers are based on finding motifs with predefined length.

Closed Motifs for Streaming Time Series Classification – **Chapter 3**

- Data stream classification:
  - Many data stream classifiers do not work well with high dimensional data.
  - Most data stream classifiers do not adapt well to different types of concept drifts.

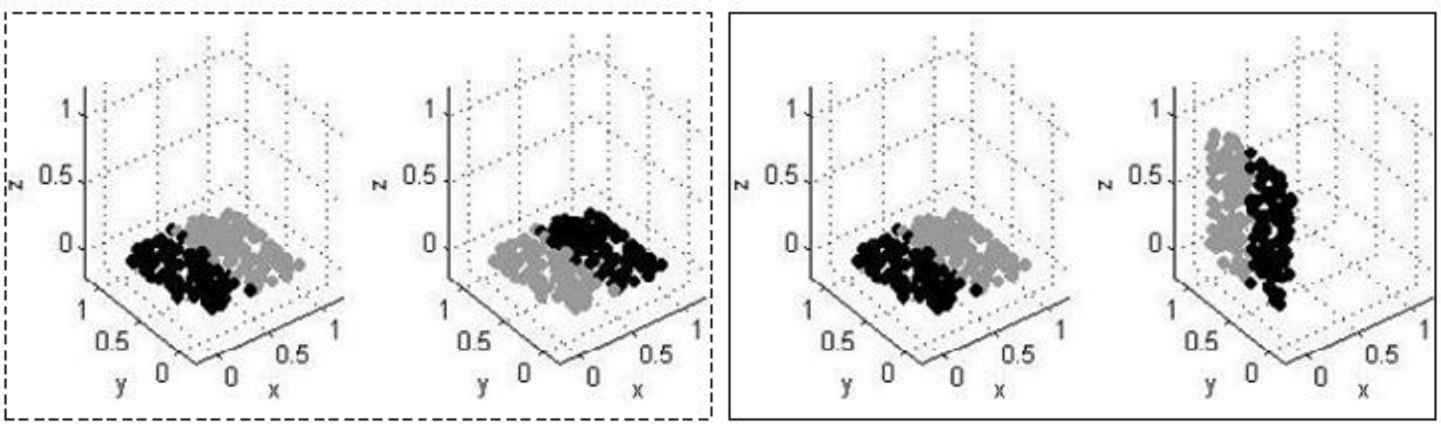
Ensemble Learning of Feature Drifts – **Chapter 4**

  - Most data stream algorithms focus on a single, stand-alone mining task.
  - Most data stream classifiers do not work well with very sparsely labeled datasets.

Concurrent Semi-supervised Learning – **Chapter 5**

Feature Drifts in Data Streams

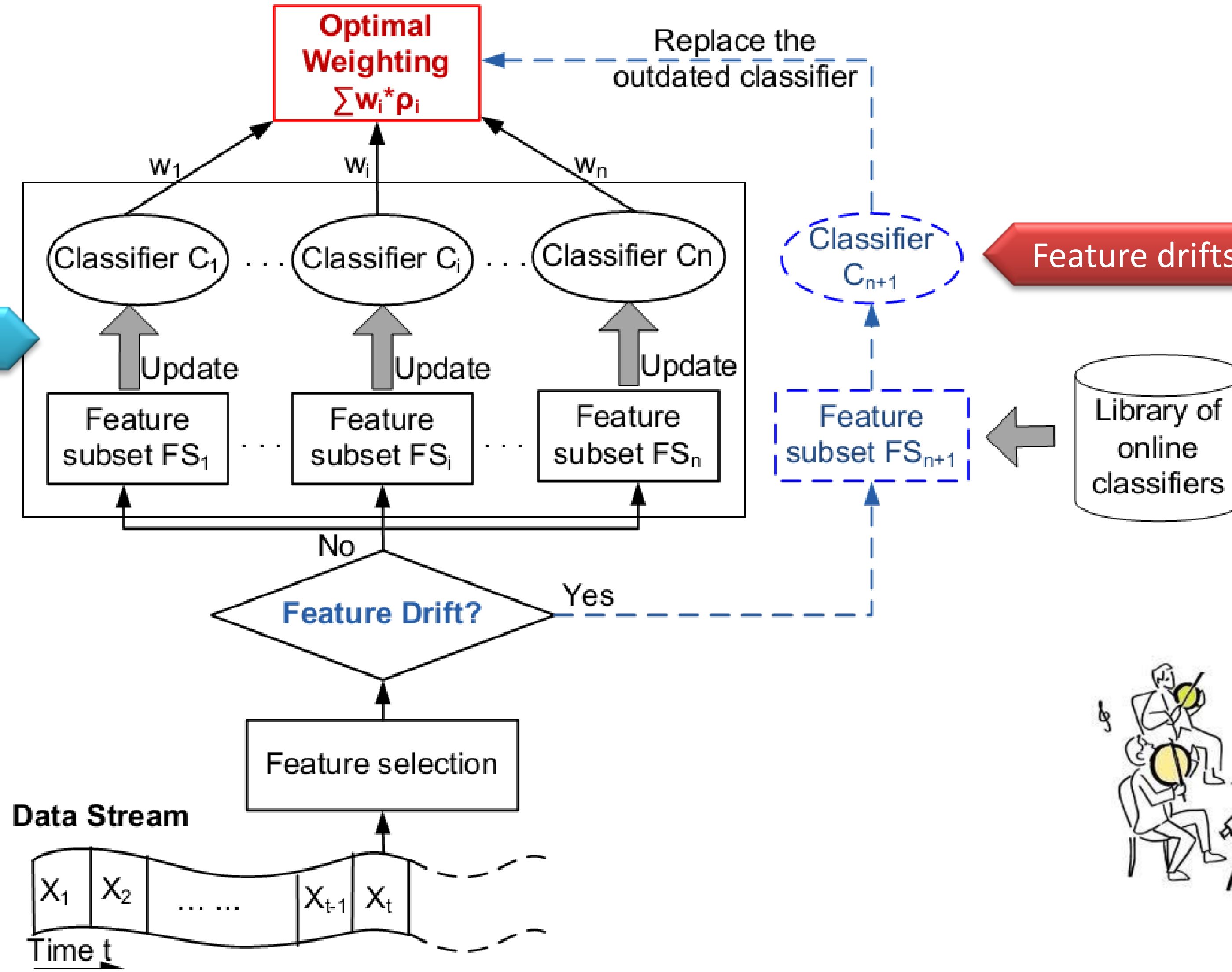
- Given a feature space F , at time point t , we can always select the most discriminative feature subset $F_t \subseteq$ F. If for any two time points i and j , $F_i \neq F_j$, we say that there is a feature drift.
- For example:
  - Personalized news filtering: the users' interests might change over time. Therefore, the set of discriminate words change over time , too.
- Feature drifts occur at a slower rate than concept drifts.



(a) Concept drift (b) Feature drift

Ensemble Learning of Feature Drifts

Gradual drifts



Feature drifts

Ensemble Learning of Feature Drifts

- Sliding window version of FCBF

$$SU(X,Y) = 2\left[\frac{H(X) - H(X|Y)}{H(X) + H(Y)}\right] = 2\left[\frac{I(X,Y)}{H(X) + H(Y)}\right]$$

  - When feature subsets of two consecutive windows are different, we state that a feature drift has occurred.

- Optimal weighting method

$$w_k = (E_{add}^k + \alpha)^{-1}\left[\sum_{m=1}^{N}(E_{add}^m + \alpha)^{-1}\right], \alpha \approx 0.001$$

Experimental Results

- HEFT-Stream:
  - Classifier members are CVFDT & Online Naive Bayes
  - HEFT-Stream-noFS: without feature selection
- Competitors: AWE(NB), AWE(C4.5), Bagging(OnlineNB), and Bagging(CVFDT)

Dataset	AWE(NB)		AWE(C4.5)		Bagging(OnlineNB)		Bagging(CVFDT)		HEFT-Stream-noFS		HEFT-Stream	
	Acc	Time	Acc	Time	Acc	Time	Acc	Time	Acc	Time	Acc	Time
SEA	88.08	6.61	88.12	26.08	87.91	2.9	89.12	21.47	88.95	21.46	**89.28**	22.65
HYP	87.94	19.42	72.72	27.40	86.92	8.07	88.90	40.41	86.95	39.04	**89.18**	12.42
LED	73.91	74.33	72.13	40.34	73.93	26.21	73.79	83.00	73.74	72.04	**74.07**	28.02
KDD'99	95.09	280.33	94.68	281.33	92.95	230.3	**97.75**	209.6	96.89	246.47	96.37	142.0
MNIST	9.87	2054.0	78.49	1246.7	9.87	1286.00	21.13	1456.00	19.65	1727	**79.36**	439.00
CRYST	53.70	40.38	83.30	147.00	54.28	57.63	76.18	101.33	53.97	10.11	**83.52**	37.10
Average	68.10	412.51	81.57	294.80	67.64	268.53	74.48	318.65	70.03	317.69	**85.30**	113.53

Limitations

- Time series classification:
  - Most of the time series classifiers do not work in a streaming manner.
  - Many time series classifiers are based on finding motifs with predefined length.

Closed Motifs for Streaming Time Series Classification – **Chapter 3**

- Data stream classification:
  - Many data stream classifiers do not work well with high dimensional data.
  - Most data stream classifiers do not adapt well to different types of concept drifts.

Ensemble Learning of Feature Drifts – **Chapter 4**

  - Most data stream algorithms focus on a single, stand-alone mining task.
  - Most data stream classifiers do not work well with very sparsely labeled datasets.

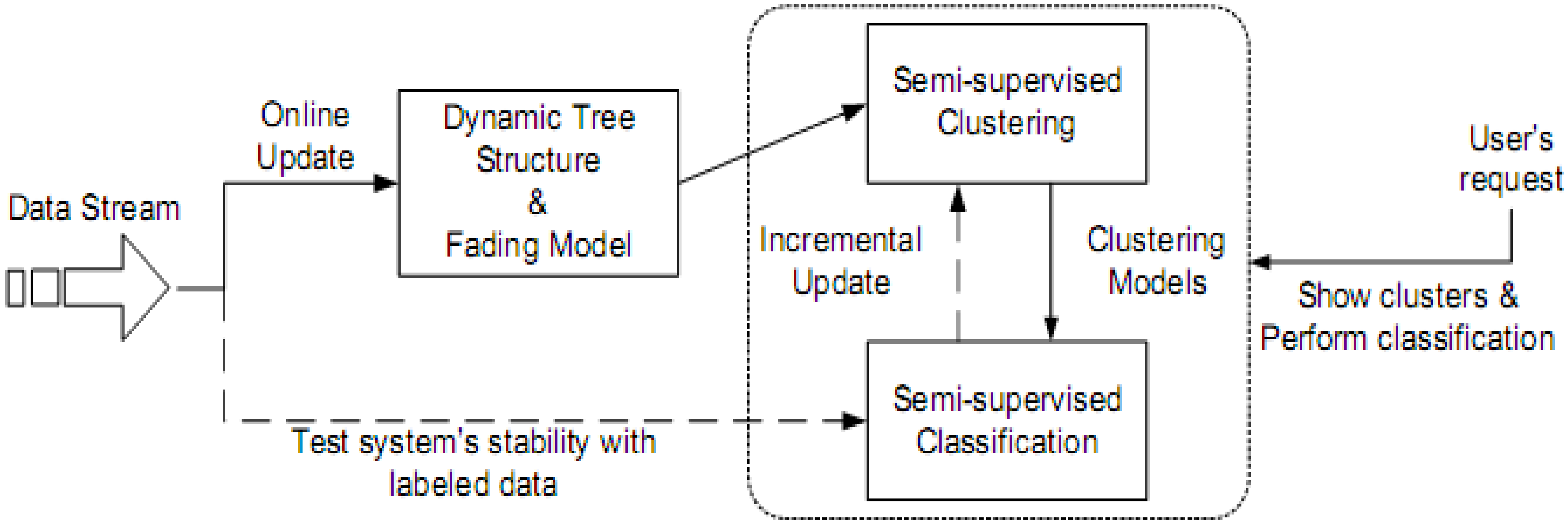
Concurrent Semi-supervised Learning – **Chapter 5**

Concurrent Semi-supervised Learning

- Why do we need concurrent mining?
  - To better exploit the data streams and maximize the system 's throughput
  - The knowledge gained by one mining task may also be useful to other mining tasks.

- Examples:
  - Web click streams: we want to cluster web-pages into similar topics & classify different kinds of users.
  - Health monitoring data: we want to group patients and predict diseases.

Concurrent Semi-supervised Learning



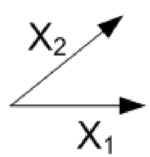
Bounded Memory	Single-pass	Real-time	Concept -drift
√	√	√	√

Dynamic Tree Structure

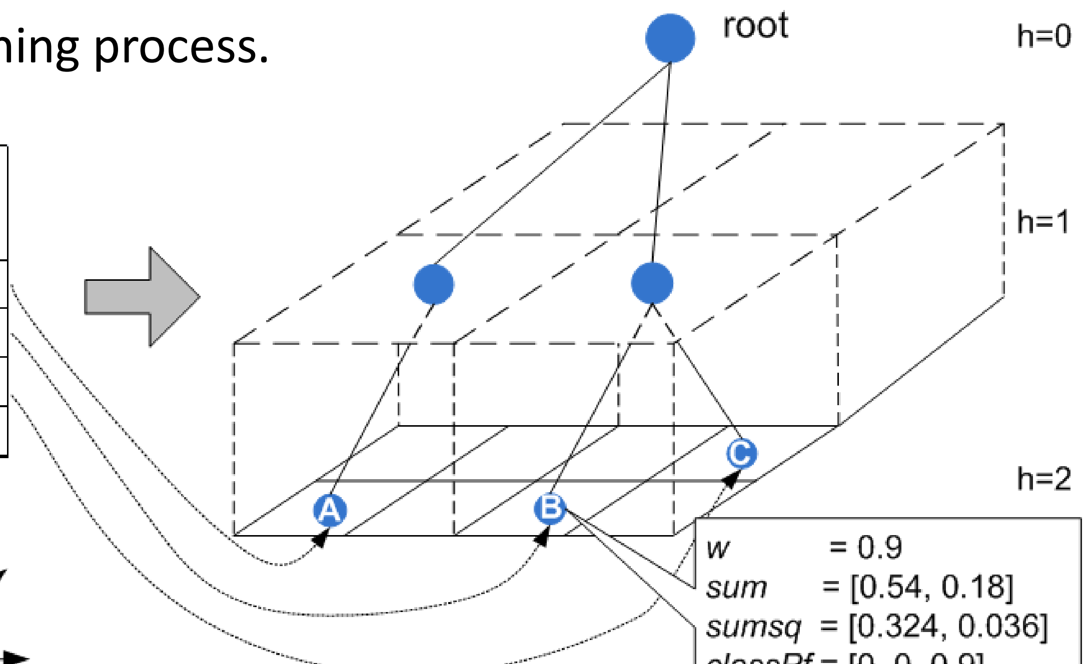
- Tree nodes store statistical summary of its belonging instances:
    - *weight, sum of coordinates, sum of squared coordinates, and class profile vector.*
- Pruning and merging processes: to remove redundant tree nodes.
    - Accelerate the mining process.

TID	X_1	X_2	label	Time stamp
tid_01	0.1	0.2	0	1
tid_02	0.6	0.2	2	2
tid_03	0.8	0.4	1	3
...	...	...	...	...

Fading parameter $\lambda = 0.9$



(a)

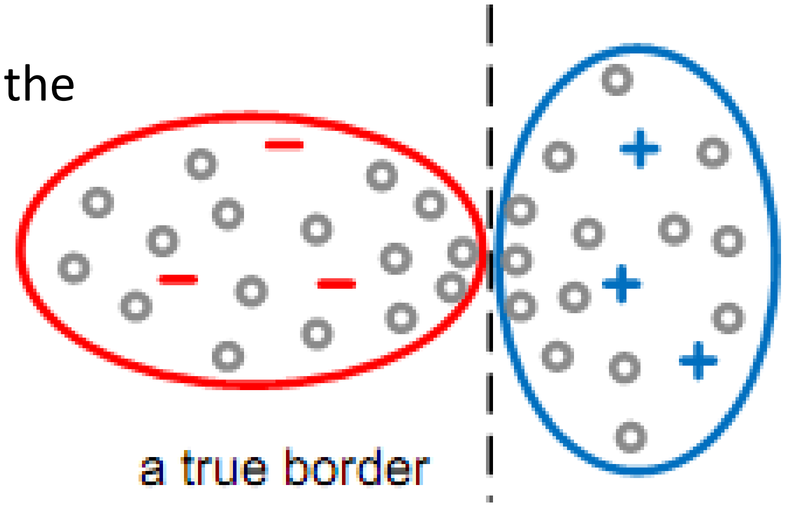


(b)

The synopsis tree for the 2-dimensional data stream with tree height = 2

Concurrent Semi-supervised Learning

- Density-based clustering:
  - The node’s weight is used as density. There are three types of nodes: dense nodes, sparse nodes, and transitional nodes.
  - Neighbor dense nodes are considered to put in the same cluster.
- Semi-supervised clustering:
  - Each cluster is labeled with the dominant class label.
  - Two clusters are merged if they have the same label.

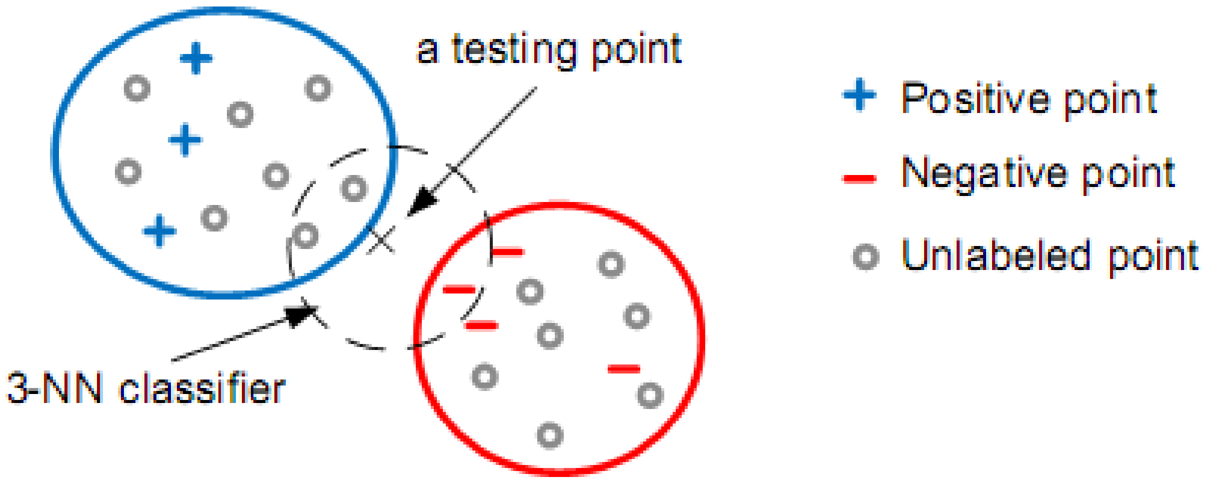


Concurrent Semi-supervised Learning

- Semi-supervised classification: exploits the clustering 's results
  - Find the closest cluster
  - If it is in the acceptable range of the cluster, assign the cluster's label to the testing instance.

distance (x, μ) ≤ θε,

where θ is a statistical range, mean $\mu = sum/w$, variance ε = $\sqrt{sumsq/w - \mu^2}$



Experimental Settings

- Datasets:

Name	#Instances	#Attributes	#Classes
RBF	100,000	10	5
HYP	100,000	10	2
LED	100,000	24	3
SHUTTLE	580,000	9	7
KDD'99	494,022	34	5
COVERTYPE	494,022	54	7

- Clustering:
  - Compare to D-Stream[1] and Alone-Clustering (CSL-Stream – class labels)
  - Quality measure: B-Cubed
- Classification:
  - Compare to SmSCluster[2] and Alone-Classification (CSL-Stream – clustering)
  - Quality measure: Accuracy

1. Chen, Y. and L. Tu (2007). Density-based clustering for real-time stream data. ACM SIGKDD.
2. Masud, M. M., G. Jing, et al. (2008). A Practical Approach to Classify Evolving Data Streams: Training with Limited Amount of Labeled Data. ICDM.

Experimental Results

- Clustering comparisons

	CSL-Stream		Alone-Clustering		D-Stream	
	Time	B-Cubed	Time	B-Cubed	Time	B-Cubed
RBF(10,0.001)	$\underline{10.64_{\pm 1.56}}$	$\mathbf{37.67_{\pm 4.93}}$	$17.22_{\pm 1.72}$	$36.27_{\pm 2.32}$	$17.37_{\pm 1.36}$	$17.39_{\pm 2.41}$
HYP(10,0.001)	$\underline{13.37_{\pm 1.23}}$	$\mathbf{65.24_{\pm 10.66}}$	$23.67_{\pm 1.66}$	$57.14_{\pm 4.01}$	$33.37_{\pm 1.46}$	$55.54_{\pm 4.66}$
LED	$\underline{53.9_{\pm 1.68}}$	$\mathbf{68.38_{\pm 11.74}}$	$205.61_{\pm 2.34}$	$19.1_{\pm 0.58}$	$203.8_{\pm 2.94}$	$19.3_{\pm 0.63}$
SHUTTLE	$\underline{1.37_{\pm 0.16}}$	$\mathbf{93.46_{\pm 1.04}}$	$1.37_{\pm 0.16}$	$89.07_{\pm 1.83}$	$1.45_{\pm 0.17}$	$88.1_{\pm 0.93}$
KDD'99	$39.78_{\pm 0.62}$	$\mathbf{76.89_{\pm 27.83}}$	$\underline{38.68_{\pm 0.84}}$	$76.88_{\pm 27.83}$	$53.79_{\pm 1.06}$	$73.5_{\pm 31.24}$
COVERTYPE	$\underline{130.24_{\pm 1.87}}$	$26.54_{\pm 9.68}$	$212.07_{\pm 2.45}$	$\mathbf{35.11_{\pm 10.79}}$	$152.46_{\pm 2.67}$	$12.55_{\pm 5.8}$

- Classification comparisons

	CSL-Stream		Alone-Classification		SmSCluster	
	Time	Accuracy	Time	Accuracy	Time	Accuracy
RBF(10,0.001)	$49.29_{\pm 2.35}$	$\mathbf{71.57_{\pm 6.37}}$	$53.32_{\pm 3.24}$	$44.57_{\pm 7.18}$	$\underline{41.45_{\pm 3.35}}$	$30.1_{\pm 12.38}$
HYP(10,0.001)	$\underline{15.78_{\pm 1.68}}$	$\mathbf{87.88_{\pm 1.88}}$	$16.17_{\pm 2.03}$	$70.66_{\pm 2.09}$	$40.18_{\pm 2.65}$	$76.05_{\pm 2.61}$
LED	$\underline{34.17_{\pm 2.16}}$	$\mathbf{72.73_{\pm 1.82}}$	$98.85_{\pm 3.12}$	$10.24_{\pm 1}$	$85.69_{\pm 3.87}$	$54.70_{\pm 3.45}$
SHUTTLE	$\underline{2.56_{\pm 0.35}}$	$\mathbf{98.3_{\pm 0.3}}$	$2.64_{\pm 0.91}$	$98.28_{\pm 0.31}$	$20.35_{\pm 2.61}$	$97.50_{\pm 0.49}$
KDD'99	$\underline{83.06_{\pm 2.47}}$	$\mathbf{98.06_{\pm 8.29}}$	$87.57_{\pm 2.13}$	$98.25_{\pm 8.24}$	$565.02_{\pm 3.87}$	$85.33_{\pm 33.39}$
COVERTYPE	$\underline{183.75_{\pm 3.05}}$	$\mathbf{81.63_{\pm 10.43}}$	$194.41_{\pm 3.46}$	$78.96_{\pm 9.39}$	$320.65_{\pm 3.02}$	$49.23_{\pm 15.42}$

Future work

- Cloud Computing
  - Cloud computing aims to provide computation power, data access, software, and storage services that are available anywhere, anytime, and on demand.
  - We plan to migrate our algorithms to cloud computing platforms.
- Mining Social networks
  - Social networks generate a large amount of different type data streams, such as, text data, multimedia data, and interactions.
  - Several research issues has been arisen, e.g., event detection, community detection, evolution analysis.
- Privacy-Preserving Data Mining
  - For streaming data, the privacy-preservation problem has not been effectively addressed as it is difficult to perform expensive privacy transformations, such as encryption or randomization.

Q&A