

Shapelets and Parallel Coordinates Based Automated Query Generation for Complex Event Processing

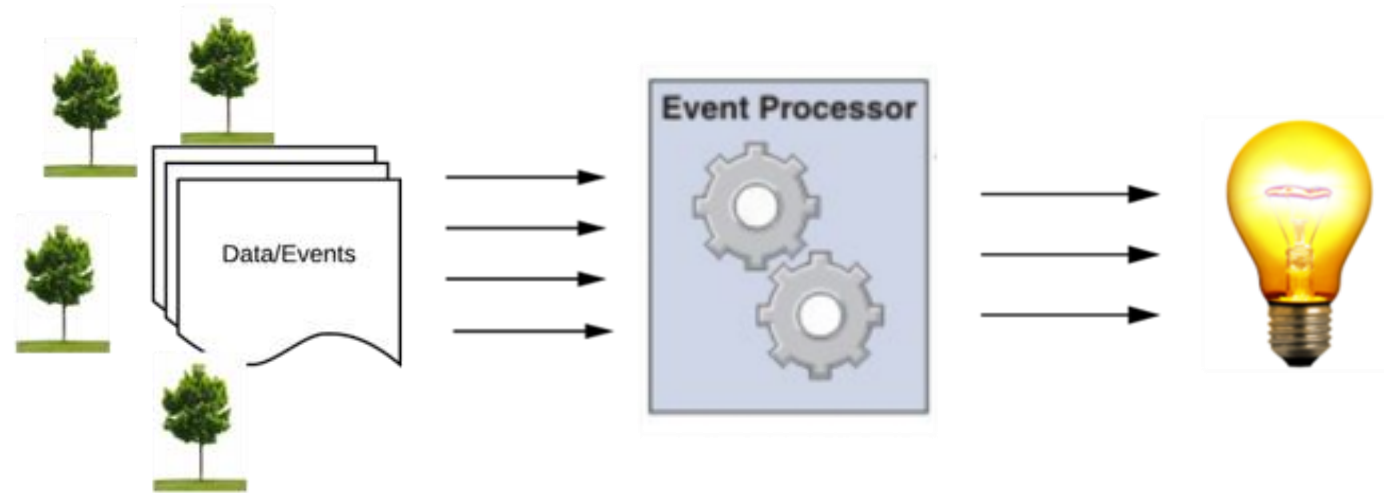
R.N. Navagamuwa, K.J.P.G. Perera, M.R.M.J. Sally, L.A.V.N. Prashan, and H.M.N. Dilum Bandara

(randika.12, pravinda.12, jaward.12, prashan.12, dilumb)@cse.mrt.ac.lk

Department of Computer Science and Engineering
University of Moratuwa
Katubedda, Sri Lanka

Motivation

Complex Event Processing



- Given a dataset consisting of temperature & smoke readings, how to generate a fire alert?

```
SELECT * WHERE TEMPERATURE ≥ 40 AND SMOKE > 800 WITHIN 0 ≤ TIME ≤ 5
```

CEP Query Processing

```
SELECT * WHERE TEMPERATURE ≥ 40 AND SMOKE > 800 WITHIN 0 ≤ TIME ≤ 5
```

- Which values should attributes carry?
- Do they need to appear in a specific, temporal order?
- How query can be optimized?
- When we a prepare query, above should be answered

But Domain Experts have to be there to answer ...

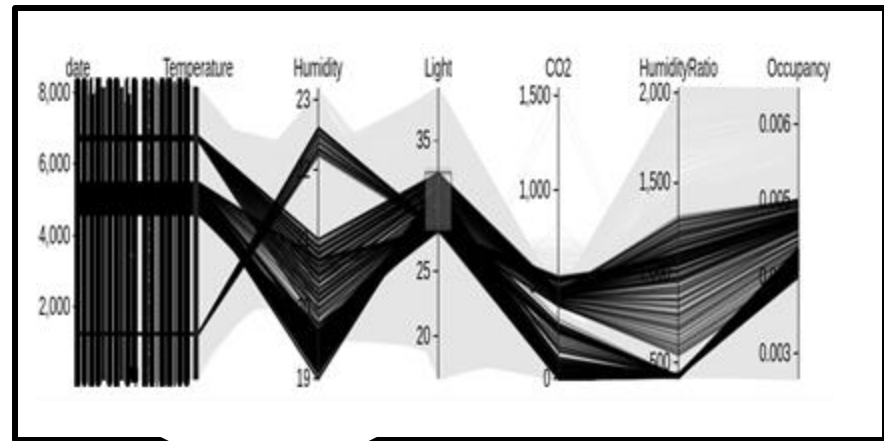
We Automate This Process

Proposed Solution

Architecture

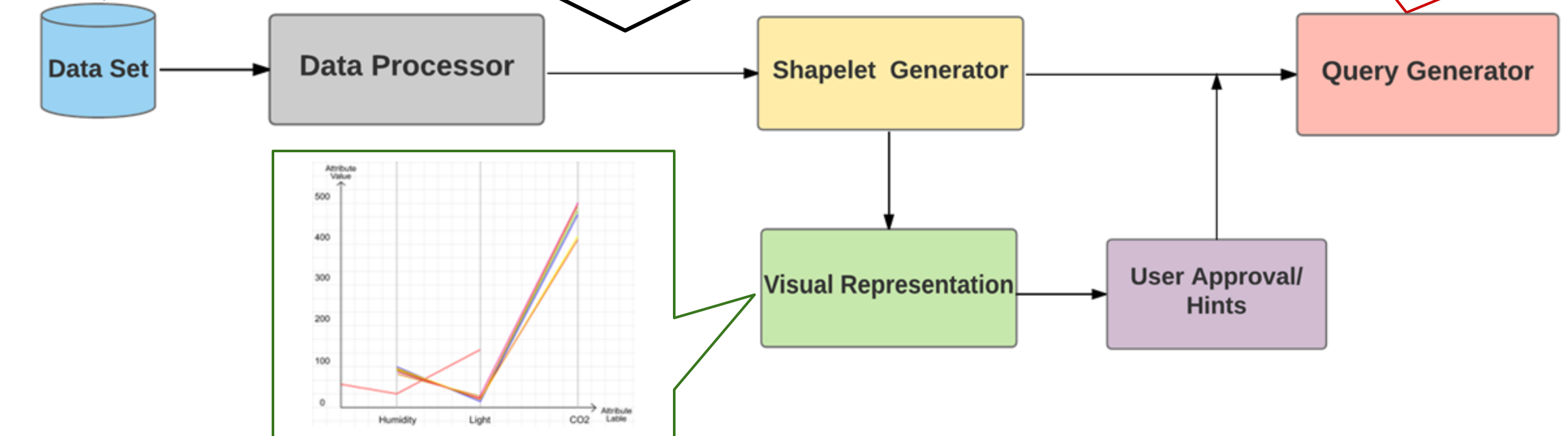
```

75,10634740109348464,26253,-1998,-156,65275,1074769,2564,6881,4635,9999,15,76
07,10634741571599146,20678,-581,191,124892,604942,-213,461,-3987,-4432,8930,785
06,10634742358491499,27811,-558,72,161687,1351392,-2988,-891,9501,5289,-5063,6638
03,1063474364818012,10219,-893,443,72231,73242,5443,-2400,6161,9919,1355,371
75,10634745876273309,26252,-1998,-156,75571,644954,51,8691,4844,9203,3733,1566
09,1063474559632893,27238,-481,1506,106878,609991,4375,1493,-8866,8973,7101,298
07,10634746480366437,26679,-582,187,140382,602466,2827,-30,-9395,9917,4201,-1011
05,106347479889949386,27488,-5849,161,37112,1200195,5291,-7071,-4683,8356,-5332,1336
06,1063474723043778,27811,-558,71,114779,1020279,-3542,-494,9331,5758,-5758,5693
07,1063474848037666,26259,-893,443,78332,975068,6943,-136,7176,9168,3897,-528
75,10634750844798352,26256,-1999,-162,79748,525563,-1258,9191,3732,6411,7635,771
09,1063475036056916,27238,-481,1518,67619,833931,-5863,5422,-9333,-7216,6609,1855
07,10634751465173986,26679,-582,187,132337,519334,2656,-589,-9023,9667,5152,488
07,10634751877841538,27488,-5847,154,163998,3111721,2532,6622,-7852,2219,9584,-1793
04,10634751971986811,28125,-1665,-228,52995,568329,556,-134,9979,-6799,-2439,7902
06,10634752182436822,27811,-558,72,188335,1460882,-756,758,9942,9158,-658,3897
05,10634753452646031,28219,-892,447,101683,568386,3694,-2367,8968,7979,-4452,4662
75,10634753812324236,26253,-1999,-157,75878,692212,445,9323,3817,8777,4763,-490
08,1063475588837792,27238,-481,1514,56179,242866,1991,6984,-7244,-1148,9148,3864
44,1063475554714829,25744,-1447,293,99742,1479951,-3302,-7284,8082,-9904,-367,759
59,10634755550373943,27859,-4108,1335,23036,2087700,-8348,2081,3896,-8887,9187,-1293
08,10634755873994189,25983,-477,389,23957,731019,3529,778,-9323,-2499,7267,-8388
07,10634758322381533,26888,-583,183,158627,730175,3833,-476,-9523,9690,1914,-1522
07,10634758649938124,27489,-5847,152,164169,1002357,4589,5554,-7868,5341,8827,-1942
106,10634756819918576,25352,-2752,275,111387,839988,1231,-5115,-8584,684,-9883,-4277
04,10634758822381533,26888,-583,183,158627,730175,3833,-476,-9523,9690,1914,-1522
06,1063475761478385,27811,-559,72,188892,1513717,-236,-1186,9926,9351,4628,3754
03,1063475717903878,27858,-130,-119,178302,2700132,3614,-9332,-1892,4149,-8999,1339
09,10634757470845739,27892,-4438,22,86584,1277958,-4849,-7167,5197,-2389,-9688,-1217
05,10634758350661935,26256,-892,446,65743,288493,-128,-1322,9411,-9471,3364,2169
08,1063475928688844,27238,-481,1514,51899,742866,1875,6528,-7339,-1188,9148,3864
75,1063479988188361,26248,-1098,-162,142124,1722577,-3790,9194,-1042,-4389,8837,-156
44,10634780315481787,25744,-1446,292,128859,1453876,437,-9599,2893,-954,-9824,1483
59,10634788414146099,27859,-4118,1161,93266,2767953,-8988,2081,4138,-9183,2726,-2937
68,10634790799662719,25981,-474,314,178688,2876539,-3549,9309,-1223,-4227,6644,-1973
    
```



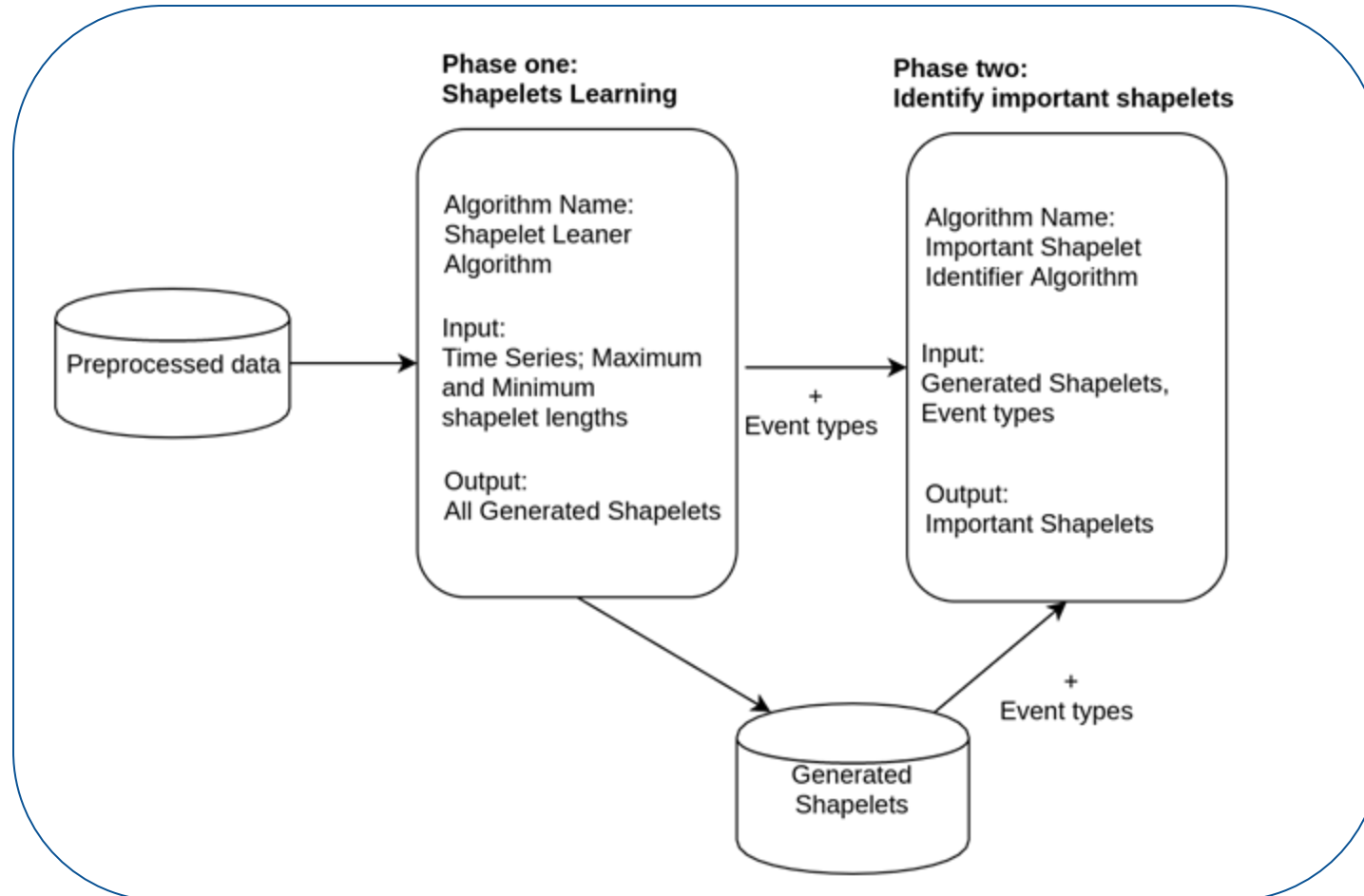
```

SELECT {*}
WHERE {attr1 ≥ a and attr2 < b}
WITHIN {t1 ≤ time ≤ t2}
    
```



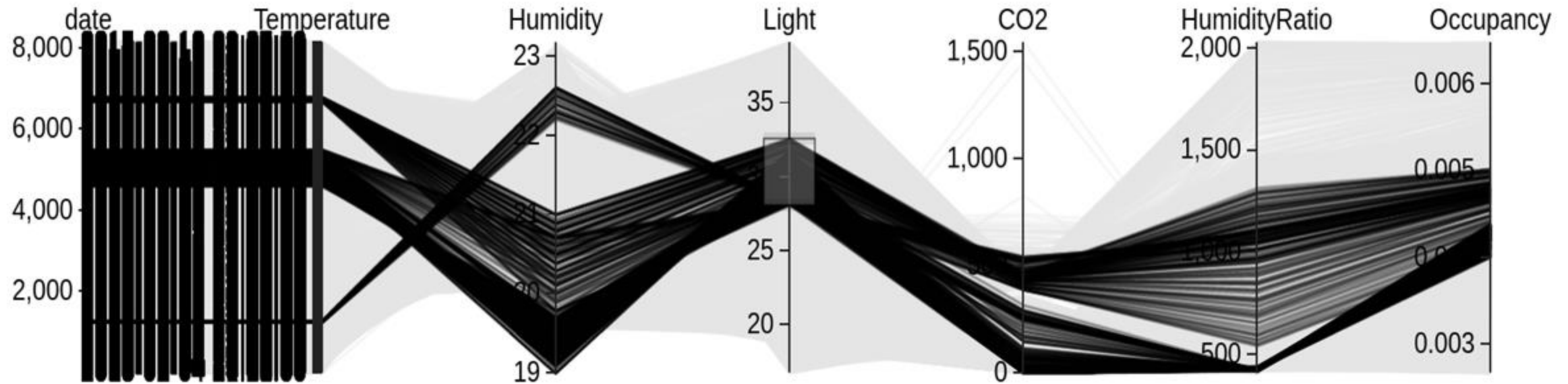
Research Methodology

Expanded Shapelet Generator

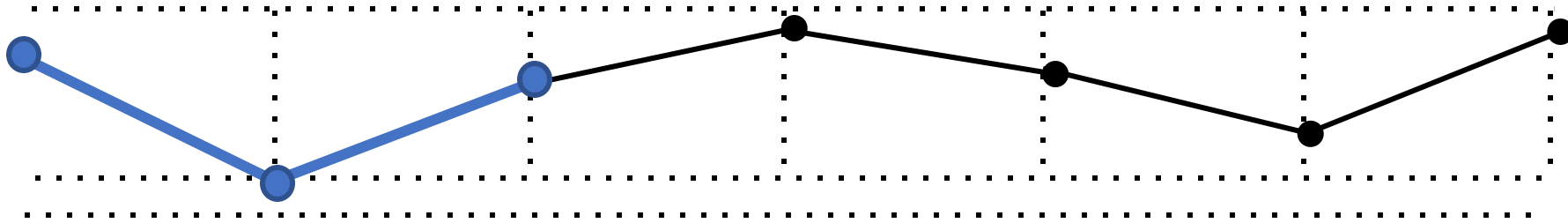


Shapelet Generation

- Parallel coordinates are used to generate all possible shapelets



Shapelet Learner



```
"Shapelet" : {  
    "seriesID" = 4,  
    "startPos" = 1,  
    "content"  =[0.8,0.1, 0.7],  
    "informationGain" = 0.6546,  
    "eventType" = 2  
}
```

Important Shapelet Identification

infoGain=0.901
eventType=1

infoGain=0.780
eventType=1

infoGain=0.654
eventType=0

infoGain=0.510
eventType=0

infoGain=0.480
eventType=0

infoGain=0.354
eventType=0

infoGain=0.189
eventType=1

Event distribution – Total Dataset
eventType 1 = 60%
eventType 0 = 40%

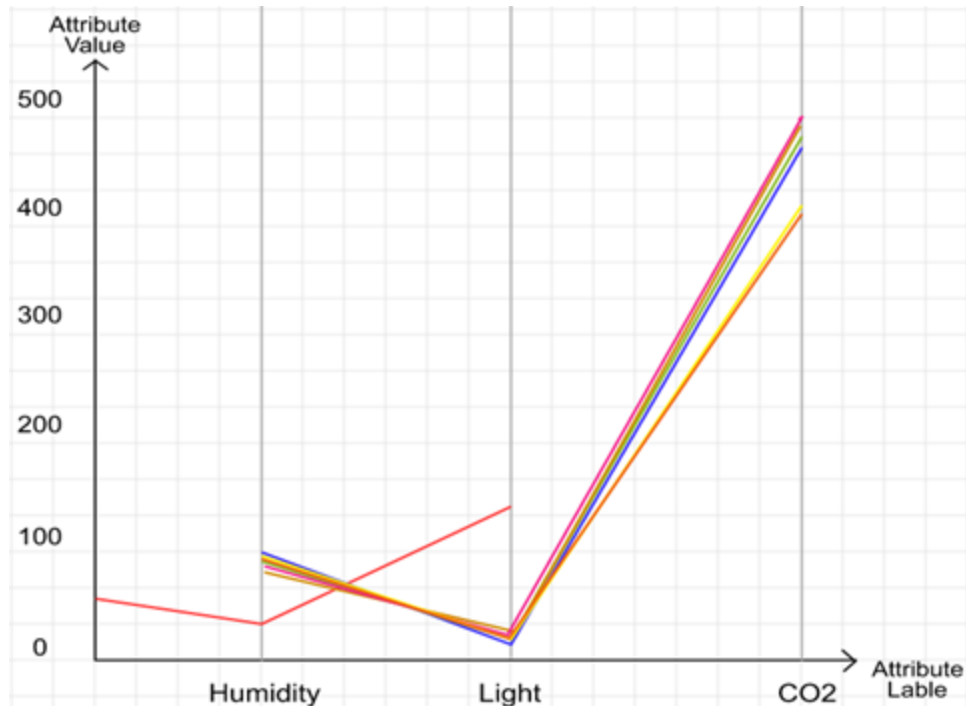
Probability of event 1 = 100%
Probability of event 0 = 0%

Probability of event 1 = 0%
Probability of event 0 = 100%

Probability of event 1 = 33%
Probability of event 0 = 66%

Visual Representation & Query Generation

- Generated important shapelets will be shown to the user
- User can simply approve them or select another set of shapelets as the important ones



Dataset Description

Name: Occupancy Detection Data Set

Number of instances: 20560

Number of attributes : 7

Data Set Characteristics: Multivariate, Time-Series

Source: <https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+>

Results

Dataset 1

Dataset Description

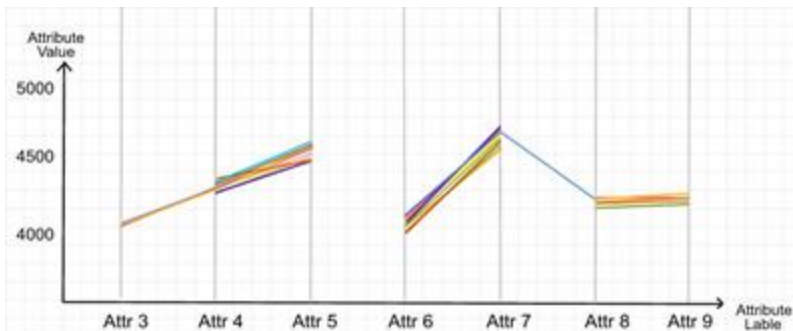
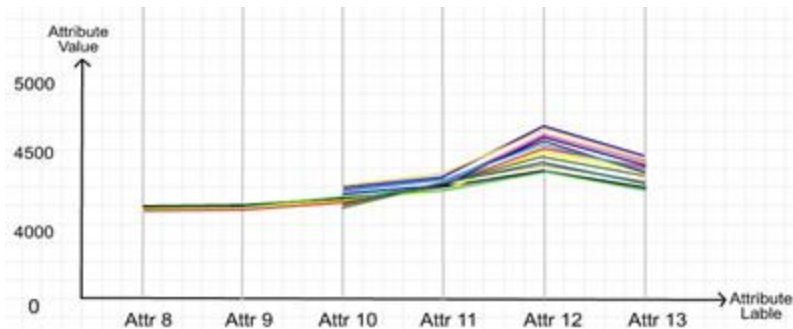
Name: Occupancy Detection Data Set

No of instances: 20,560

No of attributes: 7

Data Set Characteristics: Multivariate, Time-Series

<https://archive.ics.uci.edu/ml/datasets/Occupancy+Detection+>



Event	Metric	Value
Not Occupied	No of events in dataset	291
	No of events detected using CEP query	286
	Recall	98.3%
	Precision	100.00%
	False positives	0
	False negatives	5 (1.7%)
Occupied	No of events in dataset	196
	No of events detected using CEP query	196
	Recall	100.00%
	Precision	84.5%
	False positives	36 (18.4%)
	False negatives	0

Dataset 2

Dataset Description

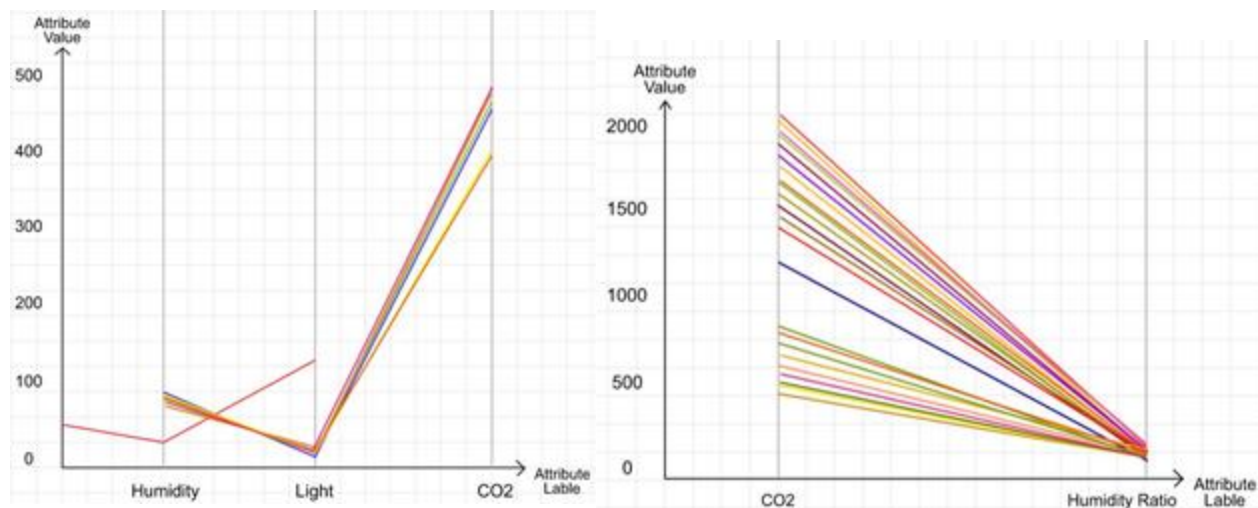
Name: EEG Eye State Data Set

No of instances: 14,980

No of attributes: 15

Data Set Characteristics: Multivariate, Sequential,
Time-Series

<http://archive.ics.uci.edu/ml/datasets/EEG+Eye+State>



Event	Metric	Value
Eye open	No of events in dataset	652
	No of events detected using CEP query	635
	Recall	97.39%
	Precision	100.00%
	False positives	0
	False negatives	17 (2.67%)
Eye closed	No of events in dataset	69
	No of events detected using CEP query	68
	Recall	98.55%
	Precision	100.00%
	False positives	0
	False negatives	1 (1.45%)

Computational Complexities

Algorithm	Time Complexity
Algorithm 1 - Shapelet Learner Algorithm	$O(nm^3)$
Algorithm 2 - Shapelet Merger Algorithm	$O(nm^2)$
Algorithm 3 - Important Shapelet Filter Algorithm	$O(n^{3/2}m^3)$

Summary

- A method to automate the query generation for CEP that combines
 - Parallel coordinates
 - Shapelets
 - Information gain
- Proposed methodology has high precision and recall
- It also has low computational and memory complexity
- In future we plan to extend proposed methodology to work with unannotated datasets

Q & A

Thank you!

Contact: pravinda.12@cse.mrt.ac.lk