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


(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
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 prepare query, above should be answered

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

We Automate This Process
Proposed Solution
Architecture

SELECT {},
WHERE {attr_1 ≥ a and attr_2 < b}
WITHIN {t_1 ≤ time ≤ t_2}
Research Methodology
Expanded Shapelet Generator

Phase one: Shapelets Learning
- Algorithm Name: Shapelet Learner Algorithm
- Input: Time Series; Maximum and Minimum shapelet lengths
- Output: All Generated Shapelets

Phase two: Identify important shapelets
- Algorithm Name: Important Shapelet Identifier Algorithm
- Input: Generated Shapelets, Event types
- Output: Important Shapelets
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**

- 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+

<table>
<thead>
<tr>
<th>Event</th>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not Occupied</td>
<td>No of events in dataset</td>
<td>291</td>
</tr>
<tr>
<td></td>
<td>No of events detected using CEP query</td>
<td>286</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>98.3%</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>False positives</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>False negatives</td>
<td>5 (1.7%)</td>
</tr>
<tr>
<td>Occupied</td>
<td>No of events in dataset</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>No of events detected using CEP query</td>
<td>196</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>84.5%</td>
</tr>
<tr>
<td></td>
<td>False positives</td>
<td>36 (18.4%)</td>
</tr>
<tr>
<td></td>
<td>False negatives</td>
<td>0</td>
</tr>
</tbody>
</table>
## 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

<table>
<thead>
<tr>
<th>Event</th>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye open</td>
<td>No of events in dataset</td>
<td>652</td>
</tr>
<tr>
<td></td>
<td>No of events detected using CEP query</td>
<td>635</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>97.39%</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>False positives</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>False negatives</td>
<td>17 (2.67%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Event</th>
<th>Metric</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye closed</td>
<td>No of events in dataset</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>No of events detected using CEP query</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Recall</td>
<td>98.55%</td>
</tr>
<tr>
<td></td>
<td>Precision</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>False positives</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>False negatives</td>
<td>1 (1.45%)</td>
</tr>
</tbody>
</table>
## Computational Complexities

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Time Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm 1 - Shapelet Learner Algorithm</td>
<td>$O(nm^3)$</td>
</tr>
<tr>
<td>Algorithm 2 - Shapelet Merger Algorithm</td>
<td>$O(nm^2)$</td>
</tr>
<tr>
<td>Algorithm 3 - Important Shapelet Filter Algorithm</td>
<td>$O(n^{3/2}m^3)$</td>
</tr>
</tbody>
</table>
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