Composite Event Detection and Identification for Wireless Sensor Network using General Hebbian Algorithm

Kamran Ali*, Talha Anwar*, Ijaz Haider NaqviT, M. H. JafryT

*TDepartment of Electrical Engineering, LUMS School of Science and Engineering, Pakistan
*Department of Computer Science and Engineering, Michigan State University, USA
TDepartment of Electrical Engineering, Rensselaer Polytechnic Institute, New York, USA
Outline

• Introduction
  - Outlier, Event Detection & Event Identification in WSNs
• Online & Adaptive Clustering for Outlier detection
• Event Detection
• Event Identification using General Hebbian Algorithm
• Event Report Packet format
• Complexity analysis
• Simulation and Experimental Results
Introduction

• The data obtained from WSN can be classified as:
  - Normal
  - Outliers-Anomalous data due to noise or sensor faults
  - Events

• The purpose is to separate outliers & events and determine the attributes involved in the event (Event Identification).

• Characteristics of detection and identification techniques in Harsh and Non-stationary environments:
  - Unsupervised, online, distributed and adaptive
  - Multivariate data
  - Spatio-Temporal and attribute correlations
  - Insusceptible to non-stationarity & dynamic network topology
Outlier Detection using Online and Adaptive Clustering

• A WSN with strong temporal, attribute & spatial correlations, where each node collects a total of $k$, $d$ dimensional data samples $X_k$.

• For each of the newly arrived $(k+1)$th data sample, each node updates the mean and covariance of data set $X_k$ using following equations (Eqs. 5 and 6 in paper).

$$m_{k+1,\lambda} = \lambda m_{k,\lambda} + (1 - \lambda)x_{k+1}$$

$$S_{k+1}^{-1} = \frac{kS_k^{-1}}{\lambda(k-1)} \times$$

$$\left[ I - \frac{(x_{k+1} - m_{k,\lambda})(x_{k+1} - m_{k,\lambda})^T S_k^{-1}}{k^{-1}} + \frac{(x_{k+1} - m_{k,\lambda})^T S_k^{-1}(x_{k+1} - m_{k,\lambda})}{\lambda} \right]$$

• Where $\lambda$ is called forgetting factor, used for environment adaptation.
Event Detection

• Detection of ‘$q$’ consecutive outlier constitutes an event. (For a sampling rate of 10 samples per second, we take $q = 20$ in our simulations)

Example of event detection
Event Identification

• General Hebbian Algorithm is used for iterative approximation of eigen vectors of covariance matrix.
• Data is then de-correlated by projecting onto the calculated eigen vectors to get weights $E_w$.
• These weights are then used to calculate event identification ratios $C_i$ for each i-th attribute.

$$C_i = \frac{E_w[i]}{\sum_{i=1}^{n} E_w[i]} \times 100$$
Event Report Packet (ERP) Format

• Efficient ERP format:
  - Minimizes communication overhead
  - Allows detection of both local and global events by in
    network aggregation of event reports.
  - Provides spatial confidence level, type and
    identification ratio of attributes of detected
    in the event
  - For a network of total 1024 nodes, 992 leaf nodes and
    32 cluster heads (assuming each node measuring 5
    attributes), each leaf and CH node needs just 51 and
    66 bits respectively to communicate the event
    information up the hierarchy.
# Complexity Analysis

<table>
<thead>
<tr>
<th>Proposed EDI approach</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Outlier Detection</td>
<td>$O(nd^2)$</td>
</tr>
<tr>
<td>Projection along $d$ dimensions</td>
<td>$O(nv)$</td>
</tr>
<tr>
<td>Clustering along $d$ dimensions</td>
<td>$O(nv)$</td>
</tr>
<tr>
<td>Decision for outlier and event</td>
<td>$O(1)$</td>
</tr>
<tr>
<td>Total computational complexity of proposed technique</td>
<td>$O(nd^2 + 2vn + imax * nv) \approx O(nd^2)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QS-SVM Techniques</th>
<th>Computational Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-QS-SVM [17]</td>
<td>$O(n^2)$</td>
</tr>
<tr>
<td>STA-QS-SVM [10]</td>
<td>$O(n^2 + nd^2)$</td>
</tr>
<tr>
<td>STA-TASV,STA-TSV [11]</td>
<td>$O(n^2 + nd^2)$</td>
</tr>
<tr>
<td>STA-CA [11]</td>
<td>$O(n^2 + nu d^2)$</td>
</tr>
</tbody>
</table>

**TABLE I.** COMPARISON OF COMPUTATION COMPLEXITY WITH QS-SVM BASED ALGORITHMS

<table>
<thead>
<tr>
<th>Proposed EDI approach</th>
<th>Communication Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum communication overhead for parent node close to base-station</td>
<td>$O(N_e)$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>QS-SVM Techniques</th>
<th>Communication Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST-QS-SVM [17]</td>
<td>$O(n)$</td>
</tr>
<tr>
<td>STA-QS-SVM [10]</td>
<td>$O(n)$</td>
</tr>
</tbody>
</table>

**TABLE II.** COMPARISON OF COMMUNICATION OVERHEAD WITH QS-SVM BASED ALGORITHMS

$N_e = \text{number of events}, \ n = \text{number of outliers}$
Simulation Results

Node 1 Outlier Detection

Node 3 Outlier Detection

Event identification results for Node 1

Event identification results for Node 3
## Simulation Results

<table>
<thead>
<tr>
<th>Technique</th>
<th>Detection Rate</th>
<th>False Positive Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed EDI approach [17]</td>
<td>98.88%</td>
<td>0.01%</td>
</tr>
<tr>
<td>ST-QS-SVM [17]</td>
<td>16.67%</td>
<td>10.85%</td>
</tr>
<tr>
<td>STA-QS-SVM [10]</td>
<td>91.67%</td>
<td>0.5%</td>
</tr>
<tr>
<td>STA-TASV [11]</td>
<td>92.45%</td>
<td>0.48%</td>
</tr>
<tr>
<td>STA-TSV[11]</td>
<td>99%</td>
<td>0.9%</td>
</tr>
<tr>
<td>STA-CA [11]</td>
<td>91.67%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

**Table III.** Comparison of Detection and False Positive Rates with QS-SVM based algorithms
Hardware Implementation and Results

• Two CC430 based WizziMotes running DASH7 protocol were used.
• Temperature and Illumination attributes were tested.
• Tested using white LED and fluorescent lights.

- **Scenario 1** (turning off the lights of the room in which the motes were deployed) **1.286%** and **98.814%** contribution for temperature and illumination respectively were reported.

- **Scenario 2** (illuminated two incandescent bulbs in a close vicinity of both the motes) **74.87%** and **25.13%** contribution for temperature and illumination respectively were reported.
Conclusion and Future work

• Proposed scheme gives accurate identification ratios in cases where both independent as well as correlated attributes contribute towards an event.

• Future work entails:
  - Implementation in a larger WSN framework
  - More thorough evaluations of our proposed schemes in terms of required memory resources and the overall network lifetime.