Complex Event Detection and Prediction in Traffic

Blaž Kažič, Dunja Mladenić, Luka Bradeško

AI Lab, Jozef Stefan Institute,
Jozef Stefan International Postgraduate School
Ljubljana, Slovenia
Motivation

- Develop Complex Event Processing system (CEP) that could detect (and predict) complex events in traffic

- Background knowledge required:
  - From domain expert
  - Using data mining techniques

![Diagram of Complex Event Processing system](image)
Complex events in traffic

- Most traffic is caused by predictable events such as working schedules
  - Known bottlenecks, School schedules, Holidays,…

- Research how bigger social events affects on traffic nearby – Extract rules
Outline

• Data description
  o Description of used data sources

• Data preprocessing:
  o Complex event detection
  o Database of extracted complex events

• Data mining results:
  o Analyze how social events affect nearby traffic
  o Find correlations between different complex events
  o Find rules to predict complex traffic events
Data description

• Traffic Counters
  o Flow (vehicles/hour)
  o 5 min stream
  o 2011-2013

• Social events
  o 50 major social events
  o Stadium Stozice
  o Demand (1-4)

• Parking sensors
  o Parking availability count
Complex event detection

Traffic data

- Complex event is considered as the start of anomaly
- Local averages are used to described “normal” traffic
- Anomaly: if RMSE is larger than certain threshold

Traffic complex event: “t – 60”
Complex event detection

Parking data

• Parking sensor did not have specific daily pattern

Traffic complex event: “t – 60”

• Complex event in parking sensor is considered as start of “downward trend” in parking availability

Parking complex event: “t – 90”
Complex event detection

11.9.2013: lat-ukr(14:30), bel-srb(17:45), litu-fra(21:00)
Traffic Data

18.9.2013: srb-esp(17:30), slo-fra(21:00)
Traffic Data

22.9.2013: esp-cro(17:30), fra-lit(21:00)
Traffic Data

11.9.2013: lat-ukr(14:30), bel-srb(17:45), litu-fra(21:00)
Parking Data

18.9.2013: srb-esp(17:30), slo-fra(21:00)
Parking Data

22.9.2013: esp-cro(17:30), fra-lit(21:00)
Parking Data

11.9.2013: lat-ukr(14:30), bel-srb(17:45), litu-fra(21:00)
Complex Events

18.9.2013: srb-esp(17:30), slo-fra(21:00)
Complex Events

22.9.2013: esp-cro(17:30), fra-lit(21:00)
Complex Events
**Extracted data set**

<table>
<thead>
<tr>
<th>Event Description</th>
<th>Date</th>
<th>Hour</th>
<th>Visitors</th>
<th>Demand</th>
<th>Parking Sensor</th>
<th>Traffic Sensor</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLO - UKR</td>
<td>21.09.2013</td>
<td>21:00</td>
<td>10000</td>
<td>...</td>
<td>4</td>
<td>t-90</td>
</tr>
<tr>
<td>ESP - CRO</td>
<td>22.09.2013</td>
<td>17:30</td>
<td>6050</td>
<td>2</td>
<td>t-90</td>
<td>t-30</td>
</tr>
<tr>
<td>FRA - LTU</td>
<td>22.09.2013</td>
<td>21:00</td>
<td>10000</td>
<td>4</td>
<td>t-30</td>
<td>t-90</td>
</tr>
<tr>
<td>Elton John</td>
<td>11.11.2011</td>
<td>21:00</td>
<td>8000</td>
<td>3</td>
<td>?</td>
<td>t-60</td>
</tr>
</tbody>
</table>

- **Event demand attribute**: 4 possible values (1-4)
- **Parking sensor attribute**: 5 possible values (“no” – “t-90”) 
- **[Target] Traffic sensor attribute**: 5 possible values (“no” – “t-90”) 

- **Number of instances**: 50 (number of events)

---

*UCI Machine Learning Repository – Dogers Loop Sensor Data Set (81 events)*

Decision tree

- Decision Tree (*Wekas J48) - Pruned

- Target attribute: Traffic
- Evaluation:

Correctly Classified Instances: **76.6 % (36)**
Incorrectly Classified Instances: **23.4 % (11)**

Baseline Classification (ZeroR): 38.3%

--- Confusion Matrix ---

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>--- classified as</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>a = no</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>b = t-30</td>
</tr>
<tr>
<td>c</td>
<td>1</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>c = t-60</td>
</tr>
<tr>
<td>d</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>d = t-90</td>
</tr>
</tbody>
</table>

* Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
**Decision tree**

- Decision Tree (Wekas J48) - Unpruned

![Decision Tree Diagram]

**Evaluation:**

Correctly Classified Instances: **74.5 % (35)**
Incorrectly Classified Instances: **25.5 % (12)**

Baseline Classification (ZeroR): 38.3%
Rule learner

• JRIP Rules - Pruned

JRIP rules:
============

(demand = 4) => traffic=t-90
(demand = 2) => traffic=t-30
(demand = 3) => traffic=t-60
=> traffic=no

• Evaluation:

Correctly Classified Instances: 76.6% (36)
Incorrectly Classified Instances: 23.4% (11)

Baseline Classification (ZeroR): 38.3%
Rule learner

• JRIP Rules - Unpruned

```
JRIP rules:
=--------
(demand = 4) and (parking = t-30) => traffic=t-90
(demand = 4) and (parking = t-90) => traffic=t-90
(demand = 2) => traffic=t-30
(demand = 3) and (parking = t-90) => traffic=t-60
(demand = 3) => traffic=t-60
=> traffic=no
```

• Evaluation:

Correctly Classified Instances: 74.5 % (35)
Incorrectly Classified Instances: 25.5 % (12)

Baseline Classification (ZeroR): 38.3%
Conclusions

- Extracted rules:
  - demand 4 -> traffic “t-90”
  - ...

- More data sources, more complex rules
  - Weather, traffic reports, ...

- Can lead to automated rule generation
  - Extracting rules for large number of sensors
Thank You
any questions?
ailab.ijs.si