Complex Event Detection and Prediction in Traffic

Blaž Kažič, Dunja Mladenić, Luka Bradeško

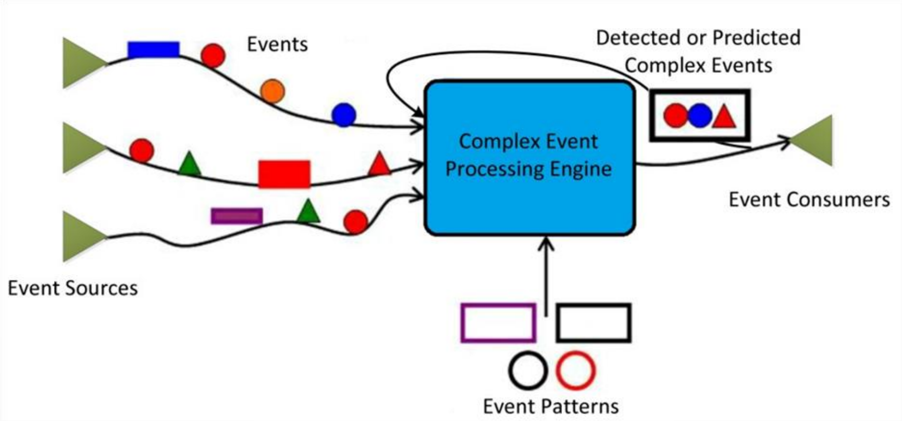
AI Lab, Jozef Stefan Institute,
Jozef Stegan International Postgraduate School
Ljubljana, Slovenia

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

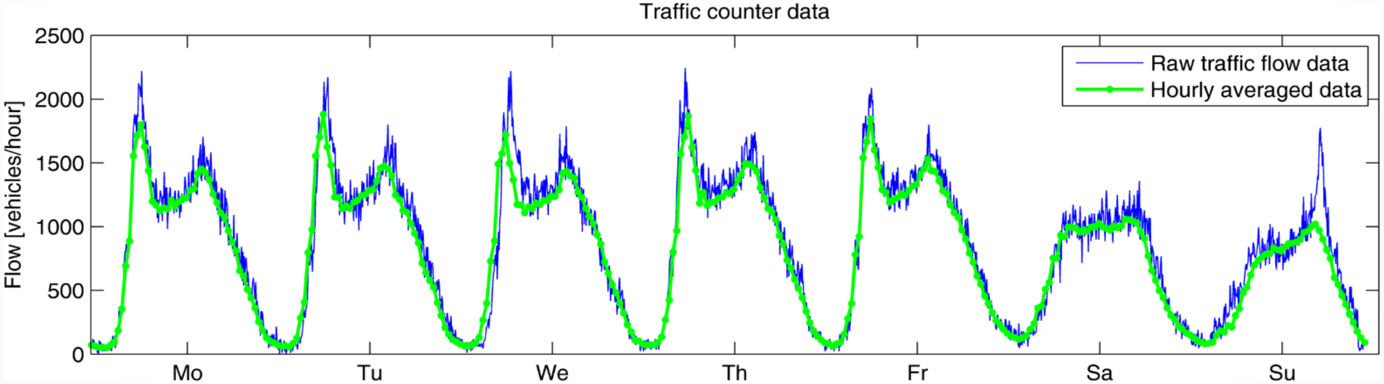
Real-time data streams



Detect meaningful events

Complex events in traffic

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



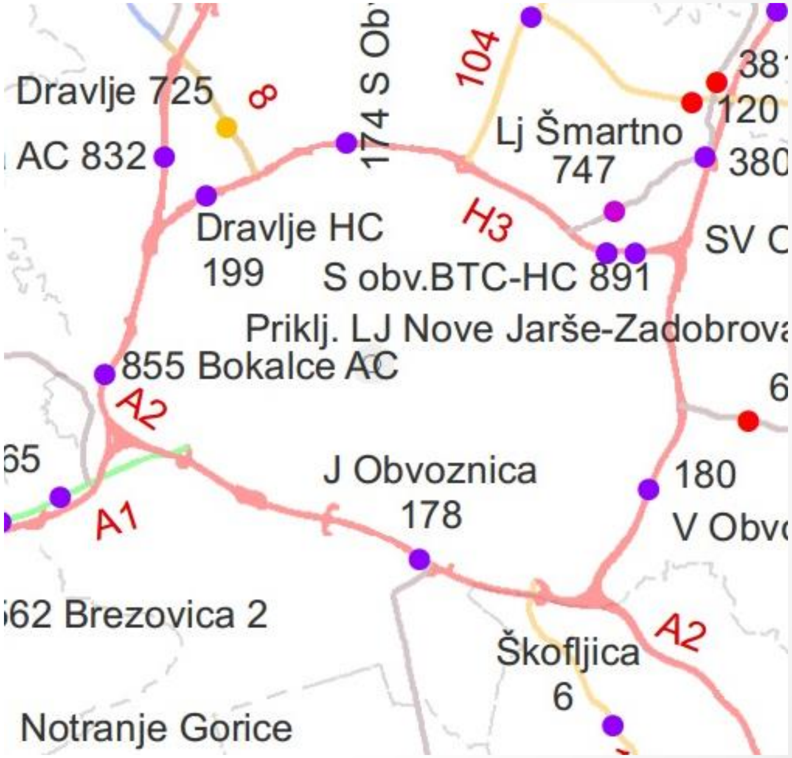
- Research how bigger social events affects on traffic nearby – Extract rules

Outline

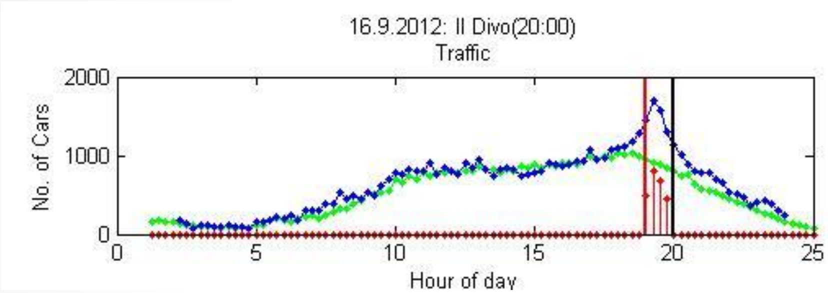
- Data description
    - Description of used data sources
- Data preprocessing:
    - Complex event detection
    - Database of extracted complex events
- Data mining results:
    - Analyze how social events affect on nearby traffic
    - Find correlations between different complex events
    - Find rules to predict complex traffic events

Data description

- Traffic Counters
  - Flow (vehicles/hour)
  - 5 min stream
  - 2011-2013
- Social events
  - 50 major social events
  - Stadium Stozice
  - Demand (1-4)
- Parking sensors
  - Parking availability count



Complex event detection
Traffic data



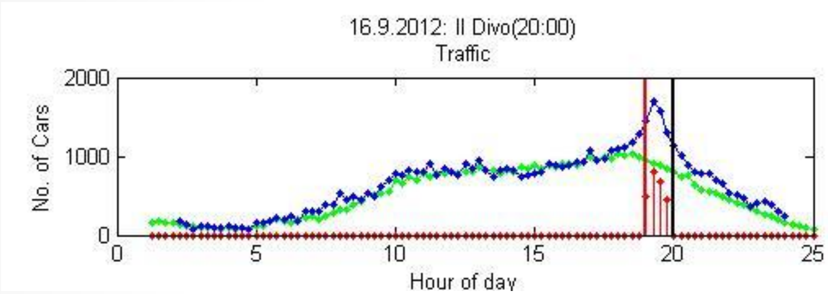
Traffic complex event: **"t – 60"**

- Complex event is considered as the start of anomaly

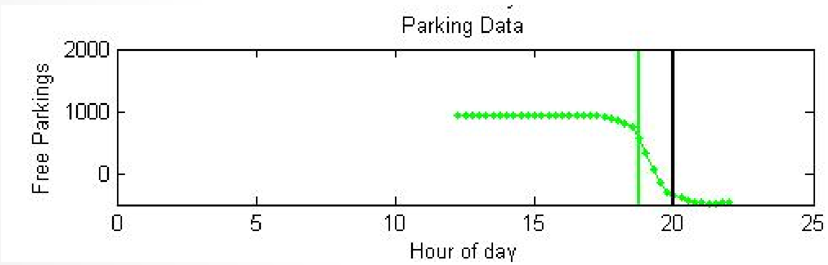
- Local averages are used to described "normal" traffic

- Anomaly: if RMSE is larger than certain threshold

Complex event detection
Parking data



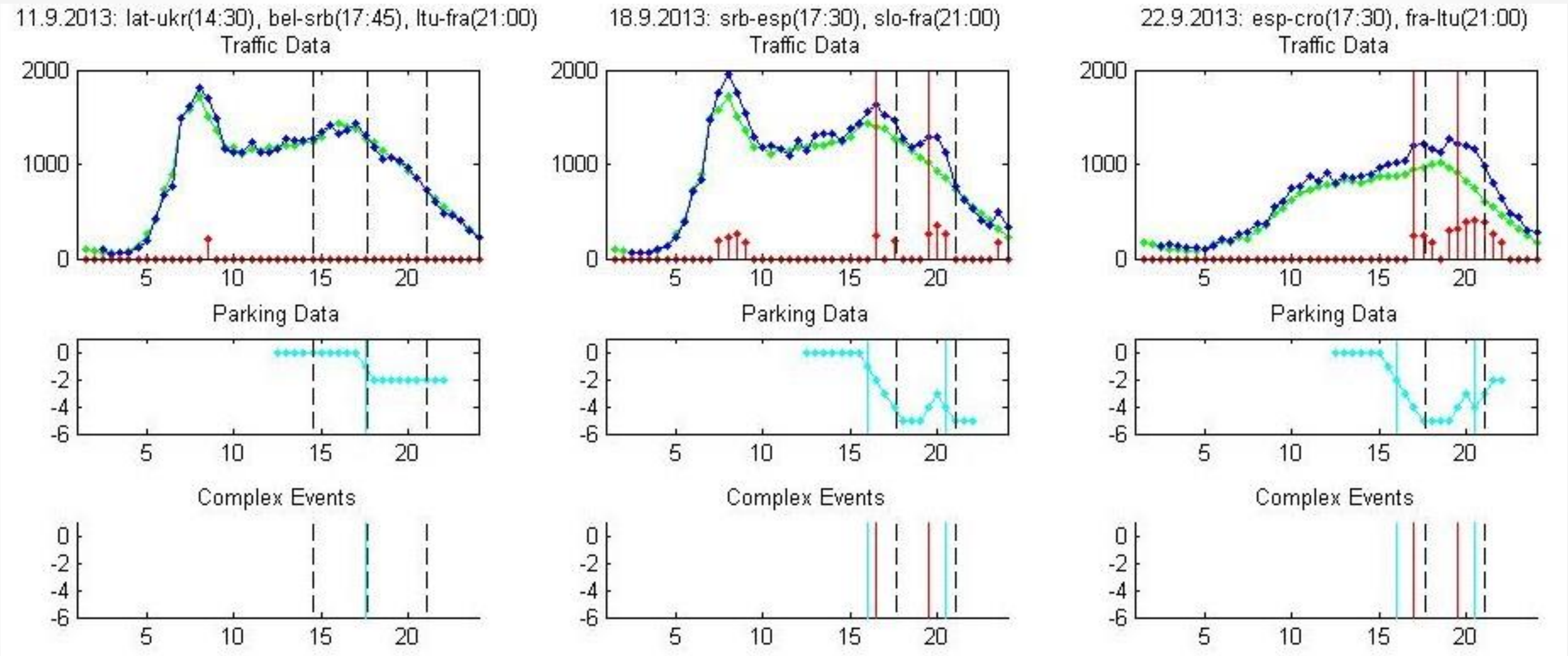
Traffic complex event: **"t – 60"**



Parking complex event: **"t – 90"**

- Parking sensor did not have specific daily pattern
- Complex event in parking sensor is considered as start of "*downward trend*" in parking availability

Complex event detection



Extracted data set

Event Description	Date	Hour	Visitors	Demand	Parking Sensor	Traffic Sensor
...	...	...	...	...	...	...
SLO - UKR	21.09.2013	21:00	10000	4	t-90	t-90
ESP - CRO	22.09.2013	17:30	6050	2	t-90	t-30
FRA - LTU	22.09.2013	21:00	10000	4	t-30	t-90
Elton John	11. 11. 2011	21:00	8000	3	?	t-60
...	...	...	...	...	...	...

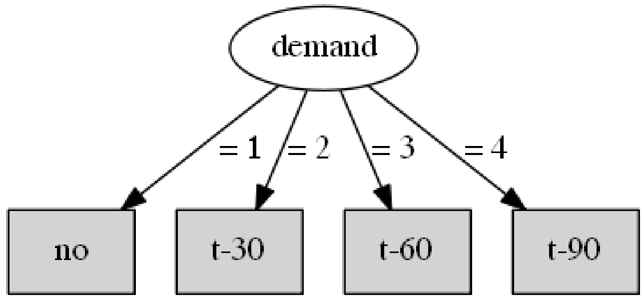
- **Event demand attribute**: 4 possible values (*1-4*)
- **Parking sensor attribute**: 5 possible vales (*"no" – "t-90"*)
- **[Target] Traffic sensor attribute**: 5 possible vales (*"no" – "t-90"*)

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

* UCI Machine Learning Repository – Dogers Loop Sensor Data Set (81 events)
* "Adaptive event detection with time-varying Poisson processes" A. Ihler, J. Hutchins, and P. Smyth. Proceedings of the 12th ACM SIGKDD Conference (KDD-06), August 2006.

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%

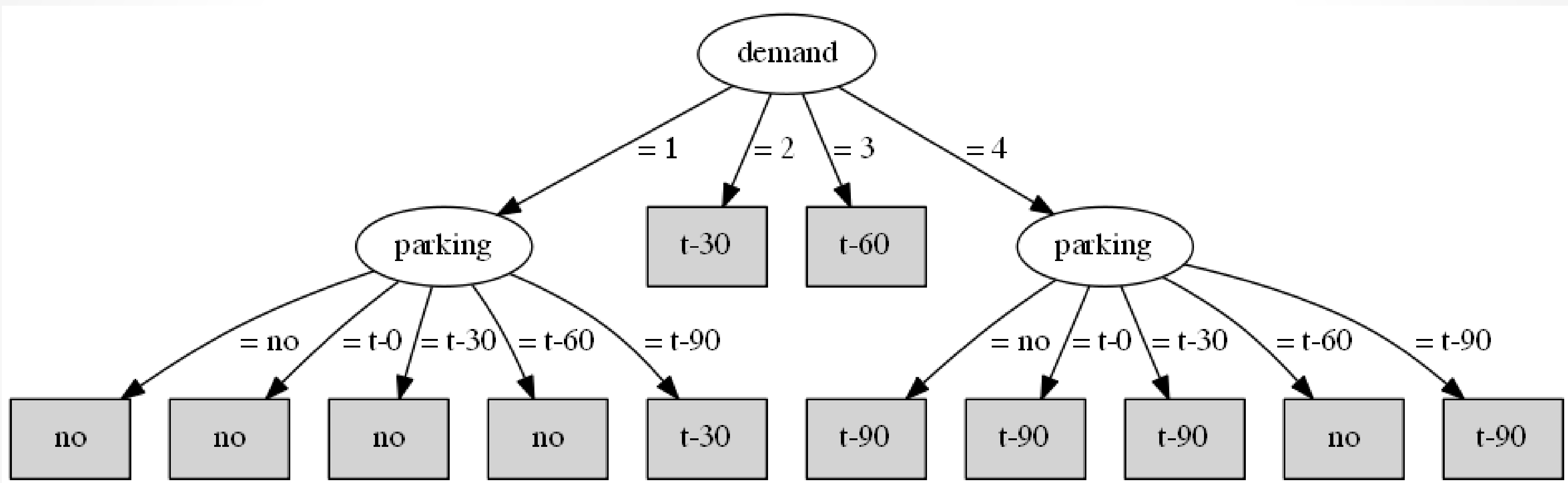
```
=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 17  0  0  1 |  a = no
  3  4  1  1 |  b = t-30
  1  1  8  2 |  c = t-60
  0  0  1  7 |  d = t-90
```

* 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



- Evaluation:

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

Baseline Classification (ZeroR): 38.3%

```
=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 17  0  0  1 |  a = no
  3  4  1  1 |  b = t-30
  1  2  7  2 |  c = t-60
  0  0  1  7 |  d = t-90
```

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%

```
=== Confusion Matrix ===

  a  b  c  d   <-- classified as
 17  0  0  1 |  a = no
  3  4  1  1 |  b = t-30
  1  1  8  2 |  c = t-60
  0  0  1  7 |  d = t-90
```

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)**

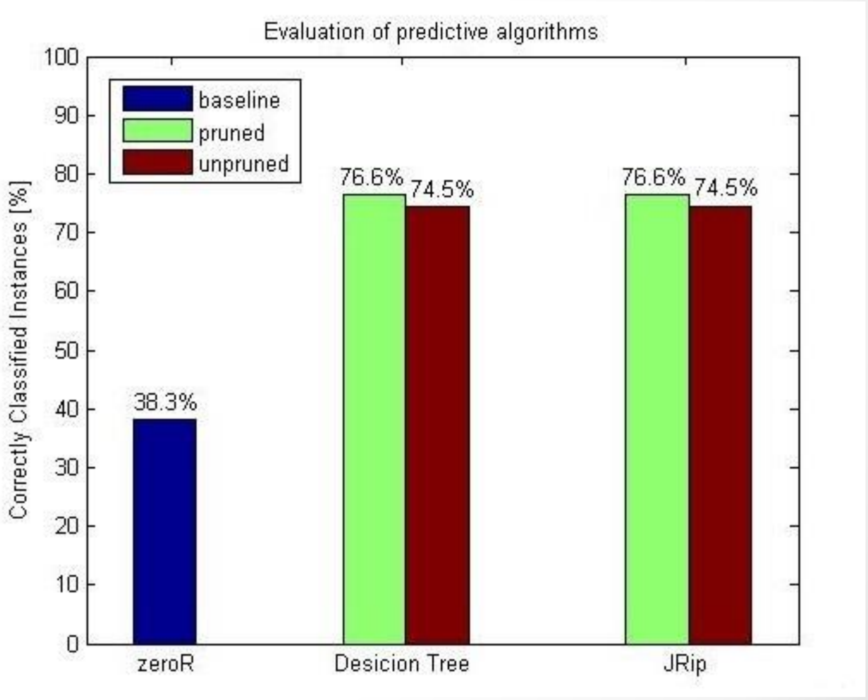
Baseline Classification (ZeroR): 38.3%

=== Confusion Matrix ===

a	b	c	d		<-- classified as
18	0	0	0	|	a = no
3	4	1	1	|	b = t-30
2	1	8	1	|	c = t-60
2	0	1	5	|	d = t-90

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