Anomaly Detection Challenges

Winter semester 2016/17
Anomaly Detection

Anomaly detection (also outlier detection, novelty detection) is the identification of items, events or observations which do not conform to an expected pattern or other items in a dataset [1].

Anomalies are...

- Rare
- Harmful
- Confusing
- Not just noise...

https://www.withfriendship.com/images/d/18083/Outlier-picture.gif
Anomalies

- Anomalous behaviors are often indicative of a compromise to a system or service
  - Abnormal transactions on credit cards could be fraudulent
  - Suspicious MRI images are possibly indicative of malicious existence of tumor
  - Anomalous measurements in network traffic might indicate hacking
  - Anomalous program behavior might be from malware activity
  - Unusual noises in motorcycle may refer to engine damage

"We need to detect, analyze and interpret anomalies!"
- The Instructors
Anomaly Detection - applications

- Medical anomalies
- Image processing
- Malware detection
- Banking fraud detection
- Spam detection
- Device fault detection
- Stock market events
- ...

Anomaly Detection - I20 Activities

- Detect malware from various large-scale data (millions of malware samples):
  - Static code analysis (PE Header, Rich Header, instruction sequences)
  - Behavioral traces: system calls, network traffic (IP, DNS, payload)
- Behavior-based authentication
- Adversarial Learning
- Anomaly Detection under constraints (parameter set, feature acquisition...)
- Spam Detection
**General Information**

<table>
<thead>
<tr>
<th>Type</th>
<th>Practical Course (Praktikum)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Credits</td>
<td>6 SWS / 10,0 ECTS-Credits</td>
</tr>
<tr>
<td>Time</td>
<td>Tu, 16:30 to 18:00 Uhr</td>
</tr>
<tr>
<td>Start</td>
<td>18.10.2016</td>
</tr>
<tr>
<td>Where</td>
<td>Room 01.06.011</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>Chair Website</td>
<td><a href="https://www.sec.in.tum.de/">https://www.sec.in.tum.de/</a></td>
</tr>
<tr>
<td>Course web page</td>
<td>./anomaly-detection-challenges</td>
</tr>
</tbody>
</table>
Goals

- Practical course related to Machine Learning
  - Primarily targeted towards Master students, others can apply
- Apply the knowledge learned from the lecture courses
- Provide a fun and competition learning environment!
  - Try to outsmart your fellow students
  - Learn from your peers
  - Have fun!

"It's in the anomalies that nature reveals its secrets."
- Johann Wolfgang Von Goethe
Prerequisites

● Important:
  ○ Basic Machine Learning knowledge
  ○ Be comfortable with the math related to ML: Linear algebra, probability, statistics, optimization
  ○ Python programming (if not, learn it during summer :))
  ○ Have your own computer for tutorials/presentations

● Nice to haves:
  ○ IT Security (we will provide a basic intro)
  ○ Scala and Spark (although only if desired by the student)
Course contents

- **Tutorials** in various topics:
  - Anomaly detection using machine learning methods
    - SVM, Ensemble Learning, Semi-supervised learning, Adversarial Learning...
  - Domain-specific tutorials: Malware Analysis, Network Intrusion Detection...
- **Challenges** based on real-world datasets from multiple domains
- **Homework** assignments related to the course tutorials
Challenges

● Team of 1-2 students
● Kaggle platform ->
  1. Download input data
  2. Train model
  3. Test performance
  4. Get ranking
● Discuss the solutions at the end of each challenge
Deliverables

- For each challenge:
  - Report about your solution
  - Send us your code
  - Presentation

- Homework: 7 are provided

- Grading: $C = 0.3 \times T + 0.4 \times R + 0.2 \times P + 0.1 \times B$
  - $T$: Talk for the results
  - $R$: Report for the results
  - $P$: Performance in class
  - $B$: Benchmarks (ranking) on Kaggle
How to apply?

● Send a short (1 page) CV until 06.07.2016. to kolosnjaji@sec.in.tum.de
  ○ Provide info in your CV that highlights your knowledge and interests that are relevant to the course
  ○ In your e-mail you can suggest what kind of datasets you would like to analyze

● Apply on matching system
  ○ If you cannot use the matching system (e.g. exchange students) let us know!
## Contact Us

<table>
<thead>
<tr>
<th>Name</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bojan Kolosnjaji</td>
<td><a href="mailto:kolo@sec.in.tum.de">kolo@sec.in.tum.de</a></td>
</tr>
<tr>
<td>George Webster</td>
<td><a href="mailto:webster@sec.in.tum.de">webster@sec.in.tum.de</a></td>
</tr>
<tr>
<td>Mohammad Reza Norouzian</td>
<td><a href="mailto:norouzian@sec.in.tum.de">norouzian@sec.in.tum.de</a></td>
</tr>
</tbody>
</table>
HiWi Wanted!

Looking for a student with basic knowledge about security and experiences in machine learning and anomaly detection.

Please send us your CV to norouzian@sec.in.tum.de