FSADA
an anomaly-detection approach

Viktor Jovanoski, Jan Rupnik
Jožef Stefan International Postgraduate School - Jožef Stefan Institute
September 2018

Roadmap

- Anomaly detection
  - Overview
  - Real-world experience, needs and challenges
- Paper contribution
  - General methodology/architecture for anomaly detection
- Future work
  - Incident detection, evaluation, prediction

Anomaly detection

What is anomaly

- **“Anomaly is data-point that is significantly different from the majority the data”**
- Different types - numeric, categorical, texts
- Potentially huge volumes - e.g. sensoric data
- Diverse velocities - from milliseconds to days
- Different latencies - from milliseconds to hours

Practical requirements

- Handle huge amounts of data
- Handle complex analyses
- As real-time as possible
- **Actionability**

Questions with progressive difficulty levels

- Is current state anomalous?
- Is current state critical?
- Which areas of the current state are the main suspects?
- What is the root cause?
- Which areas are causing the most damage
  - Directly, indirectly
- Which area should operators start addressing first?
- Given the current state is there a higher probability of problems in the near future?

Practical experience

- **Manufacturing**
  - Shop-floor
  - Logistics
  - Planing
- **IT infrastructure**
  - Servers
  - Services (multiple servers, elastic deployments, complex dependencies)
  - Network communications
  - Database and other storage (loads, volumes, response times)
  - Security

State-of-the-art

- KDD 2018
  - Several papers, domain-specific problems
  - Strong niche algorithms, with enhancements
    - Work in several phases
    - Built on top of prior, simpler work
  - Feature-engineering based on expert feedback
  - Strong feedback loop
    - Evaluation
    - Active learning

Paper contribution:

FSADA Architecture

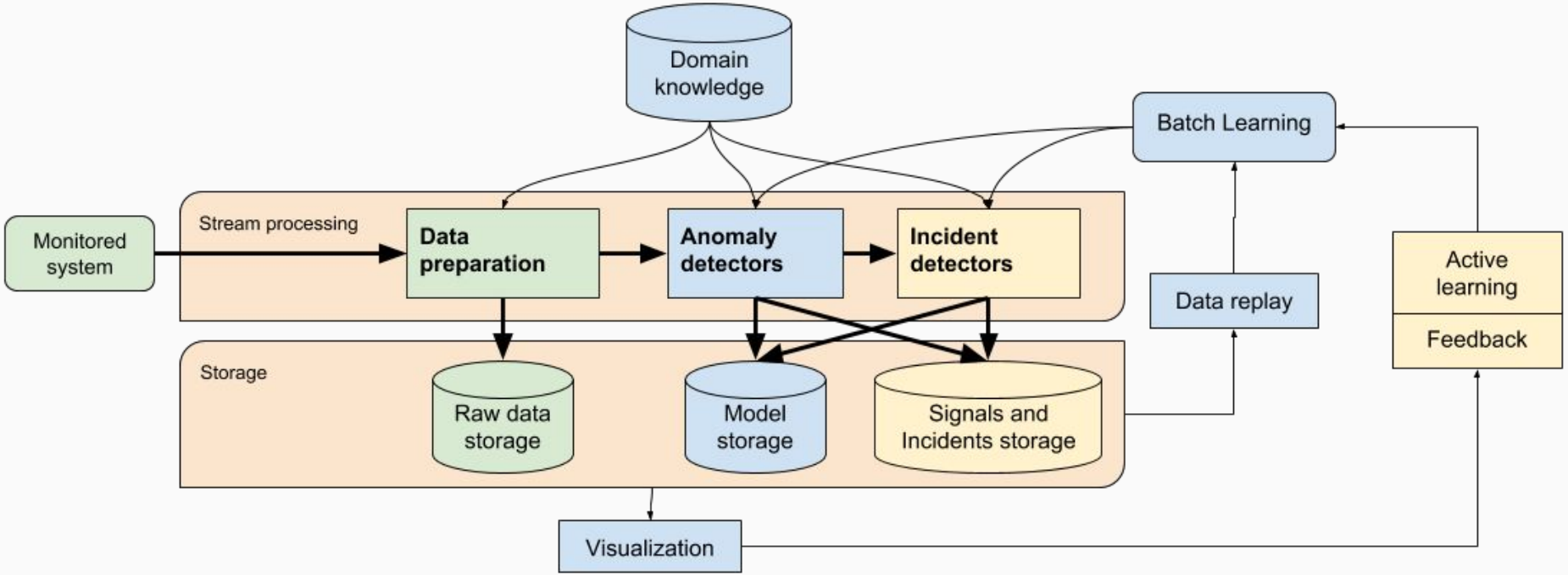
Last year’s paper

- SiKDD 2017
- Domain-specific problem - analysing logs from servers
- Need for expanded scope
  - Different algorithms, different data streams
  - Streaming scenarios
  - Alert feedback
  - Active learning
  - Cloud-enabled

Full-Spectrum Anomaly Detection Architecture

- FSADA
  - Many design iterations and inputs
- Main features
  - Streams, big data
  - Cloud-ready, scalable
  - Diverse data, diverse algorithms
  - Signals (low-level alerts) and incidents (high-level alerts)
  - Background knowledge
  - Clearly defined place for feedback and active learning

Architecture



Future work

Incidents, root causes, predictions

Further work - architecture

- System simulations and predictions
- Diverse background knowledge

Goal

- Improve analyses across data-source
  - Correlations
  - Simukations
  - Predictions
- Use more structure that is available in the data
- Use more background knowledge

Mimic some skills that people in the “operation room” use when they stare at multiple monitors showing data from diverse data sources.

Thank you