Anomaly Mining: Detection and Beyond

Leman Akoglu, Assistant Professor of Information Systems
Carnegie Mellon University
http://www.andrew.cmu.edu/user/lakoglu/
data science, anomaly mining, large-scale graph analytics

Index/keywords: anomalous events, hotspots, disinformation

Detecting anomalies or “hotspots” is one of the key unsupervised learning tasks in AI, with numerous applications in cyber-security, finance, surveillance, public health, etc. Some of the critical applications include intrusion and malware detection in computer networks, environmental or public monitoring for detecting hazardous leakage, social unrest, or disease outbreaks, video surveillance for security, a large plethora of fraud detection scenarios (tax, credit card, advertisement, auction, insurance, etc. fraud) and most recently detection of disinformation (fake reviews, false news, hoaxes).

In most scenarios, the challenges associated with the anomaly detection task involve (i) large volume of data, (ii) potentially arriving from a number of different sources, (iii) at a fast speed; where the goal is to spot the “needle in the haystack” in *an online (or real time)* fashion. Early detection is especially critical in order for one to take appropriate measures to contain the “anomaly” (application-specific) before its widespread impact on the environment, the associated system, or the population.

In this talk, I will address two different problem setups: detecting (1) anomalous events and “hotspots” in critical infrastructure and (2) false information in online media.

Anomalous events and “hotspots” in critical infrastructure: Today’s most critical infrastructures such as the power grid, roads, bridges, buildings, etc. are equipped with a variety of sensors monitoring their health. Given streams of sensor measurements (temperature, voltage, vibration, etc.) from a large number of sensors, the anomaly mining task aims to (i) *detect events:* spot significantly higher or lower measurements at any number of sensors at any given time, as well as (ii) *pinpoint “culprits”:* localize the scene of the event, i.e. the subset of sensors associated with the detected event. All in all, the goal is the real-time and early detection and localization of anomalous events, such as power line failures and traffic accidents through sensor data, which is the first step toward mitigation and timely recovery.

The above problem (i) is often formalized as a sensor-level anomaly detection task in streaming time series data. While the metric of interest quantifying the degree of anomalousness in time may differ for different applications, the general goal is to score each time point by a metric of drift from normal behavior. Given such scores for each sensor at every time point, problem (ii) then aims to further quantify the “scale” of the

event by identifying groups of sensors associated with the same event. This problem is typically formulated as a sub-graph extraction task with an objective of total anomaly score maximization (and sometimes also associated cost minimization) with connectivity constraints, where the graph edges depict relationships between the sensors (roads and intersections, proximity relations, etc.).

Note that even though the above focused on anomalous events in physical infrastructure based on physical sensors, the same techniques generalize to settings where the infrastructure is a social network (e.g., Twitter) and the sensors are "social" (human) sensors, where events could correspond to social unrest, protests, etc.

False information in online media: With the fast growth of popularity of social and information media (Twitter, Facebook, TripAdvisor, Google news, etc.) both the speed and scale of information being generated and consumed increased drastically. Today it is not only the traditional media that is generating content at a regular basis (e.g., daily newspapers), rather almost everybody (bloggers, reviewers, etc.) who is generating and also consuming information. With such speed and scale comes the challenge of verifying the truthfulness of information and containing the spread of misinformation. These challenges, combined with the anonymity that Internet provides to the content generators, have recently generated a plethora of incentives for various adversaries to ingest disinformation[1] to such online platforms. Similar to the previous part, timely detection of false information is essential in taking counter measures before its widespread impact on various target populations.

In the second part of the talk, I will focus on false information and introduce research on characterizing and detecting opinion-based and fact-based disinformation. As an anomaly detection problem this setting is arguably more challenging as it involves non-trivial adversaries; that is, while anomalous events in previous part may be due to natural occurrences (e.g. power line failures), here there exist adversaries with the intent to deceive. While disinformation aims to manipulate and bias perception and beliefs, these two types of disinformation differ in the following way. In opinion-based settings (fake reviews in Yelp, Tripadvisor, etc.), there exist no absolute truth but subjective truths. In fact-based settings (fake news), there exist singular absolute truth.

I will conclude the talk with important challenges in the above areas as well as anomaly mining at large, including adversarial robustness (how can we design detection systems that are immune to manipulation or evasion by adversaries?), information design (how can we design preventive policies to effectively counteract and mitigate such anomalies?), and human-in-the-loop detection (how can we design interactive systems to acquire necessary and timely human feedback?).

---

[1] Note that I use the terms disinformation and misinformation to convey different meanings: the latter is false information that is unintentional due to misperception or lack of understanding, while the former is false information that is deliberately intended to deceive or mislead.