STATS 798 Notes

Research Masters

Junhuang Xue

Department of Statistics

The University of Auckland

Last Modified: Monday, 02/08/2018

Lecturer Information:

**Professor Thomas Lumley**

Office: 303S-325

Extension: 64 9 923 3785

Email: t.lumley@auckland.ac.nz

Office Hour: Booking

Contents

[Forecasting: Principles and Practice - Rob J Hyndman and George Athanasopoulos (Ch 10) 3](#_Toc521591589)

[Forecasting hierarchical time series - Rob J Hyndman 9](#_Toc521591590)

[Definitions 10](#_Toc521591591)

Anomaly Detection 101 - Elizabeth (Betsy) Nichols Ph.D.

<https://www.youtube.com/watch?v=5vrY4RbeWkM>

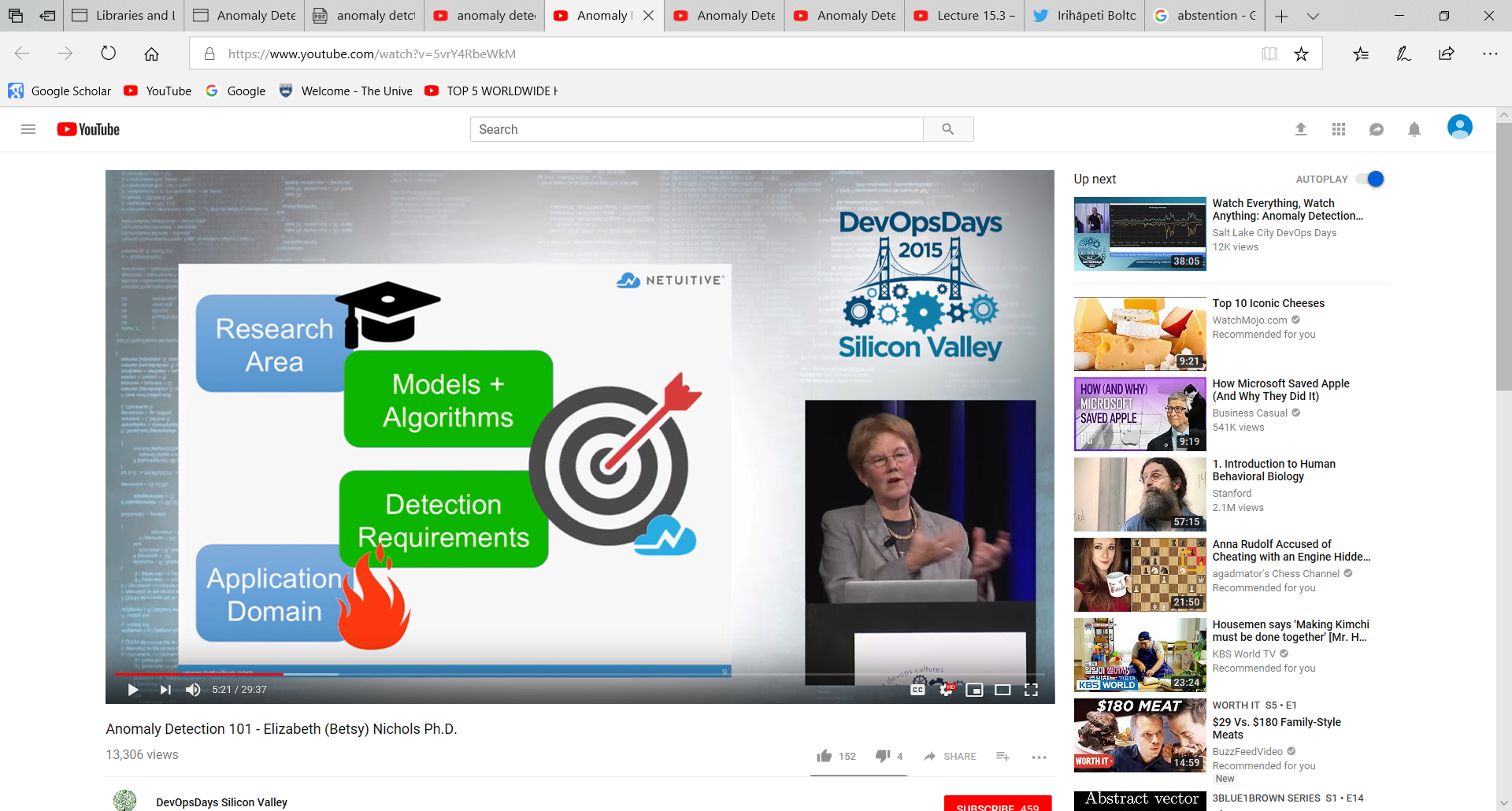
Anomaly Detection: Finding patterns that do not conform to expected behaviour.

What property are you interested in? Colour? Size? Distance?

Research turns out models that are mathematically sound

Lots of people made a good living out of anomaly detection, there are a lot of applications and some companies really depend on it.

However there is very thin sets of information about it.



**Requirements**

- What data is available?

- What types of anomaly do you want to detect

- What technique can be used, pros and con

Kinds of data

Matreics

- Most fundamental structure

- ie time series

- Needs units

Dynamic Attributes

- Slowly changing dimensions

Collections

- Virtual or physical entities

- Relationships

Types of anomaly

Point anomaly

- Unusual compare to other points

- can be hidden in a cumulative plot

Contextual anomaly

- Unusual in some context, but not number

Collective anomaly

- A collection of related data instances is anomalous to a particular set

Techniques

Deterministic

- eyeballing

Statistical

- Sudden change can be hard to detect (crying wolf effect)

- Correlation models

- Machine learning

- Correlation anomaly

Anomaly Detection Algorithms and Techniques for Real-World Detection Systems

<https://www.youtube.com/watch?v=CAvKQHHNmcY>

Hard to define, you know it when you see it

Noticeably different from the expected

Approaches

- Statistical model: is observation an outlier?

- Machine learning

Zscores

- Often used to test “extremeness”

- How many SD away

- Use moving average, and z score for new observations

- Not robust,, SD highly sensitive to outliers

Density based

- Noticeably different by space

Probalistic Density

- assume all data generated according to prob dist

- estimate underlying probalistiy density function

K Distance

- compute distance to nearest neighbour

- Lager k = more isolation

Introducing practical and robust anomaly detection in a time series

<https://blog.twitter.com/engineering/en_us/a/2015/introducing-practical-and-robust-anomaly-detection-in-a-time-series.html>

<https://www.youtube.com/watch?v=CAvKQHHNmcY>

**False positive, false negative, detection delay**

Statistical model of abnormal behaviour

Machine learning? Probably not

Need to identify anomalies in realtime? Quick and dirty methods?

z-score, how extreme an observation is , moving average? Standard deviation (not suitable for extreme values

local outlier k distance

local outlier factor

<https://instapage.com/blog/what-is-visual-hierarchy>

<https://iwringer.wordpress.com/2015/11/17/anomaly-detection-concepts-and-techniques/>

<https://www.youtube.com/watch?v=HntUY8SsYZg>

Poisson Regression Model

Large group rushed in, it will not be properly accounted for in a poisson model

<https://www.youtube.com/watch?v=sv_KXSiorFk>

poisson or not? Concept