Bruce Bassett SARAO / AIMS / SAAO / UCT





UNIVERSITY OF EAST ISWN

(Some) **Data science** Challenges for the **SKA Era**

Men of Science More or Less Agog Over Results of Eclipse Observations.

IN THE HEAVENS

LIGHTS ALL ASKEW

* * *

EINSTEIN THEORY TRIUMPHS

Stars Not Where They Seemed or Were Calculated to be, but Nobody Need Worry.

A BOOK FOR 12 WISE MEN

No More in All the World Could Comprehend It, Said Einstein When His Daring Publishers Accepted It. French Government to Open Cheap National Restaurants

PARIS, Nov. 0 .- " National res-





Dutch Reiterate on Anniversary of His Flight Their Views on Right of Asylum.





https://www.skatelescope.org/

Data ~10²⁷ B/year





Machine Learning against infinity



Machine-learning techniques used by thousands of scientists to analyse data are producing results that are misleading and often completely wrong.



Jacob Bronowski

Nadeem Ogzeer



MeerKAT, N. Oozeer, 2018



Better than Human



Danziger et al

FRI

-.

FRII

,*

When is simulated or generated data good enough?

Z. Hosenie, BB, et al, 2017

FRI FRI FRII

What do we do when human experts are not good enough?

Training on human-labelled data or on simulations inevitably leads to contamination

 \rightarrow A "new" source of **systematic human Training Error**



Weapon	99%
Gun	97%
Firearm	95%
Assault Rifle	91%
Trigger	90%
Rifle	88%
Machine Gun	81%
Gun Accessory	73%
Gun Barrel	70%

Wired



rifle_adv.png

Rotorcraft	
	66%
Aircraft	56%
Vehicle	





Anomaly Detection





A simple (hierarchical) model

 This hierarchical model can be expressed as a product of terms, each of which we can easily compute.

$$P(\tau|d_o, y_o) \propto \int dy_t \frac{P(d_o, y_o, y_t|\tau)}{P(\tau)} P(\tau)$$

$$P(\tau|d_o, y_o) \propto \int dy_t \frac{P(d_o, y_o, y_t|\tau)}{P(d_o, y_o|y_t, \tau)} P(y_t|\tau) P(\tau)$$

$$P(\tau|d_o, y_o) \propto \int dy_t P(d_o|y_o, y_t, \tau) P(y_o|y_t, \tau) P(y_t|\tau) P(\tau)$$

$$P(\tau|d_o, y_o) \propto \int dy_t \frac{P(d_o|y_t, \tau)}{P(y_o|y_t, \tau)} P(y_o|y_t, \tau) P(y_t|\tau) P(\tau)$$

What we had before... Ethan Robert, Michelle Lochner, BB – arxiv: 1902.08627

y, trais

 τ

I d test

+2



Huchra





https://www.skatelescope.org/

The need to go beyond traditional machine learning





Impact of Pointing Errors



True Sky

Recovered sky

Difference Map



Lochner et al 2014

We need to have full control of errors

- Need to propagate the full probability distributions through complex pipelines
- Currently done in cosmology using MCMC

How do we do this properly and efficiently:

(a) with ML and

(b) in very high dimensions?





방법은 그는 것이 많은 것 같은 것이 같은 것이 같은 것이 같은 것이 같은 것이 같은 것이 많은 것이 많은 것이 같이 하는 것이 같이 하는 것이 같이 없다.







Is there an optimal anomaly detection algorithm for astronomically-realistic objects?

Marginalising over all possible forms of intelligence, what ET signals should we be looking for?









