Artificial Intelligence for Epidemiology

Magnus Boman

November 23, 2018

ESCAIDE

KTH VETENSKAP OCH KONST

RISE Research Institutes of Sweden

What's on my ESCAIDE badge?

- Artificial Intelligence advanced senior expert at Research Institutes of Sweden (RISE), responsible for *Precision Medicine* and for the *Life Sciences* in RISE AI
- Professor in *Intelligent Software Services* at The Royal Institute of Technology (KTH), Stockholm
- Associate editor of *Eurosurveillance*

AI: It's...

...about learning, or it's not AI

...not only machine learning, but...

...also about learning machines

...weak or strong, but...

...actually all of it's still weak...

...so what's the fuss?

AI Systems Think (sic) Differently

"A yellow school bus parked in a parking lot"

'Collaboration on AI for the Public's Health'

University of Toronto Dalla Lana School of Public Health

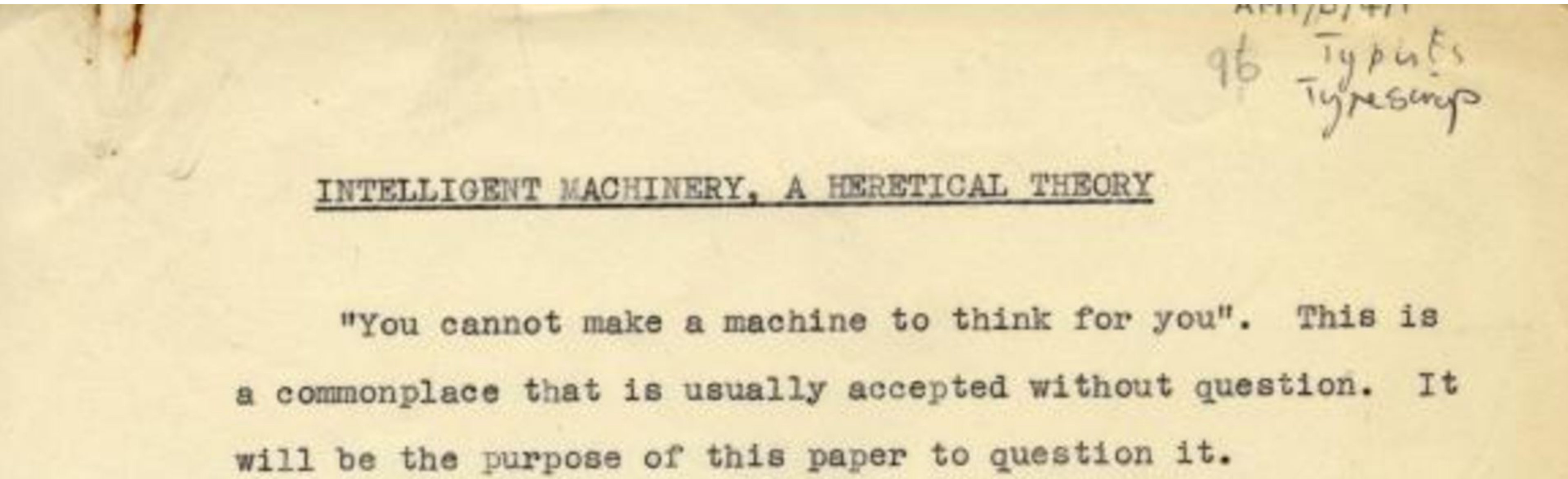
Vector Institute | Institut Vecteur

Public Health Ontario | Santé publique Ontario

Joint Centre for Bioethics University of Toronto

Population Health Analytics Laboratory

Alan Turing (Manchester, 1951)



INTELLIGENT MACHINERY, A HERETICAL THEORY

"You cannot make a machine to think for you". This is a commonplace that is usually accepted without question. It will be the purpose of this paper to question it.

AI Education: Turing's Component List

- Memory (chronological log)
- 'Indexes of experience' (event-based log)
- Special features observed in the indexes already used
- Experience of outcome (+/- weights)
- Crude rules of thumb (heuristics) for outcome classification (supervised), over time replace by sophisticated rules (unsupervised)
- Random noise

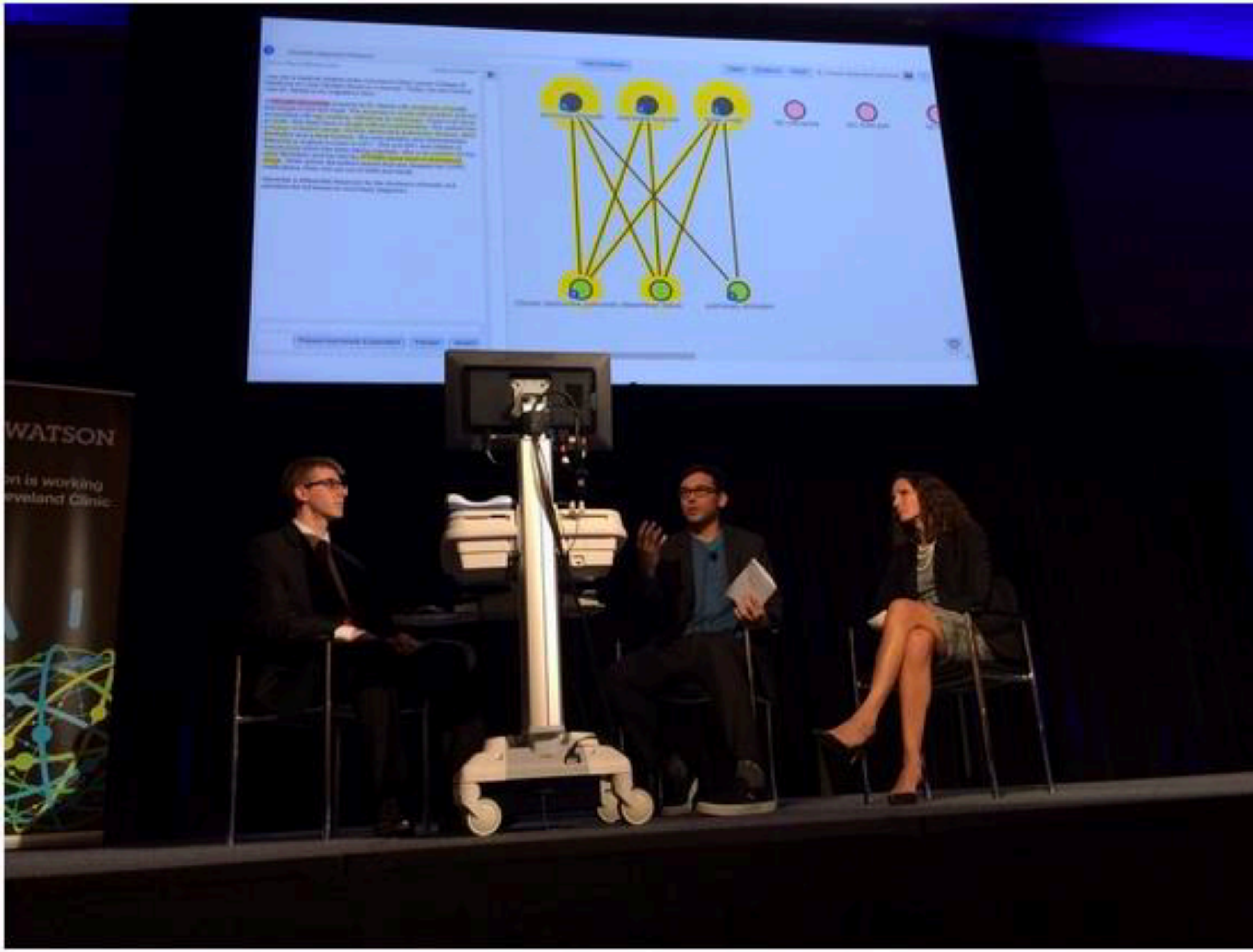


Photo by CHUCK SODER

Officials from the Cleveland Clinic and IBM demonstrated how doctors might one day use IBM's Watson technology to diagnose patients and come up with treatments. From left to right are Sean Steenberge, a student at the Clinic's Lerner College of Medicine, Michael Barborak, a senior software developer and research from IBM, and Julie Tebo, an assistant professor with the college of medicine.

What can Learning Machines Learn?

- Spatial distributions
  - People movement during an epidemic
  - Information diffusion about an epidemic
  - Geographical spread, probabilistic models
- Networks and graphs
  - Social networks
  - Preferential attachment and strata
  - Hubs, bridges, giant components,…
- How to learn
- How to communicate

Data Visualisation without AI, no probs

Boman, M.; Ghaffar, A.; Liljeros, F. and Stenhem, M. Social network visualization as a contract tracing tool AAMAS 2006

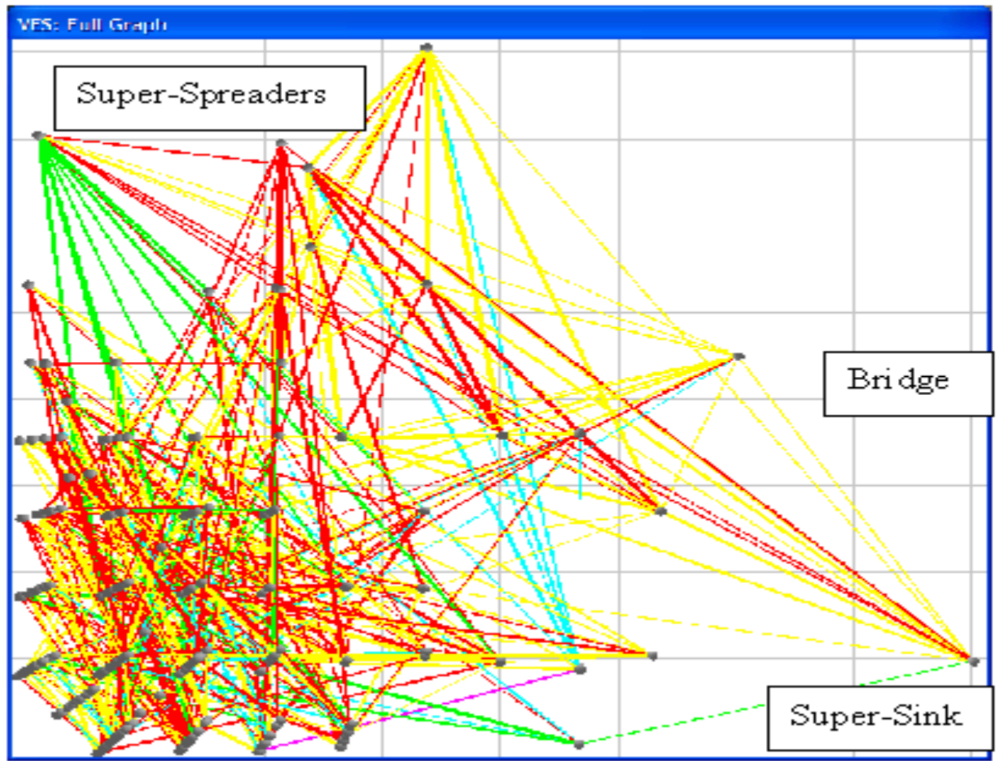


Figure 1: MRSAApp plot for MRSA Dataset Stockholm 2004, with the x-axis mapped onto In Degree and the y-axis mapped onto Out Degree, at t = 2000 days.

The AI Question (Vapnik 1990)

What must one know *a priori* about an unknown functional dependency in order to estimate it on the basis of observations?

Old answer: Almost everything

New answer: Some general properties of the set of functions to which the unknown dependency belongs

AI Prediction meets Statistics

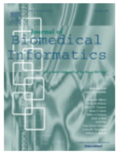
<table>
<tr><td></td><td></td><td colspan="2">True condition</td><td></td><td colspan="2"></td></tr>
<tr><td></td><td>Total population</td><td>Condition positive</td><td>Condition negative</td><td>Prevalence = $\frac{\Sigma \text{ Condition positive}}{\Sigma \text{ Total population}}$</td><td colspan="2">Accuracy (ACC) = $\frac{\Sigma \text{ True positive} + \Sigma \text{ True negative}}{\Sigma \text{ Total population}}$</td></tr>
<tr><td rowspan="2">Predicted condition</td><td>Predicted condition positive</td><td>True positive,
Power</td><td>False positive,
Type I error</td><td>Positive predictive value (PPV), Precision = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Predicted condition positive}}$</td><td colspan="2">False discovery rate (FDR) = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Predicted condition positive}}$</td></tr>
<tr><td>Predicted condition negative</td><td>False negative,
Type II error</td><td>True negative</td><td>False omission rate (FOR) = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Predicted condition negative}}$</td><td colspan="2">Negative predictive value (NPV) = $\frac{\Sigma \text{ True negative}}{\Sigma \text{ Predicted condition negative}}$</td></tr>
<tr><td colspan="2"></td><td>True positive rate (TPR), Recall, Sensitivity, probability of detection = $\frac{\Sigma \text{ True positive}}{\Sigma \text{ Condition positive}}$</td><td>False positive rate (FPR), Fall-out, probability of false alarm = $\frac{\Sigma \text{ False positive}}{\Sigma \text{ Condition negative}}$</td><td>Positive likelihood ratio (LR+) = $\frac{\text{TPR}}{\text{FPR}}$</td><td rowspan="2">Diagnostic odds ratio (DOR) = $\frac{\text{LR+}}{\text{LR−}}$</td><td rowspan="2">$F_1$ score = $\frac{2}{\frac{1}{\text{Recall}} + \frac{1}{\text{Precision}}}$</td></tr>
<tr><td colspan="2"></td><td>False negative rate (FNR), Miss rate = $\frac{\Sigma \text{ False negative}}{\Sigma \text{ Condition positive}}$</td><td>Specificity (SPC), Selectivity, True negative rate (TNR) $= \frac{\Sigma \text{ True negative}}{\Sigma \text{ Condition negative}}$</td><td>Negative likelihood ratio (LR−) = $\frac{\text{FNR}}{\text{TNR}}$</td></tr>
</table>

Wikipedia: Positive and negative predictive values

Prediction - A Recent Example

ELSEVIER

Journal of Biomedical Informatics
Volume 87, November 2018, Pages 50-59



Predicting of anaphylaxis in big data EMR by exploring machine learning approaches

Isabel Segura-Bedmar [a], Cristobal Colón-Ruíz [a], Miguél Ángel Tejedor-Alonso [b, c], Mar Moro-Moro [b]

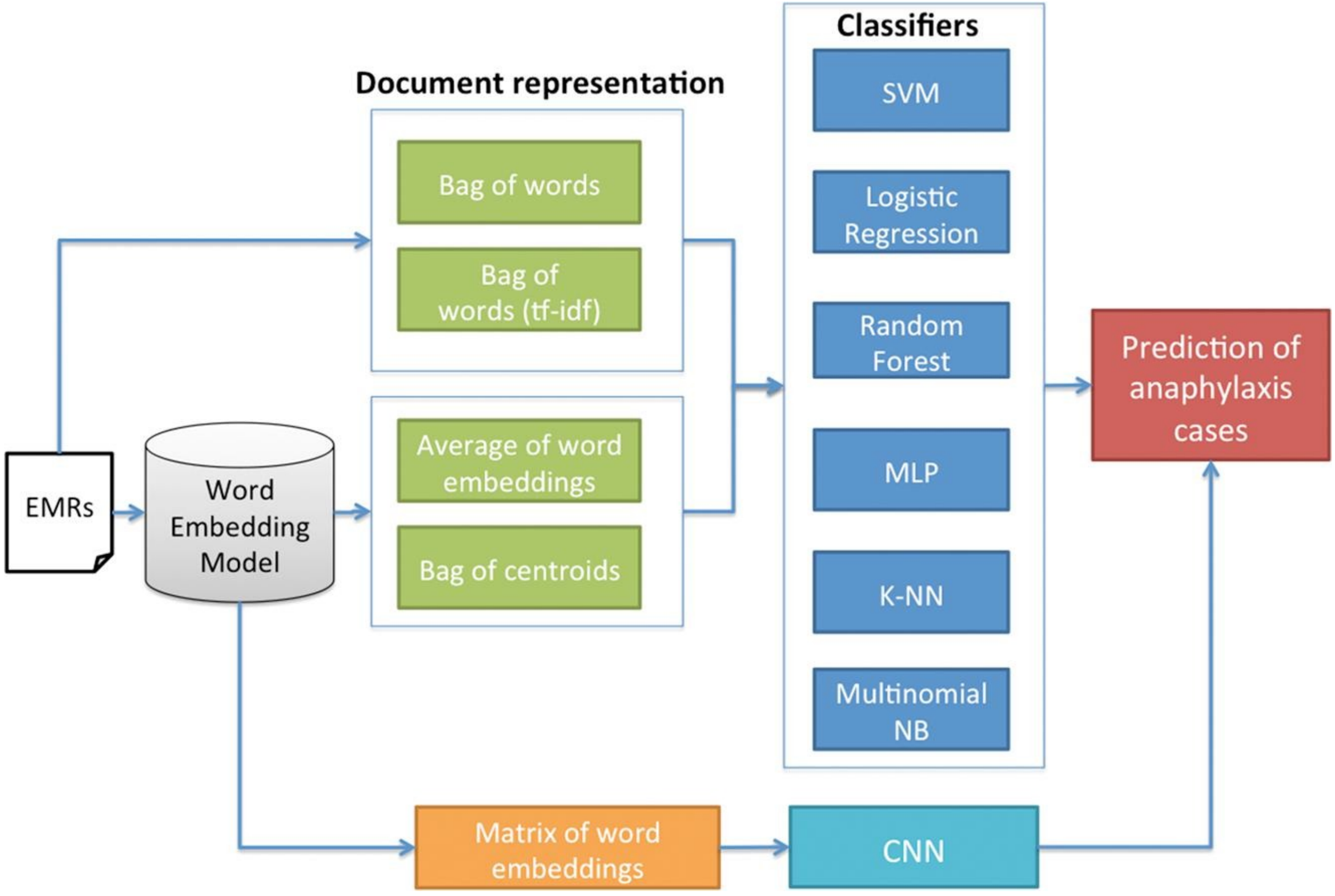
⊞ **Show more**

https://doi.org/10.1016/j.jbi.2018.09.012

Highlights

- Identification of anaphylaxis cases in records by machine learning classifiers.
- Linear classifiers achieve very high performance (F1 = 95%).
- CNN provides slightly better performance, but with higher computation time.
- The use of undersampling does not improve the results.

Representation and Classification



Perceptron Convergence Theorem (Minsky and Papert 1988)

If a set of examples is learnable (100% correct classification, cf. gold standard), the perceptron learning rule will find the necessary weights

- in a finite number of steps
- independent of the initial weights

The rule does gradient descent search in weight space, so if a solution exists, gradient descent is guaranteed to find an optimal solution for any 1-layer neural network

Sentiment Analysis

Methods

Volume 129, 1 October 2017, Pages 50-59

Large-scale machine learning of media outlets for understanding public reactions to nation-wide viral infection outbreaks

Sungwoon Choi [a, b], Jangho Lee [a], Min-Gyu Kang [c], Hyeyoung Min [d], Yoon-Seok Chang [e], Sungroh Yoon [a, f]

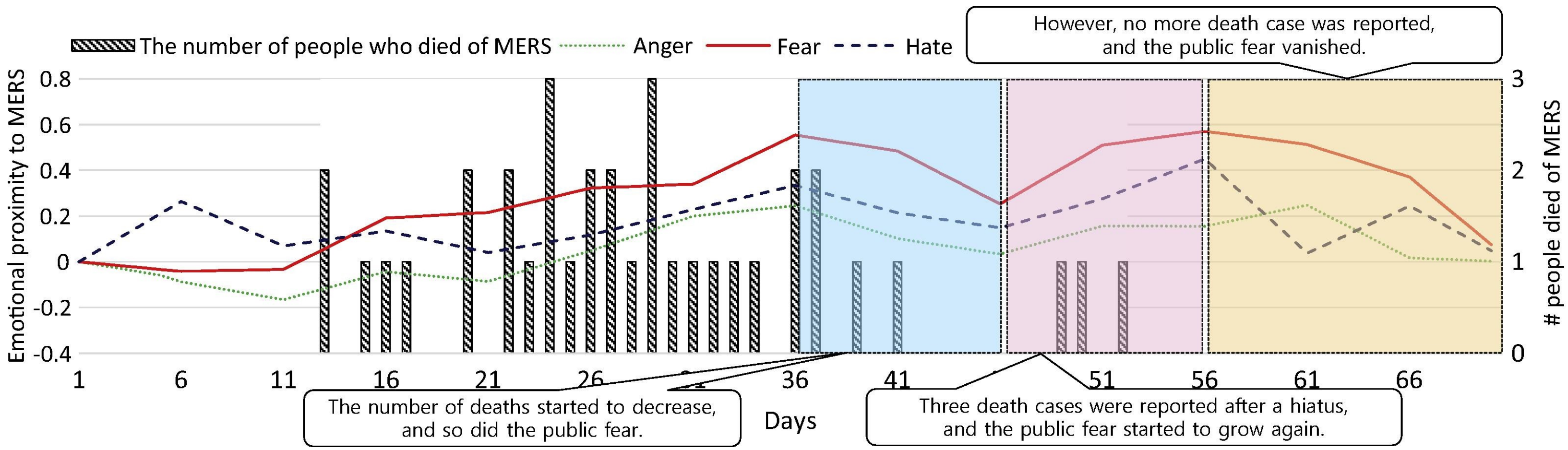
Show more

https://doi.org/10.1016/j.ymeth.2017.07.027

Highlights

- The emotional public responses to a nation-wide outbreak of Middle East respiratory syndrome (MERS) in Korea was analyzed.
- Massive media outlet data was collected during the outbreak.
- An intriguing loop of information transfers between the media and the public was discovered.
- This method would be helpful for alleviating the unnecessary fear and overreaction of the public regarding infectious diseases.

Visualization of Sentiment for MERS



Smart Foodborne Illness Tracking

npj Digital Medicine

www.nature.com/npjdigitalmed

ARTICLE OPEN

Machine-learned epidemiology: real-time detection of foodborne illness at scale

Adam Sadilek[1], Stephanie Caty[2], Lauren DiPrete[3], Raed Mansour[4], Tom Schenk Jr[5], Mark Bergtholdt[3], Ashish Jha[2,6], Prem Ramaswami[1] and Evgeniy Gabrilovich[1]

Machine learning has become an increasingly powerful tool for solving complex problems, and its application in public health has been underutilized. The objective of this study is to test the efficacy of a machine-learned model of foodborne illness detection in a real-world setting. To this end, we built FINDER, a machine-learned model for real-time detection of foodborne illness using anonymous and aggregated web search and location data. We computed the fraction of people who visited a particular restaurant and later searched for terms indicative of food poisoning to identify potentially unsafe restaurants. We used this information to focus restaurant inspections in two cities and demonstrated that FINDER improves the accuracy of health inspections; restaurants identified by FINDER are 3.1 times as likely to be deemed unsafe during the inspection as restaurants identified by existing methods. Additionally, FINDER enables us to ascertain previously intractable epidemiological information, for example, in 38% of cases the restaurant potentially causing food poisoning was not the last one visited, which may explain the lower precision of complaint-based inspections. We found that FINDER is able to reliably identify restaurants that have an active lapse in food safety, allowing for implementation of corrective actions that would prevent the potential spread of foodborne illness.

npj Digital Medicine (2018)1:36 ; doi:10.1038/s41746-018-0045-1

Data: It’s...

…never enough for machine learning

…almost never BIG

…always noisy: humans make noise

…not always best to impute

…a terrible idea to ask organisations and companies to put their data into common repositories

…about cleaning and cleaning and cleaning and:

in the end you might be able to do some AI programming

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

Induction – for general purpose intelligent machines

Isn't it very close to Epidemiology?

Search

Pattern-recognise

Learn

Plan

Induce

Case-Based Reasoning, yes?

Search

Pattern-recognise

Learn

Plan

Induce

The Strand

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

Induction – for general purpose intelligent machines

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

Induction – for general purpose intelligent machines

Search (McCulloch & Pitts 1943)

"There is the genetical or evolutionary search by which a combination of genes is looked for, the criterion being survival value. The remarkable success of this search confirms to some extent the idea that intellectual activity consists mainly of various kinds of search."

upon reading Turing (1937)

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

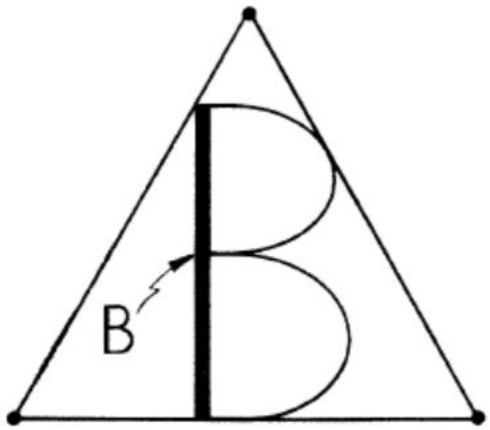
Planning – by analysing and further directing search

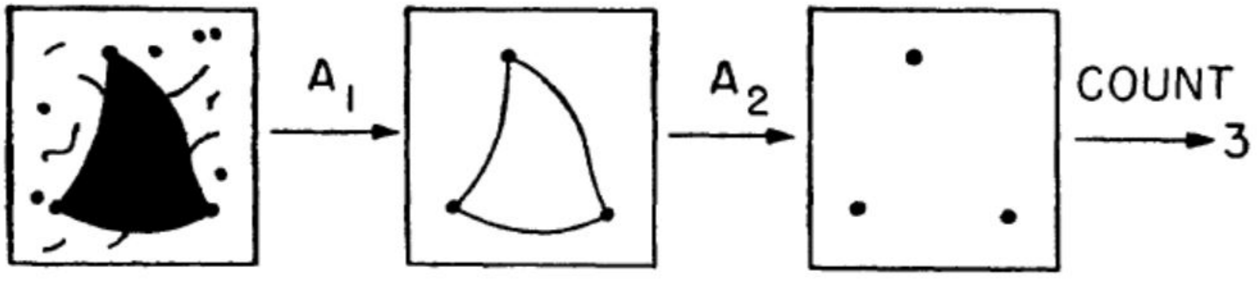
Induction – for general purpose intelligent machines

Pattern Recognition (Minsky 1960)

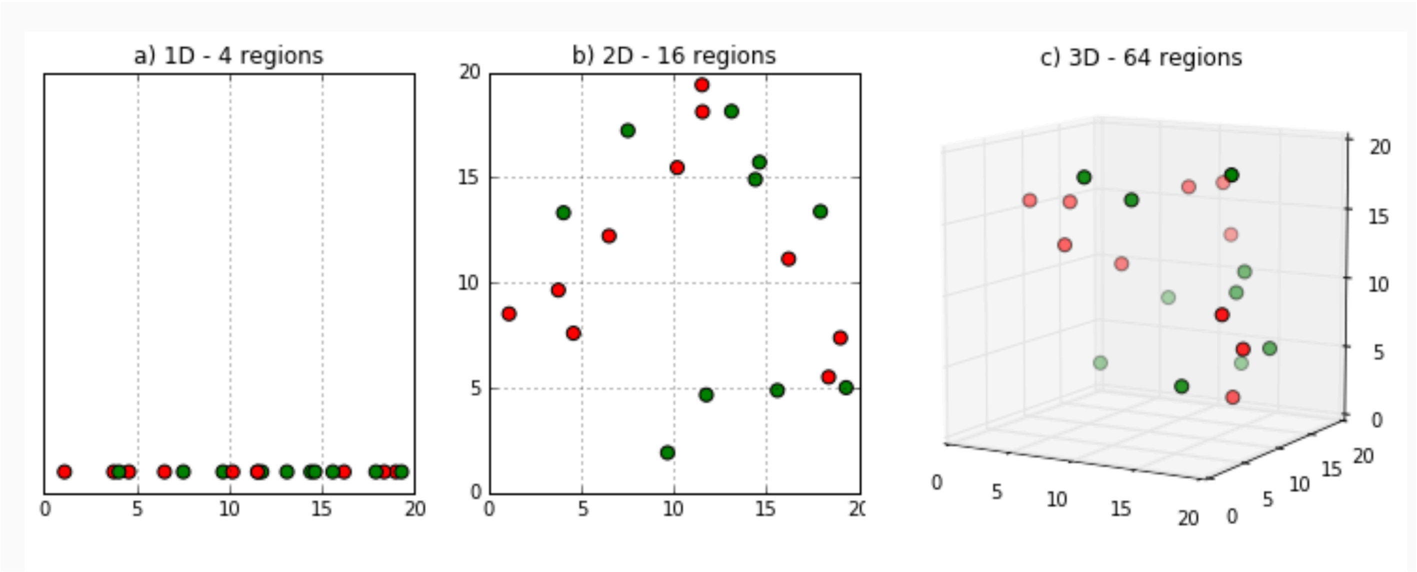
- Classify problem situations into categories associated with the machine's methods
- Extract heuristically significant features of objects
- Define useful properties and make them resistant to noise
- Combine many properties to form a recognition system

Normalisation and Dimensionality Reduction





The Curse of Dimensionality



www.kdnuggets.com/2017/n15.html

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

Induction – for general purpose intelligent machines

Learning

The basic learning heuristic is to use successful methods

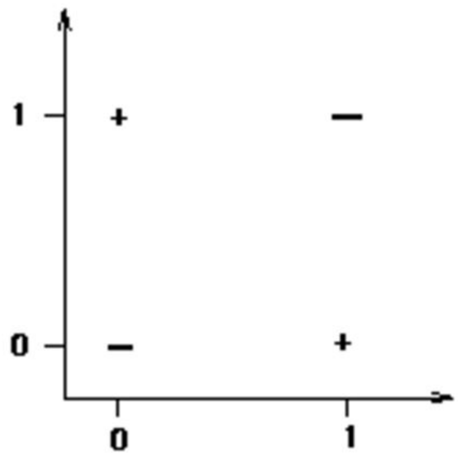
The advanced learning heuristic is meta-level learning: learning how to learn
(e.g., a program-writing program)

Success-reinforced models are averaging models and leads to stimulus-response-type reasoning

Stimulus-Response Learning

- A machine receives stimuli via its sensors or receptors
- A machine acts or performs actions via its effectors
- That an action is a response to a stimulus means they co-occur; to add a particular time, does not add to expressive power
- Since both stimulus and response may be complicated, their relation is complicated, even in this simplest case
- Reinforcement learning lets a machine mirror itself in the results of its actions

Simple Learning Limitations



XOR is a function that cannot be learned by a perceptron: let + correspond to output 1 and – to output 0, then XOR is not linearly separable

Simple learning machines are (digital) automata, whereas some organic processes are analog, like the human respiratory system (nervous response to blood CO_2 levels)

Analogue Distributed Neuromorphic Learners

PHYSICAL REVIEW E **98**, 052101 (2018)

Editors' Suggestion

Modeling reservoir computing with the discrete nonlinear Schrödinger equation

Simone Borlenghi,[1] Magnus Boman,[2,3] and Anna Delin[1,4]

[1]*Department of Applied Physics, School of Engineering Sciences, KTH Royal Institute of Technology, Electrum 229, SE-16440 Kista, Sweden*

[2]*KTH Royal Institute of Technology, EECS/SCS, Electrum 229, SE-16440 Kista, Sweden*

[3]*RISE SICS, Electrum 230, SE-16429 Kista, Sweden*

[4]*Swedish e-Science Research Center (SeRC), KTH Royal Institute of Technology, SE-10044 Stockholm, Sweden*

(Received 26 April 2018; published 1 November 2018)

We formulate, using the discrete nonlinear Schrödinger equation (DNLS), a general approach to encode and process information based on reservoir computing. Reservoir computing is a promising avenue for realizing neuromorphic computing devices. In such computing systems, training is performed only at the output level by adjusting the output from the reservoir with respect to a target signal. In our formulation, the reservoir can be an arbitrary physical system, driven out of thermal equilibrium by an external driving. The DNLS is a general oscillator model with broad application in physics, and we argue that our approach is completely general and does not depend on the physical realization of the reservoir. The driving, which encodes the object to be recognized, acts as a thermodynamic force, one for each node in the reservoir. Currents associated with these thermodynamic forces in turn encode the output signal from the reservoir. As an example, we consider numerically the problem of supervised learning for pattern recognition, using as a reservoir a network of nonlinear oscillators.

DOI: 10.1103/PhysRevE.98.052101

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

Induction – for general purpose intelligent machines

Planning

- Costly to replan over and over again
- Time constraints: Real-time? Online? Batch?
- Learning how to plan, or planning how to learn?
- Can an AI system break rules?
- Norm- rather than rule-regulation?
- Weak AI vs. Strong AI

AI Programming (Minsky 1961)

Search – through some large space of solution attempts

Pattern-recognition – restricting to appropriate methods

Learning – by directing search according to experience

Planning – by analysing and further directing search

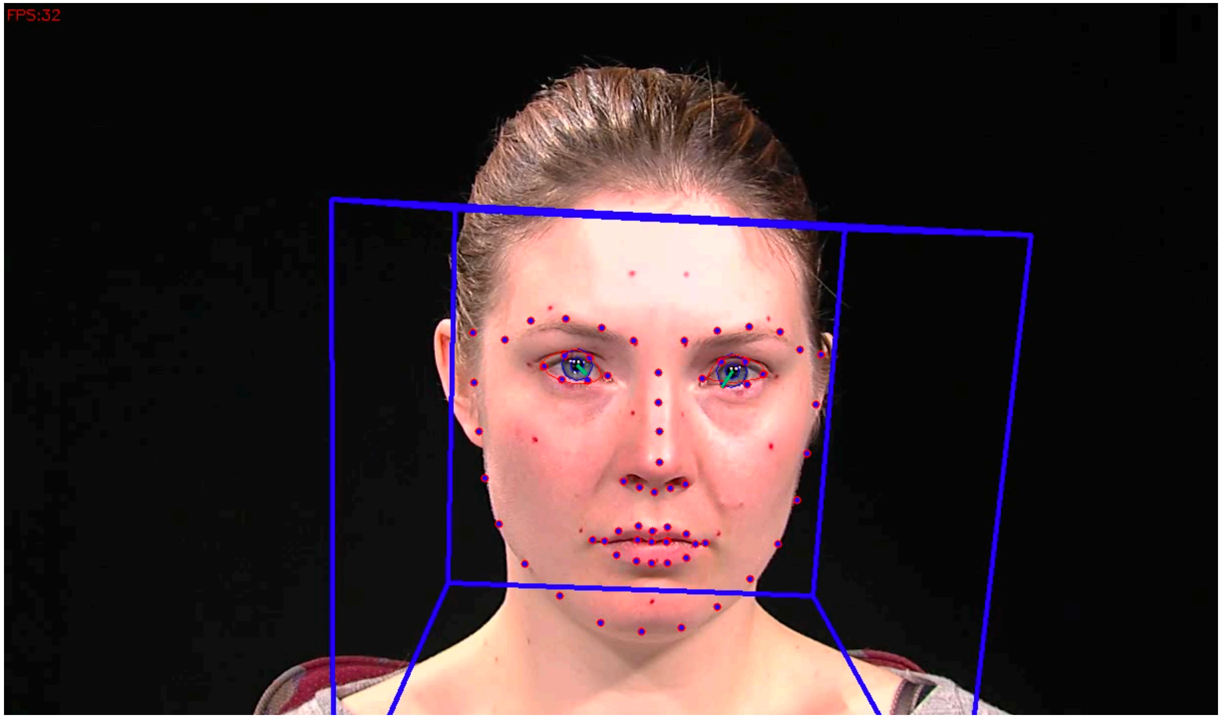
Induction – for general purpose intelligent machines

Induction (Elman 1990)

"… by themselves, the simple learning systems are useful only in recurrent situations; they cannot cope with any significant novelty. Nontrivial performance is obtained only when learning systems are supplemented with classification or pattern-recognition methods of some inductive ability".

Strategy of deep learning (Jordan 1986, Elman 1990):
Extract heuristically significant features of objects

Emotion Recognition as an Example of AI

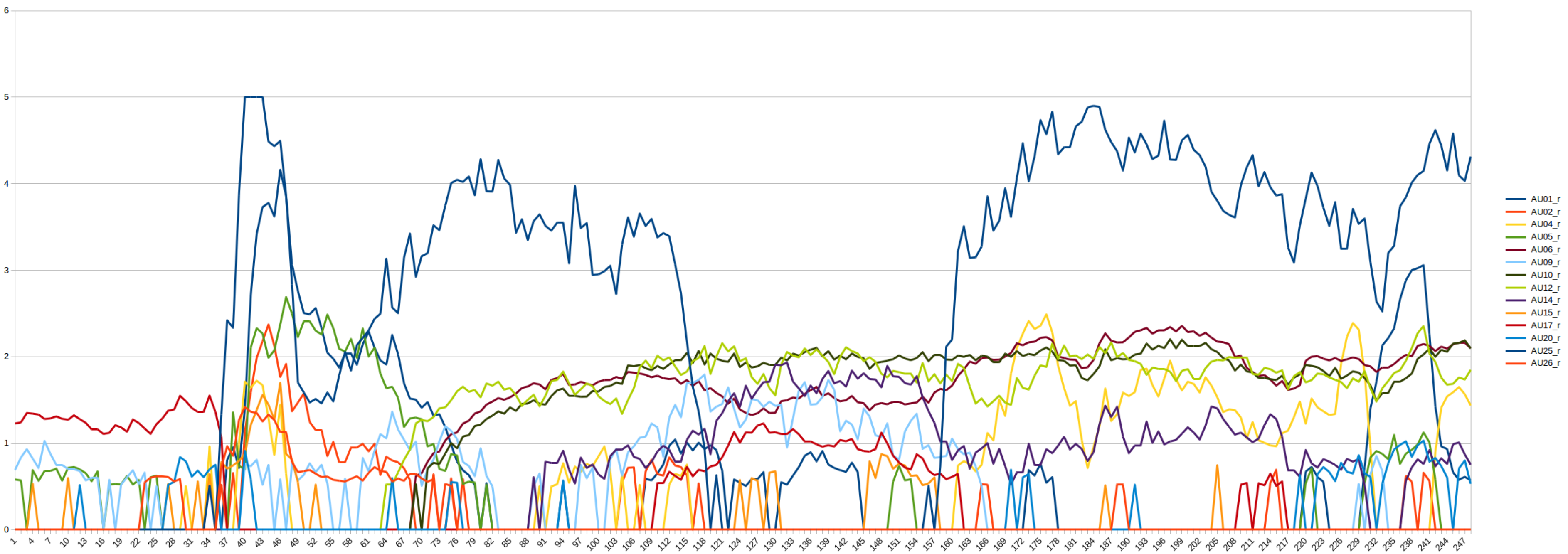


Unpublished work, with Abubakr Karali

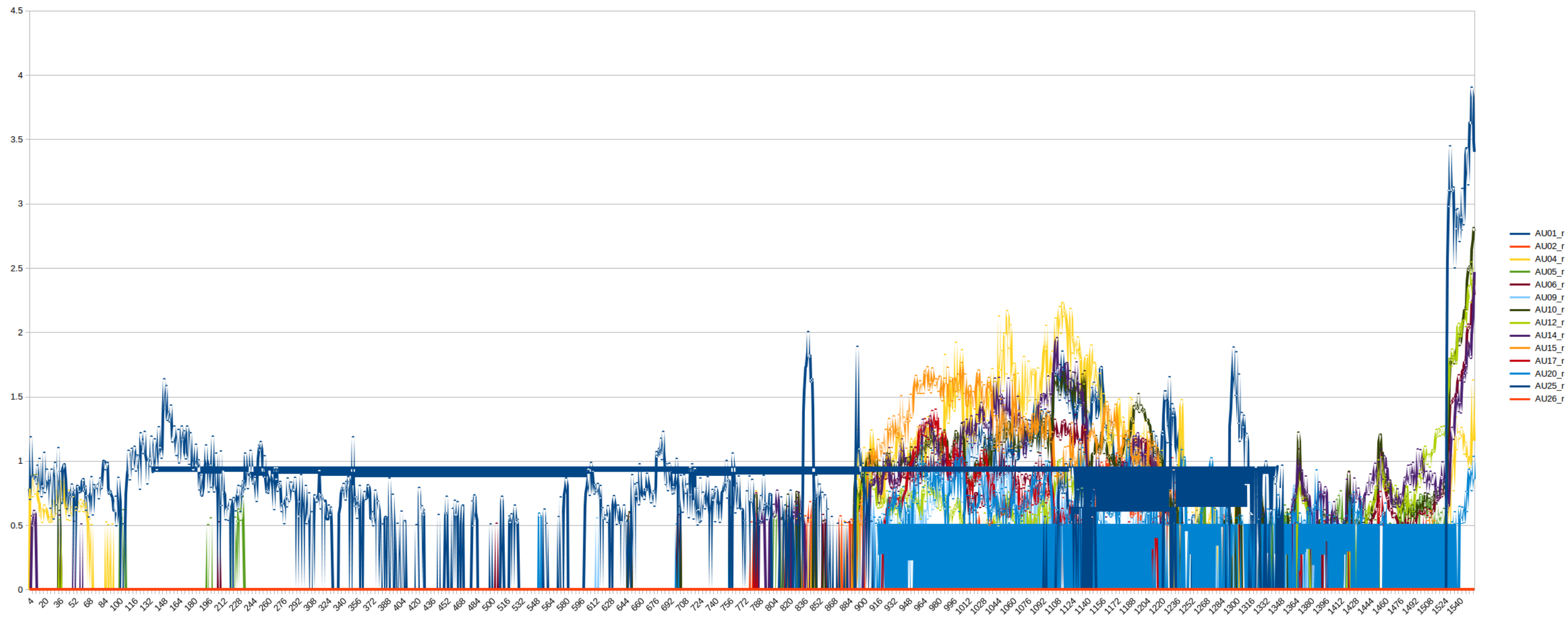
Emotion Recognition

- The Paul Ekman (FACS) paradigm vs. machine learning
- The umami of emotions
- The dream of transfer learning
- Serendipitous synergy
- Ethics, bias, and fear of AI
- *Lie to me*, not forgetting the Chinese ghost

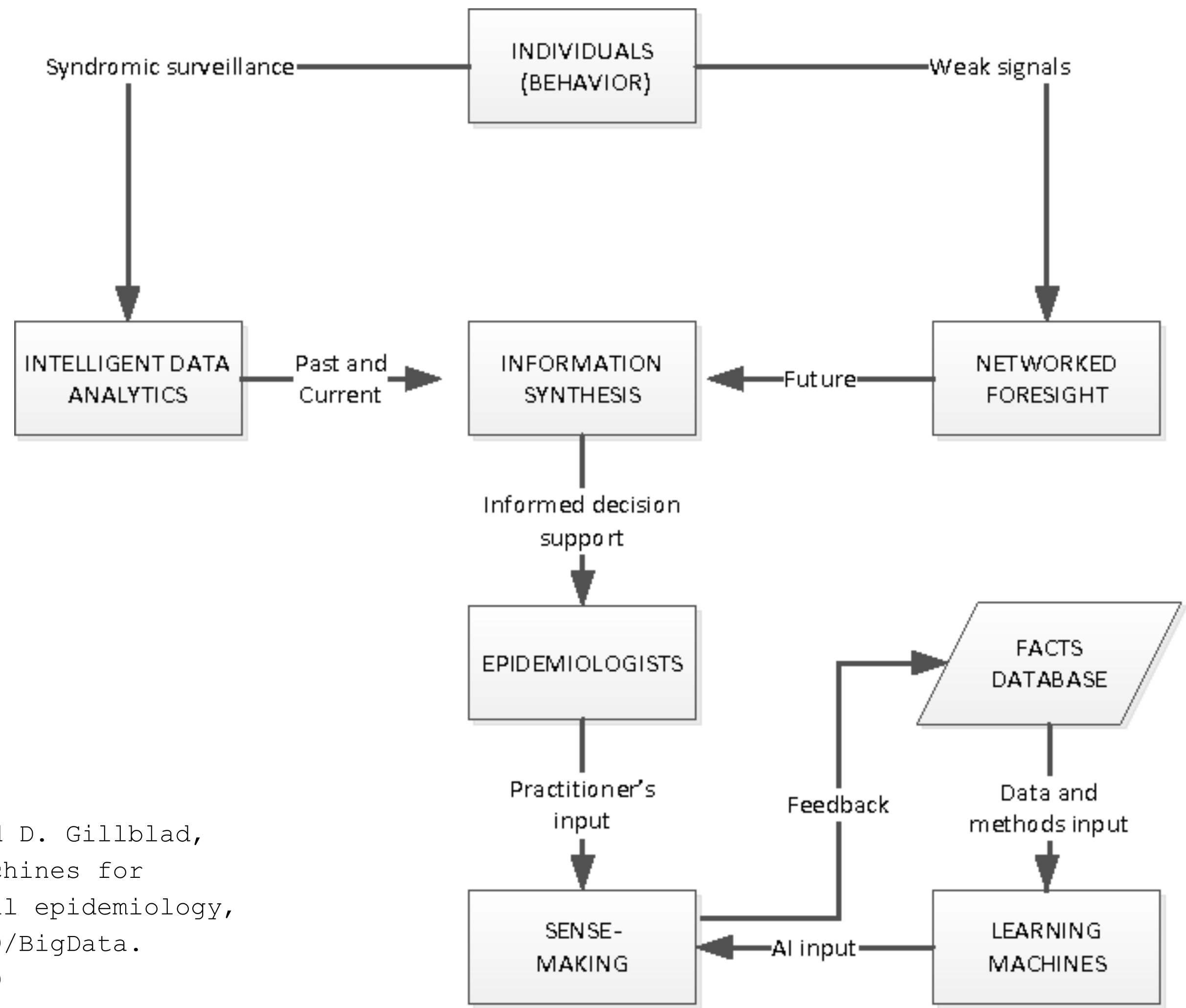
Traditional Facial Action Coding Units 2



Traditional Facial Action Coding Units 1

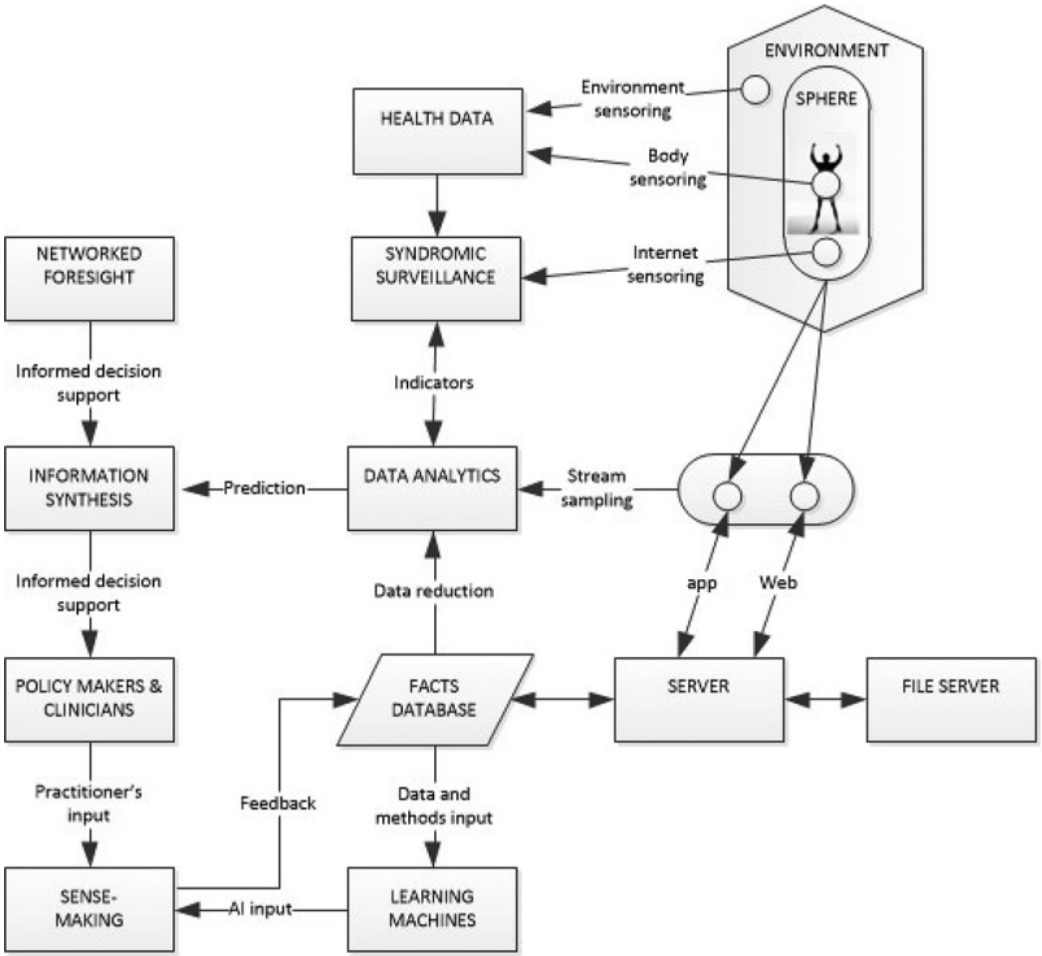


A Systemic Model of Sensemaking



M. Boman and D. Gillblad, Learning machines for computational epidemiology, doi: 10.1109/BigData.2014.7004419

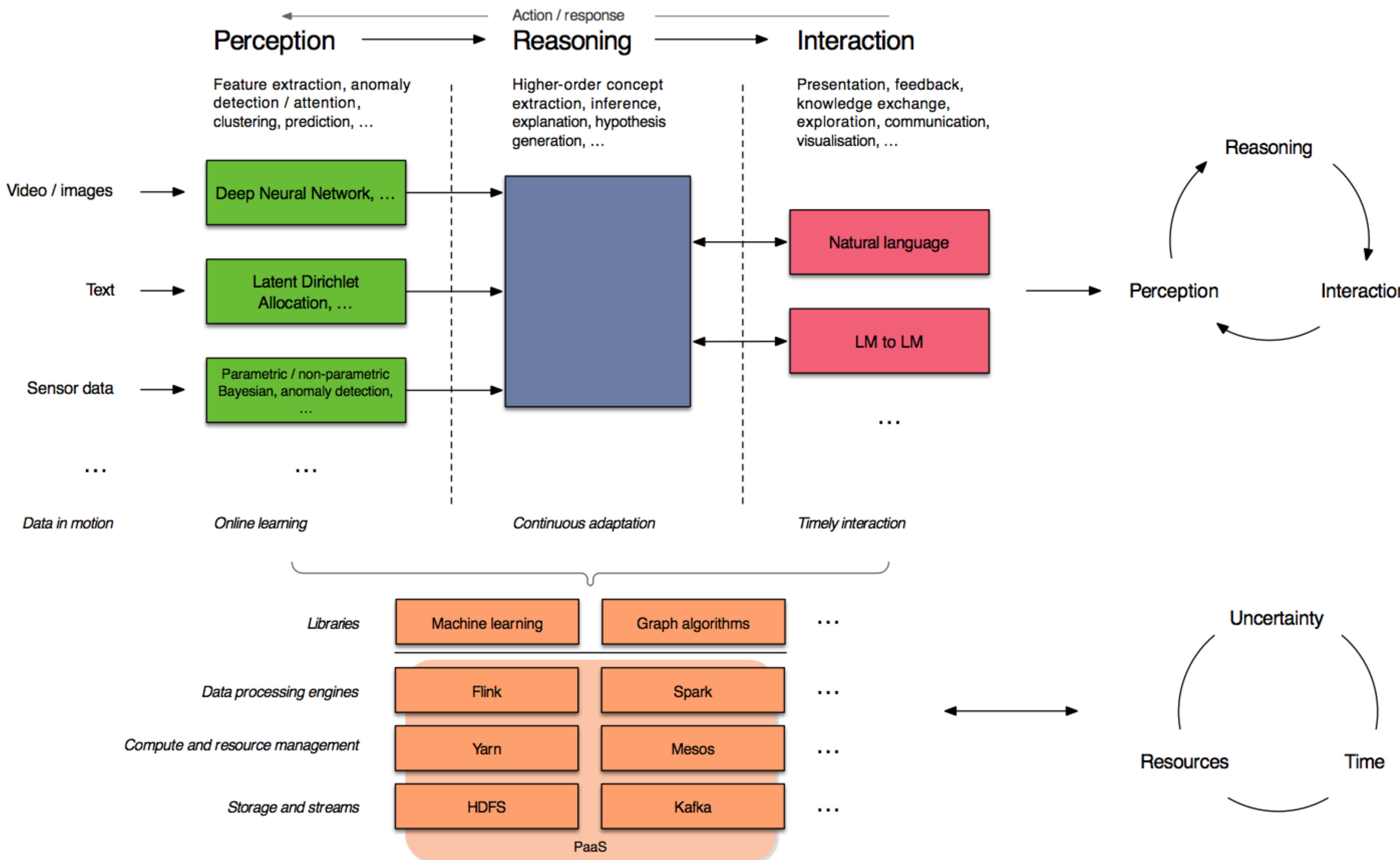
A Systemic Model of Future Sensemaking



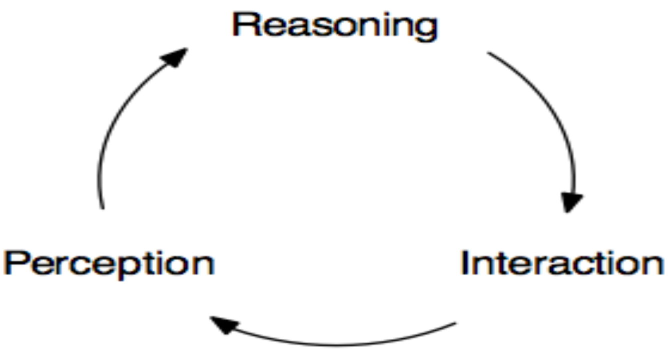
M. Boman and E Kruse, Supporting global health goals with ICT, Glob Health Action. doi: 10.1080/16549716.2017.1321904.

Learning Machines for Internet Psychiatry

Learning Machine system overview



~4.500 variables, n=5.218, complete data



Prediction, Extraction and Mining - Example

SCIENTIFIC REPORTS

OPEN

Machine Learning methods for Quantitative Radiomic Biomarkers

Chintan Parmar[1,3,4,*], Patrick Grossmann[1,5,*], Johan Bussink[6], Philippe Lambin[3] & Hugo J. W. L. Aerts[1,2,5]

Received: 02 April 2015
Accepted: 17 July 2015
Published: 17 August 2015

Radiomics extracts and mines large number of medical imaging features quantifying tumor phenotypic characteristics. Highly accurate and reliable machine-learning approaches can drive the success of radiomic applications in clinical care. In this radiomic study, fourteen feature selection methods and twelve classification methods were examined in terms of their performance and stability for predicting overall survival. A total of 440 radiomic features were extracted from pre-treatment computed tomography (CT) images of 464 lung cancer patients. To ensure the unbiased evaluation of different machine-learning methods, publicly available implementations along with reported parameter configurations were used. Furthermore, we used two independent radiomic cohorts for training ($n = 310$ patients) and validation ($n = 154$ patients). We identified that Wilcoxon test based feature selection method WLCX (stability $= 0.84 \pm 0.05$, AUC $= 0.65 \pm 0.02$) and a classification method random forest RF (RSD $= 3.52\%$, AUC $= 0.66 \pm 0.03$) had highest prognostic performance with high stability against data perturbation. Our variability analysis indicated that the choice of classification method is the most dominant source of performance variation (34.21% of total variance). Identification of optimal machine-learning methods for radiomic applications is a crucial step towards stable and clinically relevant radiomic biomarkers, providing a non-invasive way of quantifying and monitoring tumor-phenotypic characteristics in clinical practice.

Feature Selection and Classification

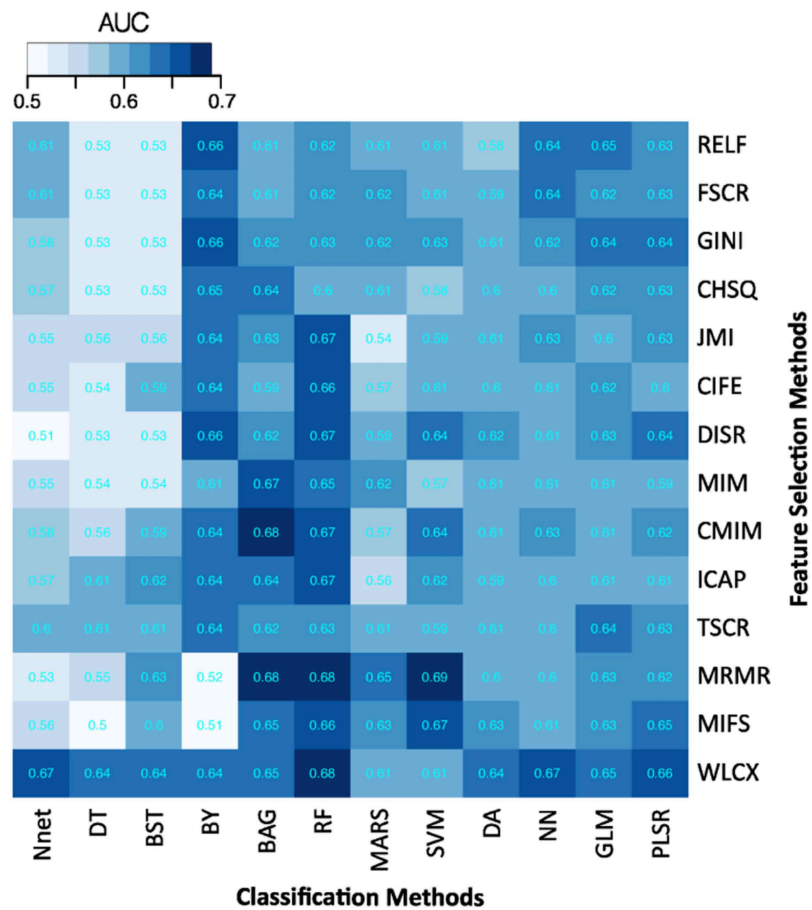


Figure 2. Heatmap depicting the predictive performance (AUC) of feature selection (in rows) and classification (in columns) methods. It can be observed that RF, BAG and BY classification methods and feature selection methods WLCX, MRMR and MIFS shows relatively high predictive performance in many cases.

Ethics

- Interpretability
- Transparency
- Accountability
- What black boxes do you use?

Interpretability

The Mythos of Model Interpretability

Zachary C. Lipton

(Submitted on 10 Jun 2016 (v1), last revised 6 Mar 2017 (this version, v3))

Supervised machine learning models boast remarkable predictive capabilities. But can you trust your model? Will it work in deployment? What else can it tell you about the world? We want models to be not only good, but interpretable. And yet the task of interpretation appears underspecified. Papers provide diverse and sometimes non-overlapping motivations for interpretability, and offer myriad notions of what attributes render models interpretable. Despite this ambiguity, many papers proclaim interpretability axiomatically, absent further explanation. In this paper, we seek to refine the discourse on interpretability. First, we examine the motivations underlying interest in interpretability, finding them to be diverse and occasionally discordant. Then, we address model properties and techniques thought to confer interpretability, identifying transparency to humans and post-hoc explanations as competing notions. Throughout, we discuss the feasibility and desirability of different notions, and question the oft-made assertions that linear models are interpretable and that deep neural networks are not.

Comments: presented at 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, NY

Subjects: **Machine Learning (cs.LG)**; Artificial Intelligence (cs.AI); Computer Vision and Pattern Recognition (cs.CV); Neural and Evolutionary Computing (cs.NE); Machine Learning (stat.ML)

Cite as: arXiv:1606.03490 **[cs.LG]**

(or arXiv:1606.03490v3 **[cs.LG]** for this version)

The Largest Problem for AI: Talent

Christian Drosten: Spend money on training and labs

John Nkengasong: 1.500 epidemiologists for 1.2B Africans

Take Home Message

Don’t believe the hype!

mab@kth.se