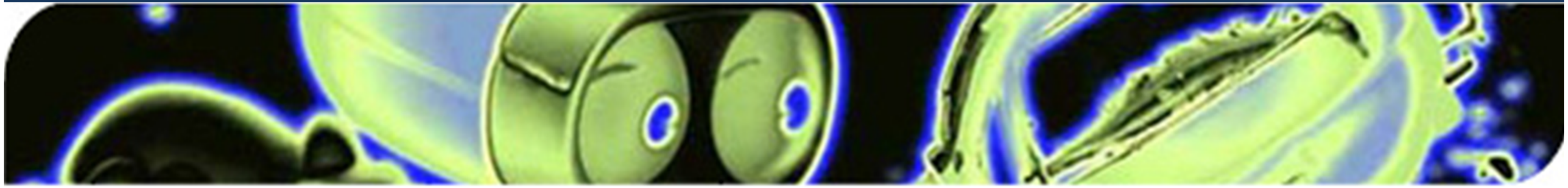
Alfredo Vellido

Mind the interpreter:

Notes on making machine learning interpretable in biomedicine and beyond

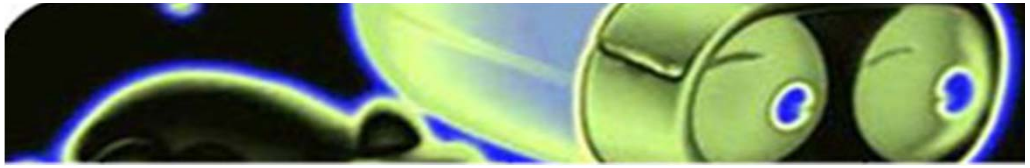
ICANN Sept 6 2016, Barcelona



SOCO
Soft Computing Research Group

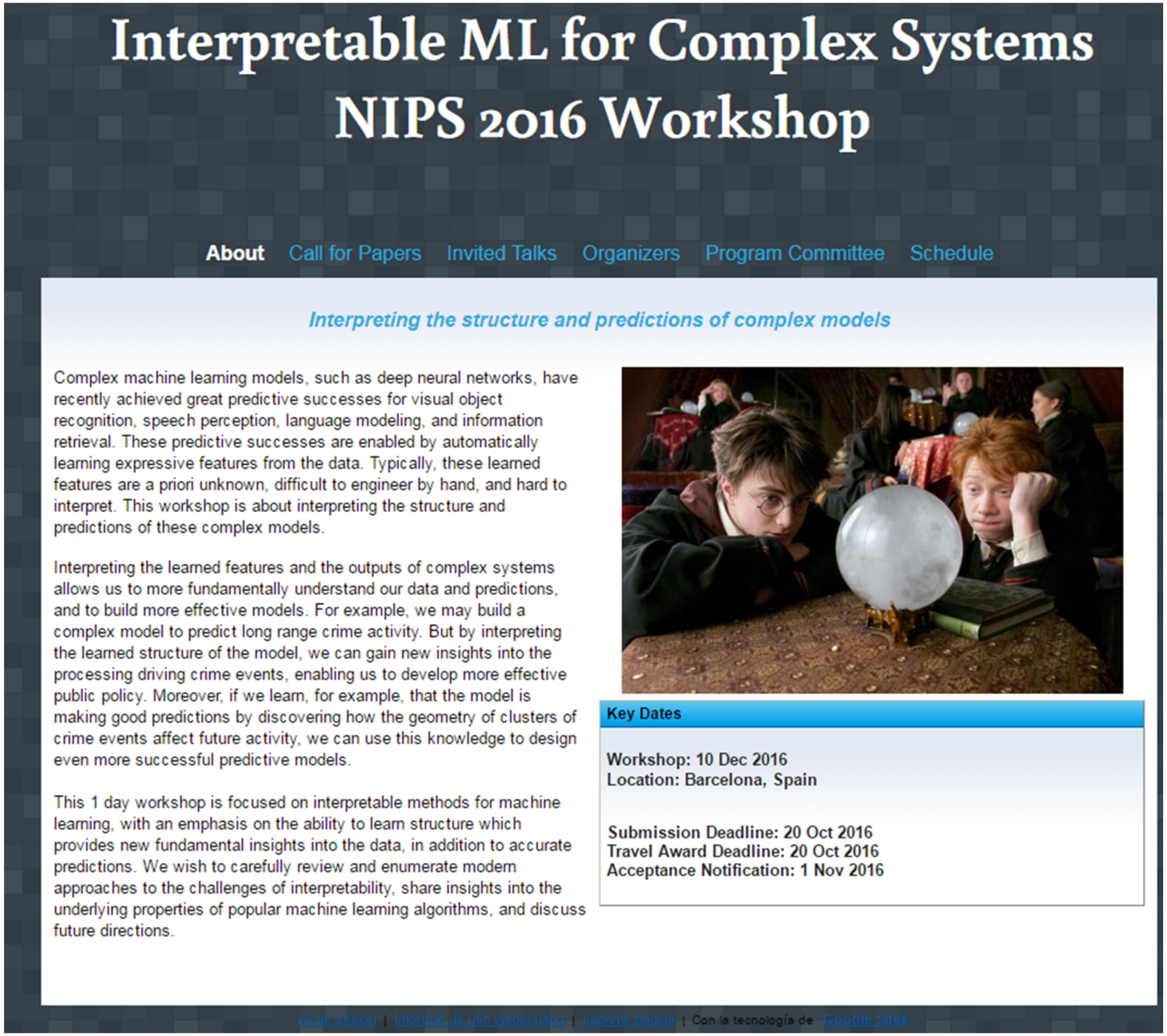
UPC
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Mind the Interpreters



Interpreting BCN

ICANN, NIPS, ICDM



Interpretable ML for Complex Systems
NIPS 2016 Workshop

Interpreting the structure and predictions of complex models

Complex machine learning models, such as deep neural networks, have recently achieved great predictive successes for visual object recognition, speech perception, language modeling, and information retrieval. These predictive successes are enabled by automatically learning expressive features from the data. Typically, these learned features are a priori unknown, difficult to engineer by hand, and hard to interpret. This workshop is about interpreting the structure and predictions of these complex models.

Interpreting the learned features and the outputs of complex systems allows us to more fundamentally understand our data and predictions, and to build more effective models. For example, we may build a complex model to predict long range crime activity. But by interpreting the learned structure of the model, we can gain new insights into the processing driving crime events, enabling us to develop more effective public policy. Moreover, if we learn, for example, that the model is making good predictions by discovering how the geometry of clusters of crime events affect future activity, we can use this knowledge to design even more successful predictive models.

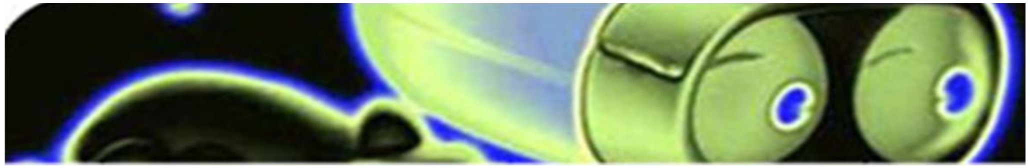
This 1 day workshop is focused on interpretable methods for machine learning, with an emphasis on the ability to learn structure which provides new fundamental insights into the data, in addition to accurate predictions. We wish to carefully review and enumerate modern approaches to the challenges of interpretability, share insights into the underlying properties of popular machine learning algorithms, and discuss future directions.

Key Dates

Workshop: 10 Dec 2016
Location: Barcelona, Spain

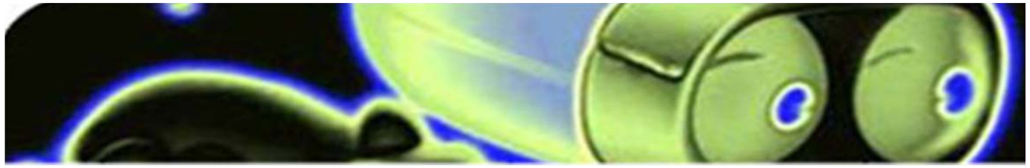
Submission Deadline: 20 Oct 2016
Travel Award Deadline: 20 Oct 2016
Acceptance Notification: 1 Nov 2016

Iniciar sesión | Informar de uso inadecuado | Imprimir página | Con la tecnología de Google Sites



Looking for a motivation

- **Data deluge.** Most fields of science + economy becoming data-intensive.
- Data of different levels of **complexity** and of ever growing **diversity** of characteristics.
- These are materials that ML/PR/CI practitioners try to **model** using an (also growing) palette of methods and tools.
- The obtained **models** are meant to be a synthetic representation of the observed data that **captures some of their intrinsic regularities or patterns**.
- This can and should be understood as a problem of **pattern recognition and knowledge discovery**, providing a door to **interpretability.**
- **In this context, can we live without model interpretation?**



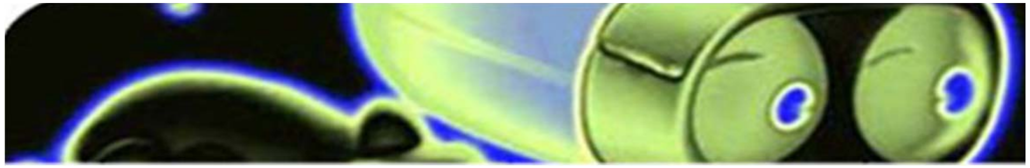
Looking for a definition

- **in·ter·pret** [in-**tur**-prit]

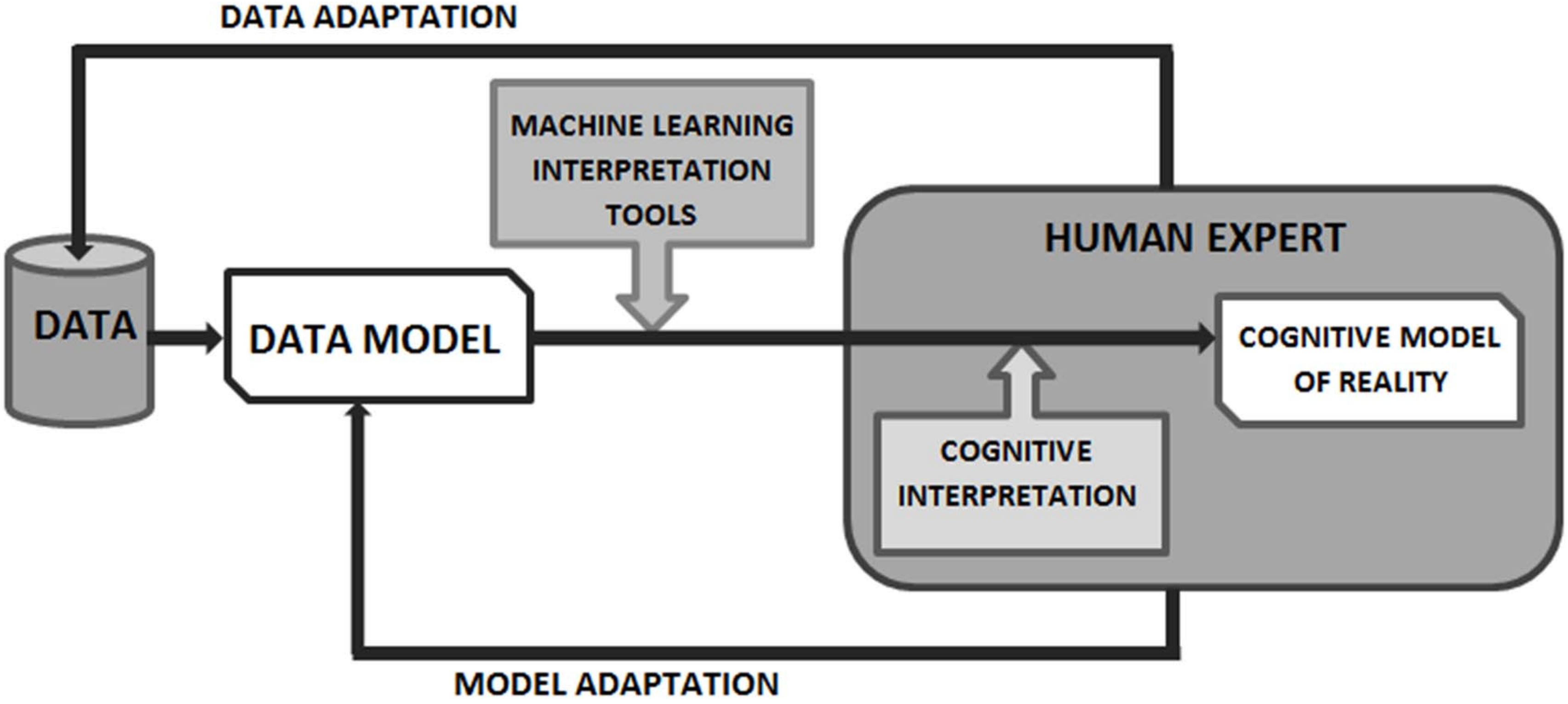
  verb (used with object)

  1. to give or provide the meaning of; explain; explicate; elucidate

- We need to acknowledge the existence of a **GAP between data modeling and knowledge extraction**.
- **Models as such can be rendered powerless in practice unless they can be interpreted**.
- In order to consider that some knowledge has been achieved from model description, we must take into account the **human cognitive factor** that any knowledge extraction process entails.
- The problem is that the process of human interpretation follows rules that sometimes go beyond technical prowess.

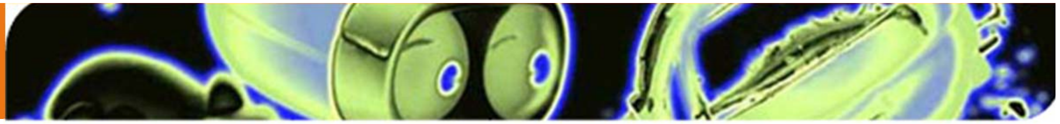


The human factor

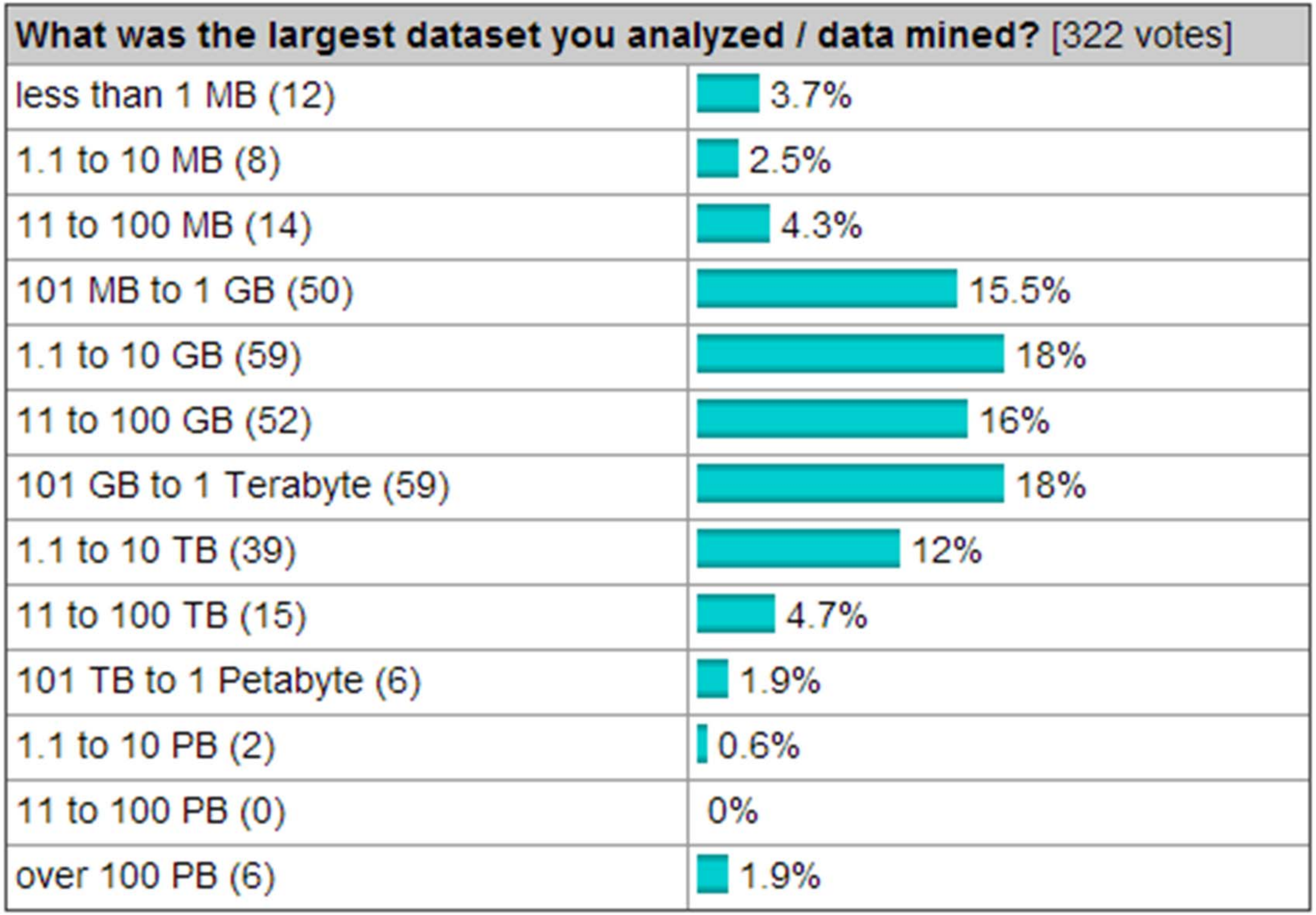


- In the end, interpretability is a paramount quality that ML and related methods should aim to achieve if they are **to be applied in practice**.

BIG DATA

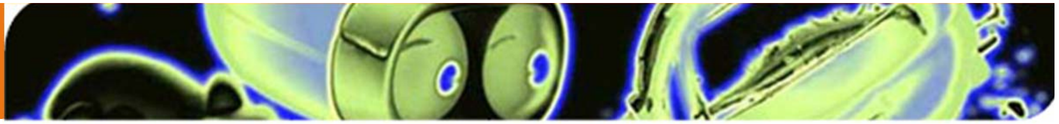


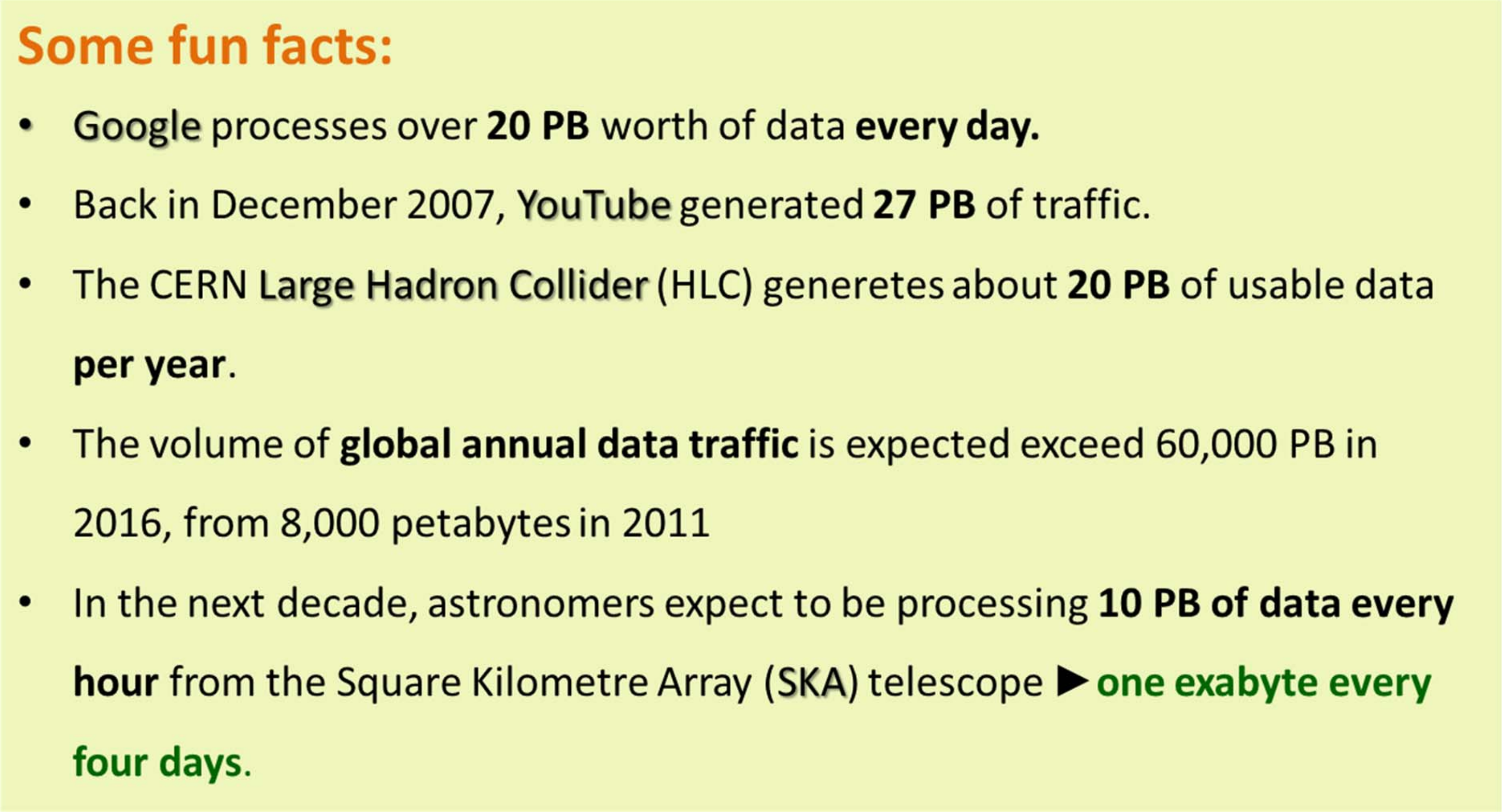
How big is yours? … (2013, KDNuggets)



What was the largest dataset you analyzed / data mined? [322 votes]	
less than 1 MB (12)	3.7%
1.1 to 10 MB (8)	2.5%
11 to 100 MB (14)	4.3%
101 MB to 1 GB (50)	15.5%
1.1 to 10 GB (59)	18%
11 to 100 GB (52)	16%
101 GB to 1 Terabyte (59)	18%
1.1 to 10 TB (39)	12%
11 to 100 TB (15)	4.7%
101 TB to 1 Petabyte (6)	1.9%
1.1 to 10 PB (2)	0.6%
11 to 100 PB (0)	0%
over 100 PB (6)	1.9%

BIG & SMALL





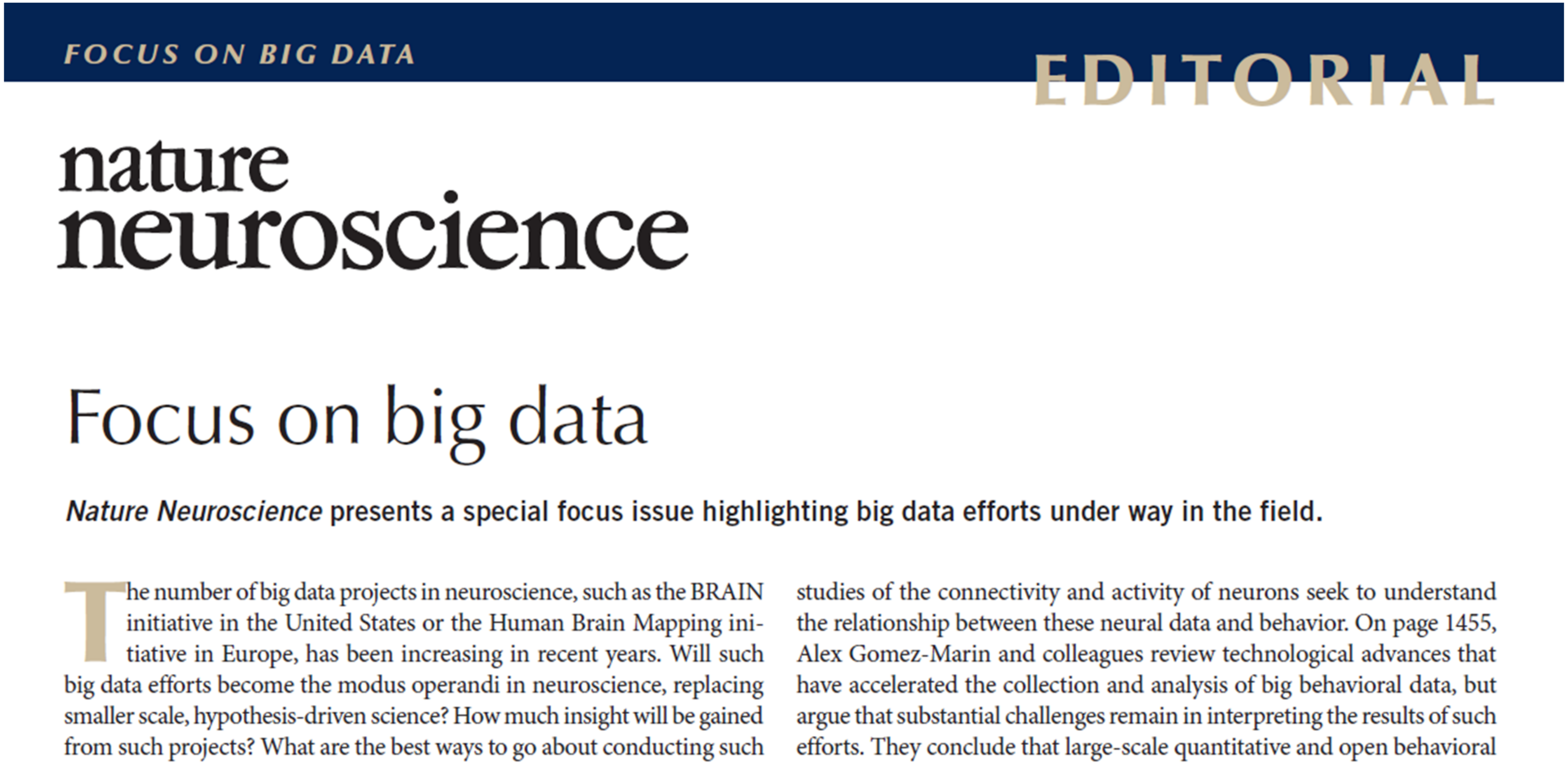
Some fun facts:

- Google processes over **20 PB** worth of data **every day.**
- Back in December 2007, YouTube generated **27 PB** of traffic.
- The CERN Large Hadron Collider (HLC) generetes about **20 PB** of usable data **per year**.
- The volume of **global annual data traffic** is expected exceed 60,000 PB in 2016, from 8,000 petabytes in 2011
- In the next decade, astronomers expect to be processing **10 PB of data every hour** from the Square Kilometre Array (SKA) telescope ▶ **one exabyte every four days**.

The Big Data Interpretation Challenge

OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …

Nov 14



FOCUS ON BIG DATA

EDITORIAL

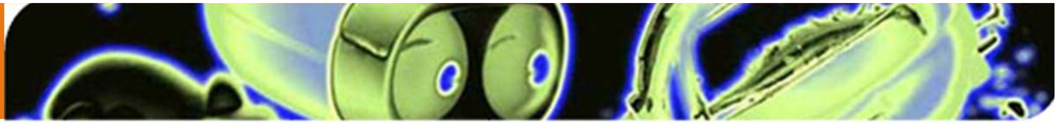
nature neuroscience

Focus on big data

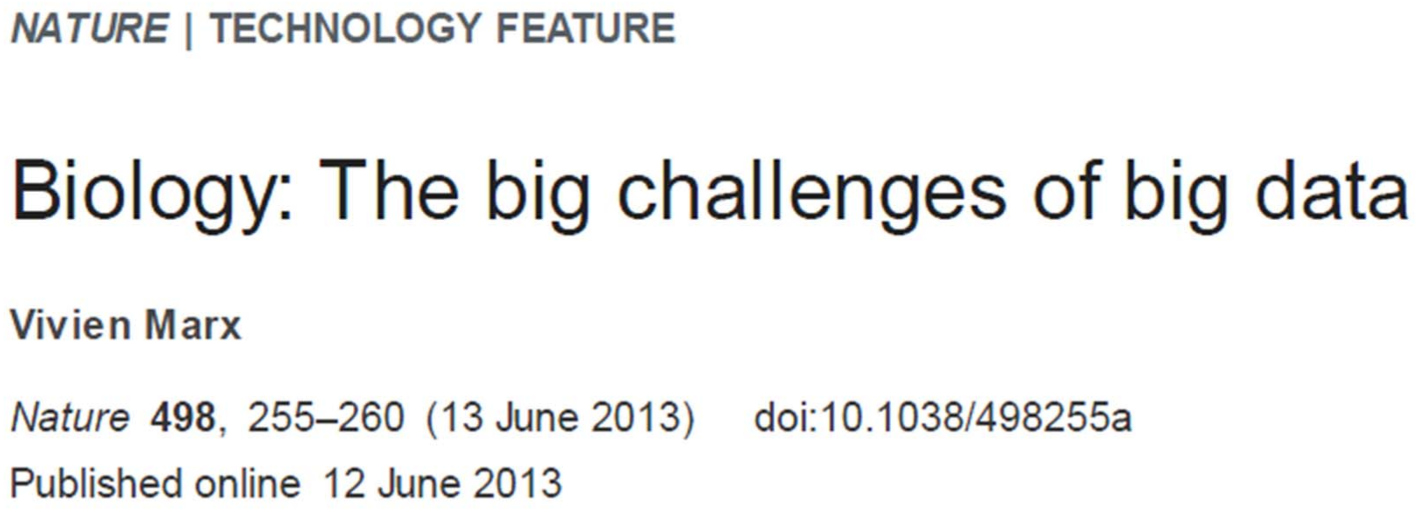
Nature Neuroscience **presents a special focus issue highlighting big data efforts under way in the field.**

The number of big data projects in neuroscience, such as the BRAIN initiative in the United States or the Human Brain Mapping initiative in Europe, has been increasing in recent years. Will such big data efforts become the modus operandi in neuroscience, replacing smaller scale, hypothesis-driven science? How much insight will be gained from such projects? What are the best ways to go about conducting such studies of the connectivity and activity of neurons seek to understand the relationship between these neural data and behavior. On page 1455, Alex Gomez-Marin and colleagues review technological advances that have accelerated the collection and analysis of big behavioral data, but argue that substantial challenges remain in interpreting the results of such efforts. They conclude that large-scale quantitative and open behavioral

BIG & SMALL



OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …



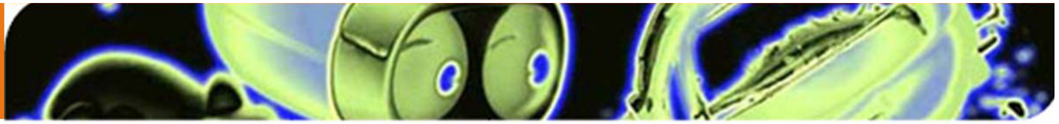
NATURE | TECHNOLOGY FEATURE

Biology: The big challenges of big data

Vivien Marx

Nature **498**, 255–260 (13 June 2013) doi:10.1038/498255a

Published online 12 June 2013

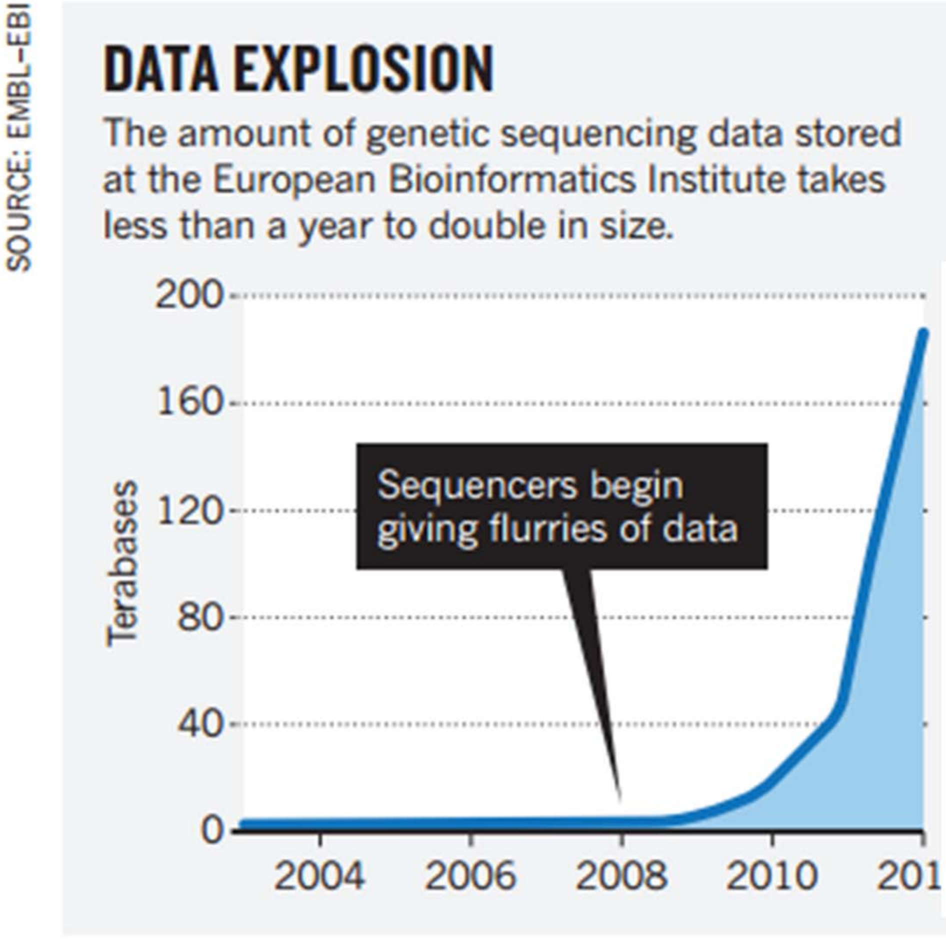


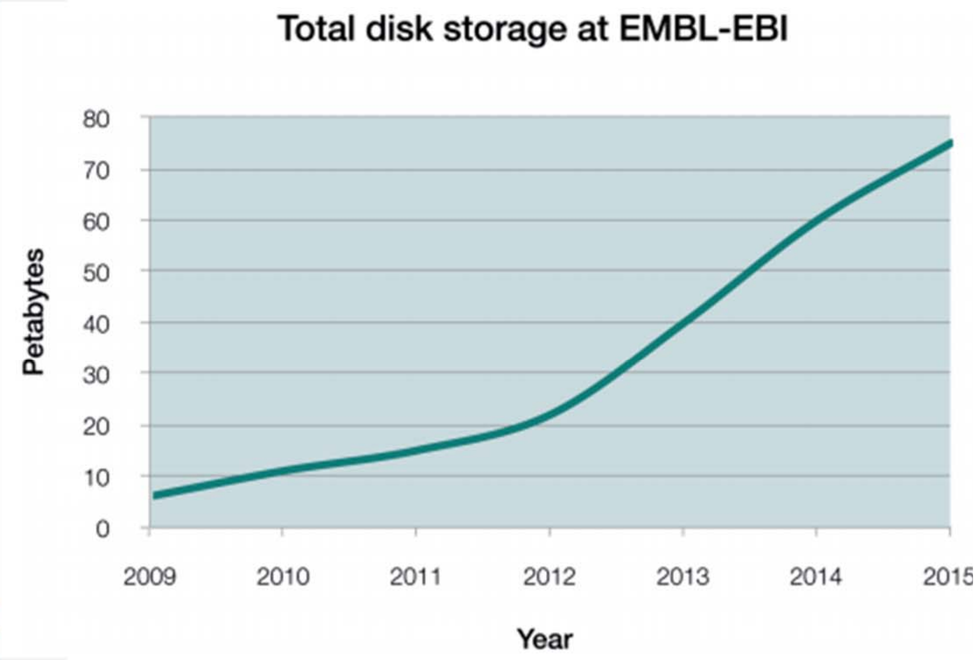
OK, but this is about the big ITCorps, la US NSA and the like, isn't it? …

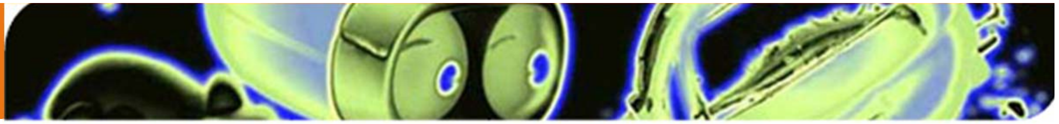
- "Biologists are joining the big-data club. With the advent of high-throughput genomics, life scientists are starting to grapple with massive data sets, encountering challenges with **handling**, **processing** and **moving** information that were once the domain of astronomers and high-energy physicists."

- "The **European Bioinformatics Institute (EBI)** in Hinxton, UK, […] one of the world's largest biology-data repositories, currently stores 20 petabytes […] Genomic data account for 2 Pb of that, a **number that more than doubles every year.**"

OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …



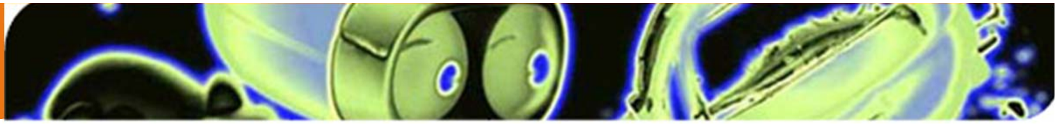




OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …

- “As **prices drop for high-throughput instruments** […], small biology labs can become big-data generators. […] **Each day** [2012], the EBI received about 9 million online requests to query its data.”

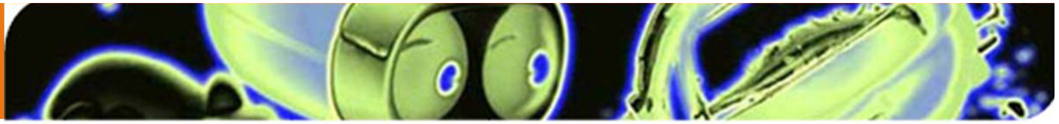
- “That means scientists have to **store** large data sets, and **analyse**, **compare** and **share** them […] Even **a single sequenced human genome is around 140 Gb** in size. Comparing human genomes takes more than a personal computer and online file-sharing applications …”



OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …

- “Much of the construction in big-data biology is virtual, focused on **cloud computing** […] “People rarely work on straight hardware anymore,” says Birney. One heavily used resource is the **Ensembl Genome Browser** […] The main Ensembl site is based on hardware in the UK, but when users in the US and Japan had difficulty accessing the data quickly, the EBI resolved the bottleneck by hosting mirror sites at three of the many **remote data centres that are part of Amazon Web Services' Elastic Compute Cloud (EC2)**.”

- “**Clouds** are a solution, but they also throw up fresh challenges. Ironically, **their proliferation can cause a bottleneck** if data end up parked on several clouds and thus still need to be moved to be shared. And using clouds means **entrusting valuable data to a distant service provider** …”

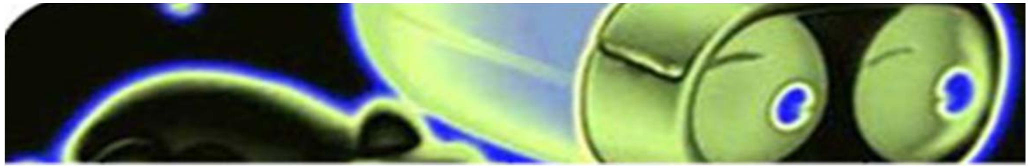


OK, but this is about the big ITCorps, la US NSA and the like, isn’t it? …

- “**Most researchers tend to download remote data to local hardware for analysis. But this method is “backward”**, says Andreas Sundquist, chief technology officer of DNAnexus. “The data are so much larger than the tools, it makes no sense to be doing that.” The alternative is to **use the cloud for both data storage and computing** […] There's no reason to move data outside the cloud. You can do analysis right there …”

SMALL DATA

Mind the Interpreters



Who is the customer?

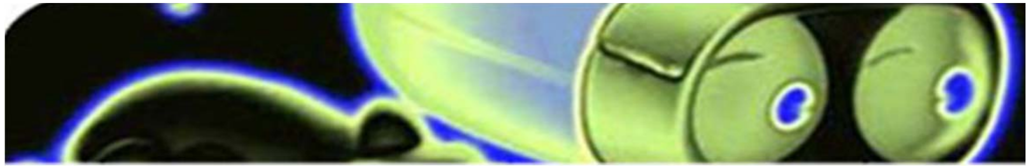
- **An example in the medical domain**. Medical experts may only accept a parsimonious outcome from a ML method, as **they require an explainable basis for their decision making** that **complies with their standard operational guidelines**, often based on simple and rigid attribute scores.

Mortality prediction due to sepsis at the ICU:

Ribas, V.J., Vellido, A., Ruiz-Rodríguez, J.C., Rello, J. (2012) Severe sepsis mortality prediction with logistic regression over latent factors. ***Expert Systems with Applications***, 39(2), 1937-1943.

- **Logistic Regression + Factor Analysis**: as complex as it gets in an application context in which **standard scores** (SOFA, APACHE) are routinely used.
- Regardless success in prediction, end-user adoption of alternative methods should not be expected.
- **Beware of existing methods of interpretation …**

Mind the Interpreters



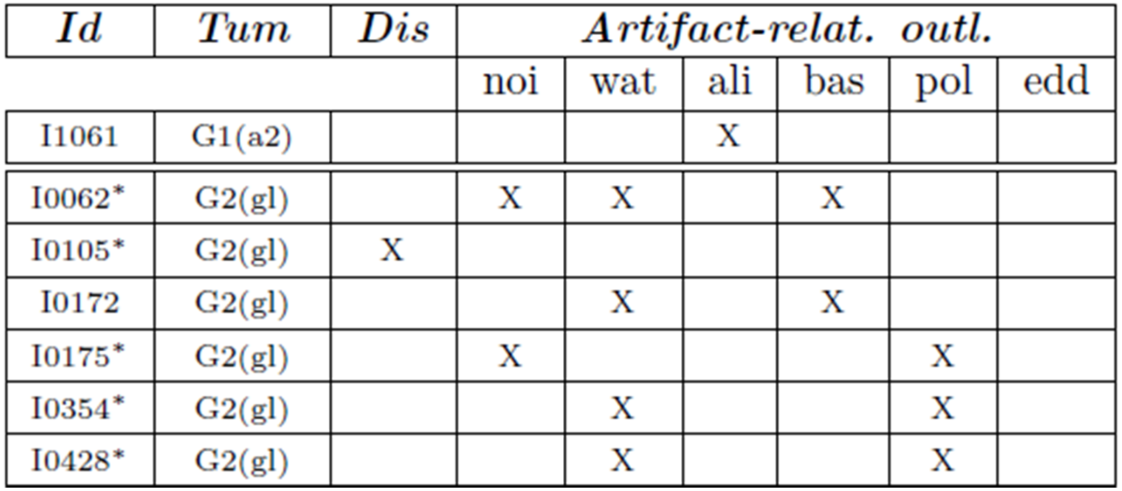
Who is the customer? (&2)

- **Another example in the medical domain**. Interpretation can also be **a matter of language**. We could be talking in a different language that domain experts might not be interested to talk.

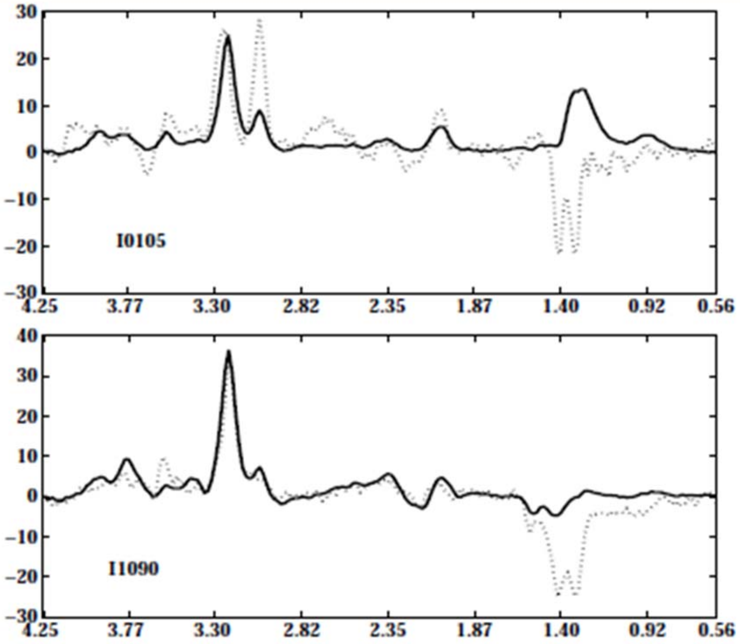
Brain tumour diagnosis for radiologists:

Vellido, A., Romero, E., González-Navarro, F.F., Belanche-Muñoz, Ll., Julià-Sapé, M., Arús, C. (2009) Outlier exploration and diagnostic classification of a multi-centre 1H-MRS brain tumour database. ***Neurocomputing***, 72(13-15), 3085-3097.

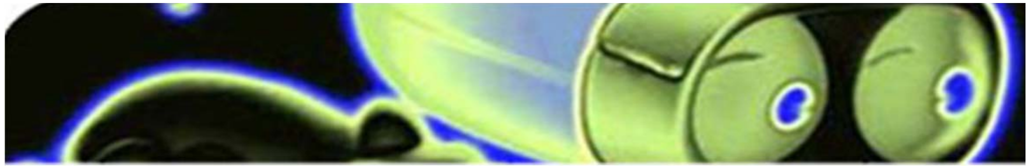
- **Beware of prior expert knowledge**



Id	*Tum*	*Dis*	*Artifact-relat. outl.*					
			noi	wat	ali	bas	pol	edd
I1061	G1(a2)				X			
I0062*	G2(gl)		X	X		X		
I0105*	G2(gl)	X						
I0172	G2(gl)			X		X		
I0175*	G2(gl)		X				X	
I0354*	G2(gl)			X			X	
I0428*	G2(gl)			X			X	



Mind the Interpreters

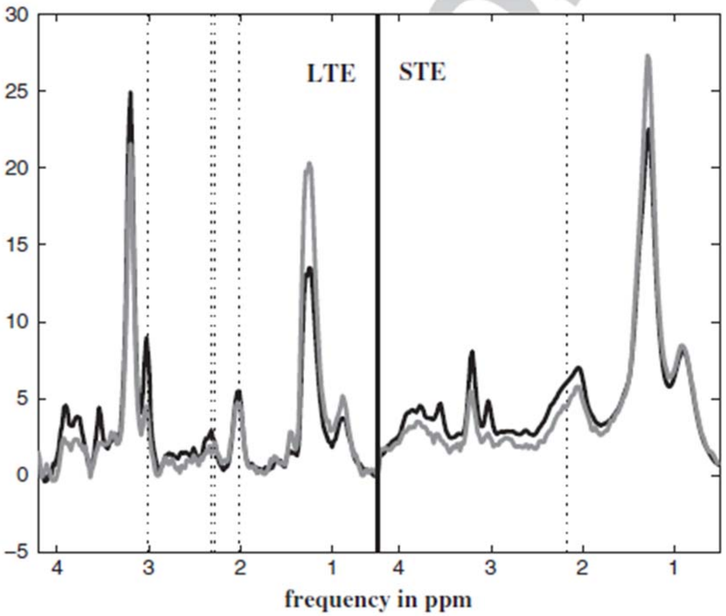


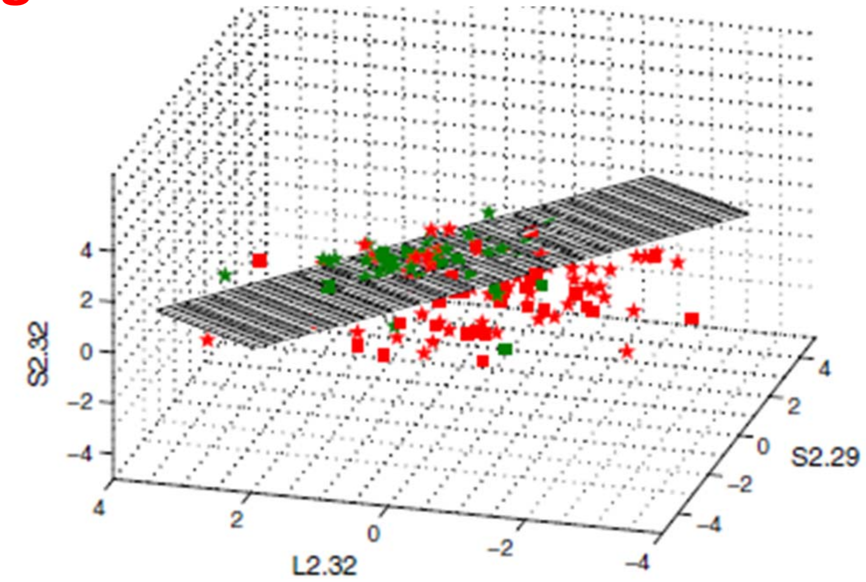
Who is the customer? (&2b)

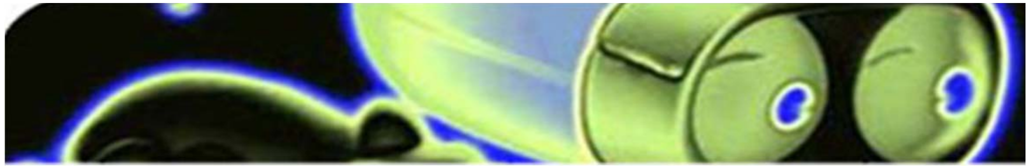
- **Another example in the medical domain.** Interpretation can also be **a matter of language**. We could be talking in a different language that domain experts might not be interested to talk.
- **Brain tumour diagnosis for radiologists:**

Vellido, A., Romero, E., Julià-Sapé, M., Majós, C., Moreno-Torres, À., and Arús, C. (2012) Robust discrimination of glioblastomas from metastatic brain tumors on the basis of single-voxel proton MRS. ***NMR in Biomedicine***, 25(6), 819-828.

- **Beware of prior expert knowledge**

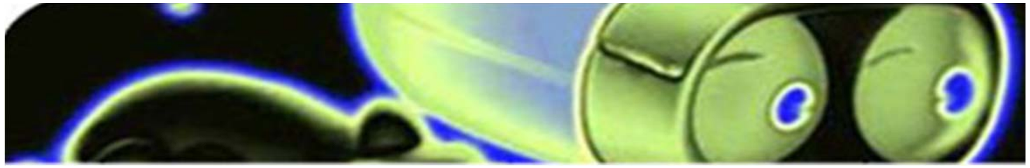






DR for Interpretation

- Problems of **high and very high dimensionality** are **becoming commonplace** in bio-fields such as, for instance, bioinformatics …
- Almost no problem is interpretable in practice if all data attributes are retained and used to provide an outcome. Furthermore, **data of very high-dimensionality are bound to show unexpected geometrical properties** that might affect their modeling and bias the interpretation of results.
- **Two main DR approaches**: **feature selection** (supervised and unsupervised) and **feature extraction**, in which new non-observable features are created on the basis of the observed ones ...
- **NOTE:** some of the most popular DR techniques in real-world applications are precisely some of the simplest ones (e.g., the ubiquitous PCA). !

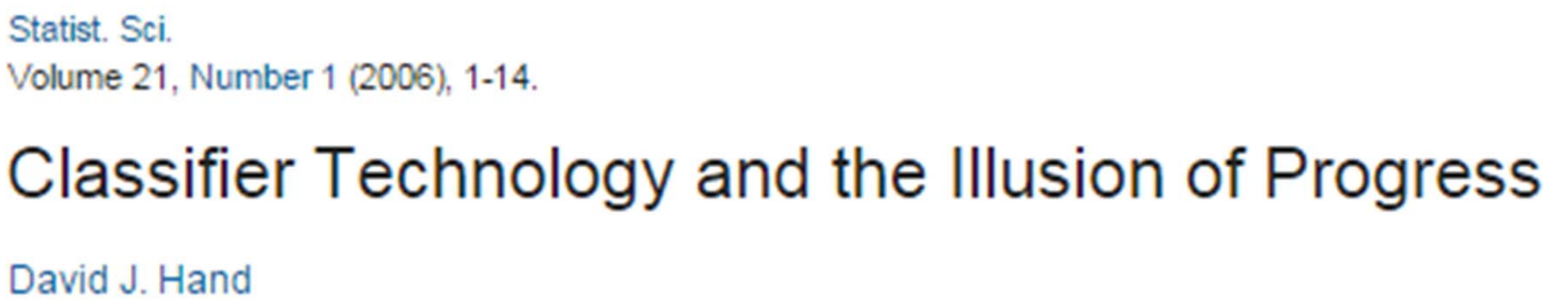


(NL)DR for Interpretation

- Many relevant ML contributions to the problem of multivariate data DR have stemmed from the **field of NLDR**.
- The **challenge of interpretability is very explicit here**: NLDR methods rarely provide an easy interpretation of the outcome in terms of the original data features, because such **outcome is usually a non-trivial nonlinear function of these features**.
- NLDR techniques usually attempt to **minimize** the **distortion** they introduce in the **mapping of the data from the observed space onto lower-dimensional spaces**.

SIMPLE MODELS x SMALL DATA

- **Politically incorrect Prof. David J. Hand**

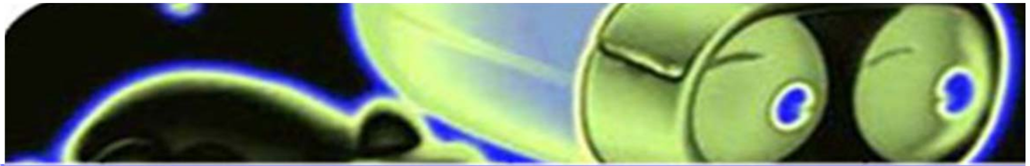


Statist. Sci.
Volume 21, Number 1 (2006), 1-14.

Classifier Technology and the Illusion of Progress

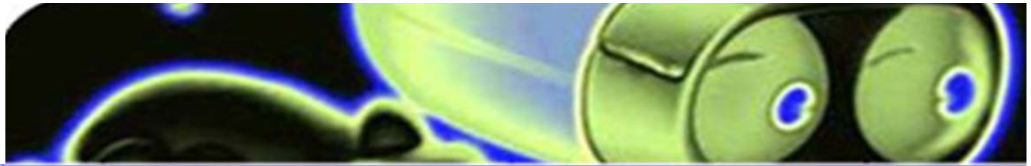
David J. Hand

Abstract
A great many tools have been developed for supervised classification, ranging from early methods such as linear discriminant analysis through to modern developments such as neural networks and support vector machines. A large number of comparative studies have been conducted in attempts to establish the relative superiority of these methods. **This paper argues that these comparisons often fail to take into account important aspects of real problems, so that the apparent superiority of more sophisticated methods may be something of an illusion**. In particular, simple methods typically yield performance almost as good as more sophisticated methods, to the extent that the difference in performance may be swamped by other sources of uncertainty that generally are not considered in the classical supervised classification paradigm.



Information visualization

- Data exploration is one of the basic building blocks, or constituting stages, of most **knowledge discovery** methodologies, either on its own or as part of a more generic phase of **data understanding**.
- We aim discover the main characteristics of usually complex **multivariate data sets**, helping bring important aspects of the data into focus (think **attention**) for study in subsequent phases of the analysis.
- The task of **data visualization** is central to **both** data exploration and **model interpretation**.
- The problem of **knowledge generation through data visualization**, is not circumscribed to data science *per se*. It can be addressed from the viewpoints of both **artificial pattern recognition (APR)** **and** **natural pattern recognition (NPR)**

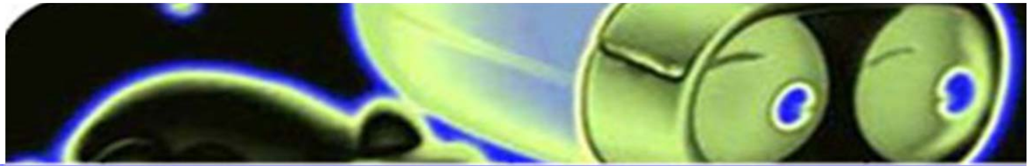


Information visualization

- **Artificial pattern recognition (APR):** Through the definition of methods/techniques/algorithms for visualization.

- **Natural pattern recognition (NPR):** Through the understanding of visualization as the **cognitive processing of visual stimuli conducted by the human brain** (which can be investigated using machine learning techniques in the context of **Computational Neuroscience**) Humans are equipped with visual NPR as a tool to understand the patterns of their natural environment and operate upon it.

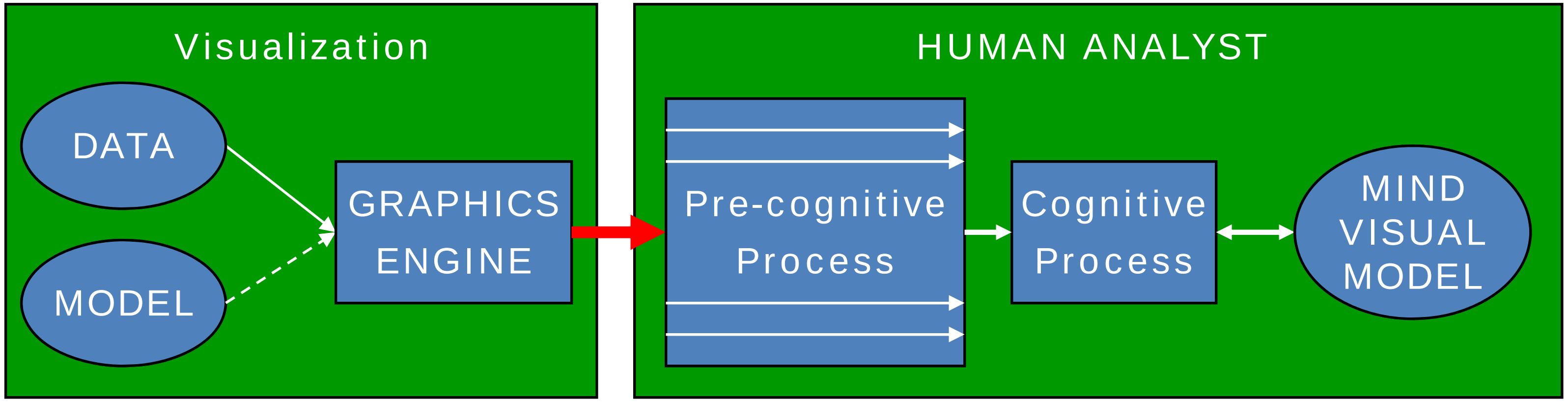
- Adequate data visualizations can help us to gain insights into a problem without the frame of a conjecture: **Out of a deductive model of research to reap the benefits of a more inductive one**.



Information visualization

- Importantly, **APR and NPR can enhance each other** in order to make data exploration a more fruitful process.
- **IV** defines processes that unify DR algorithms and visual user interfaces to allow us to explore data and interpret models using **graphical metaphors**, in a way that helps to **circumvent some of the inherent limitations of human vision**.
- To explain the success of visualization as a data exploration tool, we should not brush aside the **aesthetic aspects** involved: ***"Pretty pictures"*** are useful because they appeal to us at a very basic, non-discursive level. **Usefulness and beauty** can become indistinguishable concepts in this context: It has been argued that for a data visualization to qualify as beautiful, it must **comply with requirements of novelty, informativeness, and efficiency**.

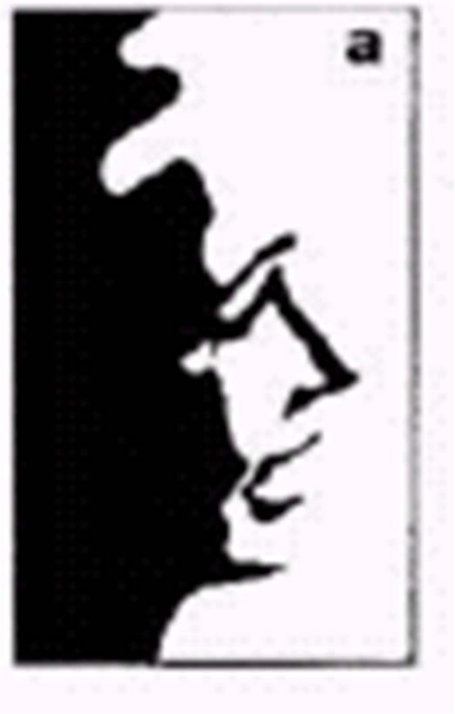
Information visualization



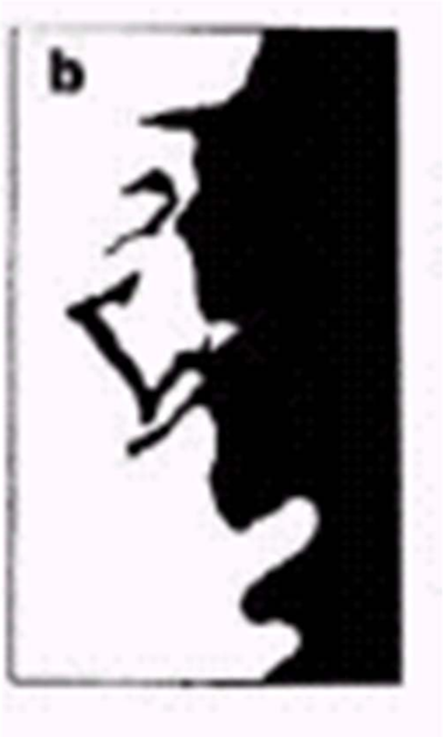
- **Pre-cognitive processing:** jigsaw pieces, or the interaction of visual elements.

- **Visualization vs. mental model:** visual patterns, illusions and *Gestalt* laws.

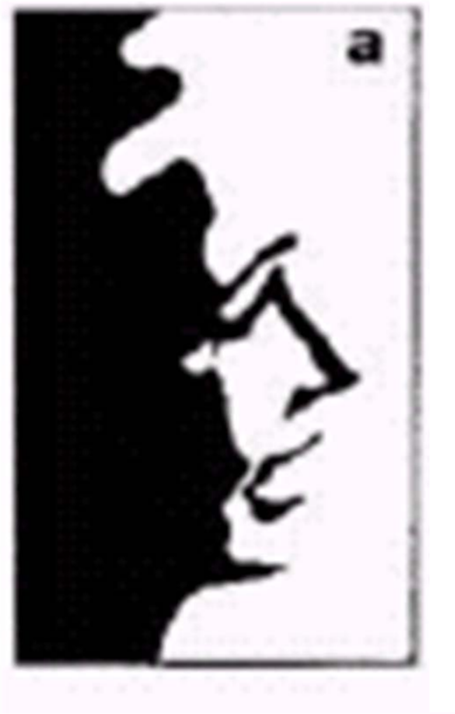
Information visualization

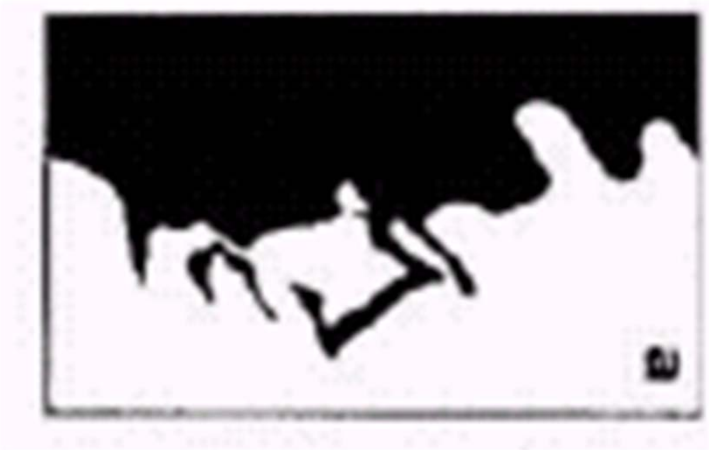


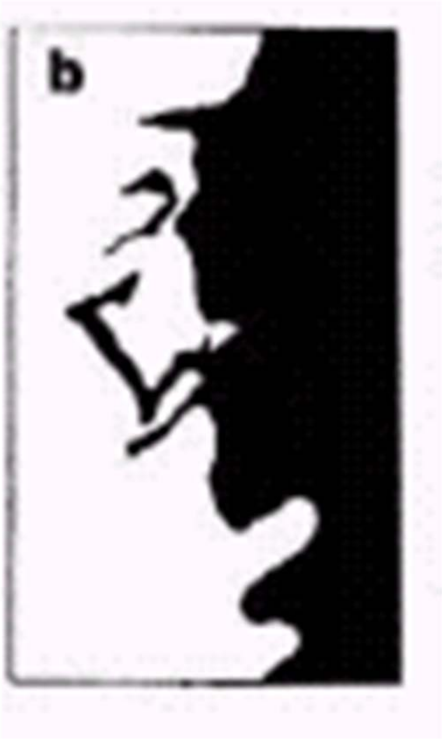
Information visualization

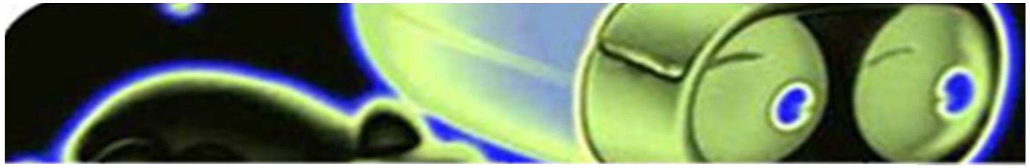


Information visualization



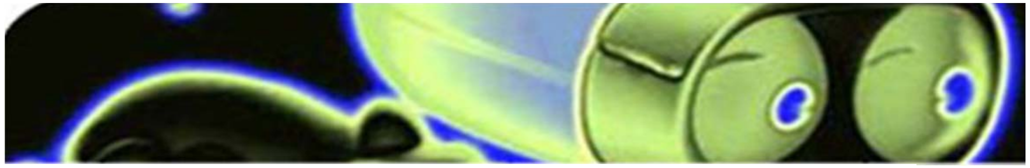






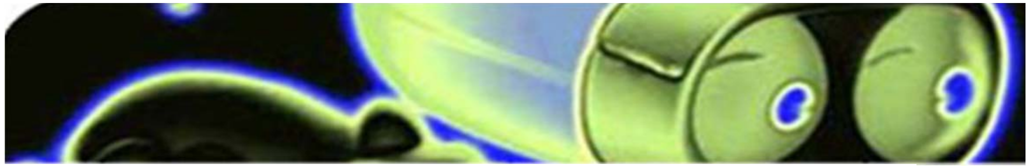
Wrap up

- The journey from data to knowledge is treacherous and **knowledge discovery processes require model interpretation**.
- **ML & related models' users** should aspire not only to integrate interpretation strategies … but also to ***negotiate*** with end-users the terms of that interpretability.
- **NLDR models** (and nonlinear models in general) are not likely to become mainstream in many app fields unless they are designed with interpretability in mind.
- **Graphical metaphors** (visualization, graphical models) are naturally suited as interpretability tools.
- When handling interpretability, we can only ignore **NPR** at our own peril.



Some reading

- Vellido, A, Martín-Guerrero, JD, & Lisboa, PJ (2012). **Making machine learning models interpretable**. In *ESANN 2012*, pp. 163-172
- Lisboa, PJ (2013) **Interpretability in Machine Learning – Principles and Practice**. *In International Workshop on Fuzzy Logic and Applications*, pp.15-21
- Turner, R (2015) **A Model Explanation System**. In *Black Box Learning and Inference NIPS Workshop*
- Schulz, A, Gisbrecht, A, & Hammer, B (2015) **Using discriminative dimensionality reduction to visualize classifiers**. *Neural Processing Letters*, 42(1), 27-54.
- Condry, N (2016). **Meaningful Models: Utilizing Conceptual Structure to Improve Machine Learning Interpretability**. *arXiv* preprint arXiv:1607.00279.
- Kim, B, Malioutov, DM, and Varshney, KR (2016) *Proceedings of the 2016 ICML Workshop on Human Interpretability in Machine Learning* (**WHI 2016**) ArXiv e-prints
- Goodman, B & Flaxman, S (2016) **EU regulations on algorithmic decision-making and a "right to explanation"**. *arXiv* preprint arXiv:1606.08813.



Some reading

- Goodman, B & Flaxman, S (2016) **EU regulations on algorithmic decision-making and a "right to explanation"**. *arXiv* preprint arXiv:1606.08813.

Abstract

We summarize the potential impact that the **European Union's new General Data Protection Regulation** will have on the routine use of machine learning algorithms. Slated to **take effect as law across the EU in 2018**, it will restrict automated individual decision-making (that is, algorithms that make decisions based on user-level predictors) which "significantly affect" users. The law will also effectively create a "**right to explanation**", whereby a user can ask for an explanation of an algorithmic decision that was made about them. We argue that while this law will pose large challenges for industry, it highlights opportunities for computer scientists to take the lead in designing algorithms and evaluation frameworks which avoid discrimination and enable explanation.

Mind the Interpreters