



Data Smashing

Zero-Knowledge Feature-free Anomaly Detection

Ishanu Chatopadhyay

Computation Institute
University of Chicago

a·nom·a·ly

noun

something that deviates from what is standard, normal, or expected.

“there are a number of anomalies in the present system”

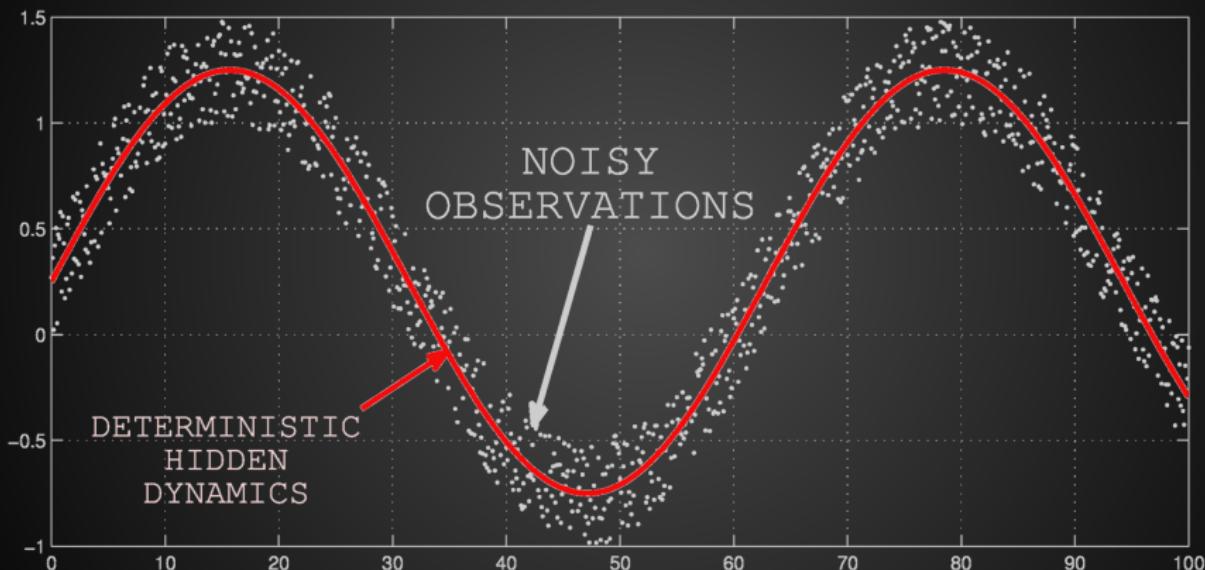
synonyms :

oddity, peculiarity, abnormality, irregularity, inconsistency, incongruity, aberration, quirk, rarity



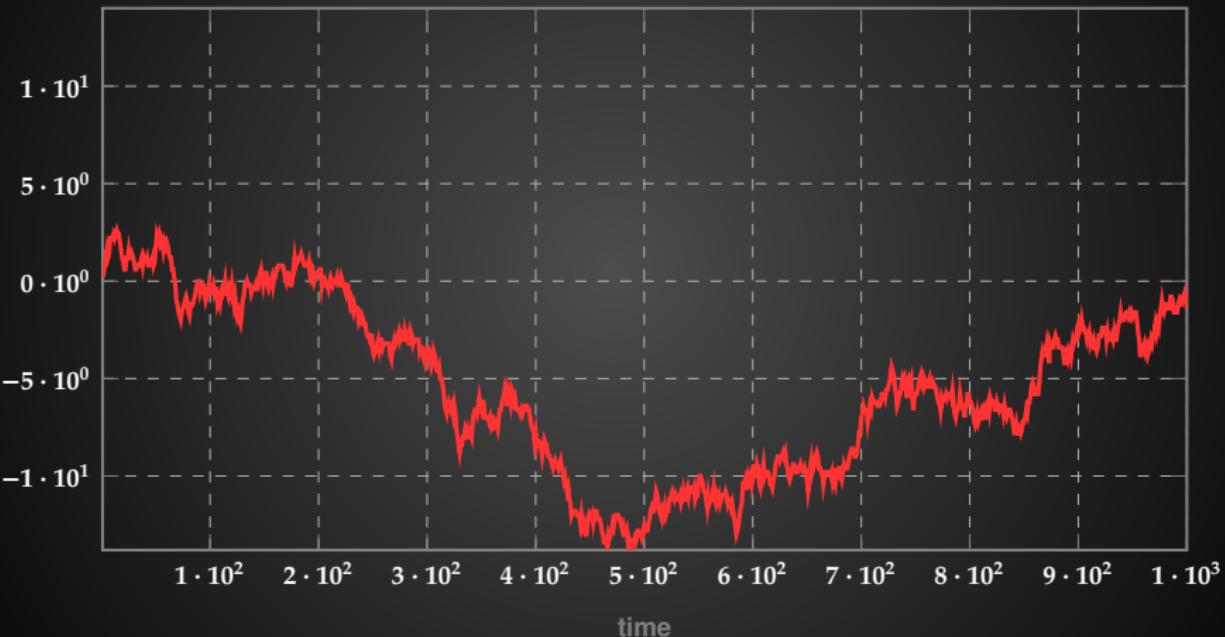
Deterministic Behavior

Specifying “Normal” Might Be Easier



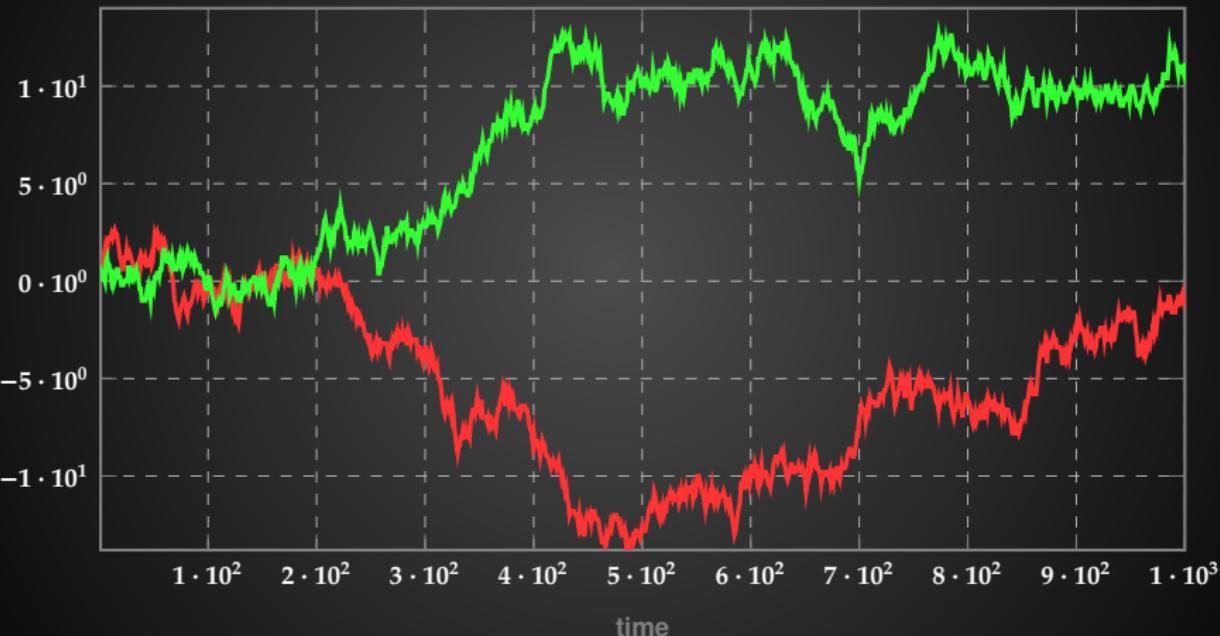


Stochastic Process \neq Deterministic + Noise



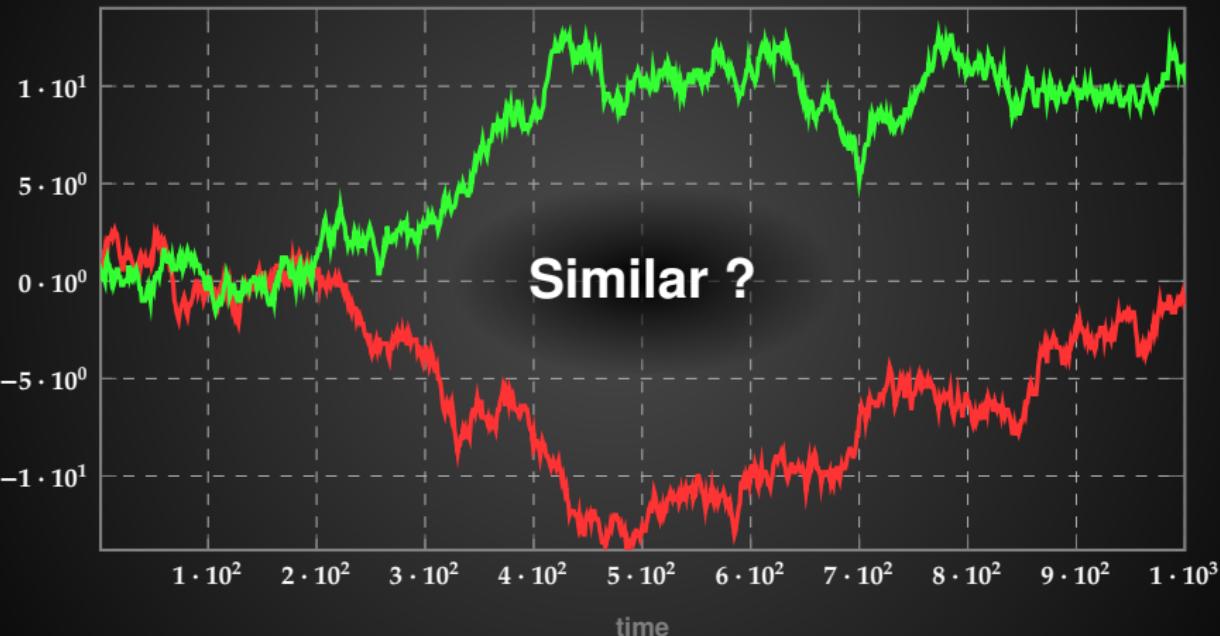


Stochastic Process \neq Deterministic + Noise





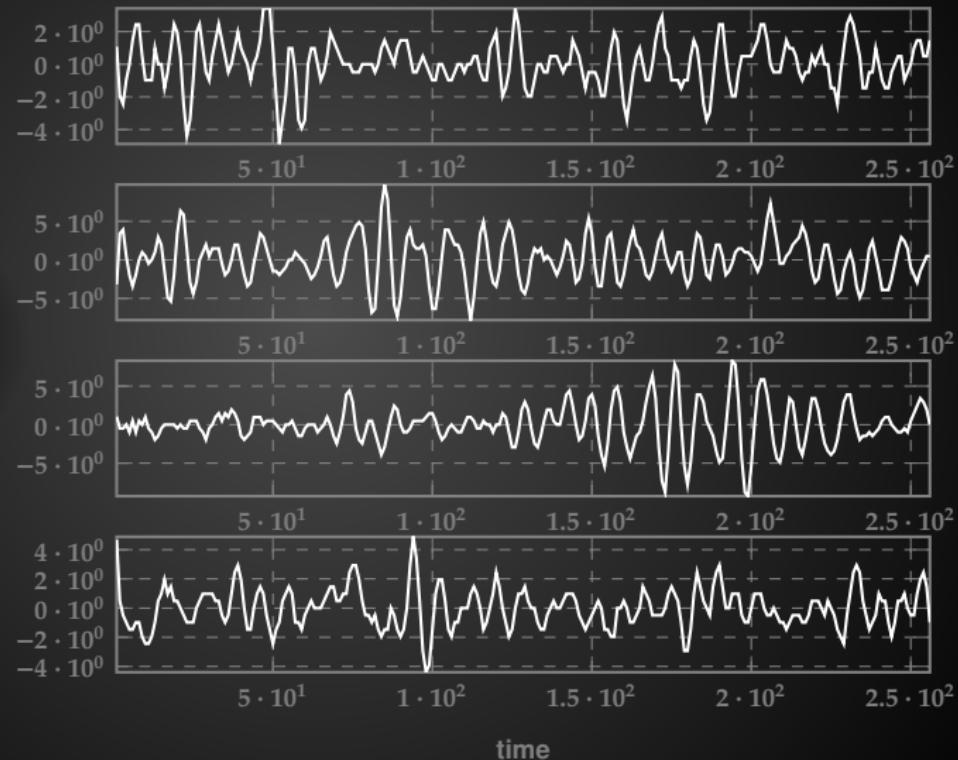
Stochastic Process \neq Deterministic + Noise





Brainwaves: Identical Stimuli

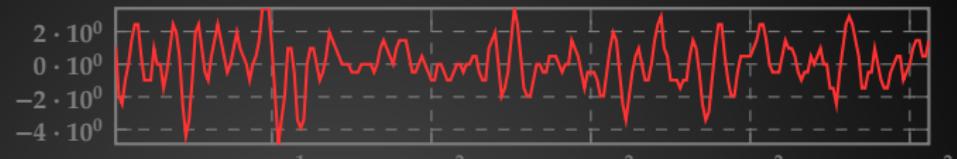
Are they
from the
same individual?



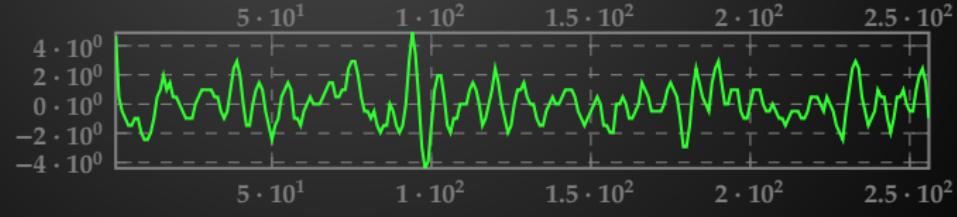
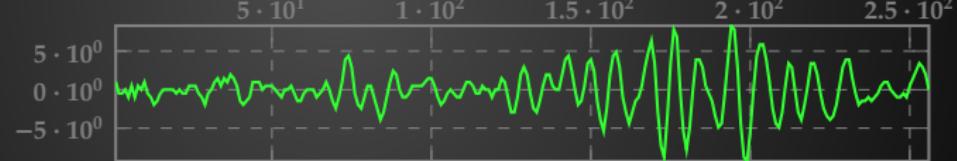
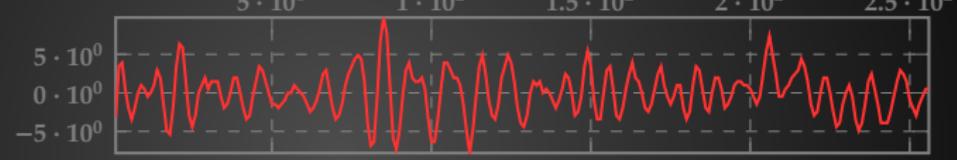


Brainwaves: Identical Stimuli

Subject A



Subject B

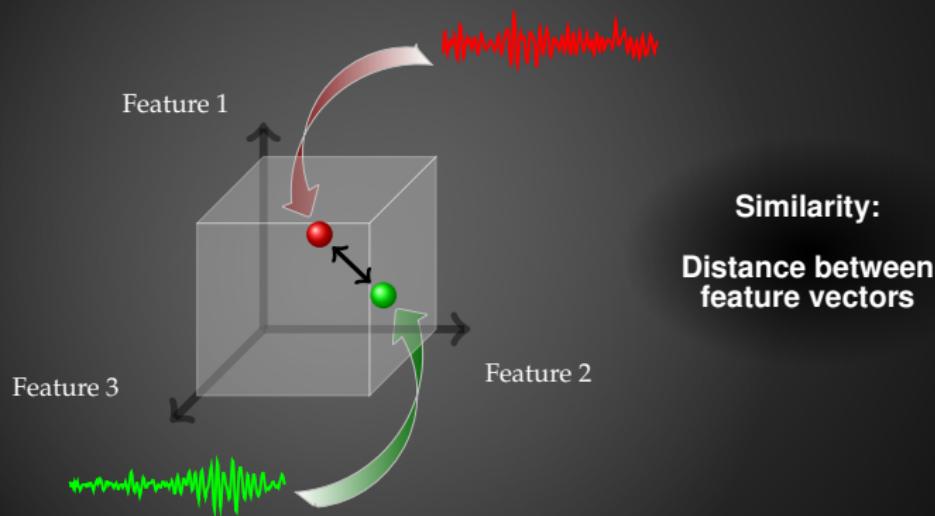


time



State of Art

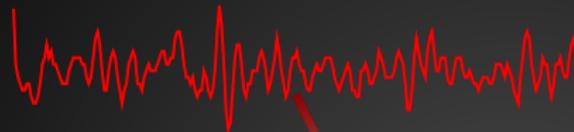
Requires Features





Data Smashing

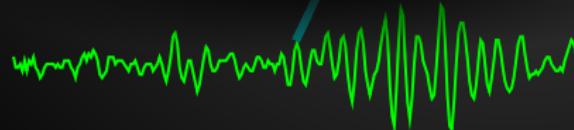
Signal 1



Quantize: bcbbbbacabcbbbbabbba...

Invert: cacacccaaabcacbcacabacaacacac...

Signal 2



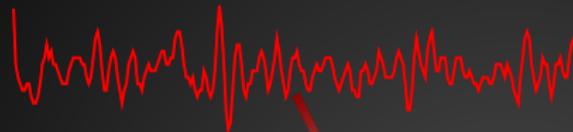
Quantize: abbababbbbabbbabcbbbbab...

cababcbababcbbacbcbcbababacacabaccbcbccab...

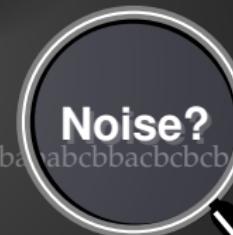


Data Smashing

Signal 1



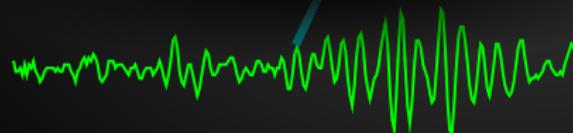
Quantize: bcbbbbacabcbbbbabbba...



Invert: cacacccaaabcacbcacabacaacacac...

Quantize: abbababbbbabbbabcbbbbab...

Signal 2





Data Smashing

I. Chattopadhyay and H. Lipson , “**Data Smashing: Uncovering Lurking Order In Data**”, Royal Society Interface, 2014 vol. 11 no. 101 20140826



Data Smashing

EIT ICT Labs

Home Innovation Areas » Education » Research Business » About us » News & Events »

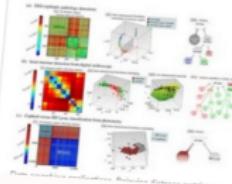
News & Events

- * News
- * Events
- * [Blog](#)
- * Press & Media Newsroom
- * In the Media
- * Success Stories
- * Newsletters
- * Publications
- * Photos
- * Videos

Blog

Data Smashing

Author - Roberto Saracco
Monday, 12 January 2015



The availability of huge data streams is pushing researchers to find innovative ways to analyse their content, compare it and derive meaning. This is not easy at all.

So far data mining approaches are based on a hunch from a human being that is then coded into an algorithm to see if it is true. Data are analysed to prove or disprove certain hypotheses. As an example you may want to see if the pattern of buying a product is related to some ads broadcasted on different media and in different places. This works pretty well and many market



Data Smashing

The image displays two web-based interfaces related to data analysis and mining.

EIT ICT Labs (Left Screenshot):

- Header:** EIT ICT Labs
- Navigation:** Home, Innovation Areas, Education, Research, Business, About us, News & events (highlighted).
- Content:** A large blue banner area, followed by a sidebar titled "News & Events" containing links like News, Events, Blog, Press & Media, In the Media, Success Stories, Newsletters, Publications, Photos, and Videos.
- Blog Post Preview:** "Data Smashing" by Roberto Saracco, Monday, 12 January 2015. It includes a thumbnail image showing various data visualization plots.

isgtw international science grid this week (Right Screenshot):

- Header:** isgtw international science grid this week
- Navigation:** ABOUT, CALENDAR, ARCHIVE, LEARN, COMMUNITY.
- Content:** A main headline "Why mine data when you can smash it?" with a subtext "SPOTLIGHT | NOVEMBER 28, 2014". Below this, a text block discusses the challenges of analyzing large data streams and how data mining approaches are based on a hunch from a human being.
- Visuals:** Several small data visualization images are shown at the bottom left.



Data Smashing

The screenshot shows the EIT ICT Labs website. At the top, there's a navigation bar with links to Home, Innovation Areas, Education, Research, Business, About us, and News & Events. A sidebar on the left is titled 'News & Events' and lists categories like News, Events, Blog, Press & Media, In the Media, Success Stories, Newsletters, Publications, Photos, and Videos. The main content area features a blog post titled 'Data Smashing' by Roberto Saracco, published on Monday, 12 January 2015. The post includes several small images illustrating data analysis and mining. To the right of the blog is a sidebar with the EIT logo and a link to the 'isgtw international s grid this week' website.

NEWS, TECHNOLOGY

'Data Smashing' algorithm might help declutter Big Data without Human Intervention

There's an immense well of information humanity is currently sitting on growing exponentially. To make sense of all the noise, whether we're talking about speech recognition, cosmic body identification or search engine results, there are algorithms that use less processing power by hitting the bull's eye or as close as possible. In the future, such algorithms will be comprised of more complex technology that gets smarter and smarter after each information parse; these algorithms and a most exciting paper detailing such a technique was recently published in the journal *ISGTW*.

Smashing data – the bits and pieces that follow a pattern – is important



Data Smashing

The screenshot shows the EIT ICT Labs website. At the top, there's a navigation bar with links for Home, Innovation Areas, Education, Research, Business, About us, and News & Events. A sidebar on the left is titled 'News & Events' and lists categories like News, Events, Blog, Press & Media, Headlines, In the Media, Success Stories, Newsletters, Publications, Photos, and Videos. The main content area features a blog post titled 'Data Smashing' by Roberto Saracco, published on Monday, 12 January 2015. The post includes a sub-headline 'Why mine data when you can smash it?' and a large circular graphic filled with binary code. Below the blog is a section titled 'A Saucerful of Data' with the text: 'This blog is about data, in a computational sense. Small and Big. Simple and Complex. And how to deal with it.' At the bottom, there are links for Friday, January 16, 2015, The Power of Anti-Data, Introduction, Pages (Home, Archives), and Web Archive.

EIT ICT Labs

Home Innovation Areas Education Research Business About us News & Events

News & Events

- News
- Events
- Blog
- Press & Media
- Headlines
- In the Media
- Success Stories
- Newsletters
- Publications
- Photos
- Videos

Blog

Data Smashing

Author - Roberto Saracco
Monday, 12 January 2015

RECENT POSTS

Why mine data when you can smash it?

SPOTLIGHT | NOVEMBER 28, 2014

A Saucerful of Data

This blog is about data, in a computational sense. Small and Big. Simple and Complex. And how to deal with it.

Friday, January 16, 2015

The Power of Anti-Data

Introduction

Pages

- Home
- Archives (9 news and tutorials)

Web Archive

NEWS, TECHNOLOGY

'Data Smashing' algorithm might help declutter Big Data without Human Intervention

There's an immense well of information humanity is currently sitting on growing exponentially. To make sense of all the noise, whether we're talking about speech recognition, cosmic body identification or search engine results, there are algorithms that use less processing power by hitting the bull's eye or as close as possible. In the future, such algorithms will be comprised of more technology that gets smarter and smarter after each information parse; these algorithms and a most exciting paper detailing such a technique was recently

Smashing data – the bits and pieces that follow a



Data Smashing

EIT ICT Labs

Home Innovation Areas » Education » Research Business News & Events About EIT

CORNELL CHRONICLE

October 11, 2014

Science, Tech & Medicine | Arts & Humanities | Business, Law & Society | Campus Life

Oct. 2, 2014

'Data smashing' could unshackle automated discovery

A little-known secret in data mining is that simply feeding raw data into a data analysis algorithm is unlikely to produce meaningful results, say the authors of a new Cornell study.

A Saucerful of Data

This blog is about data, in a computational sense. Small and Big. Simple and Complex. And how to deal with it.

Friday, January 16, 2015

The Power of Anti-Data

Introduction

Pages

- Home
- #Biggers (9 news and tutorials)

Blog Archive

NEWS, TECHNOLOGY

'Data Smashing' algorithm might help declutter Big Data without Human Intervention

There's an immense well of information humanity is currently sitting on growing exponentially. To make sense of all the noise, whether we're talking about speech recognition, cosmic body identification or search engine results, there are algorithms that use less processing power by hitting the bull's eye or as close as possible. In the future, such algorithms will be comprised of more technology that gets smarter and smarter after each information parse; these algorithms and a most exciting paper detailing such a technique was recently

Smashing data – the bits and pieces that follow a

important



Data Smashing

EIT ICT Labs

CORNELL CHRONICLE
October 11, 2014

Science, Tech & Medicine | Arts & Humanities | Business, Law & Society | Camp...

Oct. 2, 2014

'Data smashing' could unshackle automated discovery

A little-known secret in data mining is that simply feeding raw data into a data analysis algorithm is unlikely to produce meaningful results, say the authors of a new Cornell study.

This blog is about data, in a computational sense.

A

Friday, January 16, 2015

The Power of Anti-Data

Introduction

ACM TechNews

| About | Current Issue | TechNews Archive | TechNews App

Welcome to the October 10, 2014 edition of ACM TechNews, p... information for IT professionals three times a week.

Please note: In observance of the Columbus Day holiday, TechNews will not be published on Monday, Oct. 13. Publication will resume Wednesday, Oct. 15.

Updated versions of the ACM TechNews mobile apps are available for phones and tablets (click [here](#)) and for iPhones (click [here](#)) and iPads (click [here](#)).

HEADLINES AT A GLANCE

- Cyberattacks Trigger Talk of 'Hacking Back'
- High-Tech Pay Gap: Minorities Earn Less in Skilled Jobs
- Gartner Lays Out Its Top 10 Tech Trends for 2015
- Computer Engineering Degrees Pay Off Big Time
- Tim Berners-Lee, Web Creator, Defends Net Neutrality
- LAUSD Announces Sweeping Expansion of Computer Science Classes
- Paper: Great Promise for Online Voting If Security, Verification Improved
- Robotic Solutions Inspired by Plants
- Young Israeli Cyberwarriors Learn to Duel in the Dark
- Smartphone Understands Gestures
- Microsoft's RoomAlive Turns Your Room Into a Holodeck
- New Paper Examines the Significant Social Strategies in Human-Computer Interaction
- **'Data Smashing' Could Unshackle Automated Discovery**



Data Smashing

EIT ICT Labs

Home Innovation Areas » Education » Research Business

CORNELL CHRONICLE

October Science Oct. 2, 2014

'Data A little-known algorithm'

ScienceDaily®
Your source for the latest research news

Mobile: iPhone Android Web

Health PHYSICAL TECH ENVIRONMENT

Latest Headlines Health & Medicine Mind & Brain

Communications News

Featured Research

Computer Science: 'Data Smashing' Could Unshackle Automated Discovery

Oct. 10, 2014 — Computing researchers have come up with a new principle they call 'data smashing' for estimating the similarities between streams of arbitrary data without human intervention, and without access to ... [full story](#)

> Information Technology; Hacking; Encryption; Computers and Internet

from universities, journals, and other organizations

> Data Smashing To Unshackle Automated Discovery

> Gene That Drives Aggressive Brain Cancer Found

> Smartphone Understands Hand Gestures

> Economy and Internet File-Sharing Patterns

> Dynamic Encryption Keeps Secrets

> Print Thin, Touch-Sensitive Displays On Stuff

> Talking to Your Car Is Often Distracting

> Games: Reducing Scientific Fraud, Cherry-Picking

> New Web Privacy System: Safer Internet Surfing?

> Pressing the Accelerator On Quantum Robotics

< newer top stories | older top stories >

ACM TechNews

Current Issue | TechNews Archive | TechNews App

October 10, 2014 edition of ACM TechNews, p T professionals three times a week.

bservance of the Columbus Day holiday, Tech day, Oct. 13. Publication will resume Wednes

If the ACM TechNews mobile apps are available (click [here](#)) and for iPhones (click [here](#)) and

GLANCE

trigger Talk of 'Hacking Back'

Gap: Minorities Earn Less in Skilled Jobs

it Its Top 10 Tech Trends for 2015

eering Degrees Pay Off Big Time

Web Creator, Defends Net Neutrality

s Sweeping Expansion of Computer Science

use For Online Voting If Security, Verification

Inspired by Plants

Warriors Learn to Duel in the Dark

stands Gestures

o's RoomAlive Turns Your Room Into a Holodeck

• New Paper Examines the Significant Social Strategies in Human

• **'Data Smashing' Could Unshackle Automated Discovery**



Data Smashing

EIT ICT Labs

Home Innovation Areas Education Research Business

CORNELL CHRONICLE

October Science Oct. 2, 2014

ScienceDaily® Your source for the latest research news

Follow: Facebook Twitter Google+ SOCIETY EDUCATION QUERIES

KurzweilAI | Accelerating Intelligence. News

'Data smashing' could automate discovery, untouched by human hands

October 28, 2014

From recognizing speech to identifying unusual stars, new discoveries often begin with comparison of data streams to find connections and spot outliers. But simply feeding raw data into a data-analysis algorithm is unlikely to produce meaningful results, say the authors of a new Cornell study.

That's because most data comparison algorithms today have one major weakness: somewhere, they rely on a human expert to specify what aspects of the data are relevant for comparison, and what aspects aren't.

But these experts can't keep up with the growing amounts and complexities of big data.

ACM TechNews

Current Issue | TechNews Archive | TechNews App

October 10, 2014 edition of ACM TechNews, p T professionals three times a week.

bservance of the Columbus Day holiday, Tech day, Oct. 13. Publication will resume Wednes

If the ACM TechNews mobile apps are available (click [here](#)) and for iPhones (click [here](#)) and

GLANCE

trigger Talk of 'Hacking Back'
Gap: Minorities Earn Less in Skilled Jobs
it Its Top 10 Tech Trends for 2015
eering Degrees Pay Off Big Time
Web Creator, Defends Net Neutrality
s Sweeping Expansion of Computer Science
use for Online Voting If Security, Verification
Inspired by Plants
Warriors Learn to Duel in the Dark
stands Gestures

RoomAlive Turns Your Room Into a Holodeck
Examines the Significant Social Strategies in Huma

'Data Smashing' Could Unshackle Automated Discovery



Data Smashing

The collage includes the following elements:

- EIT ICT Labs**: A screenshot of a website with a navigation bar including Home, Innovation, Research, Business, and People.
- CORNELL SCIENCE**: A thumbnail of a magazine cover from October 2014, featuring the headline "'Data smashing' could automate discovery".
- KurzweilAI | Accelerating Intelligence**: A screenshot of a news article from October 28, 2014, discussing the concept of "data smashing".
- ACM TechNews**: A screenshot of a news article from the ACM TechNews archive, dated October 28, 2014, about the Columbus Day holiday.
- Data Smashing**: A large central image showing a person using a hammer to smash a computer monitor or similar electronic device on a grassy surface.
- CDs and DVDs**: A small image in the bottom right corner showing a pile of optical discs.
- Related Headlines**: A list of headlines from various sources:
 - 'Hacking Back' Is Farm Less in Skilled Jobs
 - Tech Trends for 2015
 - Big Data Pay Off Big Time
 - Net Neutrality
 - Expansion of Computer Science
 - Online Voting If Security, Verification
 - Smart Plants
 - Learn to Duel in the Dark
 - Hands Gestures
 - RoomAlive Turns Your Room Into a Holodeck
 - Examines the Significant Social Strategies in Human
 - 'Data Smashing' Could Unshackle Automated Discovery



Anti-streams

Intuitive Description

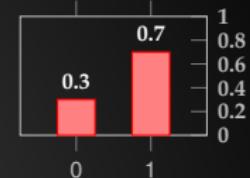
Stream $s \rightarrow$ Anti-stream s'



Anti-streams

Intuitive Description

Stream $s \rightarrow$ Anti-stream s'

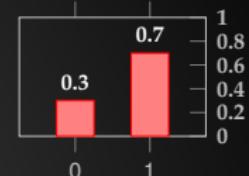




Anti-streams

Intuitive Description

Stream $s \rightarrow$ Anti-stream s'

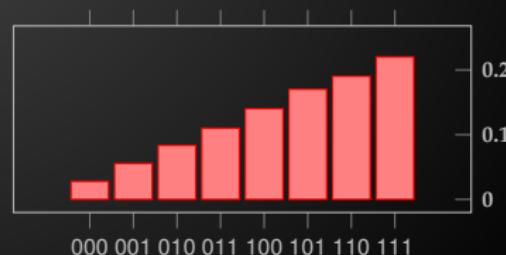
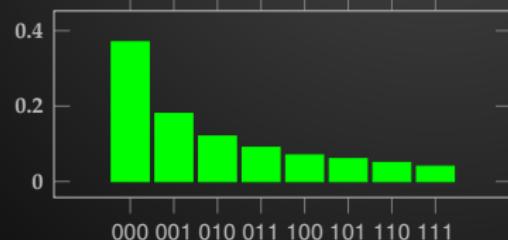
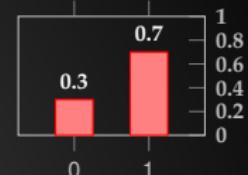




Anti-streams

Intuitive Description

Stream $s \rightarrow$ Anti-stream s'





Quantization

Mapping Continuous Data to Symbol Stream

Quantization Alphabet $\Sigma = \{0, 1, 2, 3\}$



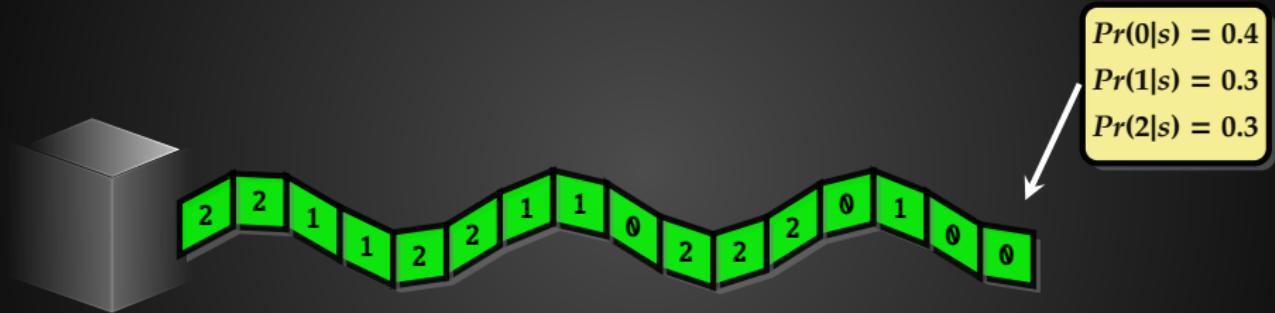
The Black Box Approach

Dynamical System As A Symbol Generator



The Black Box Approach

Dynamical System As A Symbol Generator



Ergodic Stationary Quantized Process \iff Probability Measure on Infinite Strings



System As A Symbol Generator

Probability Space $(\Sigma^\omega, \mathcal{B}, \mu)$

- Σ^ω : Set of strictly infinite strings on alphabet Σ
- \mathcal{B} : smallest σ -algebra generated by the sets $\{x\Sigma^\omega : x \in \Sigma^*\}$
- μ : Probability measure on infinite strings:

$$\mu(x\Sigma^\omega) \mapsto [0, 1]$$

$$\sum_{x \in \Sigma^*} \mu(x\Sigma^\omega) = 1$$



System As A Symbol Generator

Probability Space $(\Sigma^\omega, \mathcal{B}, \mu)$

- Σ^ω : Set of strictly infinite strings on alphabet Σ
- \mathcal{B} : smallest σ -algebra generated by the sets $\{x\Sigma^\omega : x \in \Sigma^*\}$
- μ : Probability measure on infinite strings:

$$\mu(x\Sigma^\omega) \mapsto [0, 1]$$

$$\sum_{x \in \Sigma^*} \mu(x\Sigma^\omega) = 1$$

Probability Measure on Infinite Strings \implies Equivalence Relation on Finite Strings

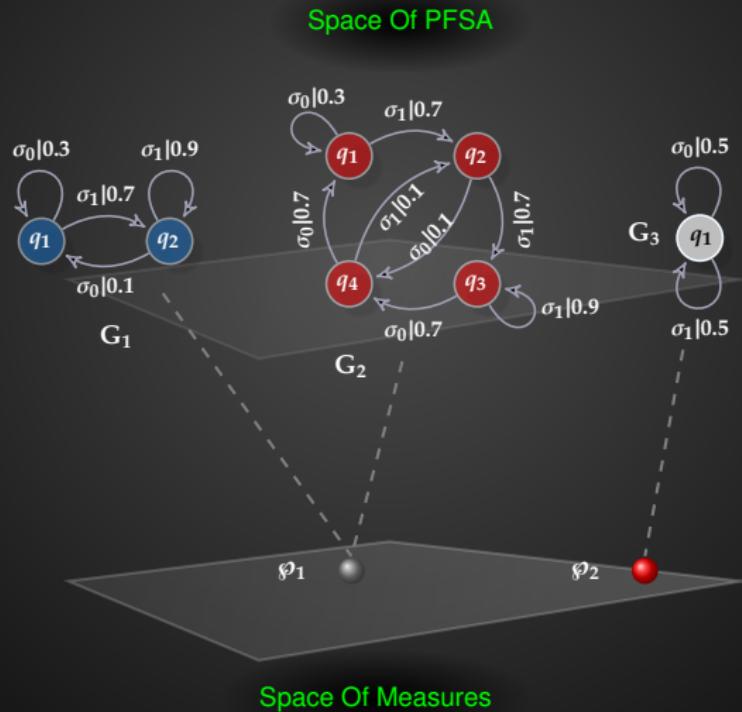
$$\forall x_1, x_2 \in \Sigma^*, x_1 \sim x_2 \\ \text{if } \forall x \in \Sigma^*, \mu(x_1 x \Sigma^\omega) = \mu(x_2 x \Sigma^\omega)$$

Equivalence Classes are causal states



Probabilistic Finite State Automata

Models For Quantized Stationary Ergodic Stochastic Processes





Adding Probability Measures

Consider two measures: $\begin{cases} \varphi_1 : \mathcal{B} \rightarrow [0, 1] \\ \varphi_2 : \mathcal{B} \rightarrow [0, 1] \end{cases}$

Define a binary operation:

$$\varphi_1 \oplus \varphi_2 \triangleq \varphi_3$$

where

$$\varphi_3(x\Sigma^\omega) = \varphi_1(x\Sigma^\omega)\varphi_2(x\Sigma^\omega) \times \text{Constant}$$

$$\sum_{x \in \Sigma^*} \varphi_3(x\Sigma^\omega) = 1$$



Adding Probability Measures

Consider two measures: $\begin{cases} \varphi_1 : \mathcal{B} \rightarrow [0, 1] \\ \varphi_2 : \mathcal{B} \rightarrow [0, 1] \end{cases}$

Define a binary operation:

$$\varphi_1 \oplus \varphi_2 \triangleq \varphi_3$$

where

$$\varphi_3(x\Sigma^\omega) = \varphi_1(x\Sigma^\omega)\varphi_2(x\Sigma^\omega) \times \text{Constant}$$

$$\sum_{x \in \Sigma^*} \varphi_3(x\Sigma^\omega) = 1$$

- Commutative
- Closed
- Unique Inverse
- Unique Identity

Abelian Group



Lifting Group Structure To PFSAs



Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group



Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

$$3 + 0 = 3$$



Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

$$3 + 0 = 3$$

We can “add and subtract” models:

$$G + H = J$$

$$G - H = K$$



Mathematical Structure Of Model Space

The Abelian Group

The Space of Models has the mathematical structure of an Abelian Group

$$1 + 2 = 3$$

$$2 - 2 = 0$$

$$3 + 0 = 3$$

We can “add and subtract” models:

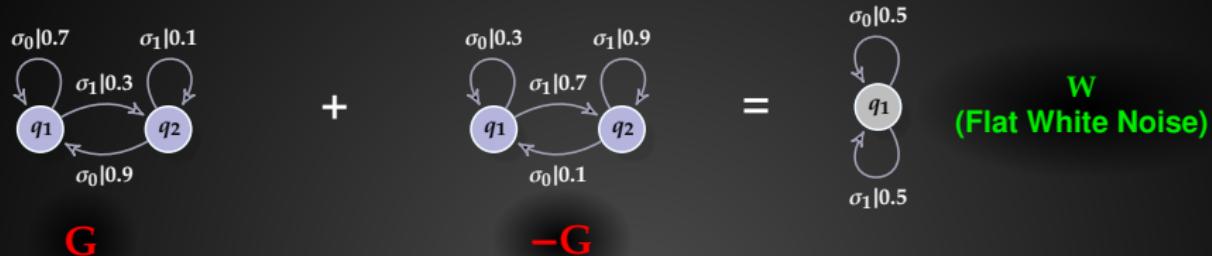
$$G + H = J$$

$$G - H = K$$

$$G - G = ?$$

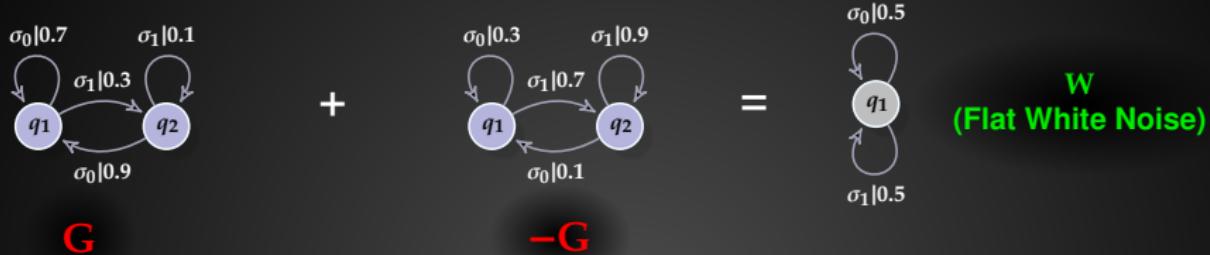


The Zero Machine

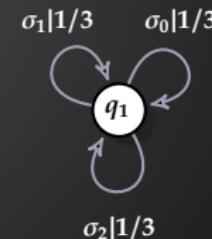




The Zero Machine



Zero PFSA
for binary alphabet



Zero PFSA
for trinary alphabet



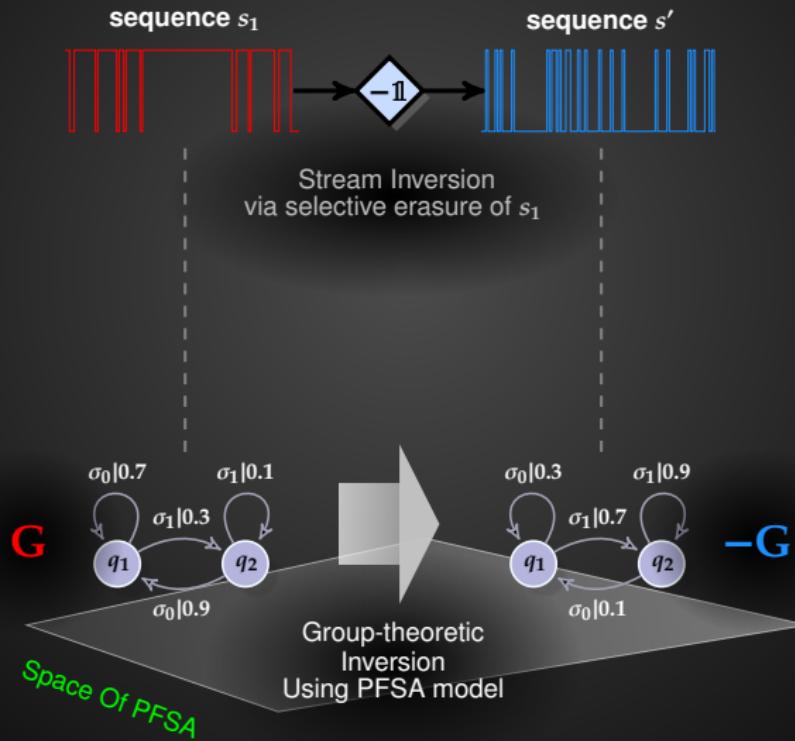
The Zero Machine

- Maximum entropy rate
- History is useless for prediction
- Encodes minimum information



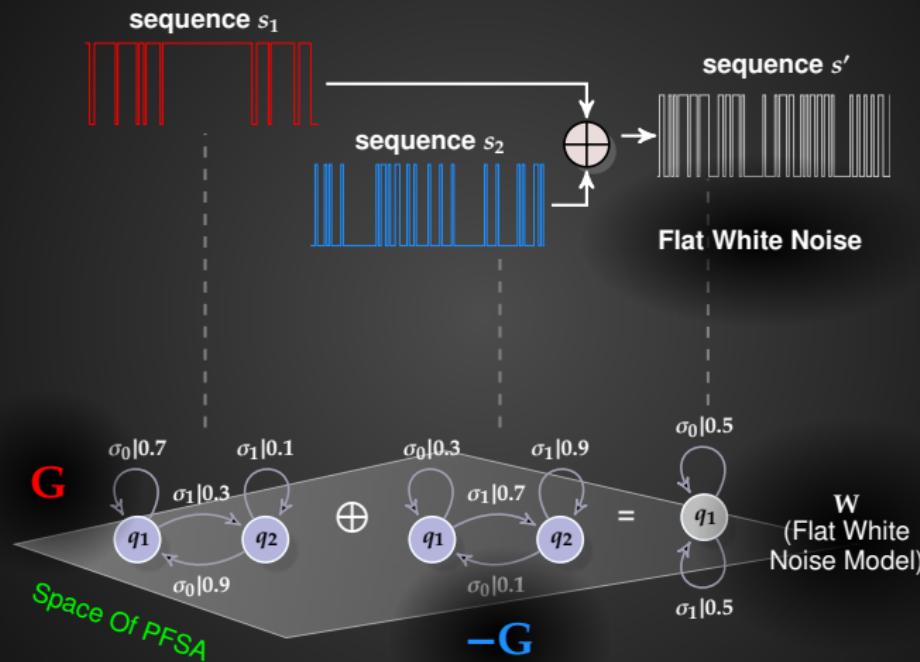
Stream Inversion

Direct Generation of Anti-streams





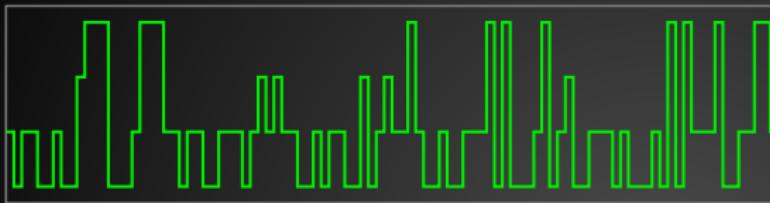
Annihilation Identity





Example Of Stream, Anti-stream, & FWN

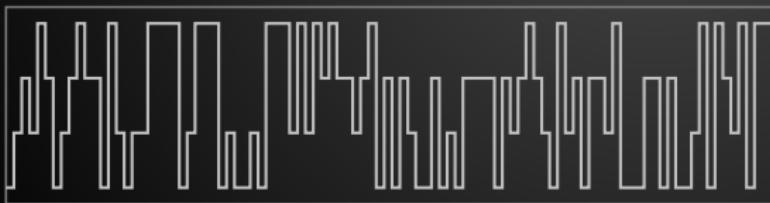
4 Letter Alphabet



Signal



Inverse Signal

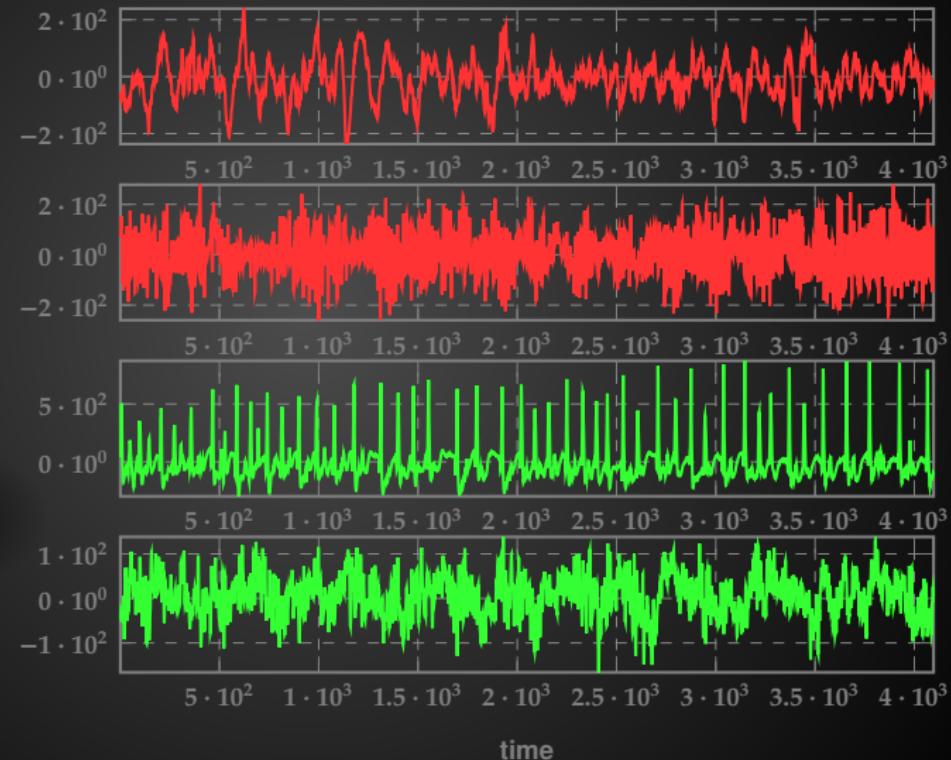


Flat White Noise



EEG - Epileptic Pathology

Eyes Open

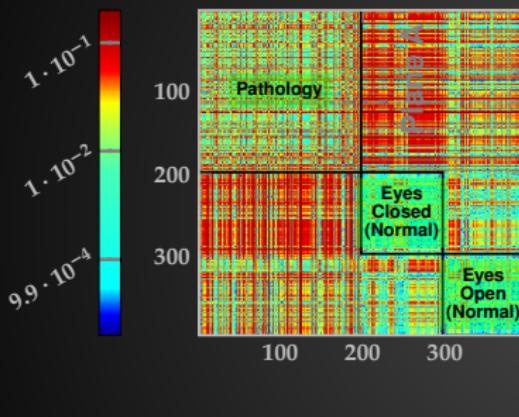


Epileptic Pathology

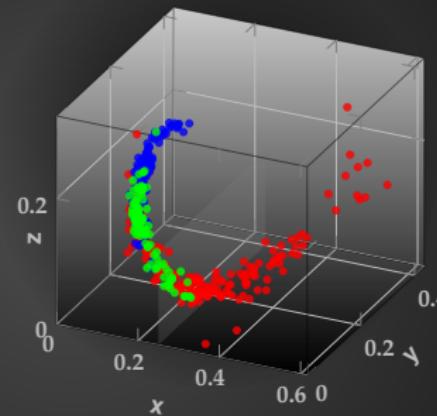


EEG - Epileptic Pathology

Pairwise Distance Matrix



3D Euclidean Embedding



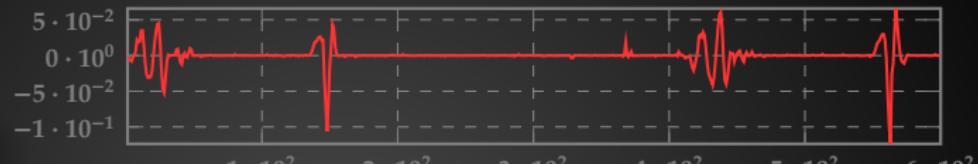
- Anomaly
- Normal (Eyes closed)
- Normal (Eyes open)



Cardiac Pathology

Disambiguate Normal Rhythm from Murmur

Normal Rhythm



Murmur

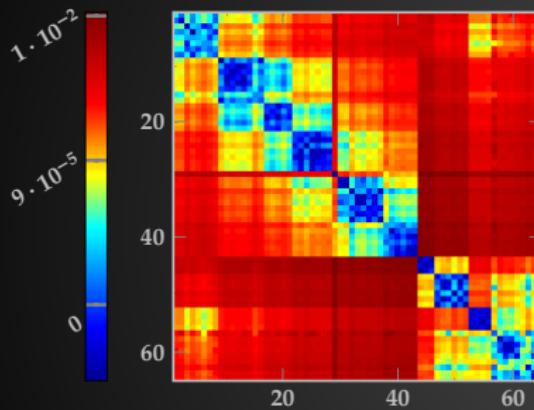


time

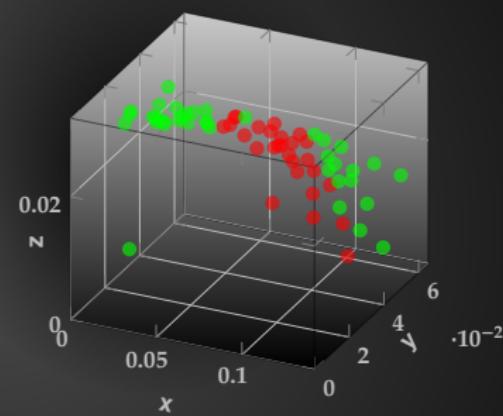


Cardiac Pathology

Distance Matrix



3D Euclidean Embeddin

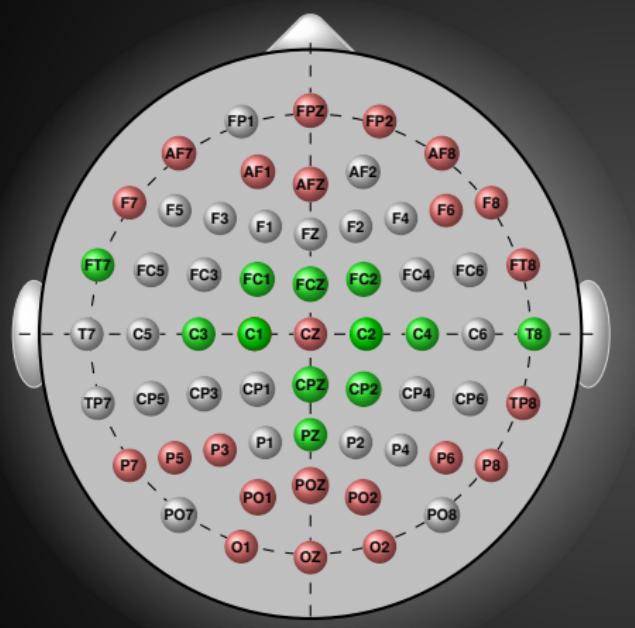


● Murmur
● Healthy

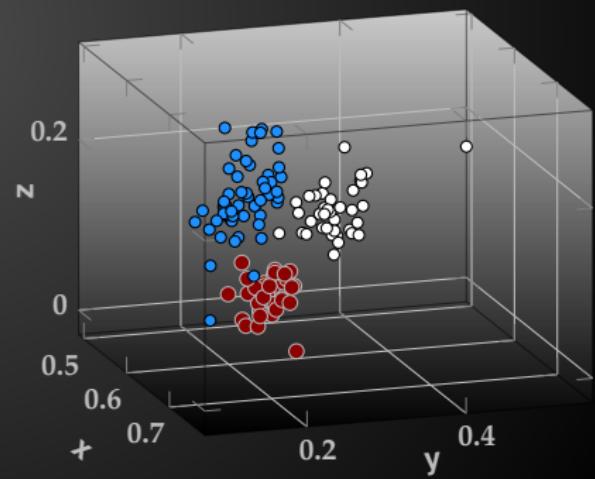


EEG Based Biometric Authentication

122 Subjects, 100% Accuracy



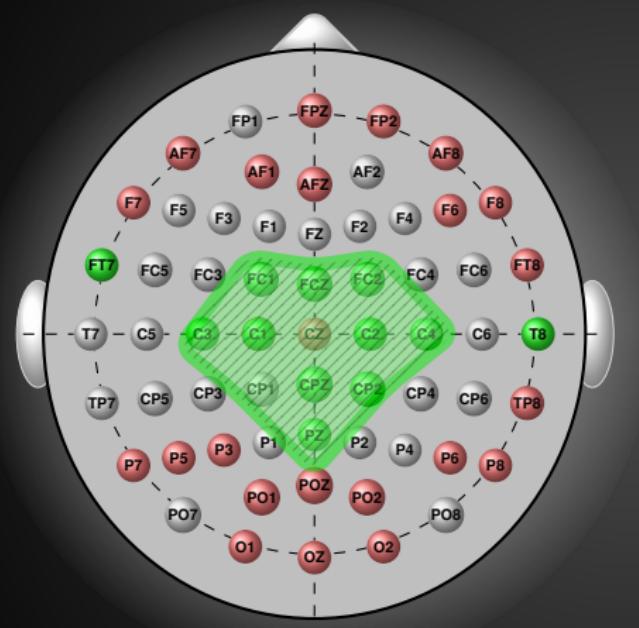
3D Euclidean Embedding
(3 random subjects)



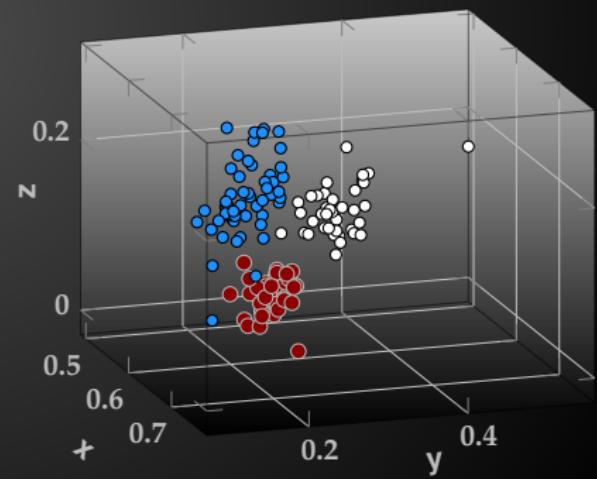


EEG Based Biometric Authentication

122 Subjects, 100% Accuracy



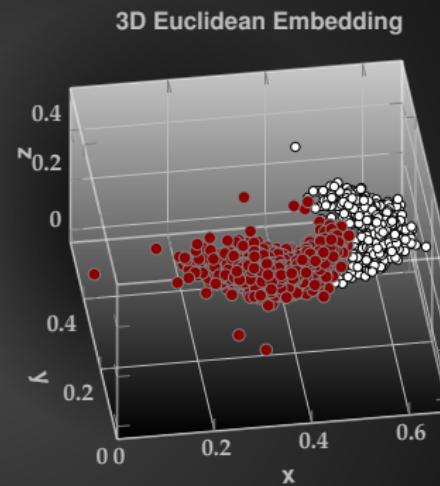
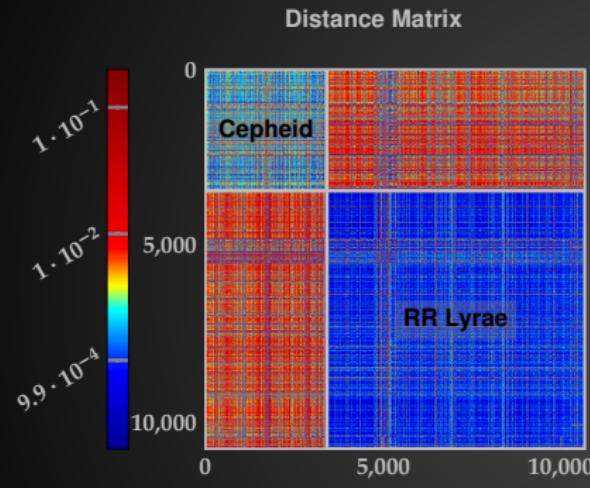
3D Euclidean Embedding
(3 random subjects)





Classification of Variable Stars

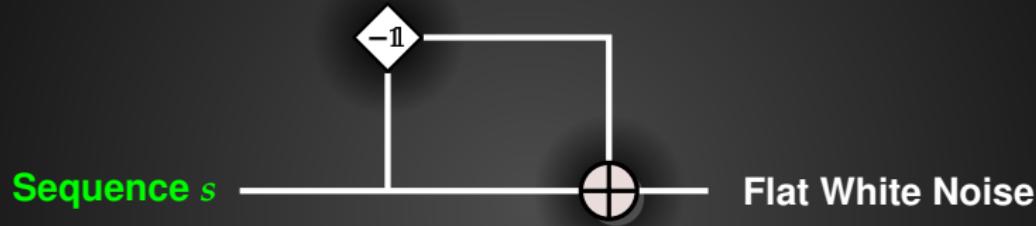
Optical Gravitational Lensing Experiment (OGLE) database



- RRL
- Cepheids

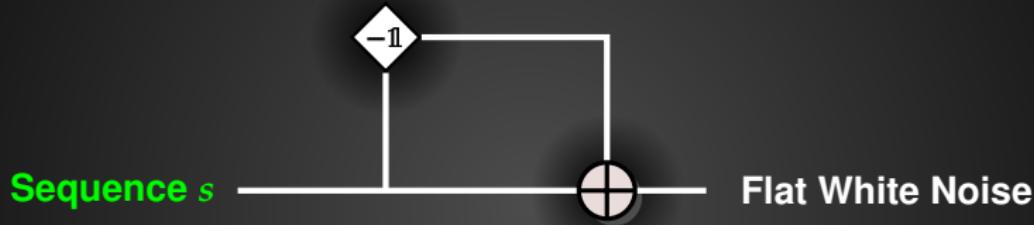


Self Annihilation Error





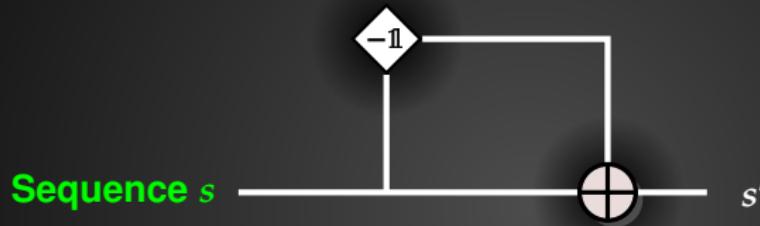
Self Annihilation Error



Only if $|s|$ is large enough !



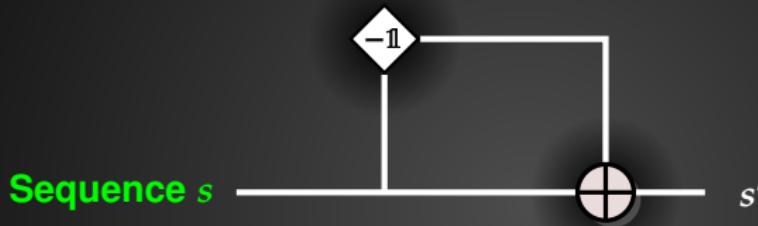
Self Annihilation Error



$$\epsilon_s = \Theta(s', W)$$



Self Annihilation Error

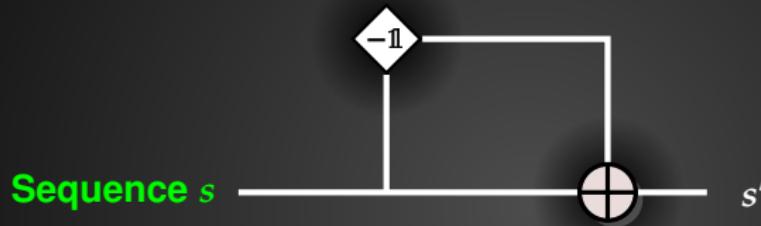


$$\epsilon_s = \Theta(s', W)$$

- $\epsilon_s \rightarrow 0$ exponentially fast with $|s|$
- Information content of stream



Self Annihilation Error



$$\epsilon_s = \Theta(s', W)$$

Auto-detect Data Sufficiency!

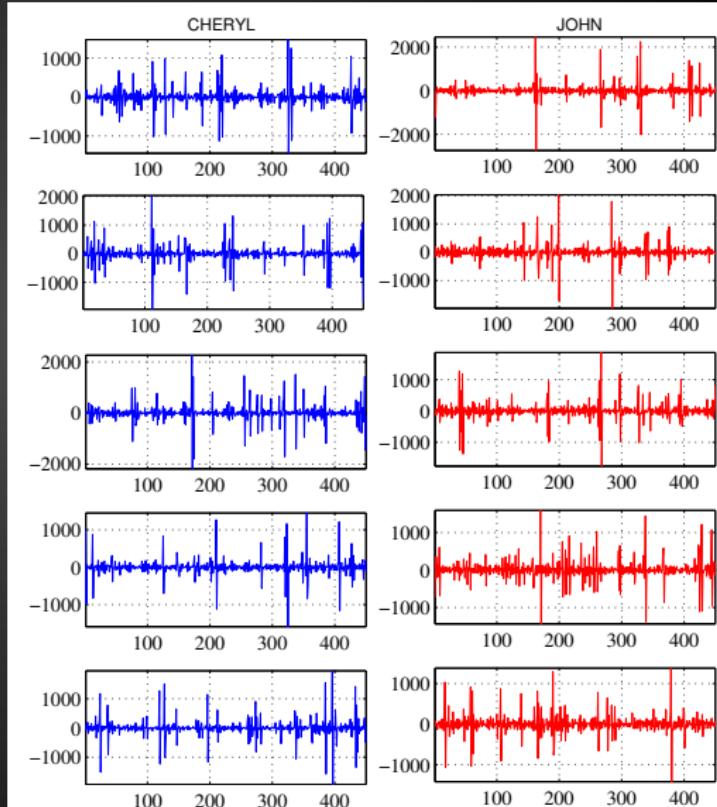
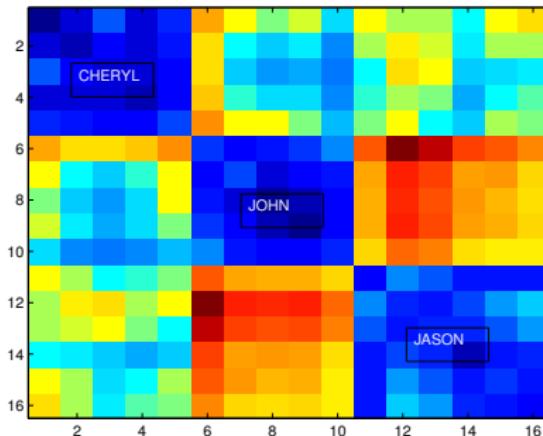


Cognitive Fingerprinting

User Authentication From Keypress Dynamics



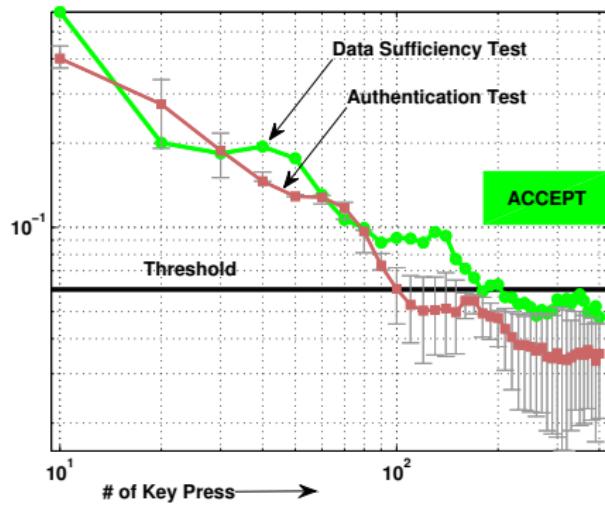
- Time-series of keypress delays
- Random text





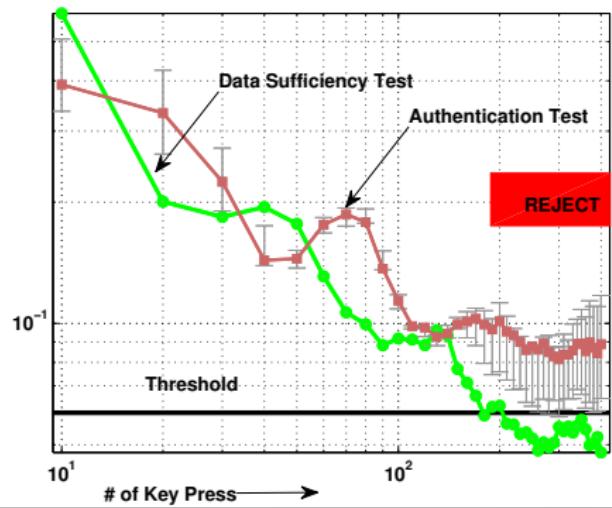
Cognitive Fingerprinting

User Authentication From Keypress Dynamics



User Authorized

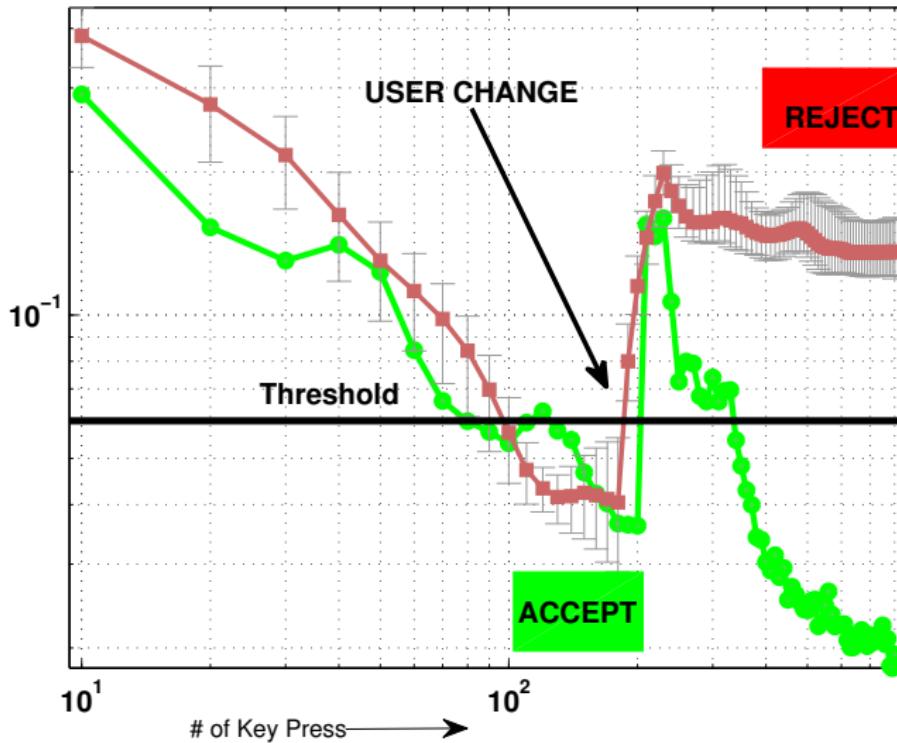
User Rejected





Cognitive Fingerprinting

User Authentication From Keypress Dynamics



Authorization
revoked
on unexpected user
change detection

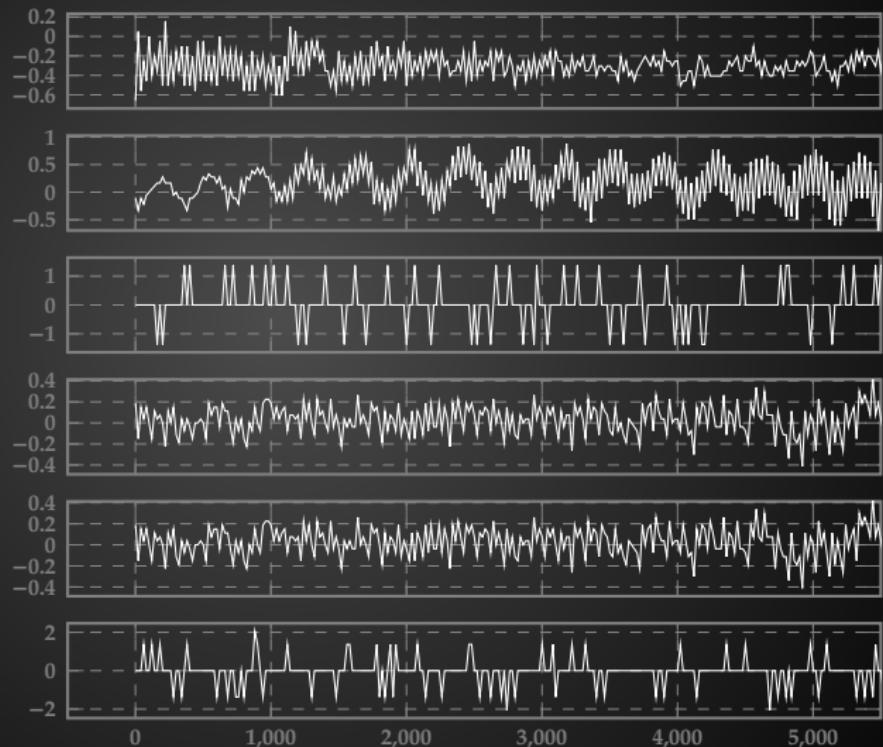


Physical Dexterity Quantification

Dynamic Regulation of Instabilities: Athletes vs Amateurs



Francisco Valero-Cuevas
Brain-body Dynamics Lab
Univ. of Southern California

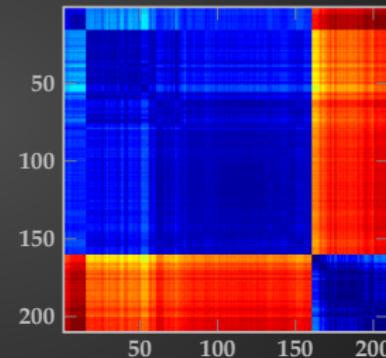




Physical Dexterity Quantification

Dynamic Regulation of Instabilities: Athletes vs Amateurs

Pairwise Distance Matrix

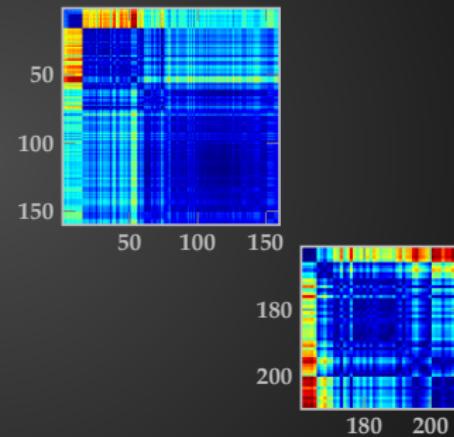




Physical Dexterity Quantification

Dynamic Regulation of Instabilities: Athletes vs Amateurs

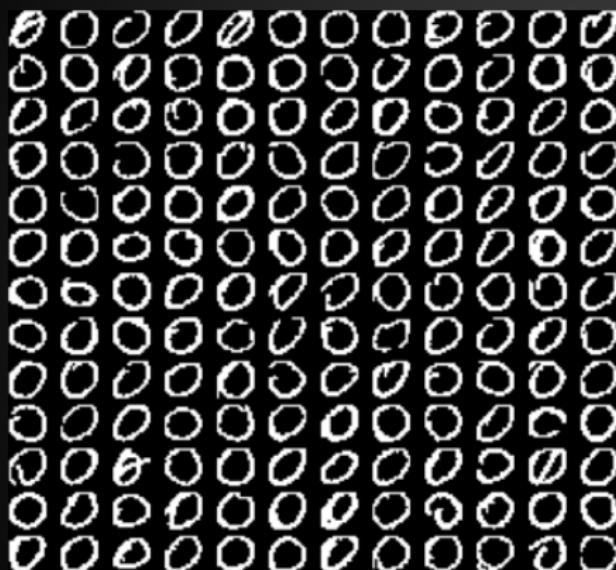
Pairwise Distance Matrix





Handwritten Digits

Recognition / Classification Problem





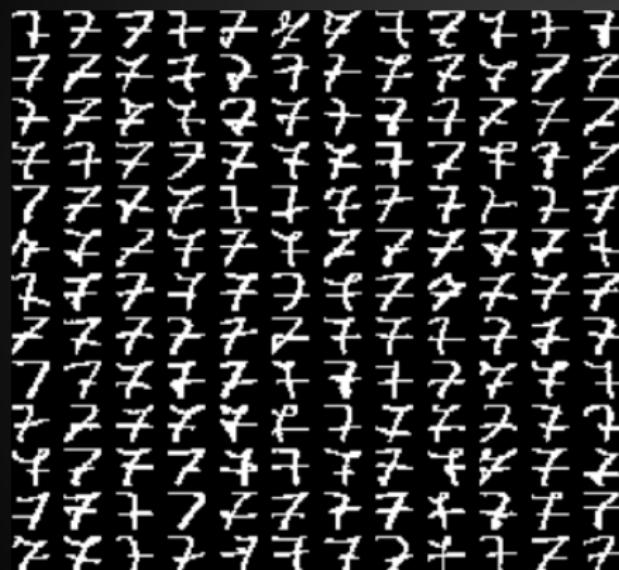
Handwritten Digits

Recognition / Classification Problem



Handwritten Digits

Recognition / Classification Problem





Handwritten Digits

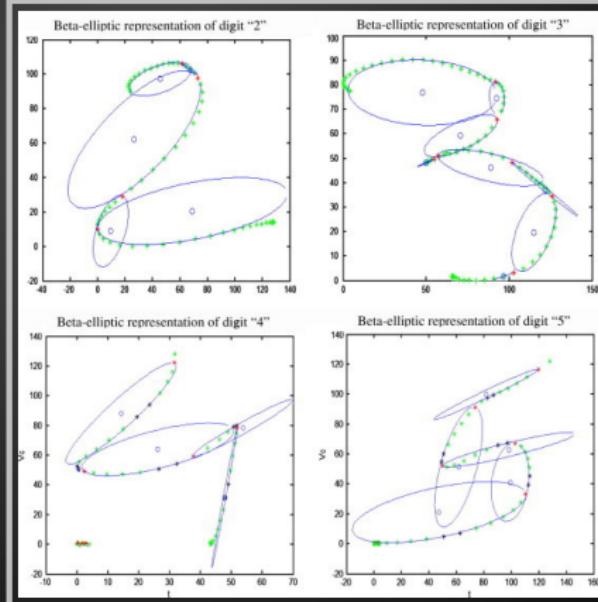
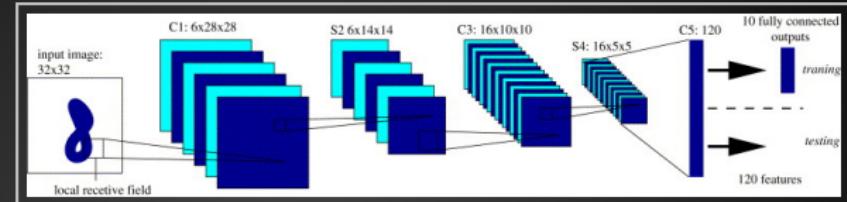
Recognition / Classification Problem

A large grid of handwritten digit '4's, arranged in approximately 20 rows and 20 columns. The digits are written in a cursive, black font on a white background.



State of Art Approaches

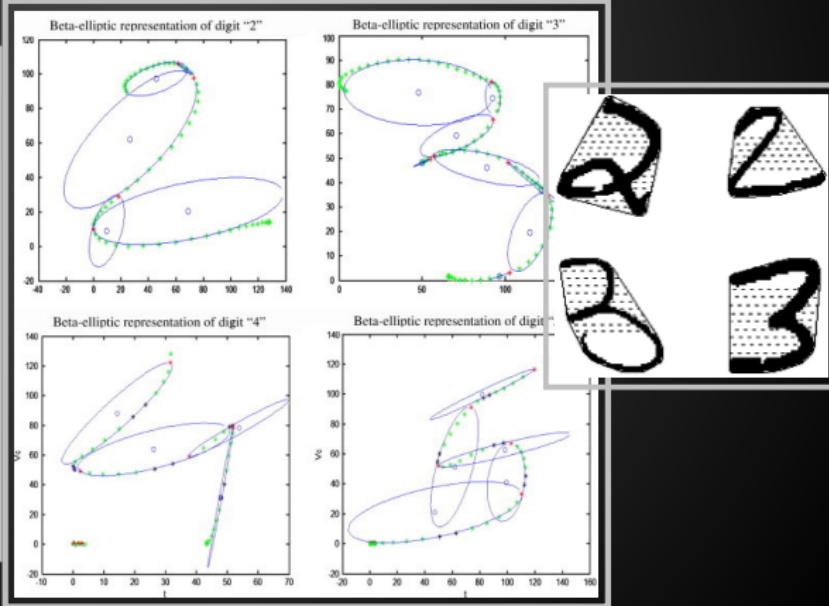
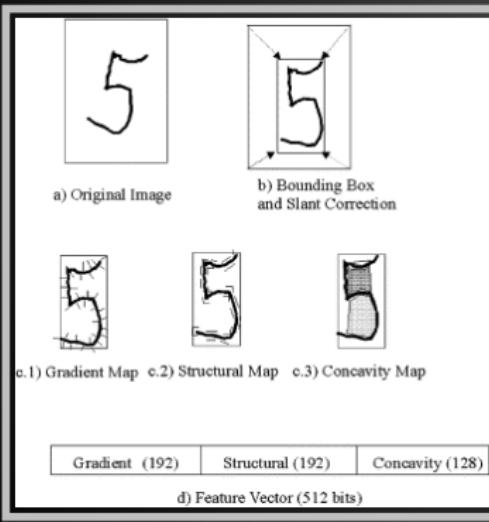
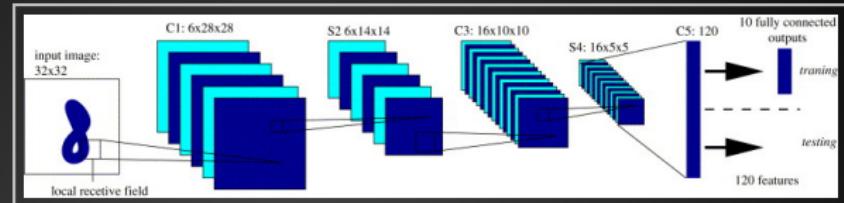
Find Clever Representations





State of Art Approaches

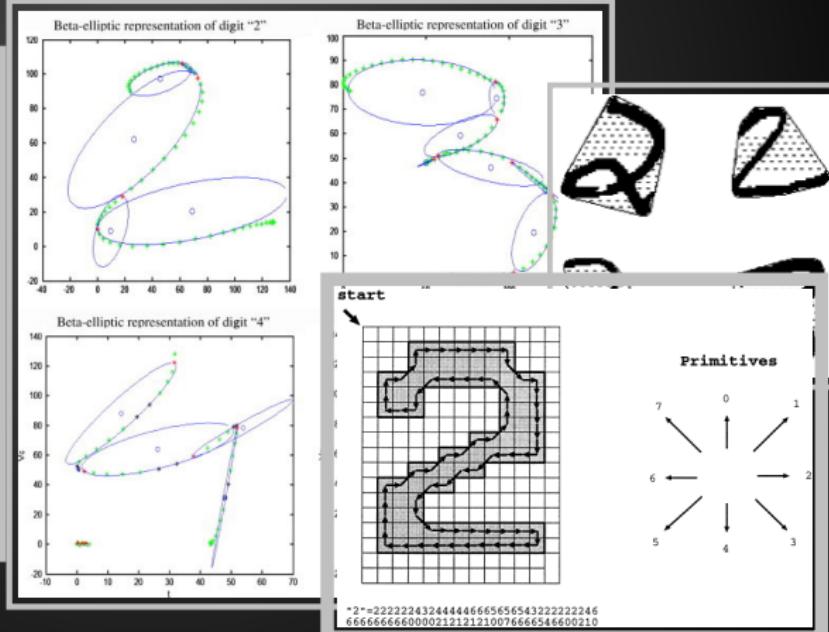
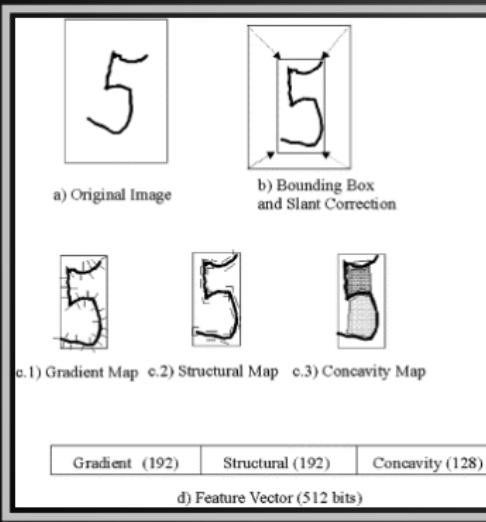
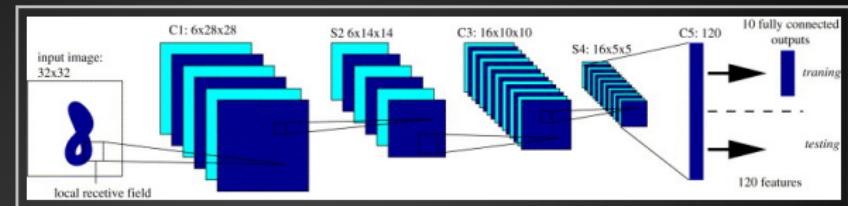
Find Clever Representations





State of Art Approaches

Find Clever Representations





State of Art Approaches

Find Clever Representations

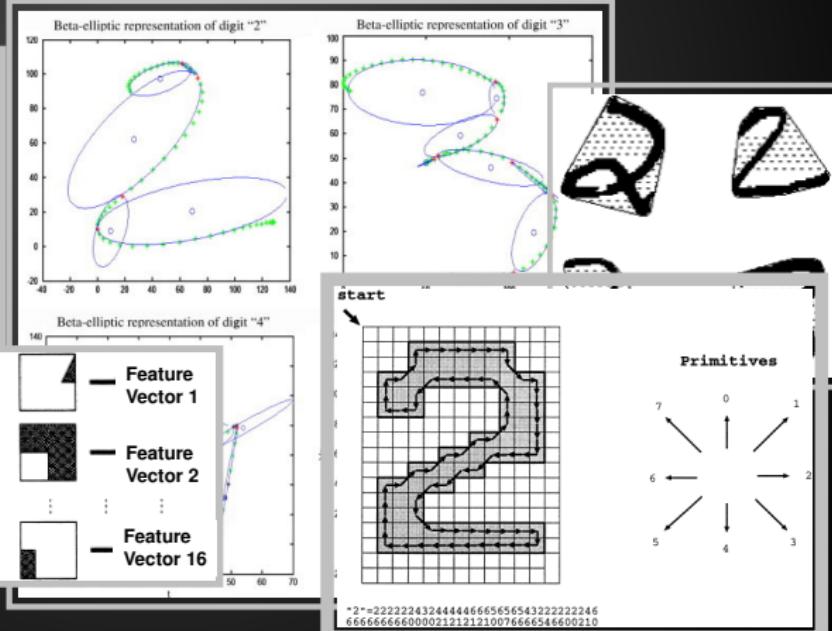
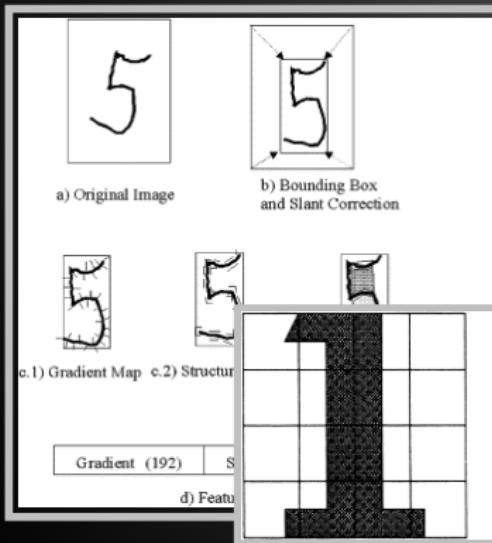
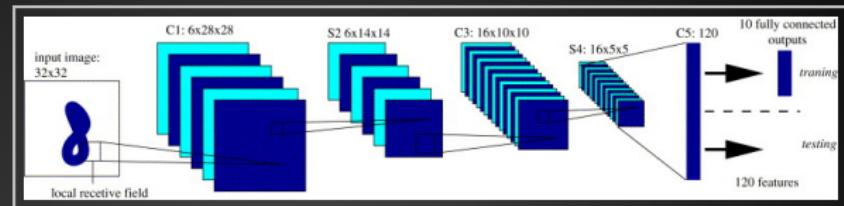




Image To Symbol Stream

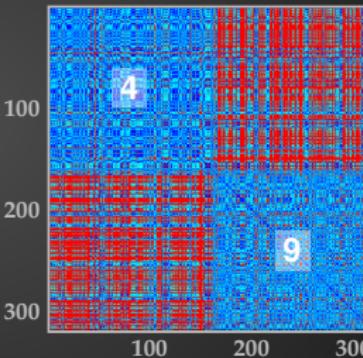
Random Walk



Handwritten Digits

Recognition / Classification Problem

Pairwise Distance Matrix





Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$	2	<input type="checkbox"/>
FWN:	0	<input type="checkbox"/>

Copy 1:



Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$	2	1
FWN:	Ø	1

Copy 1: 1



Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$	2 1 2
FWN:	Ø 1 2

Copy 1: 1 | 2



Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$	2 1 2 0
FWN:	0 1 2 2

Copy 1: 1 | 2



Memory-less Stream Manipulation



Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$

2	1	2	0	1	0	2	2	0	0	2	1	2	0	1	1	0	0	1	0	0	2	0	1	0	2
0	1	2	2	2	1	2	0	0	0	0	2	1	0	2	1	0	0	2	2	2	0	2	1	0	2

FWN:

Copy 1:

1	2	2	0	0	0	1	0	0	1	0	2	1	0	2	2	2	1	2	2	1	2	2	1	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---



Memory-less Stream Manipulation

Generating First Independent Copy for String s_1 :

$s_1:$	2 1 2 0 1 0 2 2 0 0 2 1 2 0 1 1 0 0 1 0 0 2 0 1 0 2
FWN:	0 1 2 2 2 1 2 0 0 0 0 2 1 0 2 1 0 0 2 2 2 0 2 1 0 2

Copy 1:	1 2 2 0 0 0 1 0 0 1 0 2 1 0 2 2 2 1 2 2 1 2 2 1 2 1 2 1 0 2
----------------	---

Generating Second Independent Copy for String s_1 :

$s_1:$	2 1 2 0 1 0 2 2 0 0 2 1 2 0 1 1 0 0 1 0 0 2 0 1 0 2
FWN:	0 0 2 0 0 2 0 2 2 2 2 2 1 2 1 0 1 0 0 0 2 0 2 2 2 0

Copy 2:	2 0 2 2 1 0 0 2 2 2 1 2 2 0 1 2 2 1 2 2 1 2 1 2
----------------	---



Memory-less Stream Manipulation

Generating Inverse of String s_1 :

Copy 1:

1	
2	
0	

Copy 2:

2	
0	

Inverse:

0



Memory-less Stream Manipulation

Generating Inverse of String s_1 :

Copy 1:

1	2
---	---

Copy 2:

2	0
---	---

0	1
---	---

Inverse:

0	1
---	---



Memory-less Stream Manipulation

Generating Inverse of String s_1 :

Copy 1:

1	2	2
---	---	---

Copy 2:

2	0	2
0	1	

Inverse:

0	1
---	---



Memory-less Stream Manipulation



Memory-less Stream Manipulation

Generating Inverse of String s_1 :

Copy 1:

1	2	2	0	0	0	1	0	0	1	0	2	1	0	2	2	1	2	2	1	2	2	1	2	1	2	2	1	2	1	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Copy 2:

2	0	2	2	1	0	0	2	2	2	1	2	2	0	1	2	2	1	2	1	2	2	1	2	1	2	2	1	2	1	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

0	1	1	2	2	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Inverse:

0	1	1	2	2	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Summing s_2 with Inverse of s_1

String s_2 :

2	2	1	1	2	2	1	1	0	2	2	2	0	1	0	0	2	1	2	2	2	1	0	0	2	1	2	2	1	2	2	1	2	2	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Inverse of String s_1 :

0	1	1	2	2	1	1	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Result:

1	2	1	0
---	---	---	---



Summing It Up!

- Notion of Universal Similarity
- Zero-knowledge Feature-free Anomaly Detection







Patents held by Cornell University

- 6259-01-US Stochastic Automata for Earthquake Prediction from Large Scale Surveys
- 6259-02-PC Systems and Methods for Abductive Learning of Quantized Stochastic Process
- 6024-03-PC System and Methods for Analysis of Data PCT/US13/62397
- 6998-01-US Causality Network Construction Algorithm (Application no. 62170063, EFS ID 2517508)