From rough paths to streamed data

DataSig

A rough path between mathematics and data science

Terry Lyons

SBD seminar
25 March 2021

with many others … but particularly
Cris, Patrick, James, James, Varun,
Imanol, Maud, Peter, Roly, Sam,… Tom,
Harald, Hao, Patricia

A Turing vision

DataSig

A rough path between mathematics and data science

> We channel our research around a number of ambitious challenges which represent areas in which AI and data science can have a game-changing impact for science, society, and the economy. These challenges will not be led by the Turing alone, but depend on significant collaboration and partnerships.

A public domain collaboration on detecting malware

DataSig

A rough path between mathematics and data science

DataSig

Terry

Cris

Patricia

Imanol

Peter

Maud

Roly

Thomas

Varun

ICM Korea 2014:

"Rough path theory is focused on capturing and making precise the interactions between highly oscillatory and non-linear systems. It draws on the analysis of LC Young and the geometric algebra of KT Chen. The concepts and the uniform estimates have widespread application, contribute to (Graham) automated recognition of Chinese handwriting and (Hairer) formulation of appropriate SPDEs to model randomly evolving interfaces."

DataSig

A rough path between mathematics and data science

Modelling behavior of evolving systems

ICM Korea 2014:

DataSig

A rough path between mathematics and data science

> At the heart of the mathematics is the challenge of describing a potentially highly oscillatory and vector valued stream parsimoniously and efficiently, well enough to predict the response of a nonlinear system such as
>
> $$dy_t = f(y_t)dx_t, y_0 = a$$
>
> The answer is subtle, but a core idea is the description of x over intervals via the (log) signature (Chen).

Much of the research introduced today comes from DataSıg: an EPSRC/UKRI 5-year program grant

DataSıg

A rough path between mathematics and data science

Mathematics

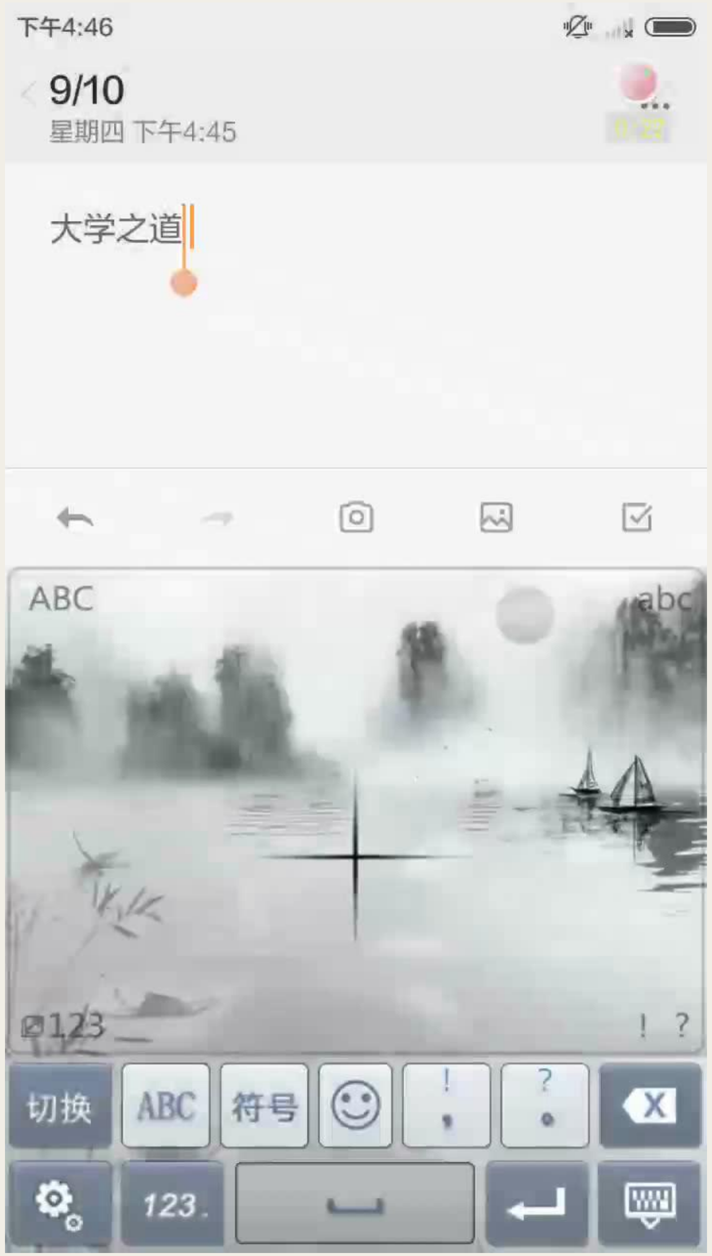
- rough path theory and signatures
- describing the interactions between complex systems from the top down
- extending the calculus of differential equations to complex contexts
- the notion of an unparameterized path captured by the order of events
- clean and minimal universal feature sets - (expected) signature

Innovation

- the notion of a neural controlled differential equation
- the notion of a pde-kernel which harnesses the numerics of PDE theory.
- a principled mathematical framework that allows further innovation (e.g. simulation)

Embedded contexts

- streamed data is everywhere; Chinese handwriting, hospital wards, event logs …



Streamed data

- a character drawn on the screen of an iPhone
- an order book
- a piece of text
- progression through hospital record
- astronomical data
- video of a person moving
- an evolving stream of emotions
- ICU data to detect sepsis
- the evolving stock position in a supermarket or computer switch

Ensembles of streamed data

- the event log of processes generated by malware
- the behaviour of crowds
- the evolution of cancer cell lines

Key questions

- understand what you have observed
- predict the distribution of what is happening next
- identify anomalies

DataSig
A rough path between mathematics and data science

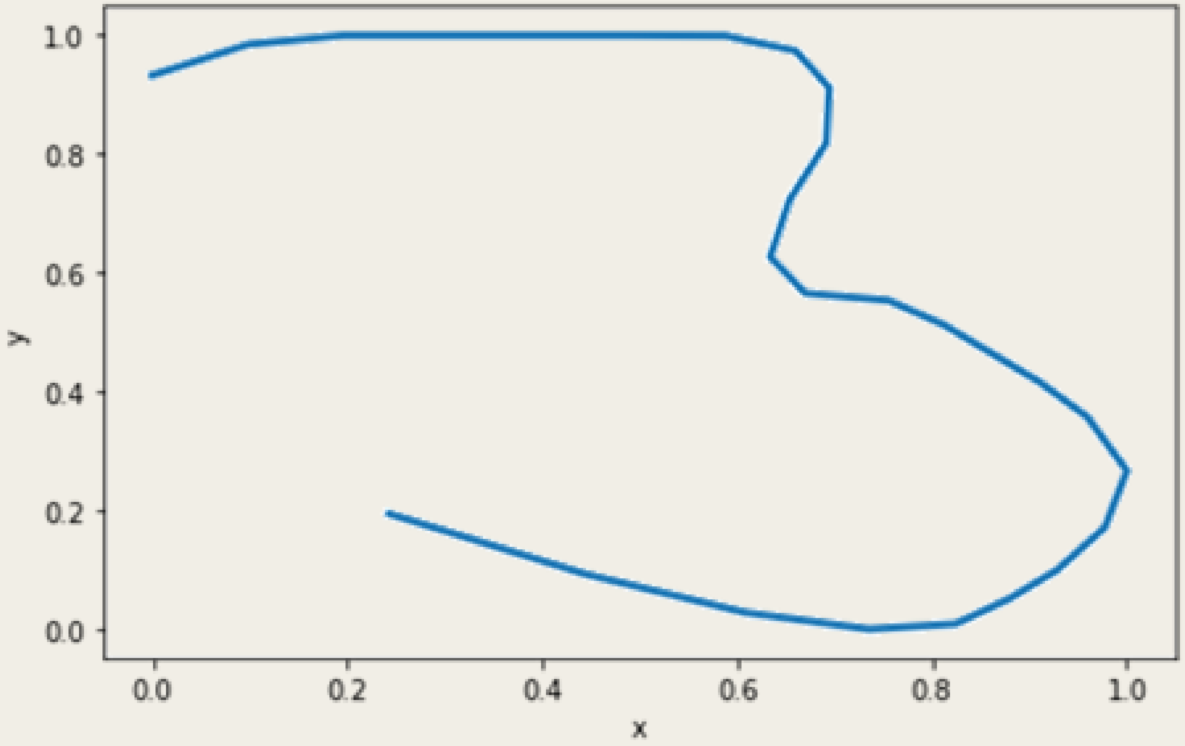
DataSig

A rough path between mathematics and data science

Some maths of evolving systems

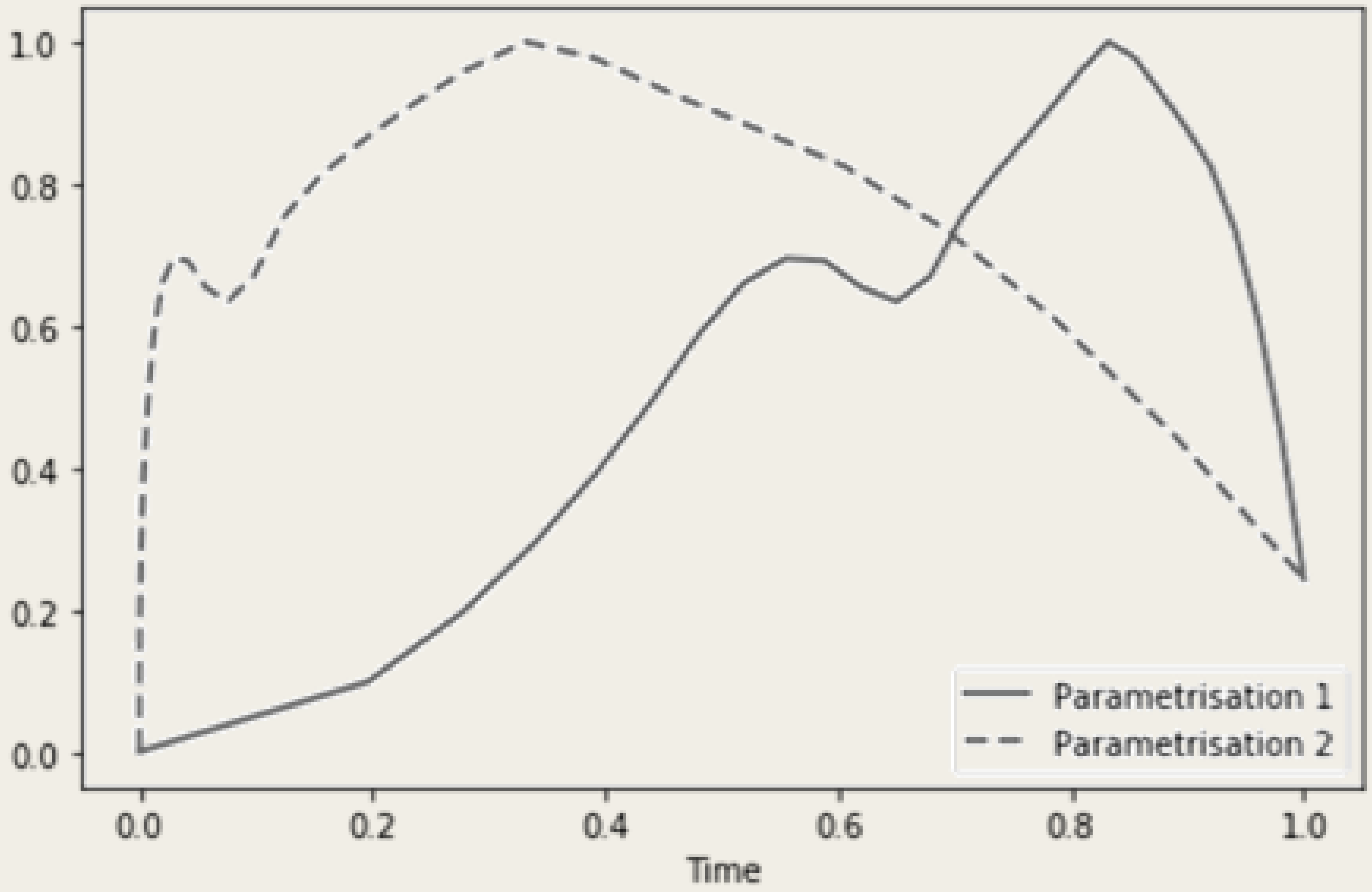
Data science does not like symmetry

- Re-parameterisation is a huge symmetry group
- Multimodal streams modulo re-parameterisation form a group
- Representing this group in the tensor algebra provides a faithful feature set and removes the symmetry
- Drawn from old mathematics, new tools, signature and log signature, and new maths describing the functions on streams



Different sampling procedures

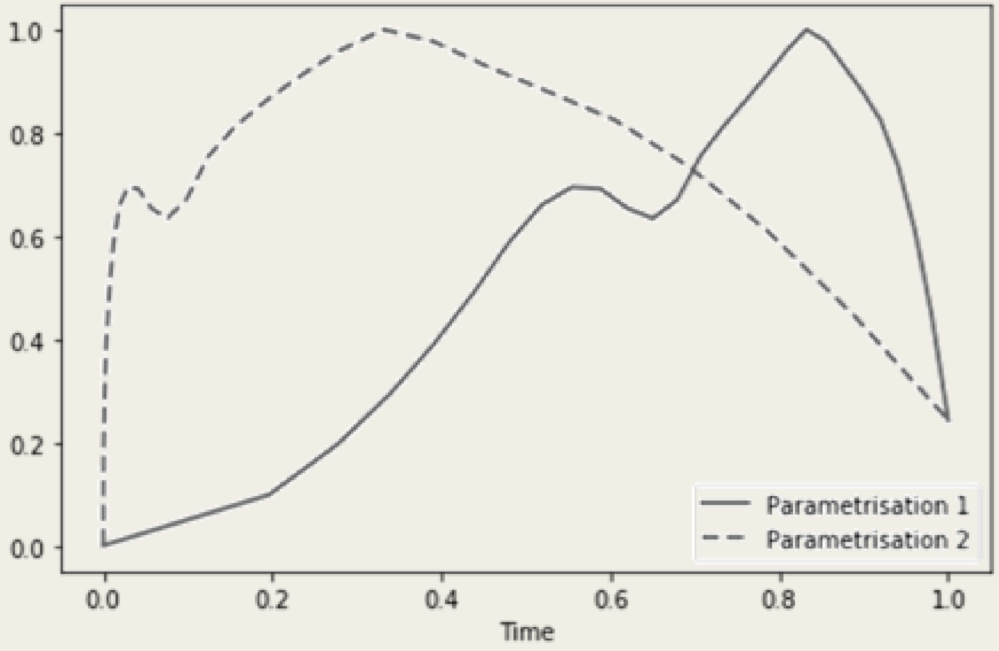
- The letter “3” is drawn from top to bottom
- The x coordinate of the evolving symbol sampled differently (at uneven speeds)

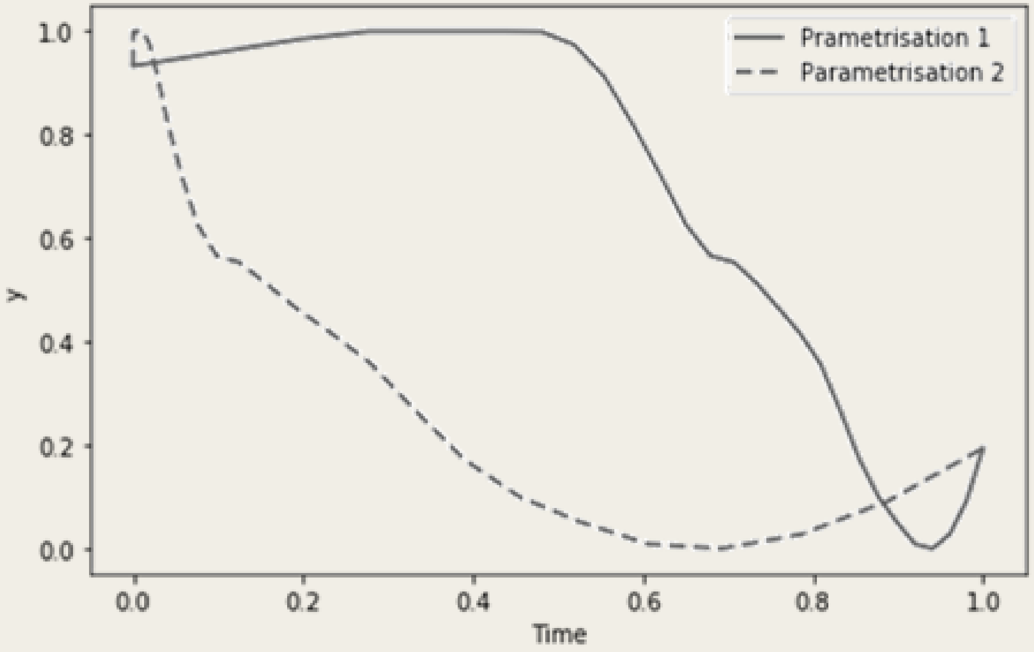


Different sampling procedures

The number “3” x, y coordinates – same picture drawn at two different speeds

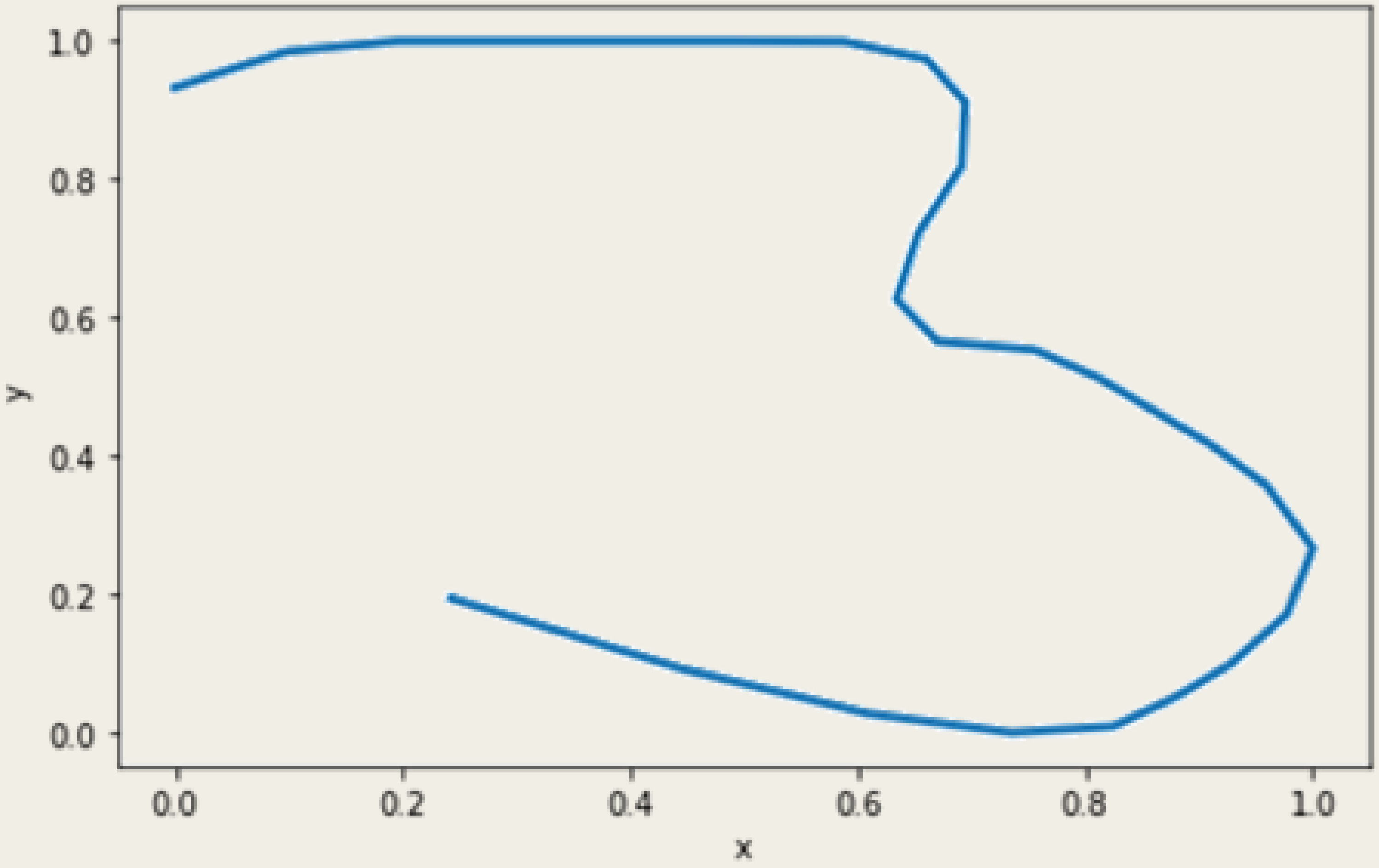
- no consistent wavelets
- reparameterisations do not form a linear space!





Different sampling procedures

- The letter “3” is drawn from top to bottom
- How does one describe the three or any path modulo the symmetry of parametrisation?



The signature of a path describes an unparameterised stream γ

Signature is a *top down* description for unparameterised paths that describes a path segment through its effects of stylised nonlinear systems

$$dS = S \otimes d\gamma$$

It filters out the infinite dimensional noise of resampling allowing prediction and classification with *much* smaller learning sets.

It gives fixed dimensional feature sets regardless of the sample points.*

* missing data/varying parameterisation not issues although inadequacy of sampling may be

The signature - faithful and universal features describing an unparameterised stream

The signature of a stream γ over $I = [s, t]$ defined by $\sum_{k=0}^{\infty} S_k$ where $S_0 = 1$ and

$$S_k(\gamma, I) := \iint\limits_{s<u_1<\cdots<u_k<t} d\gamma_{u_1} d\gamma_{u_2} \dots d\gamma_{u_k}$$

These "Fourier-like" features exactly describe the *unparameterised* stream (Hambly Lyons Annals Math 2010) up to appropriate null sets.

Coordinate iterated integrals are universal models

$$< e, S(\gamma, I) >$$

Analysis, geometry, combinatorial Hopf/dendriform/sensor algebras

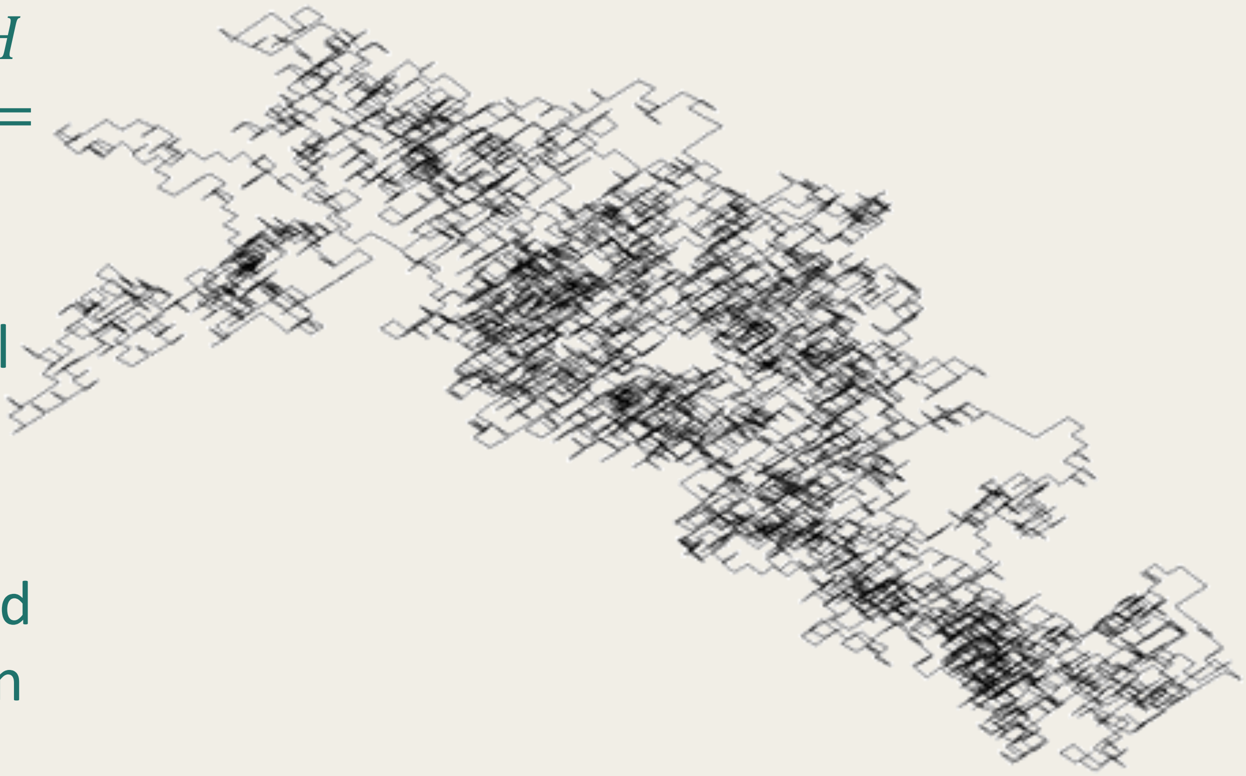
- The *Signature* is a faithful embedding of the *unparametrized* stream into a vector space
- Continuous functions on streams can be well approximated by linear functionals on signatures
- The Expected Signature describes the ensemble of paths
- The log-signature describes paths without redundancy
- There is a natural pde kernel

Analysis, geometry, combinatorial Hopf/dendriform/sensor algebras

DataSig

A rough path between mathematics and data science

- Let be x, y be unparametrized paths in H and consider the bilinear form $K(x, y) \coloneqq \langle S(x), S(y) \rangle$. Franz J. Kiraly, Harald Oberhauser; JMLR 20(31):1-45, 2019 gave a kernel trick for the tangent kernel for the truncated signature embedding.
- Salvi et al. then identified the Goursat pde as the kernel trick for the untrucated kernel and gave analytic sense for it even for rough paths:

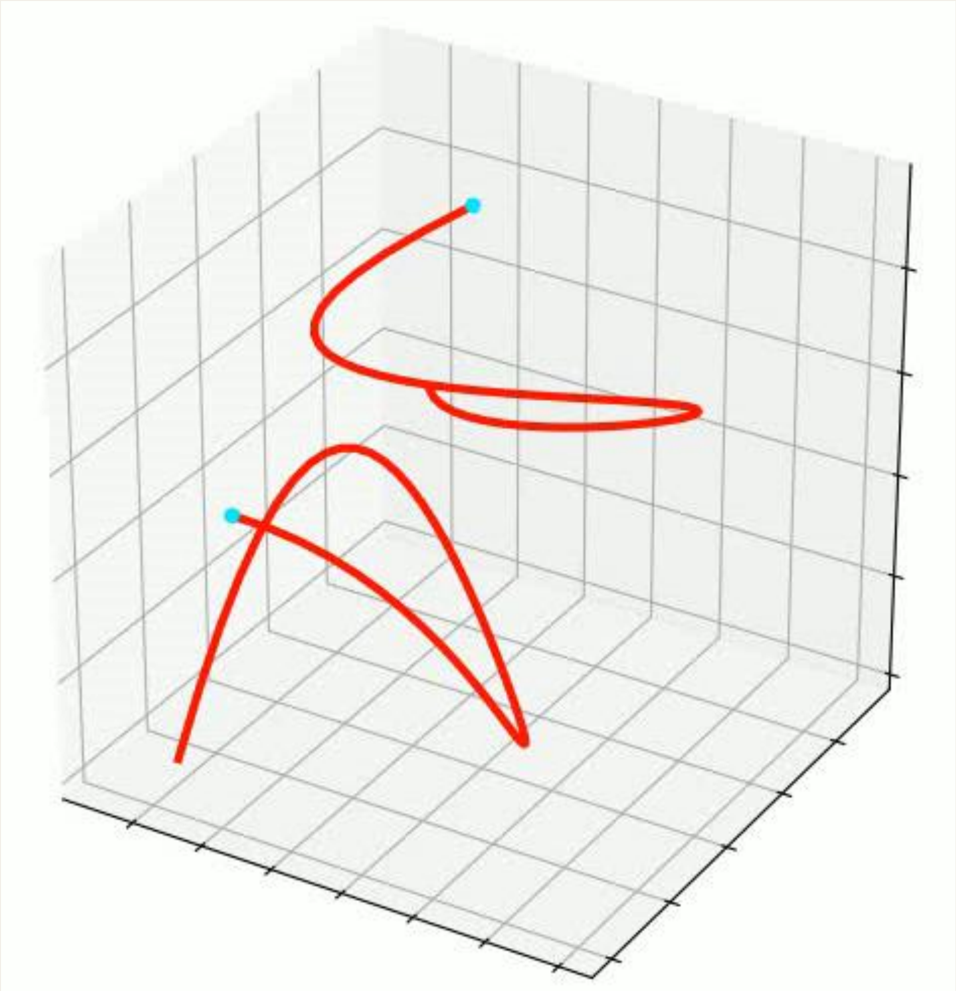


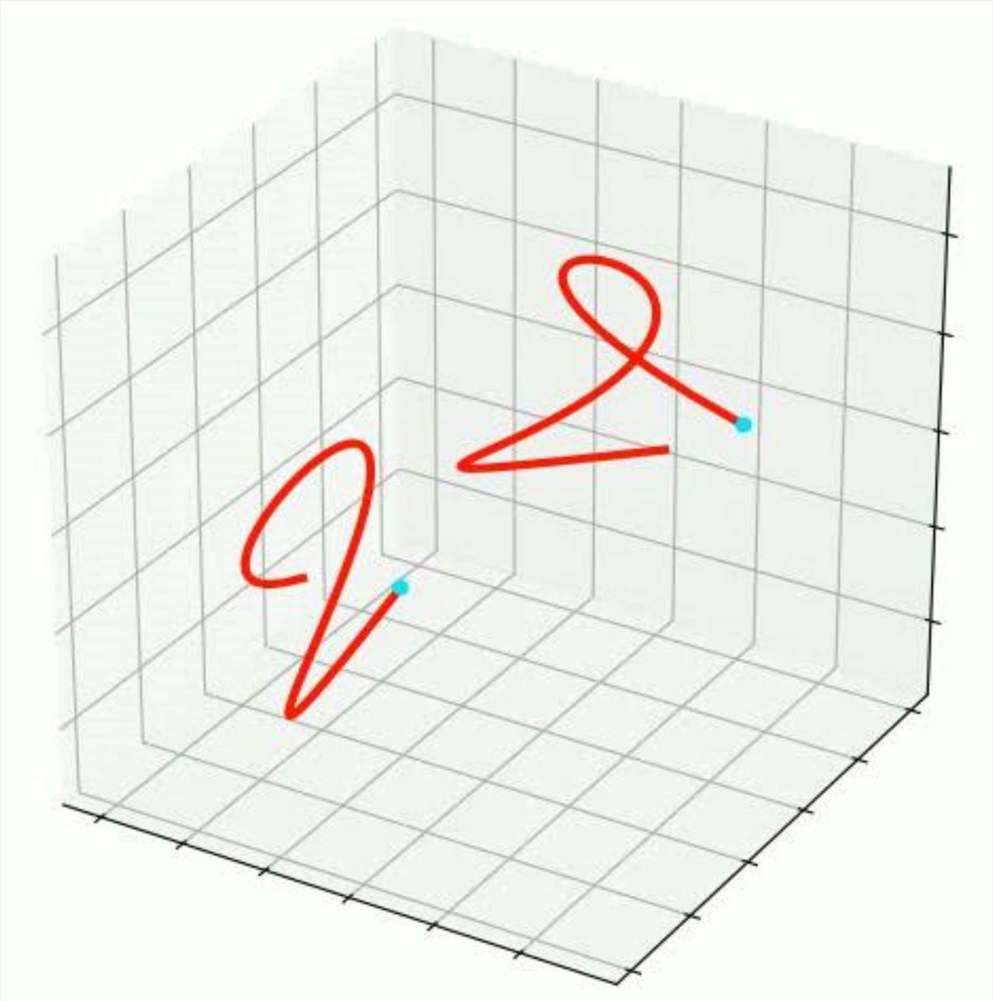
$$\frac{\partial^2 K(x|_{[u_0,u]}, y|_{[v_0,v]})}{\partial u \partial v} = \langle \dot{x}, \dot{y} \rangle K\left(x|_{[u_0,u]}, y|_{[v_0,v]}\right)$$

Recovering the curves from the signature

DataSig

A rough path between mathematics and data science

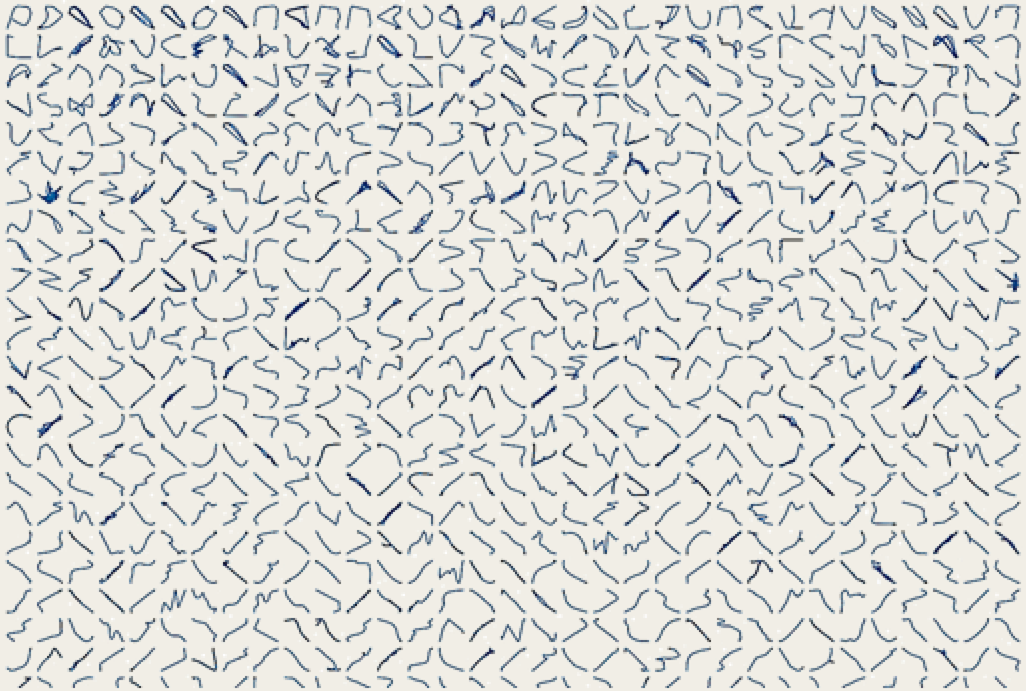




Weixin Yang, Jaiwei Chang

Fermanian

Modern data



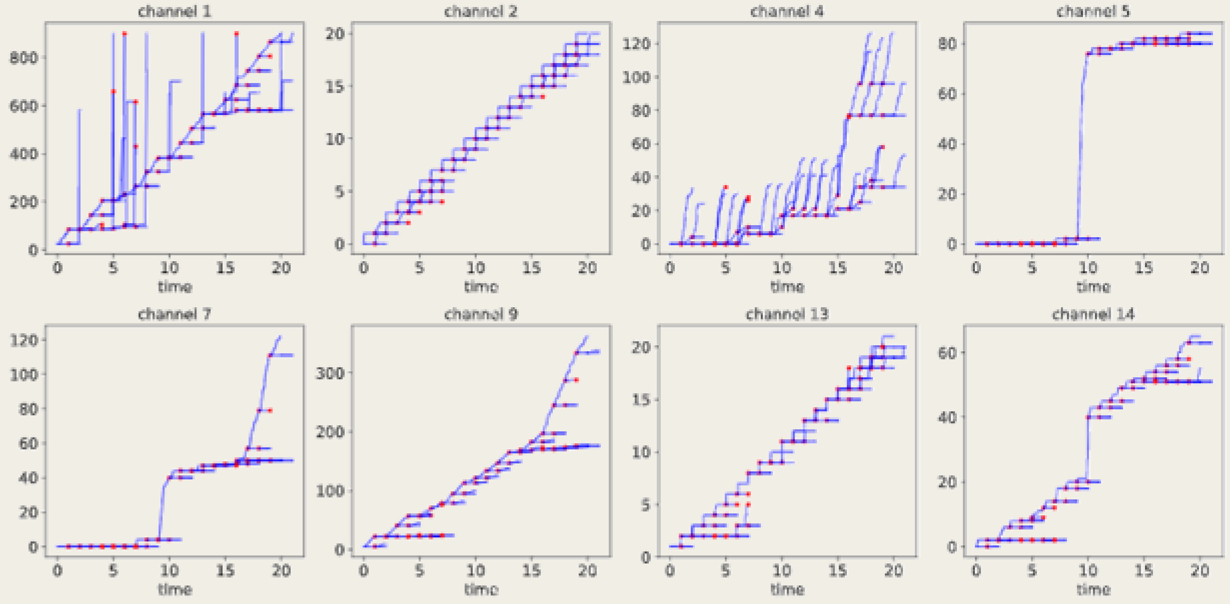


Fig. 1. Visual representation of selected channels of one single streaming tree. Each plot represents the evolution in time of the value of a given channel of the streaming tree, on its various branches. A red dot indicates a point where the currently-tracked process sets off a child process, causing the tree to branch.

Neural Controlled differential equations

DataSig

A rough path between mathematics and data science

All fields

Use controlled differential equations to model policies

Train a neural f so that

$$dy = f(y)d\gamma, y_0 = a$$

Where a represents the current state and γ determines the external policy.

Allows learning counterfactuals. (Van der Schaar).

Neural Rough Differential Equations for Long Time Series

https://arxiv.org/abs/2009.08295

James Morrill, Cristopher Salvi, Patrick Kidger, James Foster, Terry Lyons

Ensembles of paths

DataSig

A rough path between
mathematics and data science

Process tree example:
Expected signatures of clouds of paths

DataSig

A rough path between mathematics and data science

Developed a way to apply expected signature techniques by viewing processes as trees evolving over time (eg the crop yield prediction task).

Predicting the yield of wheat crops over a region from the longitudinal measurements of climatic variables recorded across different locations of the region.

Eurostat dataset containing the total annual regional yield of wheat crops in mainland France - divided in 22 administrative regions - from 2015 to 2017.

Process tree example:
Expected signatures of clouds of paths

DataSig

A rough path between mathematics and data science

Viewing a cloud of interacting paths evolving over time as an expected signature, it can be merged with other channels, and the process repeated. PDE kernels can manage dimension. The crop yield prediction task matches this model: AISTATS 2021 arxiv.org/pdf/2006.05805.pdf

The climatic measurements (temperature, soil humidity and precipitation) are extracted from the GLDAS database (Rodell et al, 2004), are recorded every 6 hours at a spatial resolution of $0.25° \times 0.25°$, and their number varies across regions. Add regional policy information, etc.

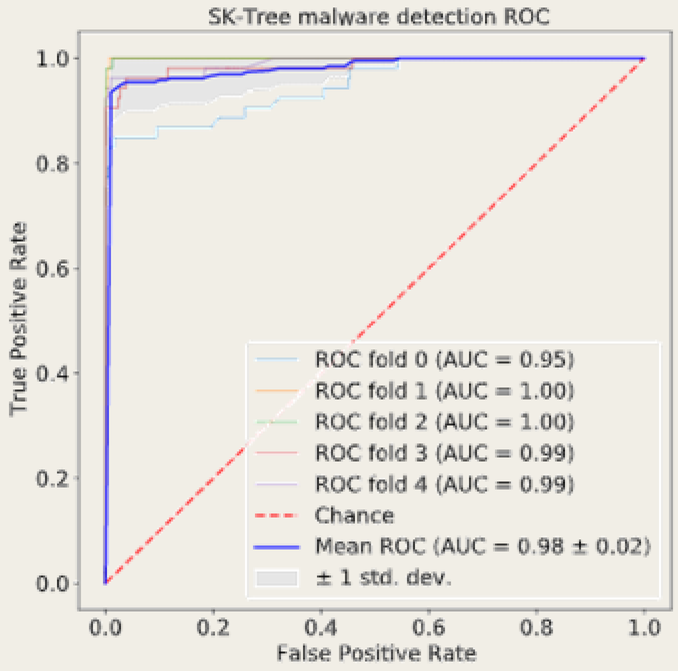
Model	MSE	MAPE
Baseline	2.38	23.31
DeepSets	2.67	22.88
DR-RBF	.82	13.18
DR-Matern32	.82	13.18
DR-GA	.72	12.55
KES	**.65**	**12.34**
SES	**.62**	**10.98**

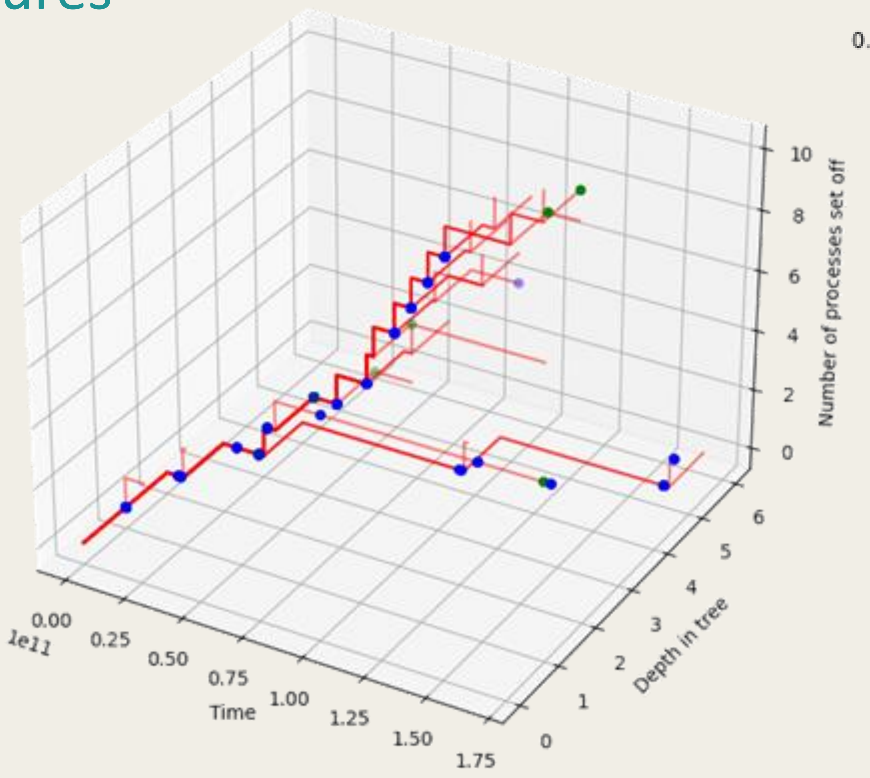
Process tree example: Expected signatures of clouds of paths

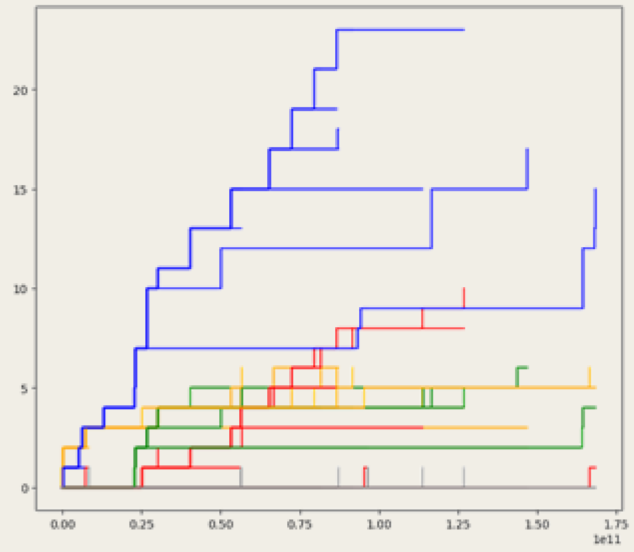
Developed SK-tree structure to apply standardised expected signature techniques to host-based event logs, by viewing processes as trees evolving over time analysed as expected signatures through a PDE kernel.
[2102.07904.pdf (arxiv.org)](https://arxiv.org/pdf/2102.07904.pdf)

We demonstrate the SK-Tree to detect malicious events on a portion of the publicly available DARPA OpTC dataset, achieving an initial AUROC score of 98% for a supervised question.

ROC evaluation of the SK-Tree binary classifier on the OpTC data







De-identified Streams

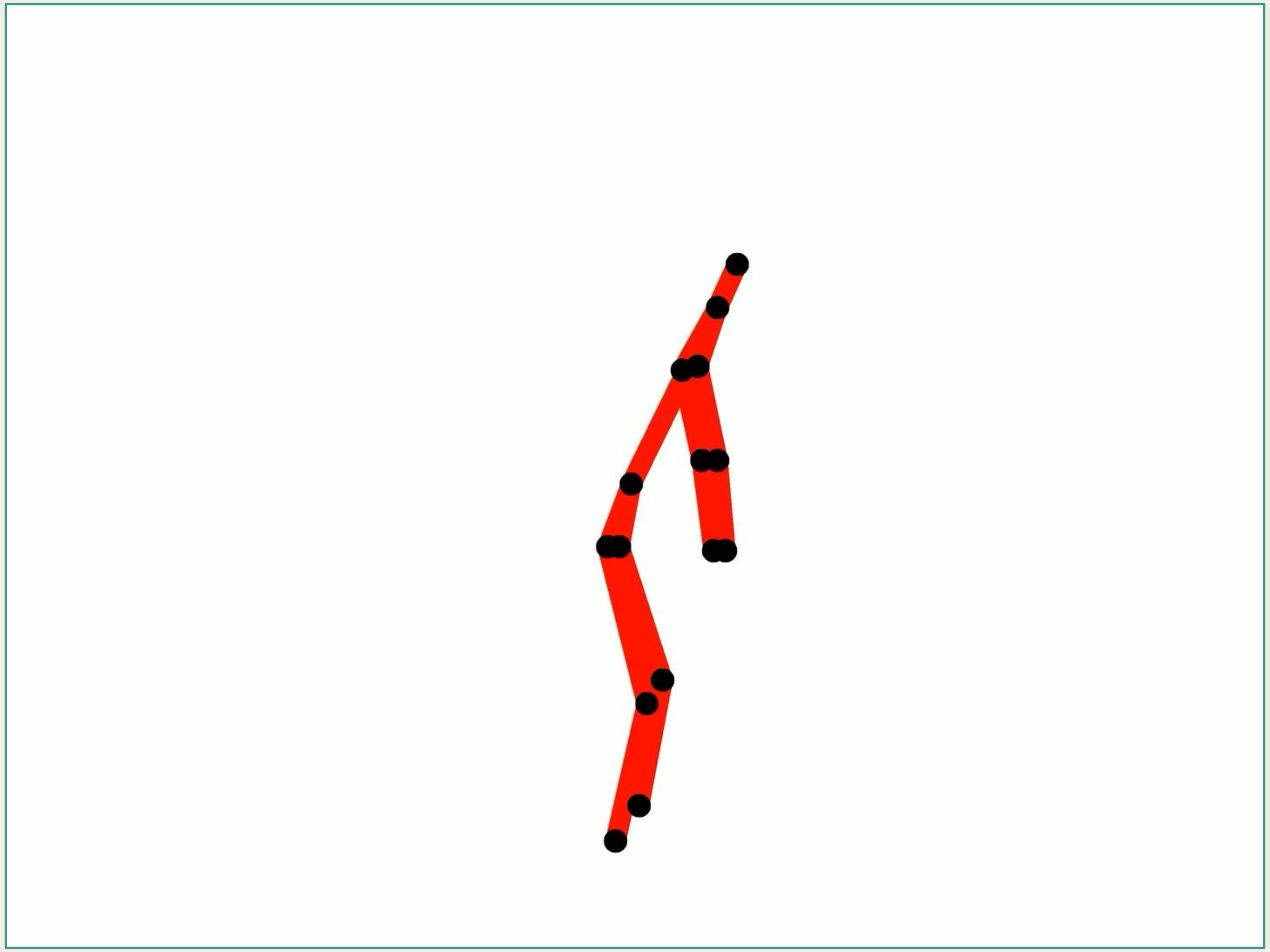
DataSig

A rough path between mathematics and data science

Evolving landmarks from video

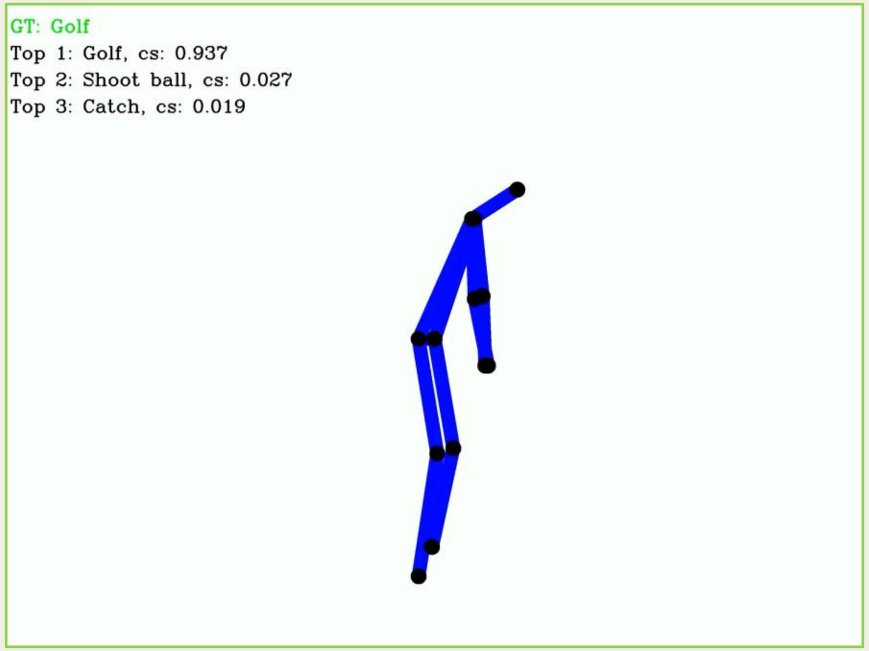
- Reducing people in video to landmarks is straightforward
- The matchstick men and women are deidentified data streams in 30-75 dimensions
- Currently working with facial landmarks and lip reading.

http://mvig.sjtu.edu.cn/research/alphapose.html

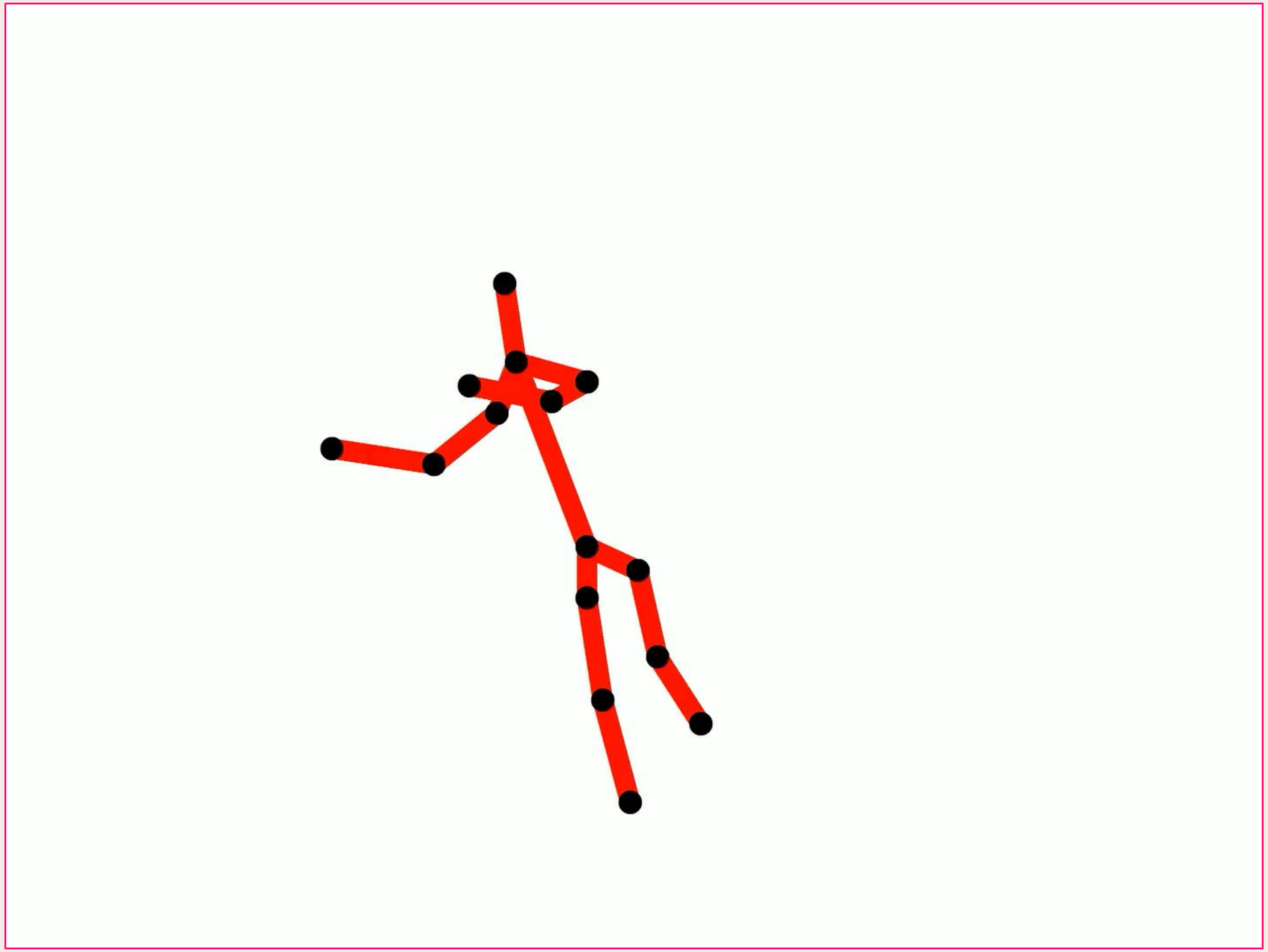


H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, “Towards Understanding Action Recognition,” in 2013 IEEE International Conference on Computer Vision, Dec. 2013, pp. 3192–3199

GOLF

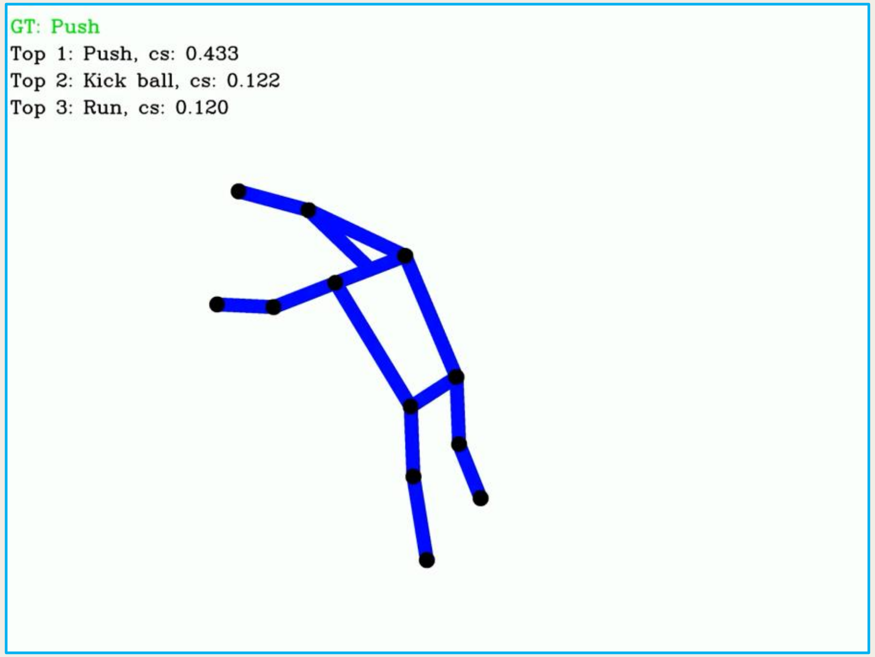


H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, “Towards Understanding Action Recognition,” in 2013 IEEE International Conference on Computer Vision, Dec. 2013, pp. 3192–3199



H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, “Towards Understanding Action Recognition,” in 2013 IEEE International Conference on Computer Vision, Dec. 2013, pp. 3192–3199

PUSH



H. Jhuang, J. Gall, S. Zuffi, C. Schmid, and M. J. Black, "Towards Understanding Action Recognition," in 2013 IEEE International Conference on Computer Vision, Dec. 2013, pp. 3192–3199

Experimental results

Method	Accuracy (%)
DT-FV [74]	65.9
P-CNN [56]	74.6
HLPF [31]	76.0
Novel HLPF [44]	79.6
Path Signature (Ours)	**80.4**

JHMDB dataset

(2D, Ground Truth Pose)

#classes: 21

#samples: 928

[74] H. Wang and C. Schmid, Action recognition with improved trajectories, In IEEE ICCV, pp. 3551-3558, 2013.

[56] G. Cheron, et al. P-CNN: Pose-based CNN Features for Action Recognition. In ICCV, 2015.

[31] H. Jhuang, et al. Towards understanding action recognition, In ICCV, pp. 3192-3199, 2013.

[44] J. Fan, Z. Zha, and X. Tian, Action recognition with novel high-level pose features, In IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1-6, 2016

Experimental results

Method	Ground Truth Pose	Estimated Pose	
		pose only	with extra clues
MST [CVPR-14]	-	-	45.3 (RGB)
HLPF [ICCV-13]	78.2	52.9	-
JointAP [CVPR-15]	-	61.2	-
P-CNN [ICCV-15]	72.5	-	66.8 (RGB)
PA-AP [arXiv-16]	-	61.5	74.7 (RGB)
Hierarchical [CVPR-16]	-	66.9	77,5 (RGB)
JDD [IJCAI-16]	81.9	-	77.7 (RGB)
RPAN [ICCV-17]	-	-	**78.6** (RGB)
Path signature (Ours)	**84.2**	**68.2**	75.7 (confidence)

Sub-JHMDB

#classes: 12

#samples: 319

Privacy

[CVPR-14] J. Wang, et al. Cross-view action modeling, learning and recognition. In CVPR, 2014.

[ICCV-13] H. Jhuang, et al. Towards understanding action recognition, In ICCV, pp. 3192-3199, 2013.

[CVPR-15] B. X. Nie, et al. Joint action recognition and pose estimation from video. In CVPR, 2015.

[ICCV-15] G. Cheron, et al. P-CNN: Pose-based CNN Features for Action Recognition. In ICCV, 2015.

[FG-16] U.Iqbal, et al. Pose for action - action for pose, In IEEE FG, 2016.

[CVPR-16] I. Lillo, et al. A hierarchical pose-based approach to complex action understanding using dictionaries of actionlets and motion poselets. In CVPR, 2016.

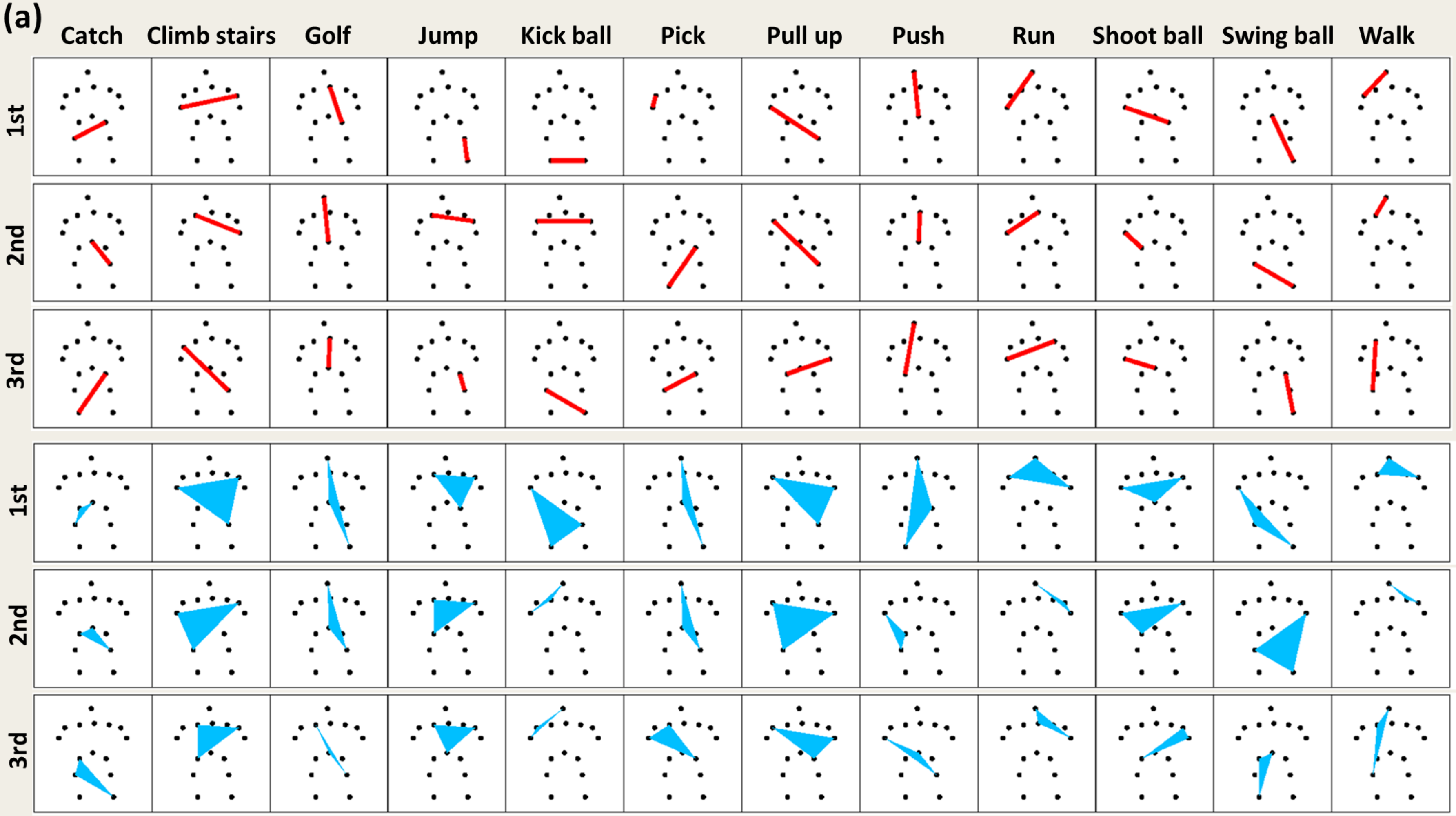
[IJCAI-16] C. Cao, et al. Action recognition with joints-pooled 3d deep convolutional descriptors. In IJCAI, 2016.

[ICCV-17] Du W, et al. Rpan: An end-to-end recurrent pose-attention network for action recognition in videos. In ICCV, pp. 3725-3734, 2017.

Toward understanding action recognition

For Spatial Structure ：

Top-3 most “important” Pairs/Triples on sub-JHMDB dataset



Experimental results

Method	Accuracy (%)
Yun et al., [32]	80.3
Ji et al., [76]	86.9
CHARM [77]	83.9
HBRNN [19] (reported by [18])	80.4
Deep LSTM (reported by [18])	86.0
Co-occurrence LSTM [18]	90.4
STA-LSTM [78]	91.5
ST-LSTM-Trust Gate [22][23]	93.3
SkeletonNet [79]	93.5
Path Signature (Ours)	**96.8**

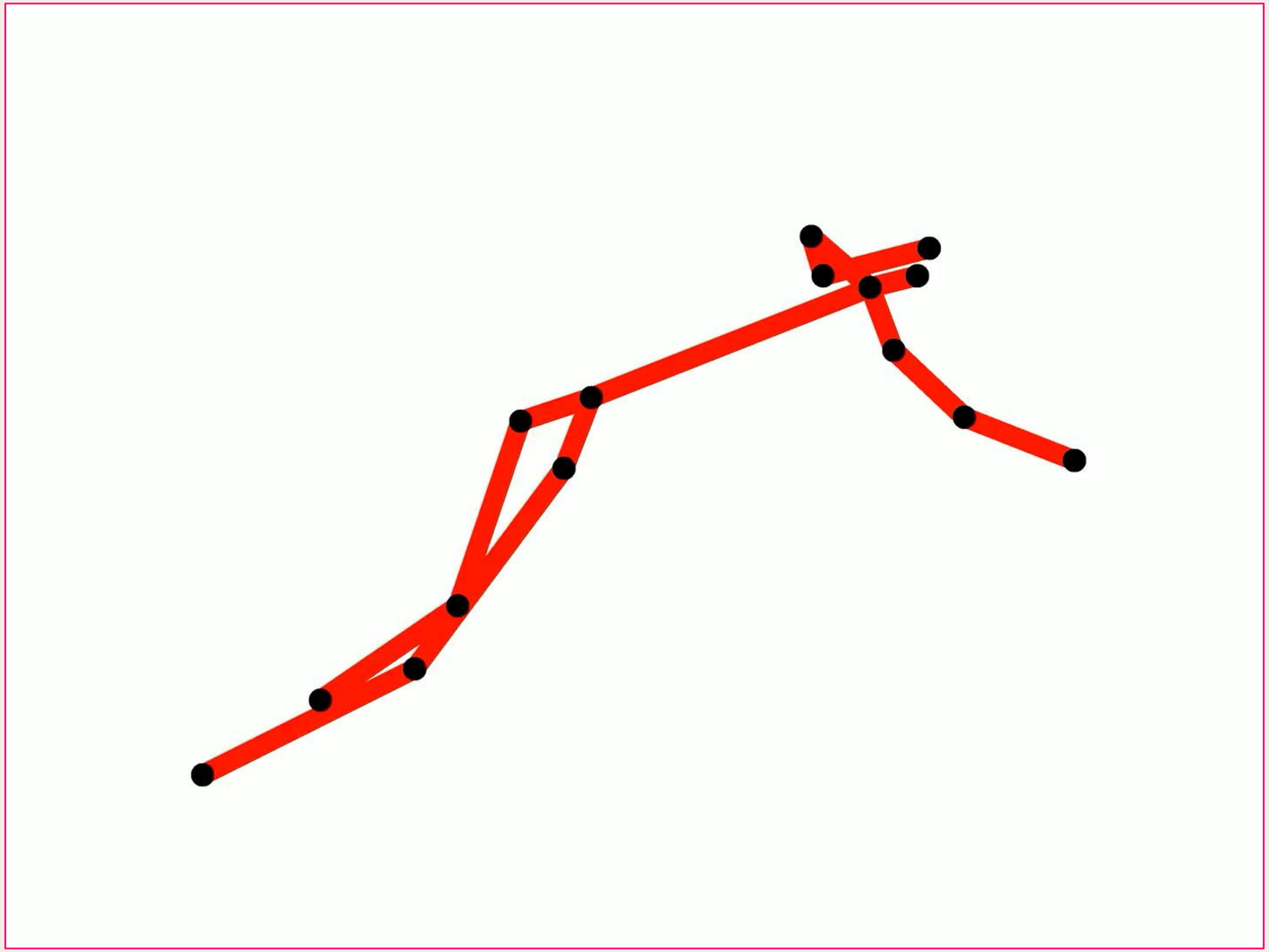
SBU Interaction Dataset (Kinect 3D)

#classes: 21

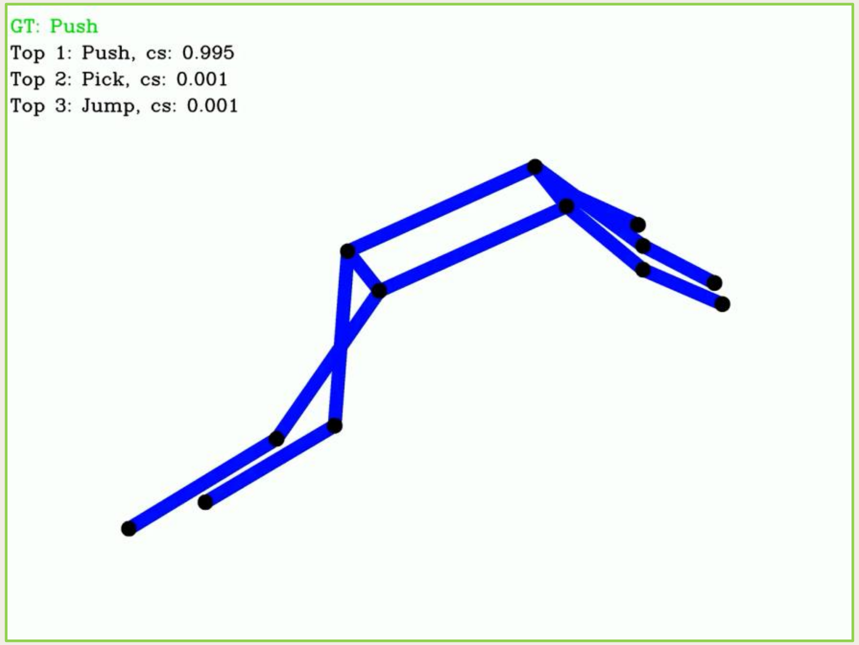
#samples: 300

Privacy

[32] K. Yun, et al. "Two-person interaction detection using body-pose features and multiple instance learning," CVPRW, pp. 28-35, 2012.
[76] Y. Ji, et al. "Interactive body part contrast mining for human interaction recognition," In ICMEW, pp. 1-6, 2014.
[77] W. Li, et al. "Category-blind human action recognition: a practical recognition system," In ICCV, pp. 4444-4452, 2015.
[18] W. Zhu, et al. "Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks," In AAAI, vol. 2, 2016.
[19] Y. Du, et al. "Hierarchical recurrent neural network for skeleton based action recognition," In CVPR, pp. 1110-1118, 2015.
[78] S. Song, et al. "An end-to-end spatio-temporal attention model for human action recognition from skeleton data," In AAAI, vol. 1, no. 2, p. 7, 2017.
[23] J. Liu, et al. "Skeleton-based action recognition using spatio-temporal lstm network with trust gates," IEEE TPAMI, 2017.
[79] Q. Ke, et al. "SkeletonNet: mining deep part features for 3-D action recognition," IEEE Signal Processing Letters, vol. 24, no. 6, pp. 731-735, 2017.
[Ours] W. Yang, T. Lyons, H. Ni, C. Schmid, L. Jin, "Leveraging the Path Signature for Skeleton-based Human Action Recognition," arXiv preprint arXiv:1707.03993, 2017.



PUSH



Communication

DataSig

A rough path between mathematics and data science

Landmark-based action recognition

To communicate our methodology, and aside from our papers, with their software we are constructing notebooks with introductory examples of what we can do.

Peter Foster has put together a simple notebook you can run that demonstrates viable approaches to recognizing these actions that can be trained on small datasets.

https://www.datasig.ac.uk/examples

Thank you!

$$dy = f(y)d\gamma$$