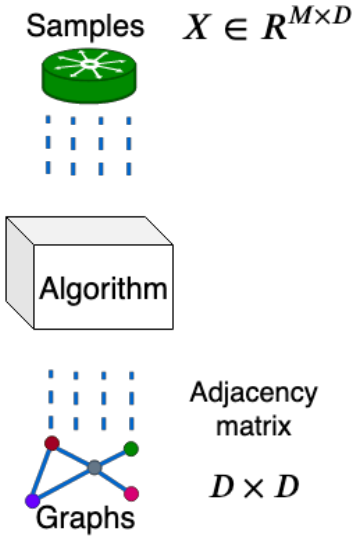
tGLAD: A sparse graph recovery based approach for multivariate time series segmentation

- ***Shima Imani, Harsh Shrivastava***

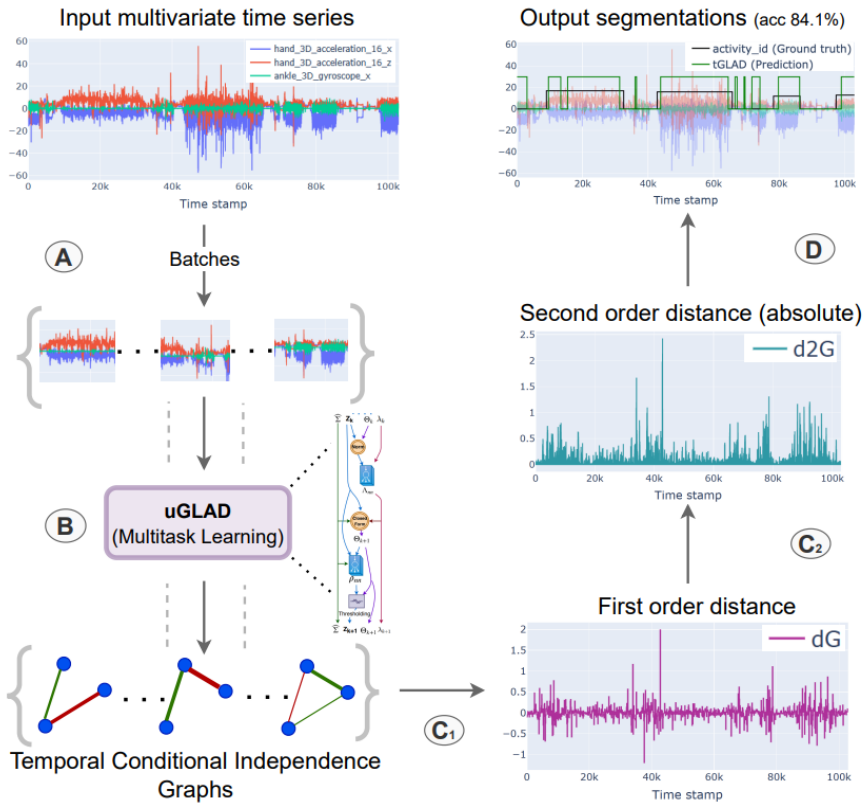
Microsoft Research

Conditional Independence Graphs

- Most formulations assume a multivariate Gaussian distribution
  - Can approximate other distributions
- Learns an undirected probabilistic graphical model, with edges corresponding to positive (green) and negative (red) correlation between variables
  - No edge between X and Y implies X and Y are conditionally independent given other variables



uGLAD: An Unsupervised deep unfolding based NN model



tGLAD framework. (**A**) The time series is divided into multiple intervals by using a sliding window to create a batch of intervals. (**B**) Run a single uGLAD model in multitask learning (or batch) mode setting to recover a CI graph for every input batch. This gives a corresponding set of temporal CI graphs. The entire input is processed in a single step as opposed to obtaining a CI graph for each interval individually. ($\mathbf{C}_1$) Get the first order distance, dG sequence, of the temporal CI graphs which captures the distance between the consecutive graphs. This is supposed to give higher values at the segmentation points. ($\mathbf{C}_2$) Again take a first order distance of the sequence in the previous step and then its absolute value to get $d2G$ sequence, which further accentuates the values at the segmentation points. (**D**) Apply a threshold to zero out the smaller values of $d2G$ and identify the segmentation blocks using an 'Allocation' algorithm.

Electricity Load and Peak Forecasting: Feature Engineering, Probabilistic LightGBM and Temporal Hierarchies

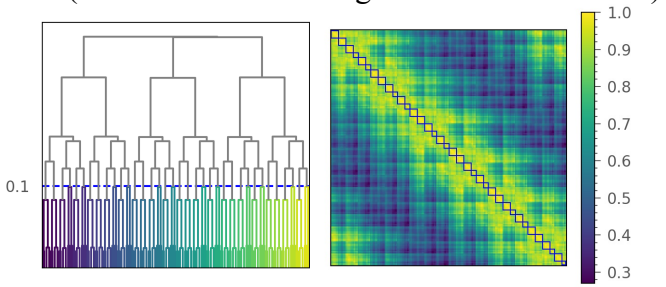
Nicolò Rubattu[1] (nicolo.rubattu@idsia.ch), Gabriele Maroni[1], Giorgio Corani[1]
[1] Dalle Molle Institute for Artificial Intelligence (IDSIA), USI-SUPSI, Lugano, Switzerland

We competed in the BigDEAL Challenge 2022, a global competition of *energy load* and *peak* forecasting.

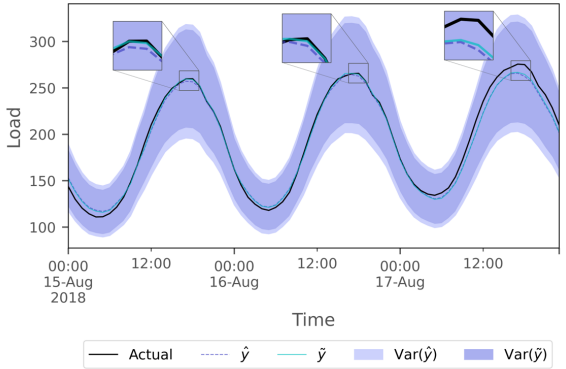
Our solution placed *3rd* in the qualifying match, and *6th* in the final match (out of 78 teams, 100+ partecipants).

1. We set a **regression** problem: $Load_t = f(X_t, \beta) + \epsilon_t$;
2. We derive a a large set of **features**:
   $X_t = \{temperatures,\ lags,\ rolling\ statistics,\ calendars,\ signal\ processing, \dots\}$;
3. We propose the *Clustered Permutation Feature Importance* (**CPFI**) method for *feature selection* and model *interpretability*;
4. We adopt Gradient Boosting (**GB**) of trees with trend modeling, *Dropout* and *distributional forecasts*;
5. We implement an approach to forecast combination known as **temporal hierarchies**, which further improves the accuracy.

CPFI (Hierarchical clustering on correlation matrix)



Reconciled distributional forecasts (LightGBM-LSS)



Full article:

VistaMilk

A World Leading SFI Research Centre

Science Foundation Ireland SFI For what's next

Do Cows Have Fingerprints?

Using Time Series Techniques and Milk Flow Profiles to Characterise Cow Behaviours and Detect Health Issues.

Changhong Jin[1] (changhong.jin@ucdconnect.ie)

John Upton (Teagasc)[2] , Brian MacNamee (University College Dublin)[1]

HOST INSTITUTION

Teagasc Agriculture and Food Development Authority

PARTNER INSTITUTIONS

DCU icbf Insight Ollscoil na Gaillimhe University of Galway Tyndall UCD UCC Walton SETU

FUNDED BY: An Roinn Talmhaíochta, Bia agus Mara Department of Agriculture, Food and the Marine; Science Foundation Ireland SFI For what's next

Motivation

- Milking cows generates unique milk flow profiles that offer high-frequency data for each cow.
- Such continuous data flow provides valuable insights into the milking performance, acting as a *‘fingerprint’* characterise cows in a herd.
- Milk flow profiles can be utilized for mastitis detection, framing the task as a time series classification problem.

Impact

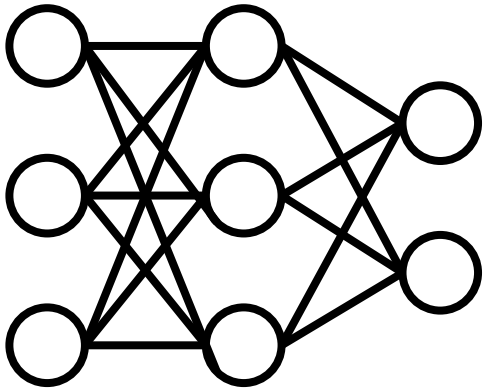
- Paper introduces innovative application of milk flow profiles for dual purposes: assessing milking performance and monitoring health issues.
- Adoption of machine learning techniques has potential to optimize data-driven decision-making in dairy farming, contributing to livestock well-being and consistent milk production.

Temporal Performance Prediction for Deep Convolutional Long Short-Term Memory Networks

Laura Fieback[1], Bidya Binayam Dash[1], Jakob Spiegelberg[1], Hanno Gottschalk[2]
[1]Volkswagen AG, Wolfsburg, Germany, [2]TU Berlin, Berlin, Germany

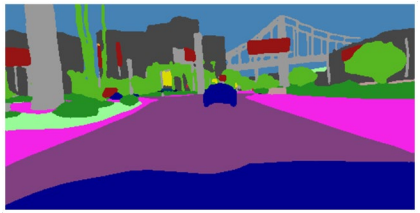
ConvLSTM
Video Frame Prediction Model

ConvLSTM Video Frame Prediction Model

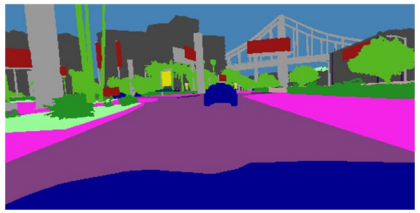


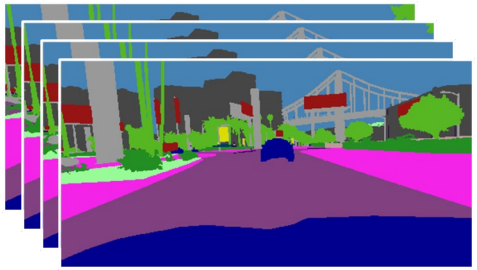
$argmax(softmax)$

Output $\hat{x}_t$



Ground Truth x_t

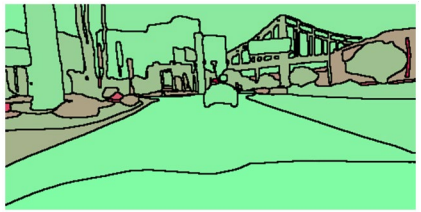




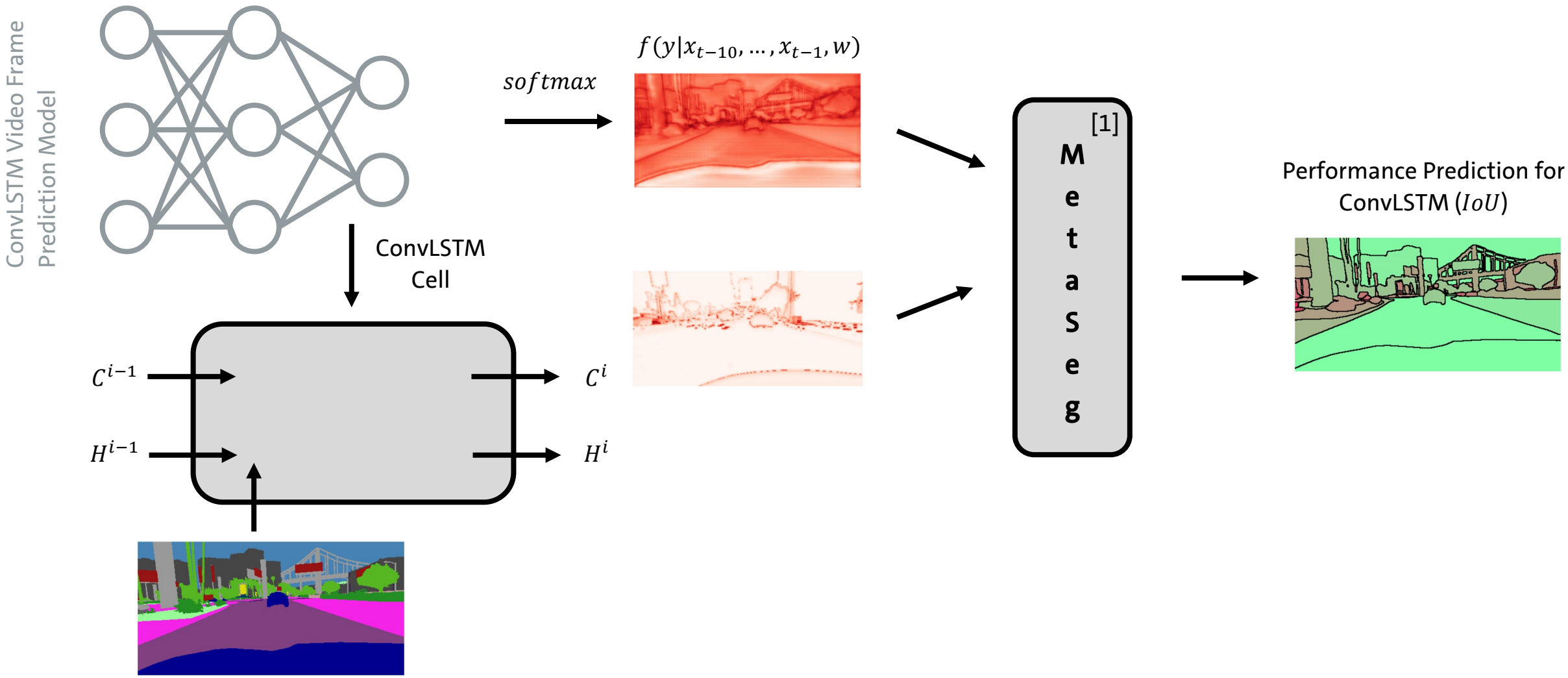
Input Sequence

$x_{t-10}, \dots, x_{t-1}$

Performance Measure for ConvLSTM: IoU



Temporal Performance Prediction in terms of Intersection over Union (IoU) for Deep Convolutional Long Short-Term Memory Networks



[1] M. Rottmann et al., "Prediction Error Meta Classification in Semantic Segmentation: Detection via Aggregated Dispersion Measures of Softmax Probabilities," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-9, doi: 10.1109/IJCNN48605.2020.9206659.

References

[1] M. Rottmann et al., "Prediction Error Meta Classification in Semantic Segmentation: Detection via Aggregated Dispersion Measures of Softmax Probabilities," 2020 International Joint Conference on Neural Networks (IJCNN), Glasgow, UK, 2020, pp. 1-9, doi: 10.1109/IJCNN48605.2020.9206659.

Contact

If you'd like to have a discussion on this topic, feel free to reach out!

laura.fieback@volkswagen.de

VOLKSWAGEN

AKTIENGESELLSCHAFT

Thank you!

Designing a New Search Space for Multivariate Time-Series Neural Architecture Search

Christopher MacKinnon
christopher.mackinnon@strath.ac.uk
University of Strathclyde, Glasgow, Scotland

Dr. Robert Atkinson
robert.atkinson@strath.ac.uk
University of Strathclyde, Glasgow, Scotland

University of Strathclyde Engineering

What is Neural Architecture Search?

The goal of Neural Architecture Search (NAS) is to find the best neural network architecture for a specific domain task. Machine learning methods optimise parameters based on the training loss whereas NAS optimises the architecture of a neural network with respect to the validation loss. The set of possible models over which the optimisation occurs is known as the search space.

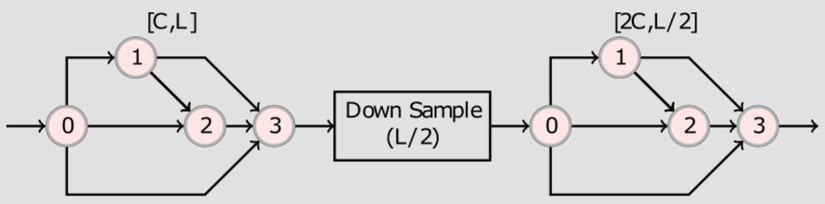
Motivation

- Deep learning has shown promising results in time-series classification tasks - particularly with multivariate data - where InceptionTime, ResNet and Transformers have found success
- The wide variety characteristics and signal lengths in time-series data makes designing a 'one-size-fits-all' architecture a challenging prospect
- Neural Architecture Search (NAS) has proven success in image classification outperforming human designed architectures
- Heuristics guiding search space design for image classifiers might not suit time-series classification; a space tailored to time-series could lead to better models

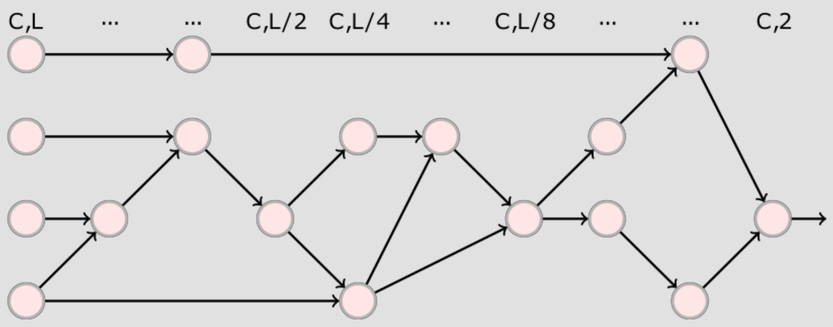
Method

A Neural Network Search-Space for Time-Series

Search space design in image classifiers focuses on deep repeating structures to extract complex features, we propose a space the produces a large set of interconnected representations while being flexible to find the optimal location and quantity for down-sampling operations.



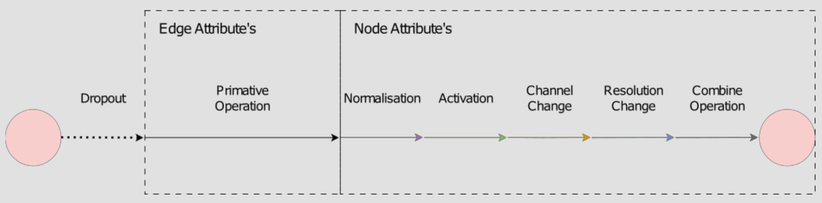
Cell-Based Model Topology



New Graph Model Topology

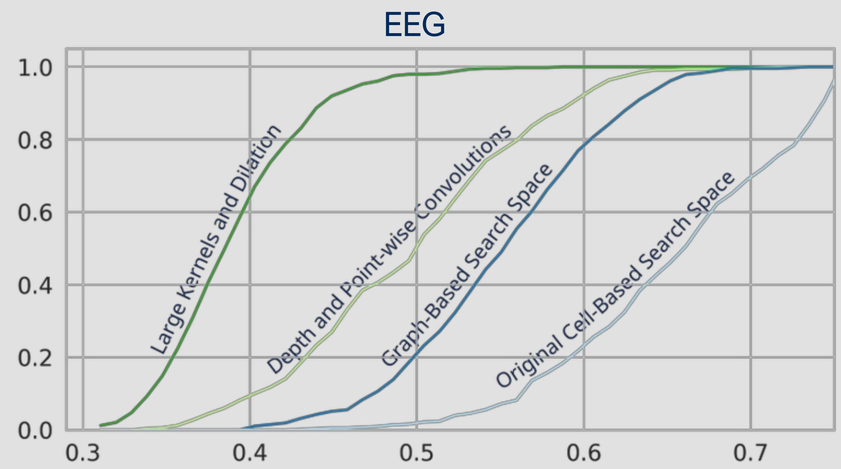
Dynamic Signal Down-Sampling and Node Attributes

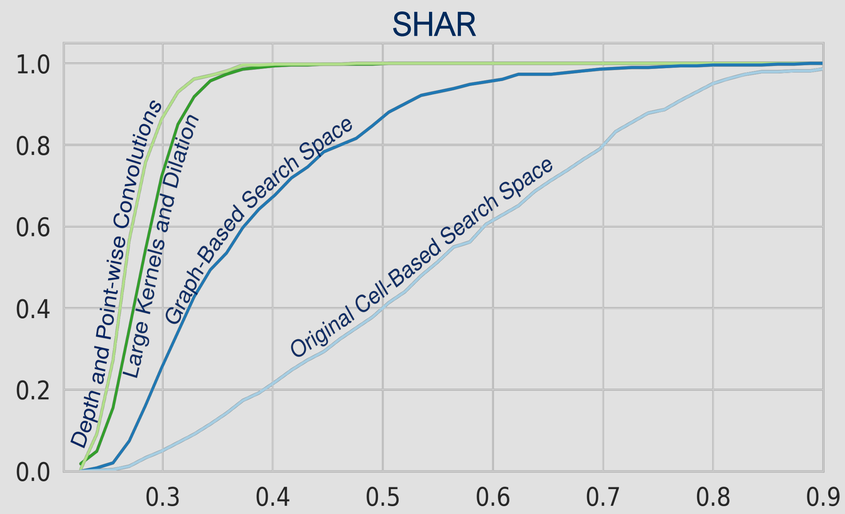
Rather than simply searching for the operation on an edge as in traditional NAS, we also search for the normalisation, activation and down-sampling as part of the search. These traits are associated with the node in the graph and are inherited by all edges which terminate there.



Evaluating and Improving the Search Space

To evaluate and compare the search spaces we randomly sample 500 architecture from each version of the search space and plot the cumulative probability of the validation error.





Breaking out of the Separable Convolution

By using the depth and point-wise convolution as our basic operations rather than the separable convolution, we further reduce the parameters

Larger Kernels and Dilation

Larger kernel sizes and dilation benefit time-series classification, as evidenced by WaveNet and ROCKET. Our model uses convolution kernels of sizes 16, 32, and 64 to capture long range dependencies. Results vary: the longer signal length EEG dataset improves notably, but the SHAR dataset slightly declines, likely because the shorter signals are already effectively

Datasets & Results

UniMib-SHAR: [1] A human activity recognition benchmark featuring daily activities like standing, walking, and sitting, plus fall incidents, recorded via waist-worn smartphones.

Method	Subject-Based Split	Random Split	Parameters
Gao, Zhang, Teng, et al. 2021[2]	-	79.03	2.40M
Mukherjee, Mondal, Singh, et al. 2020[3]	-	92.60	-
Al-qaness, Dahou, Elaziz, et al. 2023[4]	77.29	84.99	2.40M
Helmi, Al-qaness, Dahou, et al. 2023[5]	-	86.08	-
Teng, Wang, Zhang, et al. 2020[6]	-	78.07	0.55M
New Search Space	**77.63**	**95.70**	0.10M

BCI Competition IV 2a: [7] A prominent EEG dataset with recordings from 9 subjects performing 4 motor imagery tasks: left hand, right hand, both feet, and tongue

Method	Random	Parameters
ResNet	35.89	0.95M
InceptionTime (32 Channels)	49.08	0.48M
InceptionTime (64 Channels)	50.22	1.89M
New Search Space	**66.98**	0.12M

UCR Time-Series Archive: Finally we compare our approach with the SOTA [8,9] on the 4 largest multi-variate problems of equal length.

Method	FaceDetection	LSST	PenDigits	PhonemeSpectra	Average Rank
HC2	71.35	**63.70**	99.56	29.43	3.00
ROCKET	69.38	61.85	99.57	27.03	4.25
HC1	69.17	53.84	97.19	32.87	4.25
ResNet	62.97	42.94	99.64	30.86	5.25
InceptionTime	**77.24**	33.97	**99.68**	**36.74**	2.50
TapNet	52.87	46.33	93.65	-	**6.33**
New Search Space	75.01	**63.68**	99.60	29.84	2.75

Conclusion

We introduce a search space specifically designed for time-series classification tasks where, without the use of an advanced search algorithm, achieves competitive results compared with the SOTA while also producing highly efficient architectures with fewer parameters than other deep learning
approaches. We set a benchmark for further work showing that with a well designed search space NAS has strong potential as a time-series classification approach.

References

[1] D. Micucci, M. Mobilio, and P. Napoletano, "Unimib shar: A dataset for human activity recognition using acceleration data from smartphones,"Applied Sciences, vol. 7, no. 10, 2017, issn: 2076-3417.

[2] W. Gao, L. Zhang, Q. Teng, J. He, and H. Wu, "Danhar: Dual attention network for multimodal human activity recognition using wearable sensors," Applied Soft Computing, vol. 111, p. 107 728, 2021.

[3] D. Mukherjee, R. Mondal, P. K. Singh, R. Sarkar, and D. Bhattacharjee,"Ensemconvnet: A deep learning approach for human activity recognitionusing smartphone sensors for healthcare applications," Multimedia Toolsand Applications, vol. 79, pp. 31 663–31 690, 2020.

[4] M. A. A. Al-qaness, A. Dahou, M. A. Elaziz, and A. M. Helmi, "Multi-resatt: Multilevel residual network with attention for human activity recog-nition using wearable sensors," IEEE Transactions on Industrial Informatics, vol. 19, no. 1, pp. 144–152, 2023. doi: 10.1109/TII.2022.3165875.

[5] A. M. Helmi, M. A. Al-qaness, A. Dahou, and M. Abd Elaziz, "Humanactivity recognition using marine predators algorithm with deep learning,"Future Generation Computer Systems, vol. 142, pp. 340–350, 2023.

[6] Q. Teng, K. Wang, L. Zhang, and J. He, "The layer-wise training convolutional neural networks using local loss for sensor-based human activity recognition," IEEE Sensors Journal, vol. 20, no. 13, pp. 7265–7274, 2020

[7] C. Brunner, R. Leeb, G. Müller-Putz, A. Schlögl, and G. Pfurtscheller,"Bci competition 2008–graz data set a," Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, vol. 16, pp. 1–6, 2008.

[8] A. P. Ruiz, M. Flynn, J. Large, M. Middlehurst, and A. Bagnall, "The great multivariate time series classification bake off: A review and experimental evaluation of recent algorithmic advances," Data Mining and Knowledge Discovery, vol. 35, no. 2, pp. 401–449, 2021

[9] M. Middlehurst, J. Large, M. Flynn, J. Lines, A. Bostrom, and A. Bagnall, "Hive-cote 2.0: A new meta ensemble for time series classification,"Machine Learning, vol. 110, no. 11-12, pp. 3211–3243, 2021.

VistaMilk

A World Leading SFI Research Centre

Science Foundation Ireland For what's next

Back to Basics: A Sanity Check on Modern Time Series Classification Algorithms

Bhaskar Dhariyal, Thach Le Nguyen, Georgiana Ifrim
School of Computer Science, University College Dublin
VistaMilk SFI Research Centre, Ireland

bhaskar.dhariyal@ucdconnect.ie

HOST INSTITUTION
Teagasc
Agriculture and Food Development Authority

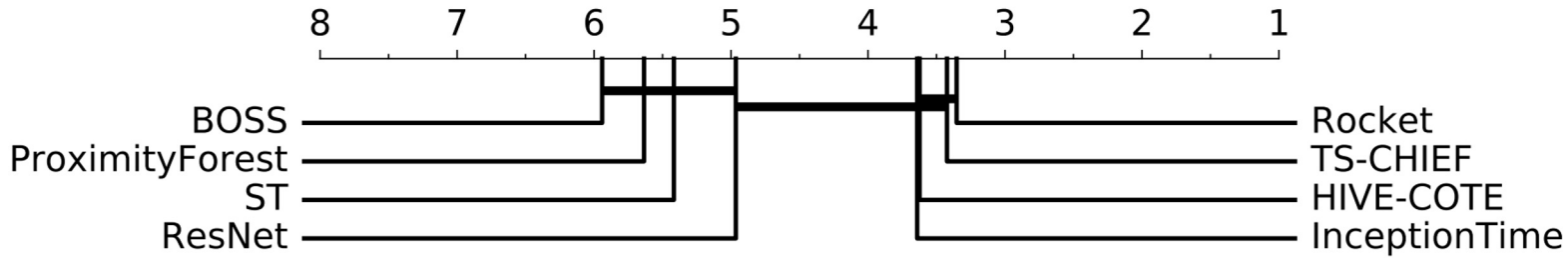
PARTNER INSTITUTIONS
DCU icbf Insight Ollscoil na Gaillimhe University of Galway Tyndall UCD UCC Walton SETU

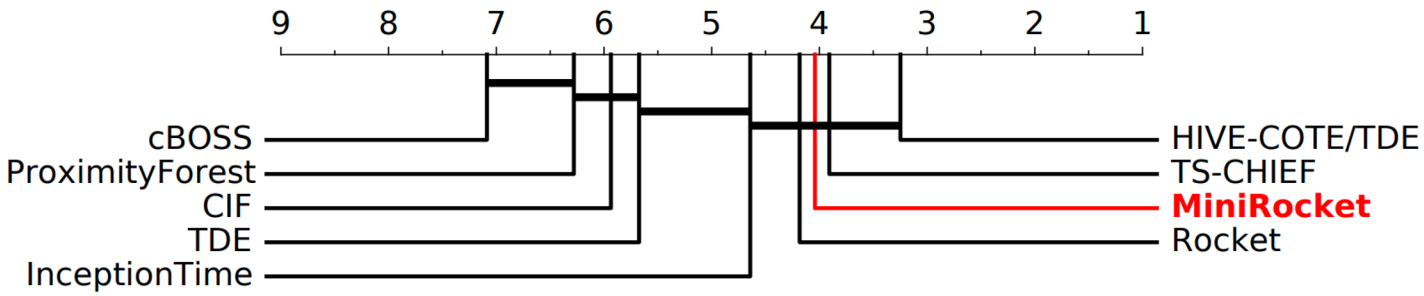
FUNDED BY:
An Roinn Talmhaíochta, Bia agus Mara Department of Agriculture, Food and the Marine
Science Foundation Ireland For what's next

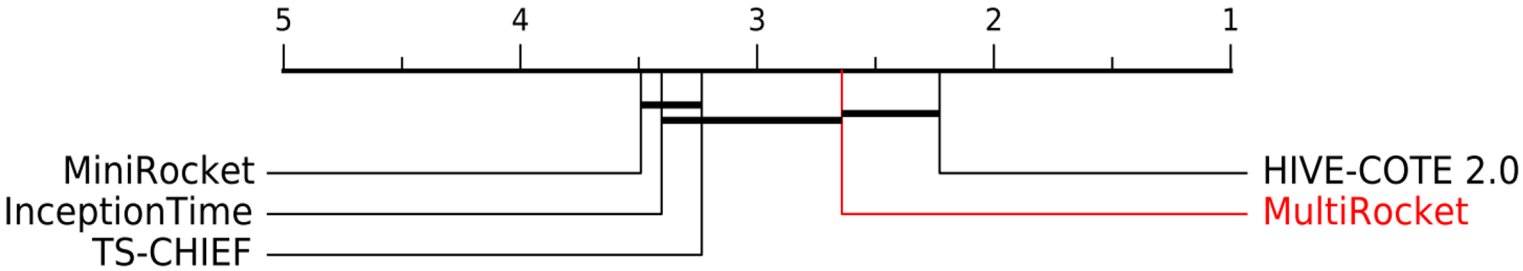
Digitalising Dairy

VistaMilk

Motivation


8
7
6
5
4
3
2
1
BOSS
ProximityForest
ST
ResNet
Rocket
TS-CHIEF
HIVE-COTE
InceptionTime


9
8
7
6
5
4
3
2
1
cBOSS
ProximityForest
CIF
TDE
InceptionTime
HIVE-COTE/TDE
TS-CHIEF
MiniRocket
Rocket


5
4
3
2
1
MiniRocket
InceptionTime
TS-CHIEF
HIVE-COTE 2.0
MultiRocket

Motivation

Highlight the importance of conducting baseline check with Tabular models.

UCR/UEA Benchmark: Univariate Time Series Classification (109 Datasets)

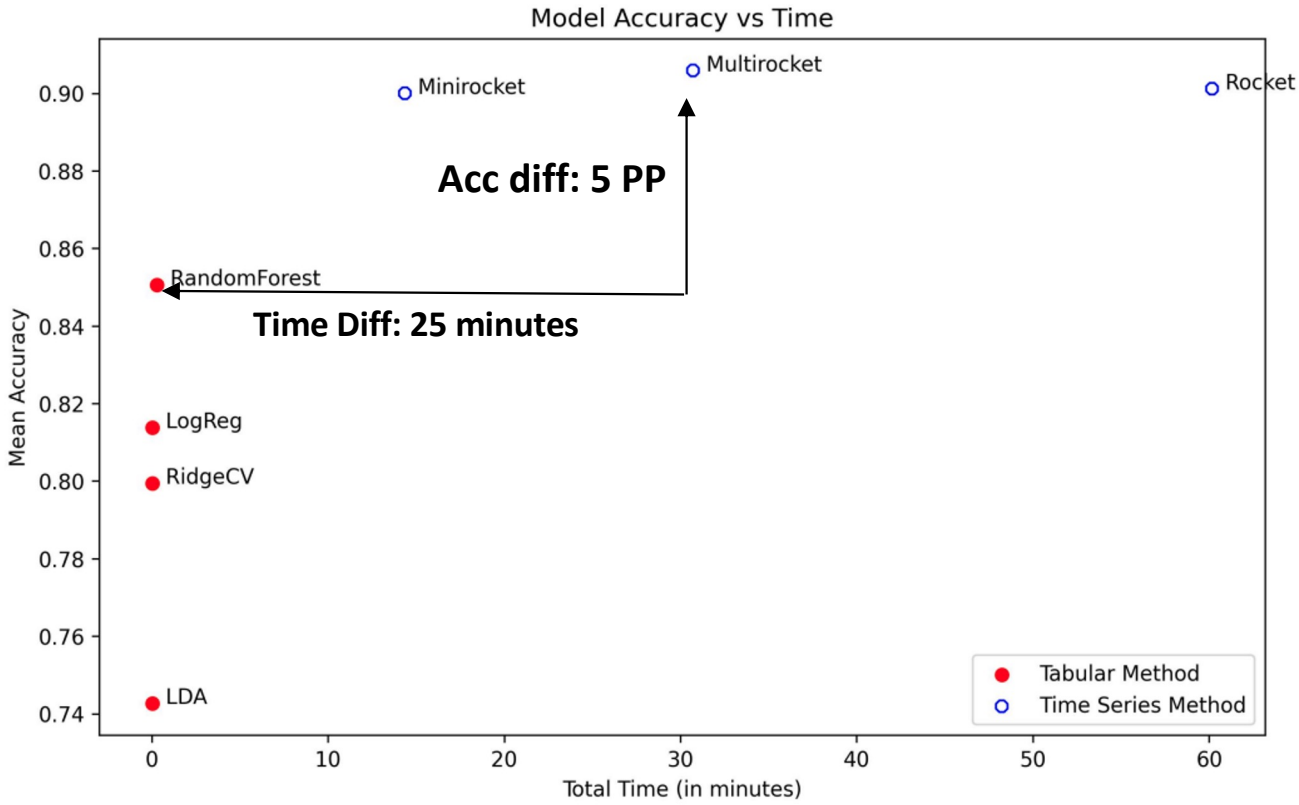
Tabular Model Better (19.2 % datasets)

Model comparable (31.1 % datasets)

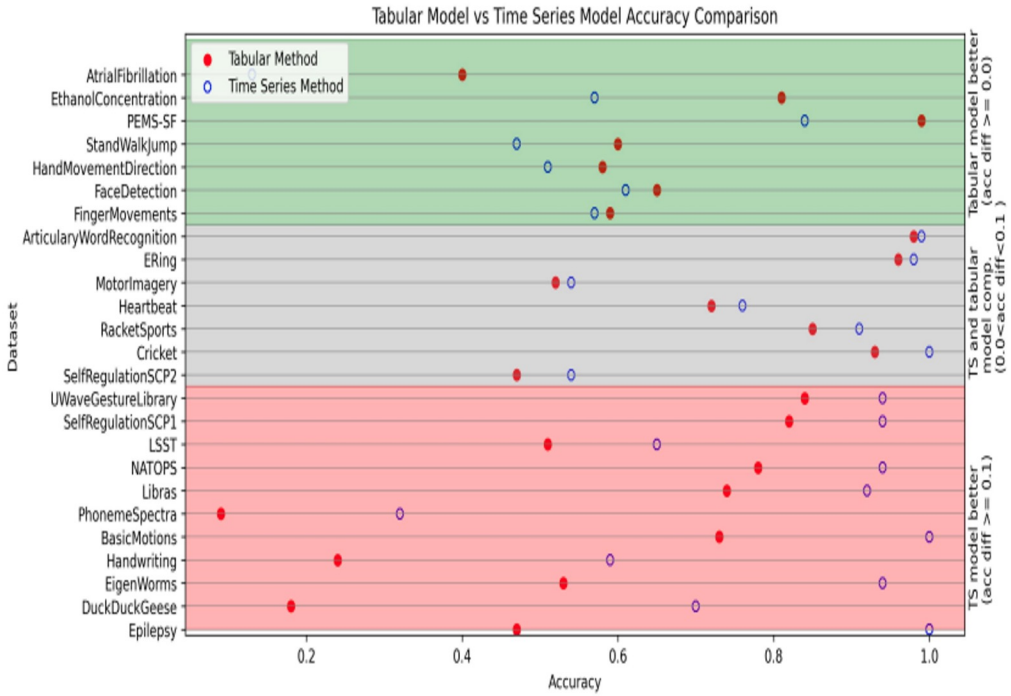
Time Series Model Better (49.5% datasets)

UCR/UEA Benchmark: Univariate Time Series Classification (109 Datasets)

Model comparable (31.1 % datasets)



UCR/UEA Benchmark: Multivariate Time Series Classification (25 Datasets)



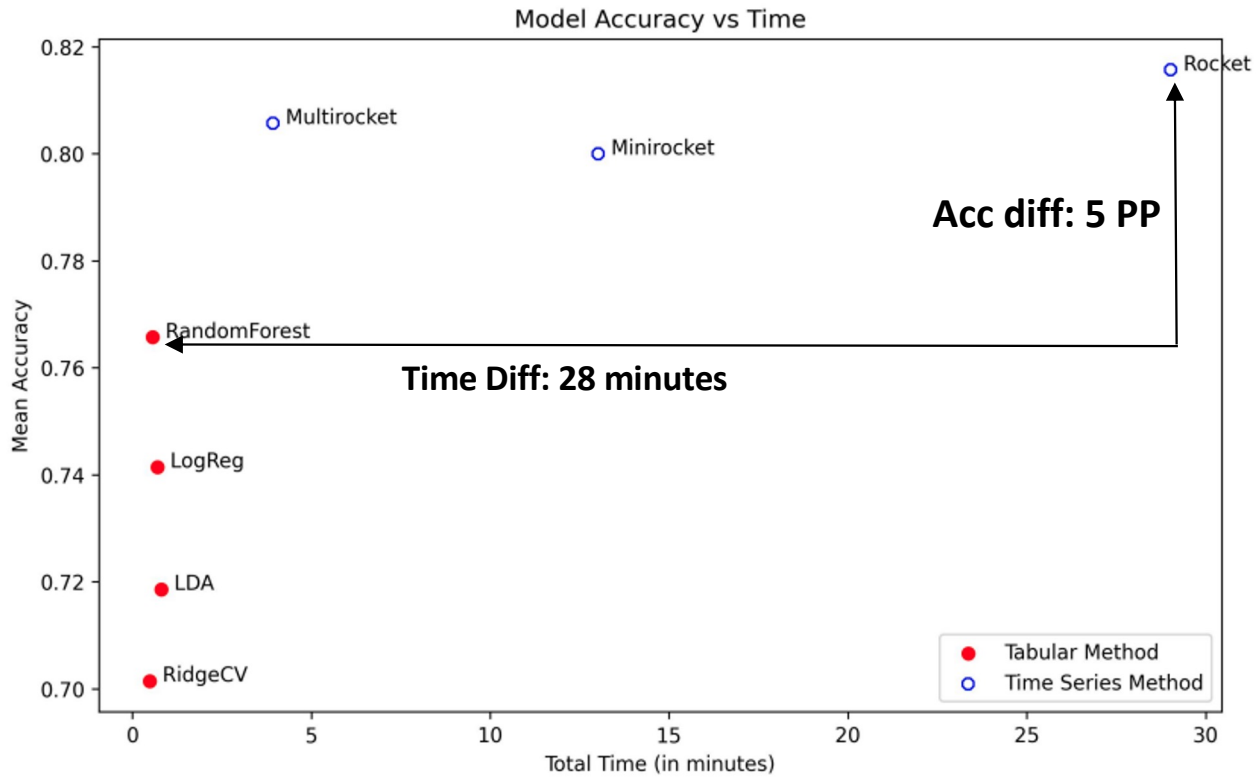
Tabular Model Better (28 % datasets)

Model comparable (28 % datasets)

Time Series Model Better (44% datasets)

UCR/UEA Benchmark: Multivariate Time Series Classification (25 Datasets)

Model comparable (28 % datasets)



Take Away

- **Baselines:** Simple tabular models perform significantly well on some datasets
- **TSC Benchmarks:** Some datasets included in the UEA/UCR archive perform much better with tabular methods (e.g., Spectroscopy). Criteria for inclusion of datasets into TSC benchmarks?
- **Trade-offs:** Significant number of datasets (grey region) TSC methods and tabular methods have similar accuracy, but tabular methods are significantly faster.

Thank You!

Details & code

https://github.com/mlgig/TabularModelsforTSC

Centre for Research Training

Science Foundation Ireland For what's next

Evaluating Explanation Methods for Multivariate Time Series Classification

Davide Serramazza, Thu Trang Nguyen, Thach Le Nguyen, Georgiana Ifrim
davide.serramazza@ucdconnect.ie

School of Computer Science, University College Dublin, Ireland

AALTD23
18/09/2023

HOST INSTITUTIONS

DCU Dublin City University

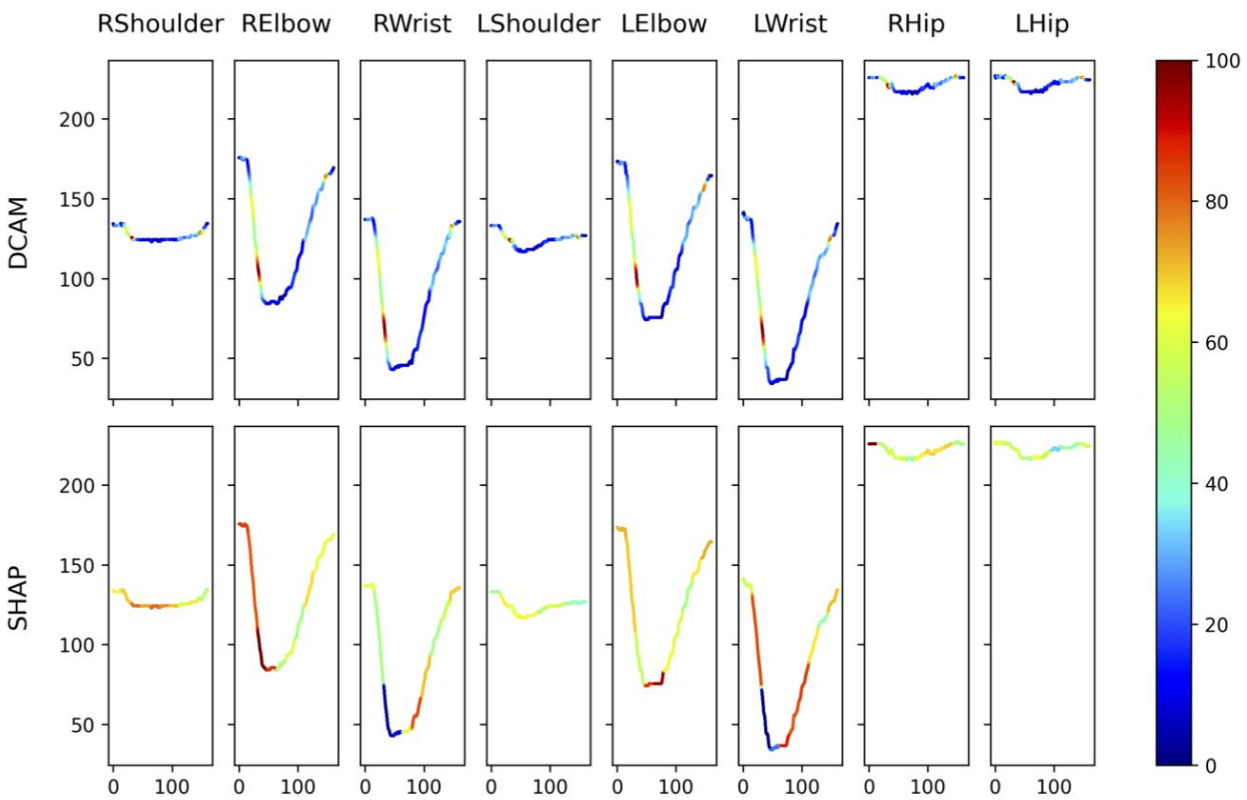
TU DUBLIN

UCD DUBLIN University College Dublin An Coláiste Ollscoile, Baile Átha Cliath

FUNDED BY: Science Foundation Ireland For what's next

Motivation

- Recent research focuses on MTS Classification, less work on **MTSC explanation methods,** i.e., providing a **2D heat map with salient data points**
- Given the same dataset and classifier, two explanation methods may disagree



State-of-the-Art

- Few **bespoke** explanation methods, tailored to deep learning classifiers (dCAM)
- Adaptations of univariate classification/explanation methods (SHAP)

HOST INSTITUTIONS
University College Dublin
An Coláiste Ollscoile, Baile Átha Cliath
FUNDED BY:

Our contribution

On 3 synthetic datasets and 2 real-world ones, we compared using a novel evaluation methodology (AMEE):

- Bespoke MTSC explanation: dRESNET + dCAM
- Univariate Adaptation: ROCKET+ SHAP
- Baseline classifier/explanation: RidgeCV
- Random explanation

ROCKET-SHAP works best among the compared explanations

If you want to know more you are welcome to our poster!

Thanks for the attention

Centre for Research Training
Science Foundation Ireland For what's next
HOST INSTITUTIONS
DCU
TU DUBLIN
UCD DUBLIN University College Dublin An Coláiste Ollscoile, Baile Átha Cliath
FUNDED BY:
Science Foundation Ireland For what's next

Time-aware Predictions of Moments of Change in Longitudinal User Posts on Social Media

Anthony Hills[1], Adam Tsakalidis[1,2], Maria Liakata[1,2,3]

[1] Queen Mary University of London [2] The Alan Turing Institute [3] University of Warwick

Introduction

- **Objective**: Predict Moments of Change in mood (MoCs) in longitudinal user posts.
- ***MoCs*** are points in time (posts) denoting a [1]:
  - **Switch**: a sudden shift in an individual's mood from negative-to-positive or vice versa;
  - **Escalation**: a gradual mood change.
- **Our work**:
  - A time-aware approach for modelling textual user posts, by transforming the LSTM hidden states over previous/future posts with self-excitation and exponential decay that varies with time.
  - We extend our approach to the bi-directional setting, outperforming non-time-aware baselines, and all teams from the CLPsych 2022 shared task [2].
  - We demonstrate the effectiveness of our approach, in an ablation study investigating (1) time-aware features and (2) bi-directionality.

Background

Hawkes Process

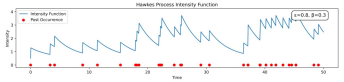
A self-exciting temporal point process, used to model sequences of events where each event can increase the probability of future events [3].

Given a set of event times $\{t_1, t_2, \ldots, t_n\}$, the Hawkes process intensity function at time t is given by:

$$\lambda(t) = \mu + \sum_{t_i < t} \epsilon \exp(-\beta(t - t_i))$$

Where:

- μ is the background intensity.
- ϵ is the excitement factor that effects the increase in intensity due to a prior event occurring.
- β is the time-decay parameter.



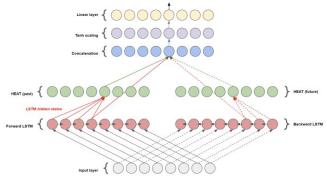
Proposed Architecture

HEAT

Encodes the dynamics of historical post representations, v, in a time-aware manner [4].

$$v_{\text{HEAT}}^{(i)} = \sum_{j:\Delta\tau_j > 0} v^{(j)} + \epsilon e^{-\beta \Delta\tau_j} \max(v^{(j)}, 0)$$

where $\Delta\tau_j = t^{(i)} - t^{(j)}$, and ϵ and β are fixed hyper-parameters reflecting the amount of self-excitation and exponential time-decay to apply to each post respectively.



We apply HEAT over BiLSTM hidden states, contrasting time-sensitive representations from the past and future to predict Moments of Change.

Datasets

We evaluate on two longitudinal datasets [1, 2] sourced from social media websites, consisting of timelines of user posts annotated with moments of change in mood:

	Reddit	TalkLife
Users	186	500
Timelines (posts)	255 (6,195)	500 (18,702)
Label distr. % (O/E/S)	77.6 / 15.8 / 6.6	84.5 / 10.8 / 4.7
Timeline Length	$\sim$ 2 months	$\leq$ 2 weeks

Results

Per-class and macro-averaged results on each dataset:

Reddit	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Majority	–	.333	.280	–	.000	.000	–	.000	.000	.724	**1.000**	.840
WResearch	.625	.579	.598	.362	.256	.300	.646	.553	.596	.868	.929	.897
UoS	.689	.625	.649	**.490**	.305	.376	.697	.630	.662	.881	.940	.909
BiLSTM-HEAT	**.706**	**.670**	**.686**	.475	**.415**	**.442**	**.741**	**.654**	**.694**	**.902**	.942	**.921**

TalkLife	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Majority	–	.333	.280	–	.000	.000	–	.000	.000	.845	**1.000**	.916
BERT(f)	.520	.554	.534	.260	**.321**	.287	.401	**.478**	.436	.898	.864	.881
BiLSTM-bert	**.621**	**.553**	**.580**	**.397**	.264	**.316**	**.568**	.461	**.508**	**.898**	.936	**.917**
BiLSTM-HEAT	.584	.552	.566	.329	.290	.308	.524	.448	.483	.897	.920	.908

Exploiting Context and Attention with Recurrent Neural Network for Sensor Time Series Prediction

Rashmi Dutta Baruah and Mario Muñoz Organero

Universidad Carlos III de Madrid, Leganes, Spain

www.claireproject.com

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

Context

Matters

Intelligent Transportation Systems: Predictive maintenance, Traffic prediction

Image credit: https://cps.es/en/transport-engineering/

Context: weather conditions and road conditions

Detecting events during oil well drilling

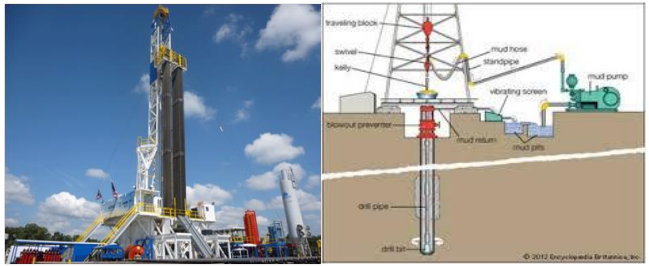
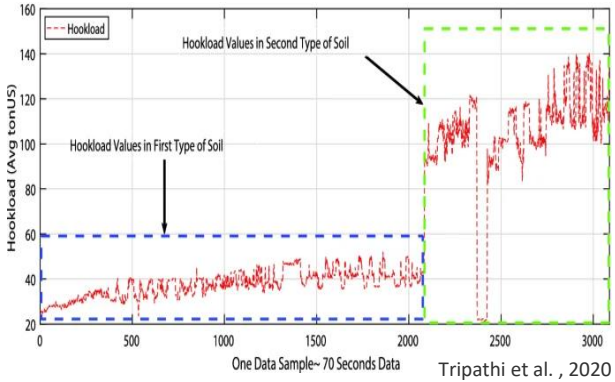


Image credit: http://www.oil-gasportal.com/drilling/introduction-to-oilgas-well-drilling/

Context: Soil type



Tripathi et al. , 2020

uc3m | Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Context

Handling

- In many sensor data applications while using sequential models (RNNs), often the contexts are ignored or incorporated as ordinary features in the model.
- We present an approach to leverage the contextual information and integrate to RNNs with attention.
- The proposed architecture uses the contextual features in two ways.
  - to weight the primary input features depending on the context
  - secondly to weight the hidden states to compute the attention.

uc3m | Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

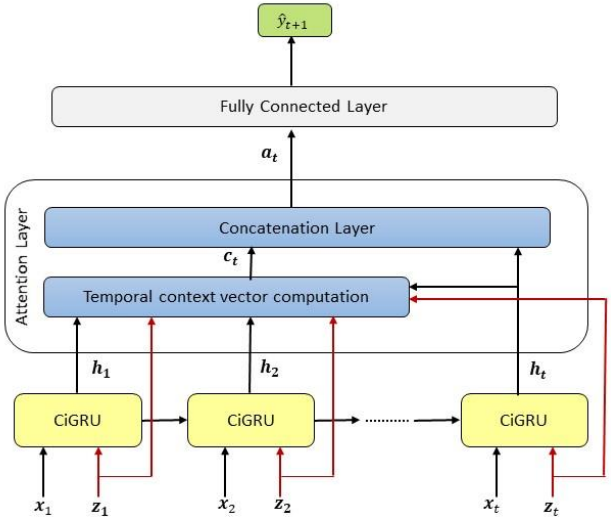
HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Proposed

Framework

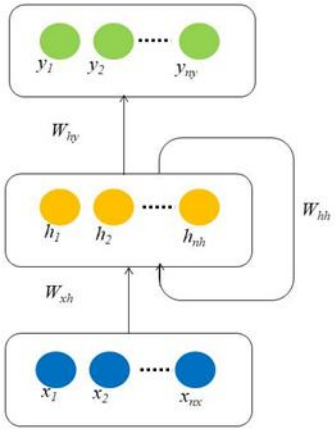
$$\alpha_{ti} = \frac{\exp(g(h_t, h_i)}{\sum_{i=1}^{t}(g(h_t, h_i)}, \quad g(h_t, h_i) = h_t^T W^a(z_t) h_i$$

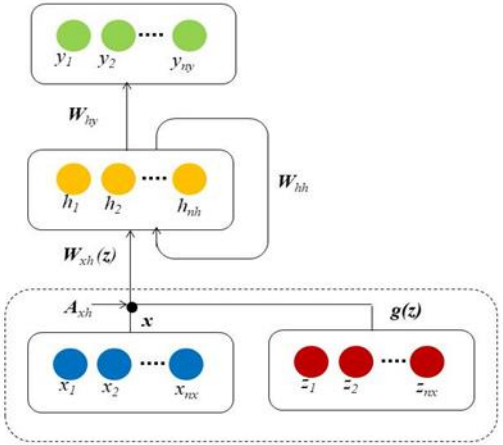


RNN vs. Context Integrated RNN (CiRNN)

$$h_t = f\left(W_{hx}^T x_t + \mathrm{W}_{\mathrm{hy}}^{\mathrm{T}} \mathrm{h}_{\mathrm{t-1}} + \mathrm{b}_{\mathrm{h}}\right) \quad (1)$$

$$h_t = f\left(W_{hx}^T(z_t) x_t + \mathrm{W}_{\mathrm{hy}}^{\mathrm{T}} \mathrm{h}_{\mathrm{t-1}} + \mathrm{b}_{\mathrm{h}}\right) \quad (2)$$





uc3m | Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

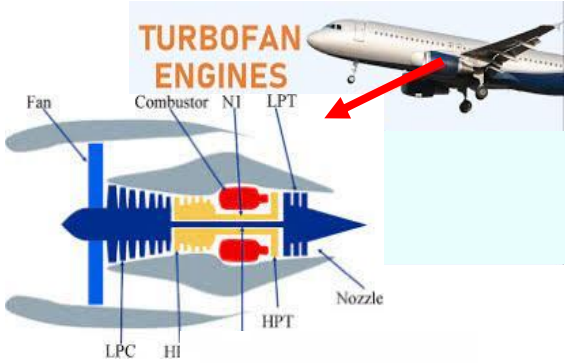
HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Dataset

Description

- Engine health prognostics- estimation of remaining useful life (NASA Turbofan Engine Degradation Simulation Data Set) (87750, 26)
  - 21 sensors (such as Total temperature at fan inlet, Total temperature at Low Pressure Compressor outlet)
  - 3 operational settings (flight altitude, Mach number, and throttle resolver angle
  - Dataset with 6 operating conditions is considered as it provides context
- Appliance energy consumption prediction (19735, 29)
  - Weather data (outside temperature, humidity, pressure etc.)
  - house temperature and humidity (temperature and humidity in kitchen area)
  - Appliances energy usage



uc3m Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

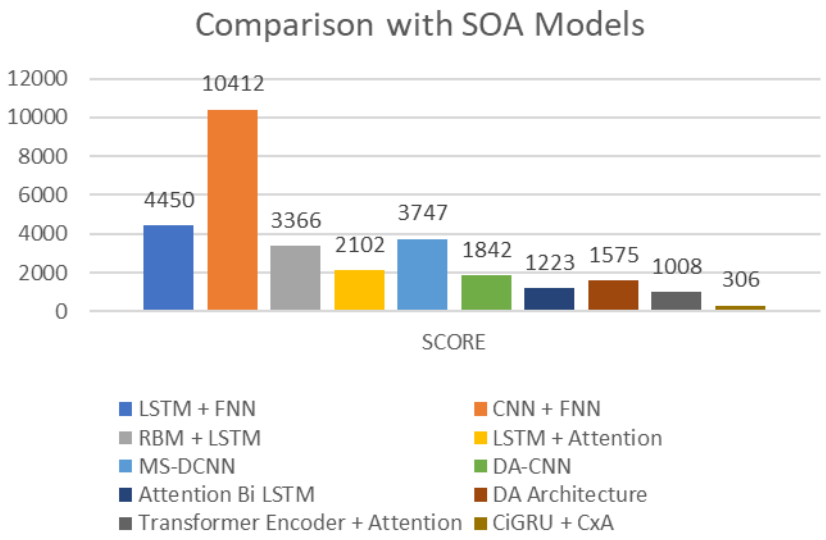
HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

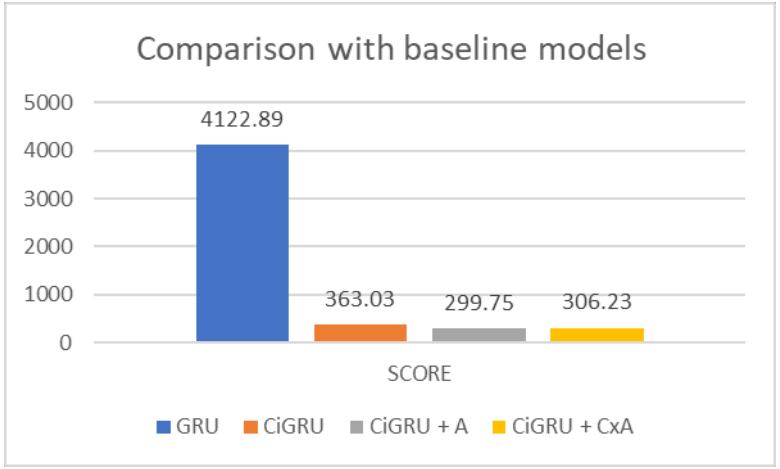
Experimental

Results

Dataset 1



decrease 25.41% and 69.62% in RMSE and score



Dataset 2- CiGRU +CxA performed better than baseline models.
Decrease in 1.46 % in RMSE when compared to best existing model

uc3m Universidad Carlos III de Madrid

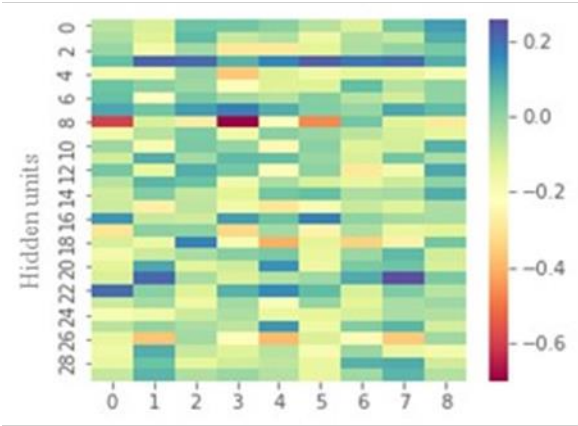
This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

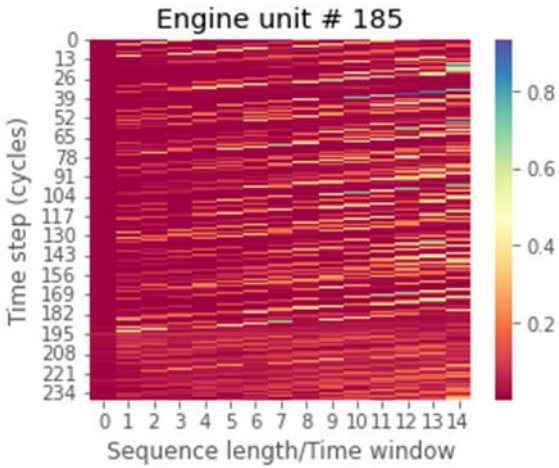
HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Experimental

Results





contextual weights associated with only one primary feature (demanded corrected fan speed), mostly positive values, indicate its influence on prediction of RUL

prediction at time steps 25 to 190 mainly relied on early as well as recent time windows (5-15), during the last time steps the network focuses at the last time window

uc3m | Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Conclusions

Future Work

- Experimental results from two benchmark datasets show that CiGRU (contextual weighting) with attention performs better than the contextual expansion approach.
- One limitation is increase in number of parameters due to introduction of context features and in turn increase in training time.
- Experiments are in progress in different domains (eg Traffic)
- Further research directions: context-based explanations, transformers with explicit contexts

uc3m | Universidad Carlos III de Madrid

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538

HR EXCELLENCE IN RESEARCH

MARIE CURIE ACTIONS

Thank

You

This project has received funding from Universidad Carlos III de Madrid and the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant Agreement No 801538