CHEN Ziyuan
CHEN Zhirong
28 July 2022

Autoencoded Quantification Of Unsupervised Representative Shapelets Net

AQOURS Net

CONTENTS

AQOURS Net

Problem Analysis

Methodology

Features

Experiment

SYMBOL TABLE.

S	Time series
t	Subsequences
v	Shapelets

We are following the convention of Time2Graph. Sorry for the counter-intuitiveness.

The phrase "t-v distance" will frequently appear.

AQOURS
Net

Problem Analysis

Methodology

Features

Experiment

The Time2Graph model is, essentially, a Finite State Machine.

with transition probabilities as weights.

constitute the nodes in a State Transition Diagram,

subsequence shapes

A finite number of

A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

Experiment

TRADITIONAL APPROACHES.

"We select the subsequences that perform well on classification tasks as potential shapelets."

Ye, 2009 (S)
Lines, 2012 (ST)
Rakthanmanon, 2013 (FS)
Wistuba, 2015 (UFS)
and almost everyone else before 2016

"We learn, from the series, shapelets that perform well on classification tasks."

Grabocka, 2014 (LS)
Ma, 2020 (ADSN)
etc.

Core idea: Good shapelets are good classifiers.

But running classification can be time-consuming.

Let's avoid this!

A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

Experiment

PROPOSAL #1.

Maybe "SELF-SUPERVISED" makes more sense……

UNSUPERVISED LABELLING

Map shapelets to subsequences.
Calculate t-v distances as labels u.

UNSUPERVISED TRAINING

Contrastively train an embedding network U to extract features.

SUPERVISED LABELLING

Feed t and v into U.
Use E.D. of outputs as labels v.

SUPERVISED TRAINING

Concatenate an MLP V after U.
Use the labels v for fine-tuning.

THE MODEL IS READY!

For each series, transfer its subsequences into possibility vectors and construct the FSM Diagram.

AQOURS Net

Problem Analysis

Methodology

Features

Experiment

PROBLEMS OF PROPOSAL #1.

At the beginning epochs, these labels for fine-tuning (v) come from the unsupervised model.

They can be unreliable!

SUPERVISED LABELLING

Feed t and v into U.
Use E.D. of <u>outputs</u> as labels v.

SUPERVISED TRAINING

Concatenate an MLP V after U.
Use the labels v for fine-tuning.

But turning to the classification task means **retreating** to the old path

A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

Experiment

SHAPELET-BASED MODEL vs. TIME2GRAPH

Finds "signature" subsequences.

Good shapelets should classify the time series accurately.

Uses shapelet transfer for embedding.

Good shapelets should express the subsequences accurately.

CONCLUSION.

"~~Selection by classification~~" is possibly not the optimal solution!

A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

Experiment

PROPOSAL #2.

AIM FOR REPRESENTATIVENESS INSTEAD OF ~~DISTINCTIVENESS~~.

Sounds rather subtle, huh……

THE MORE DIVERSE, THE MORE EXPRESSIVE.

Ma, 2020 (ADSN)
Li, 2021 (ShapeNet)
etc.

1. Use KMeans to cluster all the subsequences.
2. For each cluster, use the subsequence <u>closest to the center</u> as our "anchor".

PROPOSAL #2.

THE MORE
DIVERSE,
THE MORE
EXPRESSIVE.

1. Use KMeans to cluster all the subsequences.
2. For each cluster, use the subsequence closest to the center as our "anchor".

THE KMEANS FACTSHEET.

- KMeans is fast – it runs in O(n) time
- KMeans outperforms random picking Wistuba, 2015 (Ultra-fast) Karlsson, 2016 (Rnd. spl. forest)
- KMeans encourages representativeness
- KMeans does NOT always encourage diversity
  (Choosing centroids in the TS2Vec-embedded space would help)

Diagram of SimpleKMeans

FEATURES.
KMeans

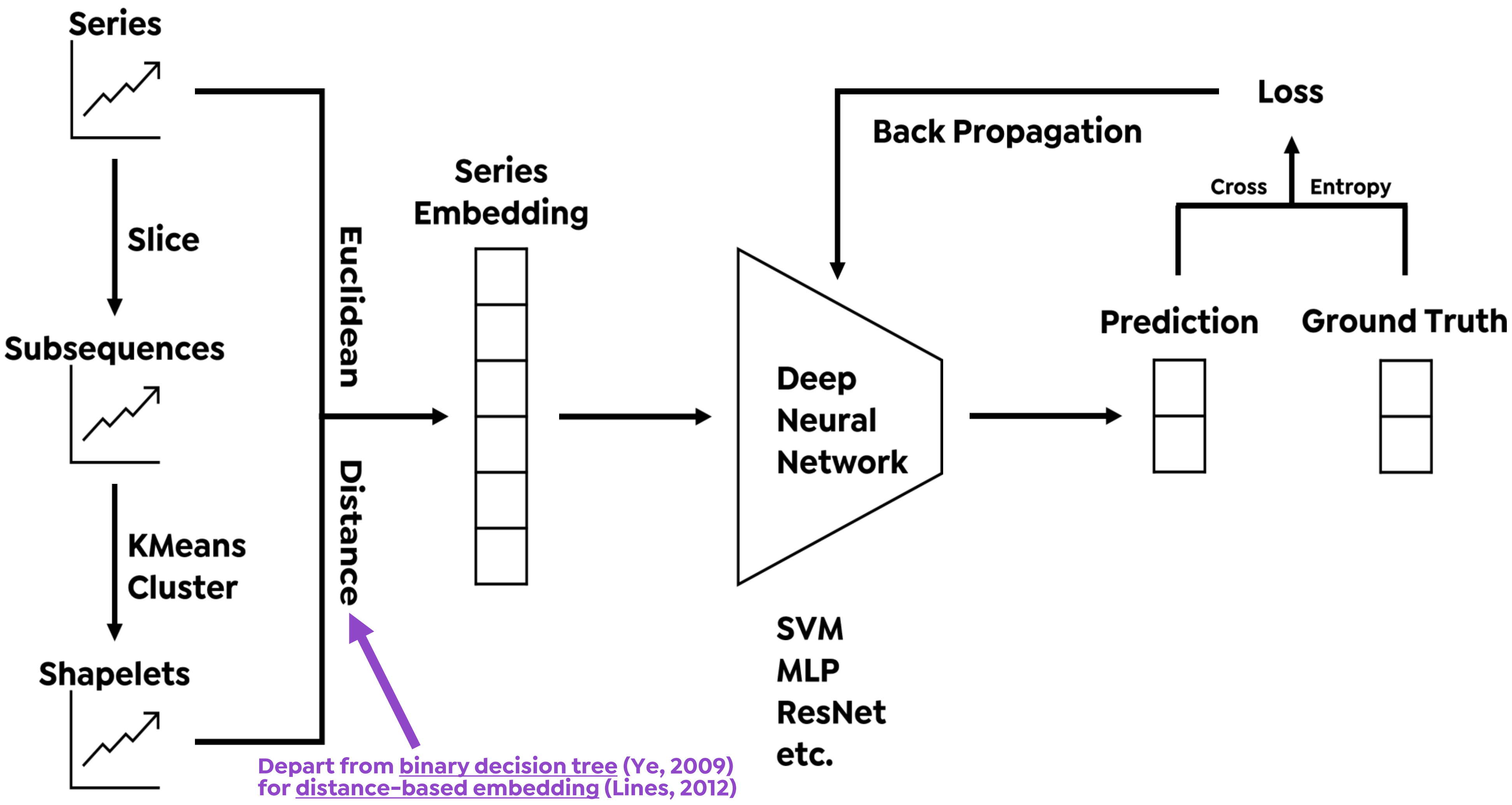
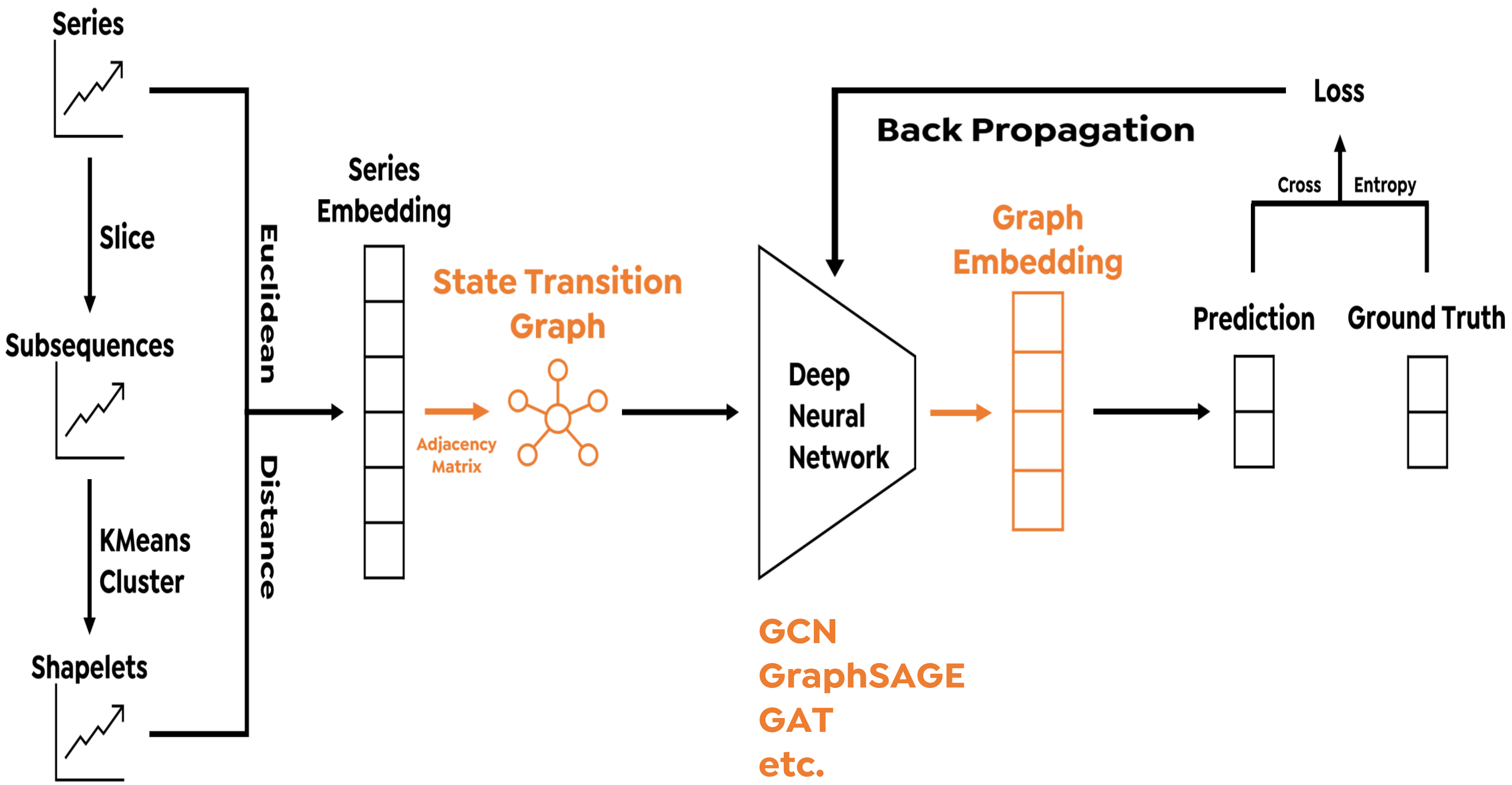


Diagram of Time2Graph

FEATURES.

KMeans

Veličković, 2017
Brody, 2021
GAT / GATv2



A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

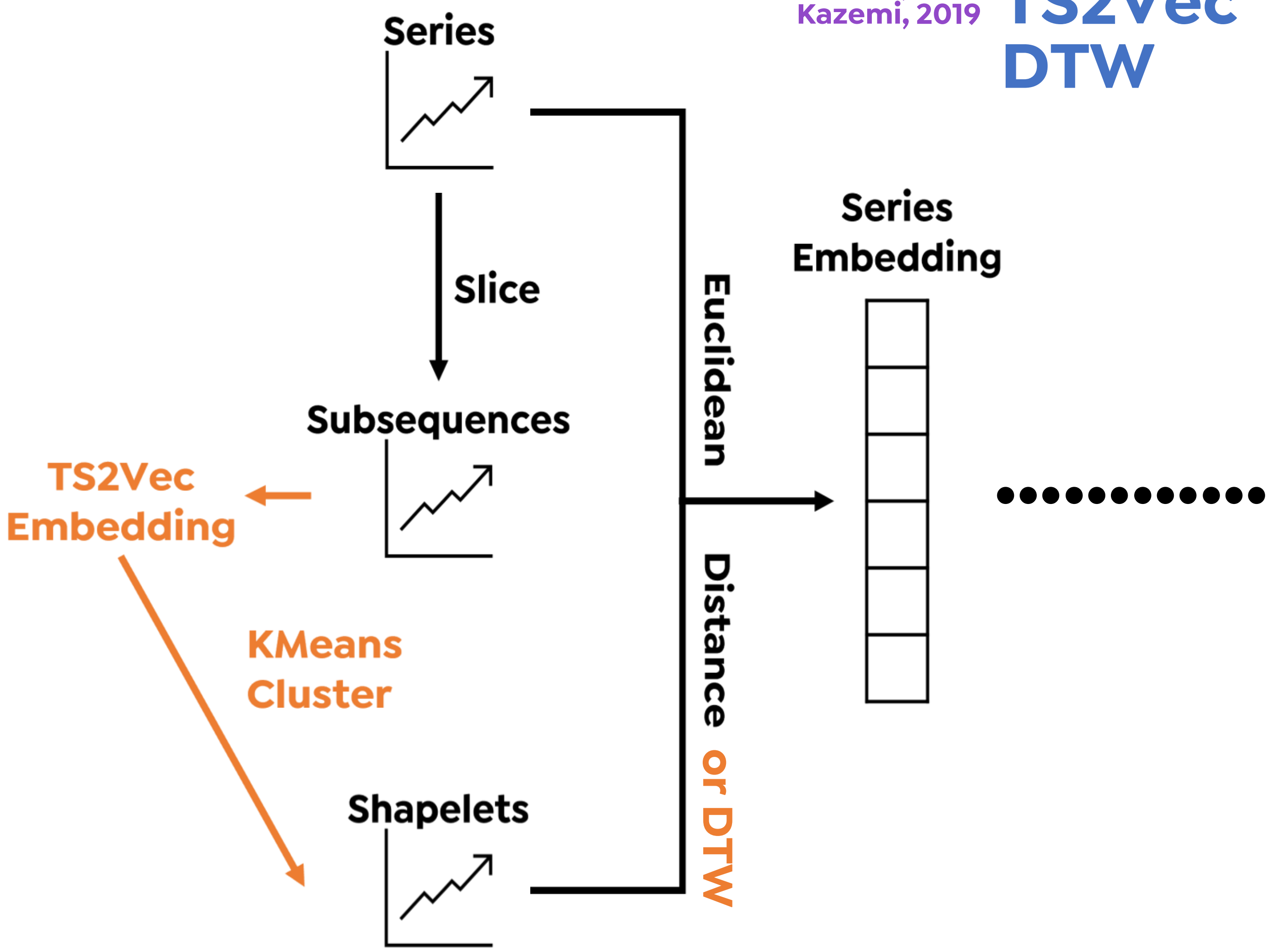
Experiment

DIAGRAM OF AQOURSNET

FEATURES.

KMeans
GAT / GATv2
TS2Vec
DTW

Meyer, 2018
Kazemi, 2019



AQOURS Net

Problem Analysis

Methodology

Features

Experiment

SimpleKMeans vs. ShapeNet

COLOR CODE: TOP 3, BOTTOM 7 AMONG A TOTAL OF 16 MODELS

Best Performance

Why do they fail???

Dataset	RotF	DTW_Rn	ST	LS	FS	SD	COTE	ELIS	ResNet	Random	BSPCOVER	Kmeans SVM	Kmeans MLP	Kmeans Res	Kmeans MLPHer	Kmeans ResHer
BeetleFly	90.00	65.00	90.00	80.00	70.00	75.00	80.00	85.00	85.00	80.00	90.00	95.00	*100.00*	95.00	50.00	*100.00*
Coffee	*100.00*	*100.00*	96.43	*100.00*	92.86	96.10	*100.00*	96.43	*100.00*	93.47	*100.00*	*100.00*	*100.00*	*100.00*	53.57	96.43
DistalPhalanxOutline	75.72	72.46	77.54	77.90	75.00	71.70	76.09	57.83	77.10	77.52	*83.17*	77.90	80.07	78.62	58.33	73.19
Earthquakes	74.82	72.66	74.10	74.10	70.50	63.60	74.82	77.64	71.20	75.49	*81.68*	74.82	74.82	78.42	74.82	74.82
ECG200	85.00	88.00	83.00	88.00	81.00	81.80	88.00	80.00	87.40	85.00	*92.00*	83.00	64.00	91.00	64.00	64.00
ECGFiveDays	90.82	79.67	98.37	*100.00*	99.77	95.30	99.88	95.45	97.50	89.95	*100.00*	96.28	93.96	97.68	—	85.02
FordA	84.47	66.52	*97.12*	95.68	78.71	77.60	95.68	67.60	92.00	90.12	96.31	75.53	52.88	65.45	51.59	65.61
Ham	71.43	60.00	68.57	66.67	64.76	61.90	64.76	63.81	75.70	71.87	*76.19*	71.43	72.38	73.33	51.43	66.67
ShapeletSim	41.11	69.44	95.56	95.00	*100.00*	67.20	96.11	*100.00*	77.90	79.25	84.44	*100.00*	82.22	96.11	53.33	87.78
SonyAIBORobotSurface1	80.87	69.55	84.36	81.03	68.55	85.00	84.53	87.85	*95.80*	80.06	88.35	92.18	92.35	78.87	57.07	68.22
SonyAIBORobotSurface2	80.80	85.94	93.39	87.51	79.01	78.00	95.17	93.17	*97.80*	79.93	93.49	91.71	84.78	92.97	61.70	75.55
Strawberry	97.30	94.59	96.22	91.08	90.27	88.40	95.14	83.85	*98.10*	89.57	94.29	93.24	86.49	96.76	64.32	83.51
ToeSegmentation1	53.07	75.00	96.49	93.42	95.61	88.20	97.37	*98.24*	96.30	81.71	96.49	92.11	90.79	94.74	55.26	80.70
TwoLeadECG	97.01	86.83	99.74	99.65	92.45	86.70	99.30	99.82	*100.00*	92.19	99.65	96.14	92.80	92.01	50.04	78.58
Wafer	99.45	99.59	*100.00*	99.61	99.68	99.30	99.98	99.43	99.90	96.49	99.81	98.91	89.21	99.43	89.21	89.21
WormsTwoClass	68.83	58.44	*83.12*	72.73	72.73	64.10	80.52	71.82	74.70	71.24	74.59	81.82	81.82	*83.12*	57.14	76.62
Yoga	82.43	84.30	81.77	83.43	69.50	62.50	87.67	83.90	87.00	81.27	*88.20*	75.23	73.30	85.07	53.57	65.33

Li, Guozhong, et al. "ShapeNet: A Shapelet-Neural Network Approach for Multivariate Time Series Classification." *AAAI Conference on Artificial Intelligence, vol. 35*, pp. 8375-8383, 2021. https://ojs.aaai.org/index.php/AAAI/article/view/17018.

AQOURS Net

Problem Analysis

Methodology

Features

Experiment

SimpleKMeans vs. Time2Graph

COLOR CODE: TOP, MIDDLE, BOTTOM

Model	KMeans SVM		KMeans MLP		KMeans ResNet	
Best Parameters	Number	Length	Number	Length	Number	Length
BeetleFly	30	30%	30	15%	20	20%
Coffee	20	10%	25	20%	20	25%
DistalPhalanxOutline	30	20%	30	30%	30	15%
Earthquakes	10	20%	25	20%	25	15%
ECG200	30	30%	20	25%	30	30%
ECGFiveDays	30	30%	25	30%	25	30%
FordA	30	20%	20	25%	20	15%
Ham	30	20%	25	20%	25	30%
ShapeletSim	10	5%	30	15%	20	15%
SonyAIBORobotSurface1	20	20%	30	25%	25	25%
SonyAIBORobotSurface2	30	10%	25	20%	30	15%
Strawberry	20	5%	20	25%	30	25%
ToeSegmentation1	30	20%	30	15%	30	30%
TwoLeadECG	20	30%	25	30%	20	30%
Wafer	30	30%	25	25%	30	30%
WormsTwoClass	30	10%	30	15%	30	20%
Yoga	30	10%	30	15%	30	20%

Method	EQS	WTC	STB
RotF	74.82	*97.30*	68.83
DTW_Rn	72.66	94.59	58.44
ST	74.10	96.22	83.12
SD	63.60	88.40	64.10
COTE	74.82	95.14	80.52
ELIS	77.64	83.85	71.82
ResNet	71.20	*98.10*	74.70
Random	75.49	89.57	71.24
BSPCOVER	*81.68*	94.29	74.59
NN-ED	68.22	62.41	95.60
NN-DTW	70.31	68.16	95.53
NN-WDTW	69.50	67.74	95.44
NN-CID	69.41	69.56	95.51
DDTW	70.79	70.92	95.60
XGBoost Origin	74.82	62.34	95.92
XGBoost Feature	75.54	64.94	*97.03*
BoP	74.80	74.42	96.45
TSF	74.67	68.51	96.27
EE	73.50	71.74	95.88
SAXVSM	73.76	72.10	*96.97*
LS	74.22	73.57	92.49
FS	74.66	70.58	91.66
LPS	66.78	74.26	96.35
MLP	70.29	59.86	96.58
LSTM	74.82	42.86	63.84
VAE	71.22	62.34	71.35
Shapelet-Seq	75.53	55.84	78.10
Time2Graph	*79.14*	72.73	*96.76*
Time2Graph+ Static	76.98	70.13	95.95
Time2Graph+	77.70	71.43	96.49
KMeans SVM	74.82	93.24	81.82
KMeans MLP	74.82	86.49	81.82
KMeans ResNet	*78.42*	*96.76*	83.12

Cheng, Ziqiang, et al. "Time2Graph+: Bridging Time Series and Graph Representation Learning via Multiple Attentions." ***Journal of IEEE Transactions on Knowledge and Data Engineering*, 2021. https://doi.org/10.1109/TKDE.2021.3094908.**

AQOURSNet

Problem Analysis

Methodology

Features

Experiment

AQOURSNet vs. ShapeNet

COLOR CODE: TOP, MIDDLE, BOTTOM

	Dataset	AQOURSNet					ShapeNet Baselines	
		Number of Shapelets	Number of Segments	Graph Embedding Dimension	Train Accuracy	Test Accuracy	ResNet	BSPCOVER
TS2Vec OFF	Earthquakes	60	10	128	88.82	75.54	71.20	81.68
	Strawberry	60	10	256	92.50	87.57	89.57	93.24
	WormsTwoClass	40	10	128	100.00	66.23	74.70	74.59
	BeetleFly	60	20	64	100.00	80.00	85.00	90.00
	Coffee	60	10	256	100.00	100.00	100.00	100.00
	DistalPhalanx	60	40	256	78.83	76.81	77.10	83.17
	ECG200	80	20	64	100.00	94.00	87.40	92.00
	ECGFiveDays	40	40	128	100.00	95.01	97.50	100.00
TS2Vec ON	Earthquakes	80	40	128	81.99	74.82	71.20	81.68
	Strawberry	80	20	128	64.27	64.32	89.57	93.24
	WormsTwoClass	80	40	128	88.95	58.44	74.70	74.59
	BeetleFly	80	20	128	100.00	65.00	85.00	90.00
	Coffee	80	20	128	50.00	53.57	100.00	100.00
	DistalPhalanx	80	40	128	79.83	74.28	77.10	83.17
	ECG200	80	40	128	100.00	92.00	87.40	92.00
	ECGFiveDays	80	40	128	100.00	97.10	97.50	100.00

A
Q
O
U
R
S
Net

Problem Analysis

Methodology

Features

Experiment

AQOURSNet vs. Time2Graph

COLOR CODE: TOP, MIDDLE, BOTTOM

Dataset	AQOURSNet				Time2Graph Baselines		
	Number of Shapelets	Number of Segments	Train Accuracy	Test Accuracy	Time2Graph	Time2Graph+ Static	Time2Graph+
Earthquakes	40	20	95.96	74.82	<u>79.14</u>	76.98	77.70
Strawberry	50	20	98.69	96.49	<u>96.76</u>	95.95	96.49
WormsTwoClass	60	20	98.90	70.13	<u>72.73</u>	70.13	71.43

FUTURE WORKS.

XGBoost

JOINT LEARNING WITH PYTORCH

MixedPrecision

FINE-TUNING OF (HYPER)PARAMETERS

Time2Vec

FUNCTIONALITY VERIFICATION

DataParallel

DISTRIBUTED ON MULTIPLE CUDAS

RECEIVE UPDATES.

AllenHeartcore/AQOURSNet_rsch22su

- Complete docs
- Full “args” interface
- Prerequisite materials
- Releases

rong-hash/Time2GraphRework

- Fine-tuned GAT
- Out-of-the-box TS2Vec
- Updated experiments

Thanks for Listening

Autoencoded
Quantification
Of
Unsupervised
Representative
Shapelets
Net