Appendix: Time-Aware Multi-Scale RNNs for Time Series Modeling

Zipeng Chen, **Qianli Ma*** and **Zhenxi Lin**

School of Computer Science and Engineering,
South China University of Technology, Guangzhou, China
zipengchencs@foxmail.com, qianlima@scut.edu.cn, zhenxi_lin@foxmail.com

A Detailed results on human motion prediction

For human motion prediction, the overall results of TP-RNN and TAMS-TP-RNN on the H3.6M data set are shown in Table 1. The comparison results show that the proposed TAMS-TP-RNN is significantly better than TP-RNN, yielding lower Euler angle errors and higher best rates, which strongly verifies the superiority of the proposed model.

B More examples of music clips

For music genre recognition, more examples of music clips are shown in Fig. 1. Music series in the yellow region has small changes, so it's no need for the model to be updated frequently and a larger scale would be chosen. On the contrary, series in the purple region varies greatly, thus the choice of scale is more flexible. Therefore, at different time steps, different temporal scales would be chosen to model the corresponding dynamic temporal patterns.

*Qianli Ma is the corresponding author.

C Ablation study on MTS data sets

We design five ablation models to verify the effectiveness of each component. The experiments are conducted on 15 MTS classification data sets and the results are shown in Table 2.

D Computational efficiency analysis

The experiments are implemented using Python 2.7 and Tensorflow 1.11.0 on a personal computer with an Intel Core i7-8700K 3.70GHz CPU, a 64GB RAM, and a GeForce GTX 2080-Ti 11G graphics card. For "FaceDetection", the number of instances for training and testing are 5890 and 3524. The proposed model spent 12.77 seconds on training for each epoch with a batch size 64 and spent 2.52 seconds on testing for the whole testing set.

	TP-RNN						TAMS-TP-RNN					
Milliseconds	80	160	320	400	560	1000	80	160	320	400	560	1000
Walking	**0.25**	0.41	0.58	0.65	0.74	0.77	**0.25**	**0.39**	**0.57**	**0.63**	**0.73**	**0.76**
Eating	**0.20**	**0.33**	**0.53**	**0.67**	**0.84**	**1.14**	0.21	0.34	0.55	0.69	0.87	1.19
Smoking	0.26	**0.47**	**0.88**	0.90	0.98	1.66	**0.25**	**0.47**	**0.88**	**0.87**	**0.93**	**1.61**
Discussion	**0.30**	**0.66**	0.96	1.04	1.39	**1.74**	0.31	**0.66**	**0.95**	**1.03**	**1.34**	1.75
Directions	0.38	0.59	**0.75**	**0.83**	**0.95**	**1.38**	**0.36**	**0.58**	0.77	0.86	0.96	**1.38**
Greeting	0.51	0.86	1.27	1.44	1.72	1.81	**0.48**	**0.81**	**1.21**	**1.37**	**1.64**	**1.75**
Phoning	0.57	1.08	**1.44**	1.59	**1.47**	**1.68**	**0.56**	**1.06**	**1.44**	**1.58**	1.49	1.73
Posing	0.42	0.76	1.29	1.54	**1.75**	**2.47**	**0.32**	**0.54**	**1.10**	**1.38**	1.80	2.74
Purchases	**0.59**	**0.82**	**1.12**	**1.18**	**1.52**	**2.28**	0.63	0.87	1.13	1.20	1.53	2.30
Sitting	0.41	0.66	1.07	1.22	1.35	1.74	**0.38**	**0.61**	**1.00**	**1.16**	**1.27**	**1.67**
Sitting down	**0.41**	0.79	1.13	1.27	1.47	1.93	**0.41**	**0.77**	**1.09**	**1.22**	**1.42**	**1.90**
Taking photo	0.26	**0.51**	**0.80**	0.95	1.08	1.35	**0.25**	**0.51**	**0.80**	**0.92**	**1.00**	**1.28**
Waiting	**0.30**	**0.60**	**1.09**	**1.31**	1.71	**2.46**	0.32	0.62	1.10	**1.31**	**1.69**	2.47
Walking dog	0.53	0.93	**1.24**	**1.38**	1.73	1.98	**0.49**	**0.85**	1.39	1.45	**1.68**	**1.90**
Walking together	**0.23**	**0.47**	0.67	0.71	0.78	1.28	**0.23**	**0.47**	**0.66**	**0.70**	**0.77**	**1.26**
Average	0.37	0.66	0.99	1.11	1.30	**1.71**	**0.36**	**0.64**	**0.98**	**1.09**	**1.27**	**1.71**
No. best	7	7	8	5	5	7	**11**	**12**	**10**	**11**	**10**	**9**
Best rate	0.47	0.47	0.53	0.33	0.33	0.47	**0.73**	**0.80**	**0.67**	**0.73**	**0.67**	**0.60**

Table 1: Euler angle errors for 15 activities on the Human3.6M data set.

Data set	LSTM +MSFD(M)+TAFM	CW-LSTM +TAFM	LSTM +MSFD(M)	LSTM +MSFD(S)	CW-LSTM	LSTM
ArticularyWordRecognition	0.973	0.967	**0.976**	0.947	0.947	0.947
AtrialFibrillation	0.400	**0.467**	0.333	0.400	0.333	0.400
BasicMotions	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.975
CharacterTrajectories	**0.994**	0.993	**0.994**	0.993	0.990	0.985
FaceDetection	**0.602**	0.596	0.595	0.562	0.594	0.573
HandMovementDirection	**0.473**	0.446	0.432	0.284	**0.473**	0.311
Heartbeat	**0.756**	0.712	0.751	0.732	0.722	0.698
MotorImagery	0.590	0.580	**0.600**	0.530	0.520	0.570
NATOPS	**0.956**	**0.956**	0.950	0.950	0.950	**0.956**
PEMS-SF	**0.890**	**0.890**	0.873	0.884	**0.890**	0.879
Pen Digits	**0.981**	0.979	0.977	0.976	0.976	0.972
Phoneme	**0.203**	0.186	0.194	0.178	0.178	0.189
SelfRegulationSCP2	**0.561**	0.544	**0.561**	0.544	0.544	0.528
SpokenArabicDigits	**0.990**	0.988	0.989	0.980	0.986	0.987
StandWalkJump	**0.400**	0.267	0.333	0.267	0.333	**0.400**

Table 2: Ablation study on 15 MTS classification data sets.

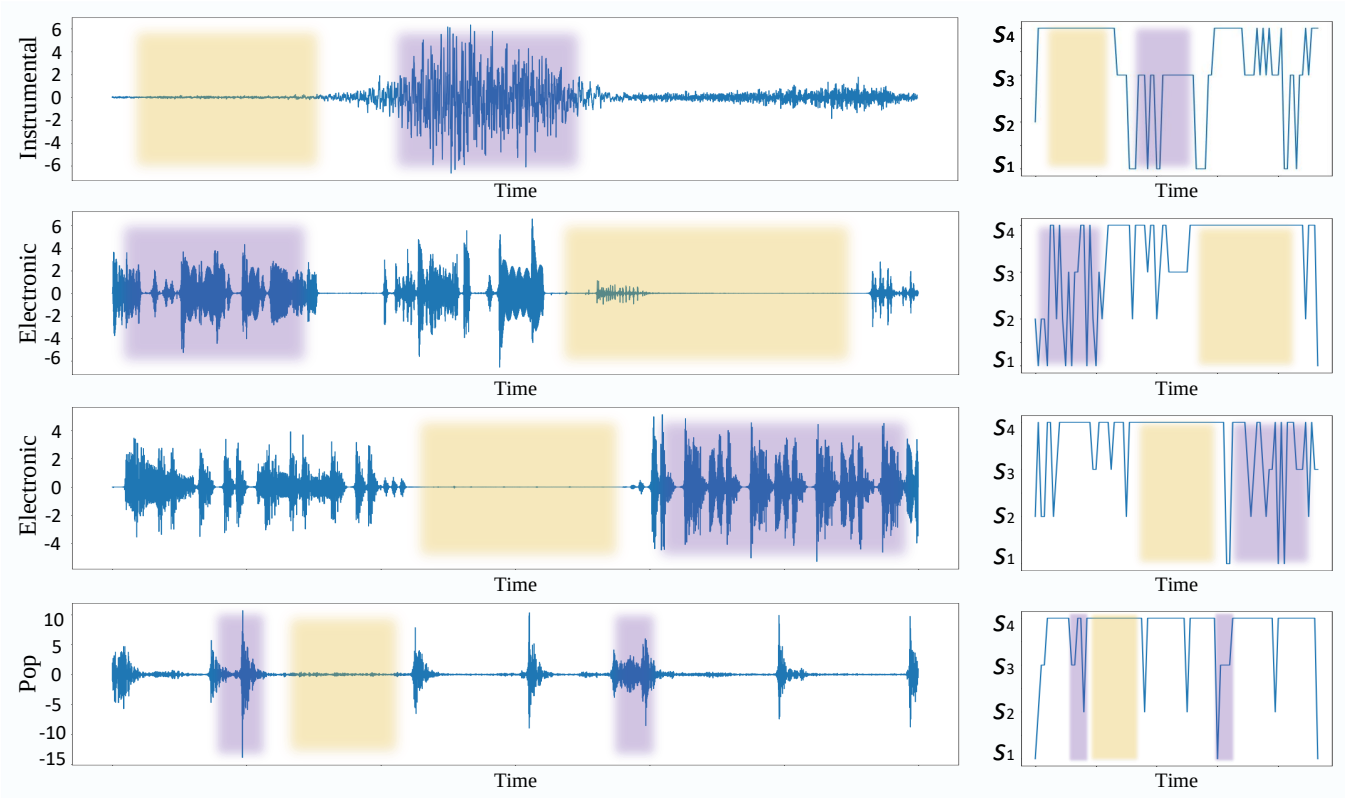

Figure 1: Five music clips and the corresponding weights assigned to the scales. Subfigures in the first column are five music clips of different genres, while subfigures in the second column are the scale with the largest weight at each time step t when t MOD $s_K = 0$ ($s_K = 64$), corresponding to music clips in the first column. Series in the yellow region prefers the largest scale due to its small changes, while series in the purple region prefers multiple scales due to its sharp fluctuations.