Supplementary Materials: Explaining Time Series Classifier through Meaningful Perturbation and Optimisation

Authors: Han Meng, Christian Wagner, Isaac Triguero

March 17, 2023

Abstract

This supplementary material provides additional information and results that complement the main paper. In this work, we analyze the performance of our generative model in predicting alternative values for certain features and compare its performance with two commonly used time series imputation methods. We provide the performance of each method on multiple datasets. Additionally, we evaluate the effectiveness of our proposed framework in explaining a Temporal Convolutional Network (TCN) classifier. The code for this work is upload into https://github.com/menghan1994/ETSC_through_Meainingful_Perturbation_and_Optimisation/.

1 The performance of our model in predicting plausible values for target features

In our work, we need to generate alternative values for certain features. Time series imputation models developed to fill missing values can be used in our framework. The scenario that these time series imputation models face is the lack of ground truth for the missing values. Therefore, the missing parts cannot be directly predicted based on the observed values. To address this challenge, these models attempt to learn the temporal dependence of time series using observed values. The missing values are then filled based on the learned temporal dependence. However, for our problem, the time series we have are complete, without any missing values. Thus, in predicting plausible values for certain observed features conditioned by their complements, we have ground truth that can be used directly to construct the loss function. We compare the performance of our model with two time series imputation models in predicting values for certain features, and the mean square losses are shown in the following Table 1. In our methods, we separately use BiRNNs and Transformer to encoder time series. The results shows that the Transformer-based model perform better than others.

Table 1: The mean square errors in predicting values for certain features

Datasets	BRITS	E2GAN	Our Method	
			BiRNN-based	Transformer-based
ArticularyWordRecognition	0.2891(0.0012)	0.3426(0.0007)	0.3142(0.0015)	**0.1163(0.0008)**
BasicMotions	0.1321(0.0030)	0.2079(0.0033)	0.1398(0.0032)	**0.1171(0.0018)**
CharacterTrajectories	0.1593(0.0008)	0.1812(0.0007)	0.1295(0.0007)	**0.0959(0.0006)**
Cricket	0.1242(0.0005)	0.2578(0.0008)	0.2758(0.0010)	**0.0721(0.0006)**
Epilepsy	**0.1969(0.0010)**	0.2522(0.0017)	0.2525(0.0012)	0.2230(0.0014)
ERing	0.3519(0.0011)	0.5574(0.0067)	0.3438(0.0004)	**0.3355(0.0045)**
InsectWingbeat	0.0956(0.0003)	0.0856(0.0003)	0.0752(0.0002)	0.0557(0.0002)
JapaneseVowels	0.3883(0.0036)	0.1826(0.0031)	0.2425(0.0017)	**0.0806(0.0010)**
Libras	0.3441(0.0072)	0.2655(0.0032)	**0.1605(0.0059)**	0.3045(0.0052)
LSST	0.1954(0.0007)	0.2046(0.0014)	0.1992(0.0007)	**0.1531(0.0006)**
NATOPS	0.1864(0.0016)	0.3306(0.0011)	0.2417(0.0014)	**0.0808(0.0004)**
PenDigits	0.4116(0.0045)	0.1333(0.0010)	**0.1024(0.0012)**	0.2985(0.0021)
PhonemeSpectra	0.1031(0.0002)	0.2308(0.0003)	0.0311(0.0000)	**0.0001(0.0000)**
RacketSports	0.1810(0.0033)	0.2165(0.0043)	0.2348(0.0031)	**0.1242(0.0019)**
SpokenArabicDigits	0.1765(0.0002)	0.1247(0.0003)	0.1884(0.0006)	**0.0500(0.0001)**
StandWalkJump	0.0401(0.0002)	0.0578(0.0001)	0.1075(0.0008)	**0.0352(0.0001)**
UWaveGestureLibrary	0.1598(0.0034)	0.3779(0.0012)	**0.1075(0.0008)**	0.2869(0.0010)

Table 2: The accuracy of two black-box classifiers on the adopted datasets

dataset	Accuracy		
	LSTM	TCN	Guess
ArticularyWordRecognition	**0.89**	**0.96**	0.04
BasicMotions	**0.73**	**0.83**	0.25
CharacterTrajectories	**0.67**	**0.99**	0.05
Cricket	**0.79**	**0.97**	0.08
Epilepsy	**0.54**	**0.94**	0.25
ERing	**0.79**	**0.93**	0.17
InsectWingbeat	**0.29**	**0.33**	0.10
JapaneseVowels	**0.94**	**0.91**	0.11
Libras	**0.71**	**0.69**	0.07
LSST	**0.62**	**0.64**	0.07
NATOPS	**0.86**	**0.81**	0.17
PenDigits	**0.99**	**0.97**	0.10
PhonemeSpectra	**0.14**	**0.20**	0.03
RacketSports	**0.76**	**0.77**	0.25
SpokenArabicDigits	**0.98**	**0.98**	0.10
StandWalkJump	**0.47**	0.40	0.33
UWaveGestureLibrary	**0.54**	**0.86**	0.13

2 The performance of our framework in explaining LSTM and TCN classifiers

In this section, the prediction accuracy of LSTM-based classifier and Temporal Convolutional Networks (TCN)[1] classifiers are provided first. Then, the deteriorate tests for these two classifiers are provided.

2.1 The predictive accuracy of these two classifiers

The accuracy of these two classifiers are shown in the Table 2.

2.2 The deteriorate tests for explanations for these two classifiers

The deteriorate tests for explanations for these two classifiers are shown in the Table 3 and Table 4. Both of these two tables shows that our method performs better than the benchmaking methods.

[1]Colin Lea, Michael D Flynn, Rene Vidal, Austin Reiter, and Gregory D Hager. *Temporal convolutional networks for action segmentation and detection.* In proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 156–165, 2017.

Table 3: The deteriorate tests for the LSTM-based classifier

Dataset	IG	SG	LIME	LIME-G	Proposed Framework	
					with BPSO	with GA
ArticularyWordRecognition	349.00(25.75) 0.02s(0.00)	401.50(20.17) 0.01s(0.00)	645.00(38.77) 5.30s(0.43)	591.65(40.38) 9.31s(0.87)	**83.50(7.23)** 27.59s(1.43)	89.00(7.50) 34.02s(1.85)
BasicMotions	86.00(6.00) 0.02s(0.00)	465.00(26.92) 0.01s(0.00)	125.00(8.32) 6.07s(0.56)	130.70(10.04) 10.45s(0.87)	73.50(4.86) 13.45s(0.90)	**60.00(3.22)** 14.08s(0.83)
CharacterTrajectories	112.00(9.81) 0.02s(0.00)	84.50(7.63) 0.01s(0.00)	108.50(7.82) 3.73s(0.20)	93.27(6.53) 6.51s(0.58)	32.50(2.55) 23.46s(1.58)	**28.00(2.75)** 24.21s(2.28)
Cricket	8.49(0.54) 0.08s(0.01)	22.23(2.18) 0.05s(0.01)	25.62(2.02) 29.26s(2.55)	29.30(2.29) 29.90s(2.50)	12.12(1.06) 142.43s(9.82)	**8.15(0.59)** 141.92s(11.15)
Epilepsy	8.00(0.48) 0.02s(0.00)	63.50(3.36) 0.01s(0.00)	7.50(0.41) 4.39s(0.29)	6.42(0.36) 6.89s(0.53)	**4.00(0.32)** 26.41s(1.89)	5.50(0.45) 29.80s(2.94)
ERing	26.50(2.64) 0.02s(0.00)	56.50(4.49) 0.01s(0.00)	16.00(1.45) 5.63s(0.36)	14.86(1.18) 10.92s(1.09)	13.00(0.83) 17.04s(1.16)	**12.00(0.67)** 7.29s(0.72)
InsectWingbeat	19.50(1.93) 0.02s(0.00)	65.50(4.29) 0.01s(0.00)	23.50(2.11) 8.23s(0.67)	44.00(2.56) 14.85s(0.98)	15.50(1.18) 33.10s(9.79)	**14.02(0.77)** 31.08s(1.59)
JapaneseVowels	84.00(6.08) 0.02s(0.00)	164.50(14.09) 0.00s(0.00)	101.50(6.32) 7.89s(0.43)	88.67(8.09) 10.67s(0.96)	38.50(1.99) 43.57s(3.06)	**31.00(2.73)** 12.30s(0.93)
Libras	7.50(0.54) 0.03s(0.00)	18.00(1.02) 0.00s(0.00)	8.50(0.76) 4.93s(0.25)	7.74(0.65) 7.67s(0.63)	**6.00(0.46)** 5.06s(0.30)	7.00(0.36) 4.80s(0.25)
LSST	12.50(0.95) 0.02s(0.00)	52.00(3.60) 0.01s(0.00)	10.50(0.71) 5.18s(0.49)	8.46(0.68) 8.46s(0.48)	13.00(1.23) 7.22s(0.66)	**9.00(0.74)** 6.58s(0.50)
NATOPS	246.00(42.97) 0.03s(0.00)	156.00(48.58) 0.01s(0.00)	222.50(73.36) 10.66s(1.01)	318.67(50.49) 18.98s(1.29)	**111.50(8.38)** 10.59s(1.05)	128.00(11.78) 9.04s(0.71)
PenDigits	5.00(0.47) 0.01s(0.00)	11.00(1.08) 0.01s(0.00)	5.50(0.50) 3.17s(0.25)	5.43(0.27) 4.27s(0.38)	**3.00(0.26)** 2.12s(0.17)	**3.00(0.20)** 2.27s(0.20)
PhonemeSpectra	77.00(6.94) 0.02s(0.00)	153.50(9.71) 0.01s(0.00)	262.00(120.66) 8.09s(0.57)	232.27(82.72) 15.24s(1.48)	79.00(5.19) 50.25s(4.40)	**71.00(3.69)** 46.93s(3.90)
RacketSports	34.50(1.94) 0.03s(0.00)	116.50(7.01) 0.01s(0.00)	44.00(2.26) 4.32s(0.43)	41.61(2.69) 6.54s(0.63)	25.50(1.46) 4.93s(0.28)	**20.50(1.25)** 5.16s(0.31)
SpokenArabicDigits	219.00(14.77) 0.02s(0.00)	935.50(75.66) 0.01s(0.00)	638.00(39.57) 4.91s(0.44)	688.94(64.33) 6.68s(0.46)	115.00(8.01) 21.41s(1.42)	**93.50(5.86)** 21.91s(1.58)
StandWalkJump	10.50(0.66) 0.08s(0.01)	25.00(1.40) 0.06s(0.00)	29.00(1.61) 32.19s(2.13)	30.58(1.81) 32.44s(1.66)	12.00(0.68) 151.23s(14.26)	**8.00(0.44)** 164.26s(14.31)
UWaveGestureLibrary	27.00(1.61) 0.02s(0.00)	29.00(1.48) 0.01s(0.00)	36.00(3.26) 4.42s(0.34)	38.26(3.65) 8.38s(0.74)	**21.00(1.81)** 52.47s(3.50)	22.50(1.23) 56.89s(3.79)
MNIST	104.00(10.06) 0.02s(0.00)	167.00(15.72) 0.00s(0.00)	90.50(8.30) 4.01s(0.30)	95.51(6.70) 7.98s(0.76)	27.50(2.36) 9.43s(0.66)	**18.00(1.56)** 10.71s(0.96)

Table 4: The deteriorate tests for the TCN-based classifier

Dataset	IG	SG	LIME	LIME-G	Proposed Framework	
					with BPSO	with GA
ArticularyWordRecognition	417.82(36.27)	346.24(33.38)	547.40(31.95)	566.35(33.49)	89.93(7.81)	**72.53(6.66)**
	0.02s(0.00)	0.01s(0.00)	6.12s(0.57)	9.98s(0.58)	30.10s(1.84)	38.80s(2.92)
BasicMotions	88.75(7.02)	536.85(29.80)	145.65(7.51)	147.71(14.08)	74.28(6.79)	**53.48(4.92)**
	0.02s(0.00)	0.01s(0.00)	6.51s(0.56)	8.18s(0.77)	14.44s(1.01)	12.07s(1.10)
CharacterTrajectories	129.77(9.75)	95.07(6.60)	122.24(8.41)	89.77(8.51)	32.19(1.95)	**30.28(2.92)**
	0.02s(0.00)	0.00s(0.00)	4.40s(0.28)	3.52s(0.30)	22.50s(1.38)	21.21s(1.93)
Cricket	9.52(0.68)	20.92(1.45)	32.01(1.97)	22.54(2.17)	9.95(0.56)	**8.99(0.77)**
	0.07s(0.01)	0.07s(0.00)	25.70s(1.38)	39.45s(3.36)	127.53s(12.51)	174.16s(14.74)
Epilepsy	7.48(0.43)	61.32(4.78)	6.61(0.55)	8.61(0.67)	**3.64(0.31)**	5.05(0.45)
	0.02s(0.00)	0.01s(0.00)	5.02s(0.37)	8.14s(0.48)	22.94s(1.43)	32.45s(2.54)
ERing	30.18(2.23)	65.10(6.30)	16.22(1.00)	15.58(1.13)	**13.08(0.91)**	13.65(0.79)
	0.02s(0.00)	0.01s(0.00)	4.78s(0.30)	7.90s(0.67)	13.66s(0.90)	7.50s(0.41)
InsectWingbeat	19.44(1.37)	53.12(4.88)	22.36(1.34)	51.67(4.55)	15.83(1.32)	**12.59(1.02)**
	0.02s(0.00)	0.01s(0.00)	9.35s(0.75)	13.61s(0.88)	131.47s(12.26)	30.92s(2.08)
JapaneseVowels	78.11(7.75)	175.71(8.79)	83.86(5.26)	86.43(7.47)	**30.95(2.13)**	33.49(3.32)
	0.01s(0.00)	0.00s(0.00)	6.51s(0.42)	11.12s(0.76)	38.19s(2.03)	14.61s(0.82)
Libras	8.52(0.56)	19.62(1.42)	9.13(0.61)	7.58(0.72)	**6.48(0.44)**	8.01(0.53)
	0.03s(0.00)	0.00s(0.00)	5.77s(0.44)	8.04s(0.44)	5.70s(0.47)	5.38s(0.32)
LSST	14.29(1.10)	52.37(4.51)	12.24(0.99)	10.44(0.81)	10.21(0.54)	**8.57(0.51)**
	0.02s(0.00)	0.01s(0.00)	5.70s(0.50)	9.04s(0.63)	6.20s(0.32)	6.42s(0.55)
NATOPS	137.82(8.85)	220.16(11.83)	261.52(11.08)	198.83(16.32)	**100.11(8.24)**	114.16(9.69)
	0.03s(0.00)	0.01s(0.00)	9.04s(0.55)	14.57s(0.74)	8.85s(0.81)	9.39s(0.62)
PenDigits	5.97(0.54)	13.04(0.73)	5.48(0.36)	5.42(0.28)	**2.72(0.25)**	3.28(0.23)
	0.01s(0.00)	0.01s(0.00)	3.13s(0.24)	5.04s(0.37)	1.72s(0.09)	2.22s(0.13)
PhonemeSpectra	87.42(7.15)	124.94(9.87)	343.19(111.36)	270.24(88.15)	**67.42(7.15)**	88.35(8.34)
	0.02s(0.00)	0.01s(0.00)	7.70s(0.74)	8.85s(0.53)	41.59s(4.10)	54.43s(3.17)
RacketSports	35.79(3.21)	134.61(9.71)	40.40(3.81)	39.86(2.79)	29.77(2.96)	**24.32(1.39)**
	0.03s(0.00)	0.01s(0.00)	3.68s(0.23)	8.26s(0.50)	5.04s(0.29)	5.02s(0.39)
SpokenArabicDigits	228.11(18.19)	838.18(45.55)	675.55(51.41)	800.00(63.09)	129.80(8.41)	**110.03(10.30)**
	0.02s(0.00)	0.01s(0.00)	4.06s(0.35)	6.04s(0.42)	18.06s(1.17)	19.16s(1.40)
StandWalkJump	8.64(0.53)	23.38(2.27)	28.68(2.73)	30.38(2.54)	14.25(1.15)	**7.74(0.54)**
	0.09s(0.01)	0.06s(0.00)	28.43s(2.02)	45.44s(2.42)	162.51s(11.19)	144.43s(12.19)
UWaveGestureLibrary	31.17(2.63)	25.78(1.97)	31.99(1.63)	39.41(2.79)	**19.72(1.75)**	25.59(2.43)
	0.02s(0.00)	0.01s(0.00)	4.87s(0.36)	7.92s(0.55)	51.82s(4.41)	48.40s(2.91)
MNIST	77.27(6.58)	175.24(10.00)	78.20(5.19)	98.64(6.95)	27.40(2.11)	**16.91(1.01)**
	0.02s(0.00)	0.01s(0.00)	3.63s(0.22)	3.83s(0.21)	9.33s(0.49)	11.52s(1.01)