Complex sequential data analysis: a systematic literature review of existing algorithms

Kudakwashe Dandajena[#], Mehrdad Ghaziasgar[*], Isabella M. Venter [#]

[#] *Department of Computer Science/ University of the Western Cape*

[1]kudadandaz@gmail.com

[2]mghaziasgar@uwc.ac.za

[3]iventer@uwc.ac.za

[*] *University of the Western Cape, Private Bag X17, Bellville, 7535, South Africa*

Annexure 1: Results of state of the art sequential frameworks

1. Attention based frameworks (At-LSTM)
2. Hybrid attention based frameworks (At-LSTM)
3. Autoregressive models (AR)
4. Hybrid autoregressive model
5. Back-propagation neural networks (BPNN)
6. Bayesian based algorithms
7. Bidirectional (Bi) based frameworks
8. Bidirectional combined with attention (Att) mechanism
9. Bidirectional combined with GRU (BiGRU) and LSTM (BiLSTM)
10. Capsule neural network (CapsNet)
11. Convolutional neural networks (CNNs)
12. Deep autoencoder (DA)
13. Deep Bayesian neural networks (BNN)
14. Deep differential privacy-inspired LSTM (DP- LSTM)
15. Deep feed forward neural network (FFNN)
16. Deep sequential spatio-temporal residual neural network (ST-ResNet)
17. Denoising autoencoder (DAE)
18. Transformer neural network
19. Transformer neural network combined with RNN and CNN
20. TrellisNet
21. Differentiable architecture (DARTS)
22. Dilated recurrent neural network (DilatedRNN)
23. Dilated temporal convolutional network (TCN)
24. Dual self-attention network (DSANet)
25. Dual-stage attention based recurrent neural network (DA-RNN)
26. Elmann recurrent neural networks (ERNN)
27. Extension GARCH (EGARCH)
28. Fast-slow recurrent neural network (FS-RNN)
29. Feed forward neural networks (FFNN)
30. Generative adversary neural networks (GAN)
31. Gated recurrent unit (GRU)
32. Gated recurrent unit with hybrid architecture
33. Gaussian models (GP)
34. General regression neural network (GRNN)
35. Generalized autoregressive conditional heteroscedasticity (GARCH)
36. Generalized linear regression (GLM))
37. Hierarchical multi-scale recurrent neural network (HM-RNN)
38. Hierarchical neural network architecture
39. Independently recurrent neural network (IndRNN)
40. Large feedforward neural network (LFNN)
41. Logistic regression (LR)
42. Long short-term memory (LSTM)
43. Memory-based ordinal regression deep neural networks (MOrdReD)

44. Momentum models (MOM)
45. Mean reversion models (MR)
46. Multilayer perception (MLP)
47. Multivariate adaptive regression splines (MARS)
48. Neural architecture search (NAS)
49. Particle filter recurrent neural networks (PF-RNNs)
50. Quasi-recurrent neural network (QRNN)
51. Radial basis neural networks (RBFNN)
52. Random Classifier (RC)
53. Random connectivity LSTM (RCLSTM)
54. Random forest (RF)
55. Recurrent highway network (RHN)
56. Recurrent neural network (RNN)
57. Rule-based regression (RBR)
58. Sequence to sequence (Seq2seq) architectures or encoder-decoder models
59. Skip recurrent neural network (SkipRNN)
60. Small feedforward neural network (SFNN)
61. Spatio-temporal long short-term network (ST-LSTM)
62. Squares support vector machine regression (LS-SVMR).
63. StockNet which uses a variational autoencoder (VAE)
64. Support vector machine regression (SVMR)
65. Support vector machines (SVM)
66. Temporal convolutional networks (TCN)
67. Transformer networks
68. TrellisNet
69. Variational LSTM

Annexure 2: Results of sequential datasets

1. Australia traffic flow data
2. Daily gas revenue data
3. Power consumption data
4. Financial stock market dataset (Cryptocurrency, S&P 500, Dow Jones Industrial Average (DJIA), NASDAQ and Russel 2000) data
5. Sentimental Natural language process dataset from social media data
6. Forecasting competitions data
7. Real time Yangtze River dissolved oxygen time series data
8. Reuters and Bloomberg sentimental data and Standard & Poor's 500 dataset from
9. S&P 500 stock price data
10. Taxicab GPS data
11. Uber Traffic dataset
12. UCI standard datasets
13. Univariate and multivariate air pollution data from GEFCom (2014)
14. Standard electricity price dataset by (Hong et al. 2016)
15. Hourly traffic dataset (2015-2016) from California Department of Transportation
16. Solar-energy data (2006) from 137 PV plants in Alabama State
17. Standard electricity consumption data (2012 to 2014)
18. Daily exchange rates data from Australia, British Canada, Switzerland, China, Japan, New Zealand and Singapore from 1990 to 2016 by (Lai et al. 2017)
19. NASDAQ stock price dataset by Qin Y. et al (2017)
20. Appliances energy prediction dataset by Candanedo L. et al (2017)
21. Air quality prediction (AIR De Vito S. et al (2008)
22. Weather dataset by Liang X, et al (2015)
23. European G´EANT traffic data points
24. Telecom datasets from Cell2Cell
25. Crowd Analytix dataset
26. Unstable social media dataset from Persian movie reviews from 2014 to 2016.
27. Standard benchmark ACL18 data for NASDAQ and NYSE markets from Jan 2014 to Jan 2016 by (Xu and Cohen, 2018)
28. Standard KDD17 dataset by (Zhang et al., 2017)
29. Stock index data (DOW 30, S&P 500 and NASDAQ)
30. Ultra-high-frequency order book data from 5 liquid U.S NASDAQ's (Google, Microsoft, Apple, Intel and Facebook) financial stocks
31. Financial stock indices dataset (S&P 500, Dow Jones Industrial Average (DJIA), NASDAQ and Russel 2000)
32. Historical financial price data from Crypto-Compare for Bitcoin, Ethereum and Monero
33. Social data from publicly available social platforms (GitHub and Reddit).
34. Standard Penn Treebank (PTB) data
35. Standard WikiText-103 (WT103) data
36. Financial news dataset from Reuters and Bloomberg on 473 Standard & Poor's 500 listed companies (Google, Amazon, Cisco, Microsoft, Apple, Intel, IMB, AMD, NVidia, Qualcomm, Walmart)
37. Sydney motorway traffic flow data of 2017
38. Financial stock dataset from Bank of China (601988), Vanke A (000002) and Kweichou Moutai (600519).
39. UCI daily grocery sales datasets
40. Univariate (Daily values for Melbourne's minimum temperature and Zurich Sunspot) datasets
41. Multi-variate (Energy production for 10 different photovoltaic power plants in California and SML2010 dataset containing internal and external measurements in a domestic house) datasets
42. Real time Yangtze River dissolved oxygen time series data automatically recorded from 2012 to 2016.
43. 4 years sequential time series Uber dataset for 8 large cities in U.S. and Canada (Atlanta, Boston,

Chicago, Los Angeles, New York City, San Francisco, Toronto, and Washington D.C)

44. Trajectory data (TaxiBJ from taxicab GPS data and meteorology data in Beijing (2013 – 2016) and Trajectory data (BikeNYC) from NYC bike system (2014)
45. Historical S&P 500 stock price data from the Yahoo Finance
46. NLP sentimental news dataset from financial domain (CNBC.com, Reuters.com, WSJ.com, Fortune.com and Wall Street Journal)
47. Daily revenue data from five gas stations companies
48. 45 datasets of different time series lengths from random real world application domains which encompass Meteorology, Astronomy, Physiology, Acoustics, and others
49. Real-world JD.com of China's (JD-demand and JD-shipment) data
50. Electricity consumption dataset for servers in a data centre by Flunkert et al.(2017)
51. Traffic flows data by Lv et al. (2015)
52. Internet traffic dataset for internet companies' by Kaggle (2017))

Annexures 3: Results of deep learning framework evaluation metrics

1. Agreement Cohen's Kappa
2. Average negative log-likelihood (NLL)
3. Computational time spent by a model
4. Copy memory loss and memory footprints
5. Correlation coefficient (R2)
6. Cosine proximity
7. Dynamic time warping (DTW)
8. Empirical correlation coefficient (CORR)
9. F-Measure
10. Hit ratio
11. Matthews correlation coefficient (MCC)
12. Max absolute percentage error (MaxAPE)
13. Mean absolute error (MAE)
14. Mean absolute percent errors (MAPE)
15. Mean absolute scaled error (MASE)
16. Mean directional accuracy (MDA)
17. Maximum error (ME)
18. Mean Error Percent (MEP)
19. Mean prediction accuracy (MPA)
20. Mean relative error (MRE)
21. Mean square error (MSE)
22. Mean squared percentage error (MSPE)
23. Mean symmetric mean absolute percentage error (SMAPE)
24. Median MASE
25. Median SMAPE
26. Normalized deviation (ND)
27. Normalized RMSE (NRMSE)
28. Normalized root mean squared error (NRMSE)
29. Precision F1 score
30. Precision jumps recall
31. Proportion of variance R2
32. Rank MASE
33. Rank SMAPE
34. Regression coefficient (R2)
35. Root mean square error (RMSE)
36. Root mean squared logarithmic error (RMLSE)
37. Root mean squared percentage error (RMSPE)
38. Root relative squared error (RRSE)
39. Symmetric mean absolute percentage error (SMAPE)
40. Trading profitability measures (cumulative return (CR), annualized return (AR), annualized volatility (AV), sharpe ratio and (SR) and draw-down (DD))

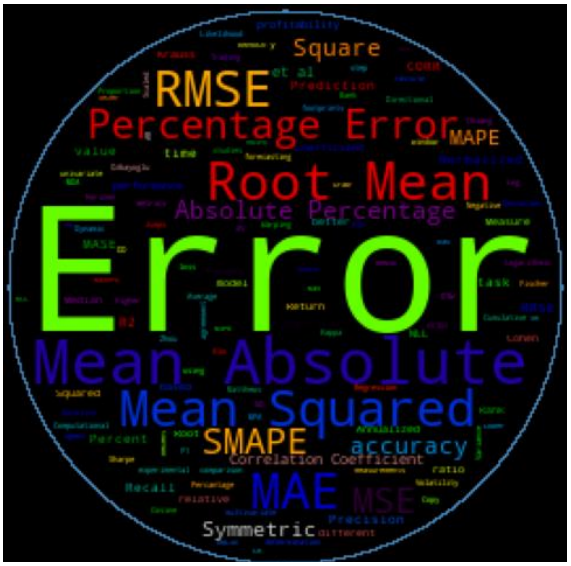

Figure 1: Framework evaluation metrics