Deep learning for forecasting model selection

ISF Presentation: 2019.6.19
Lancaster University Management School

Sasan Barak

Nikos Kourentzes, Sven F. Crone

Marketing Analytics & Forecasting | Lancaster University Management School

Agenda

1. **Model selection in Forecasting**
2. **Deep Learning for Model Selection**
3. **Experimental Evaluation**
    - Experimental Design (short and long time series)
4. **Conclusion**

CMAF

Model Selection in Forecasting

Model Selection (Fildes, 1989)

1. Aggregate Selection

2. Individual selection one particular method appropriate for each series is identified and used to forecast

A. Information Theoretic Selection
B. Empirical Accuracy Selection
C. Statisical Tests
D. Rule Based Selection

Wrapper Methodologies

Information theoretical approach: Akaike information criterion and extensions: AICc, QAIC, QAICc, AICW, Bayesian information criterion (BIC), Deviance information criterion (DIC), Focused information criterion (FIC), (Hyndman et al., 2002)

Empirical accuracy approach: Cross-validation and error measures (RMSE, MAPE, sMAPE, MASE,…) Koehler & Hyndman (2006), Fildes & Petropoulos (2015)

Filter Methodologies

Protocols: variance analysis (Gardner Jr and McKenzie, 1988), automatic identification (Vokurka et al., 1996), and rules-based forecasting (Adya et al., 2001) measure data characteristics and use them in forecasting models to generate best prediction (see Fildes et al., 2007) →doubts on efficiency of calculated characteristics / … dated?

Statistical Tests: tests for Trend (Mann-Kendall etc), Seasonality (F-test), Stationarity (DW, ADF etc.)

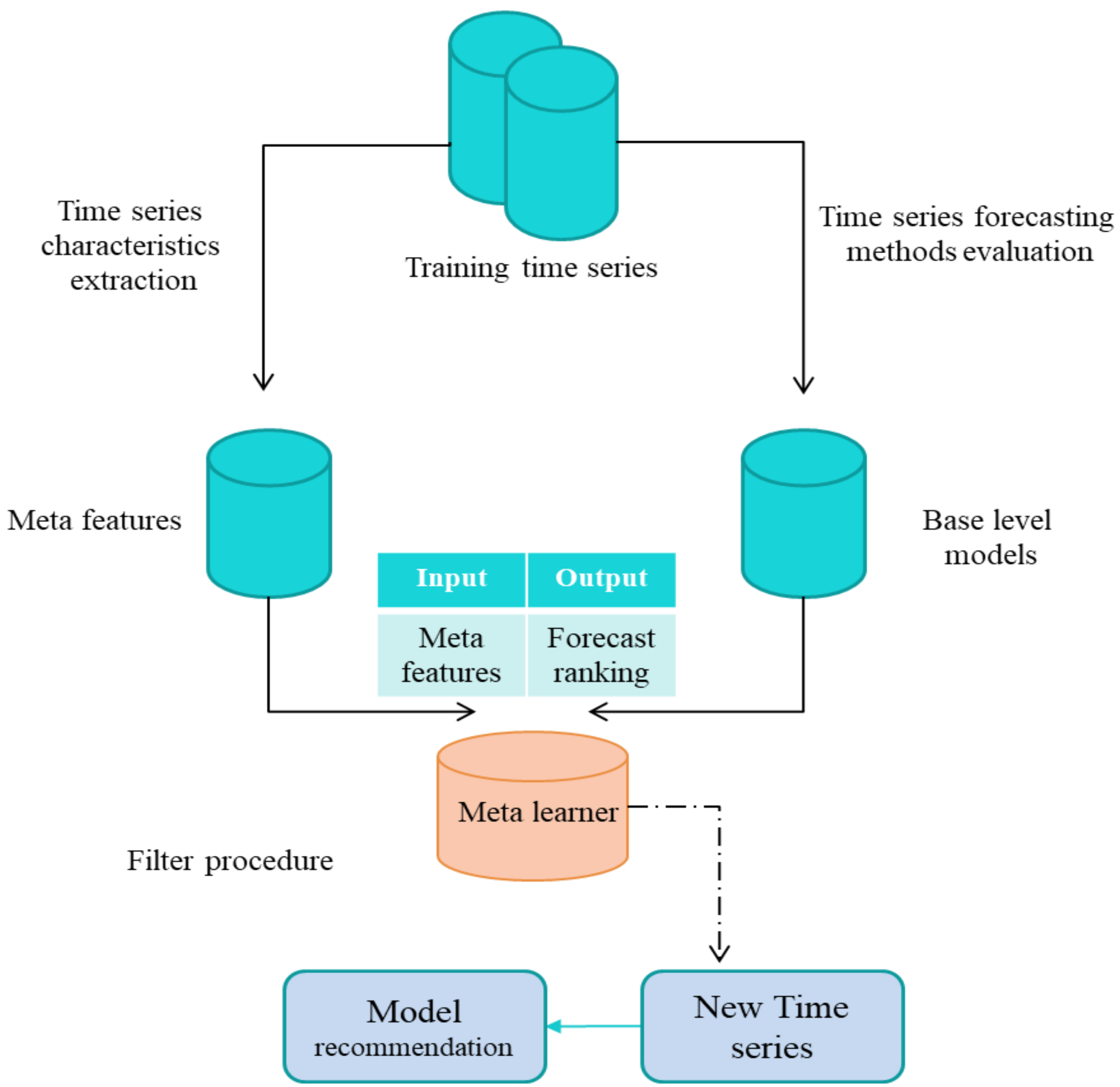
3. Model Combination is a weighted model selection (e.g., AIC weight, BIC weight, mean, median, …)

4. Judgmental selection

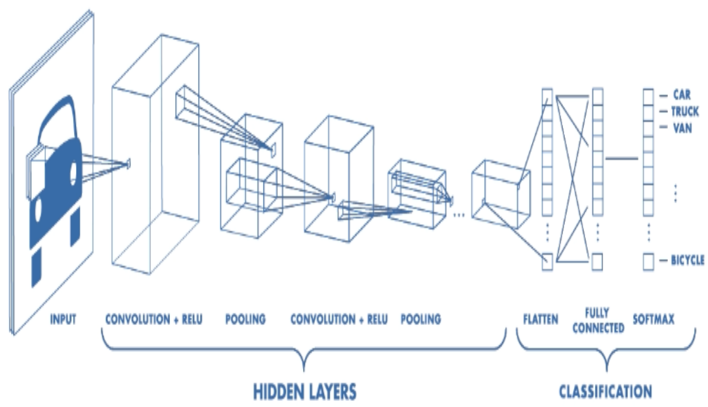
New challenges: Using Machine Learning for model selection

From Meta-learning to Deep-learning Model Selection

Meta-Learning



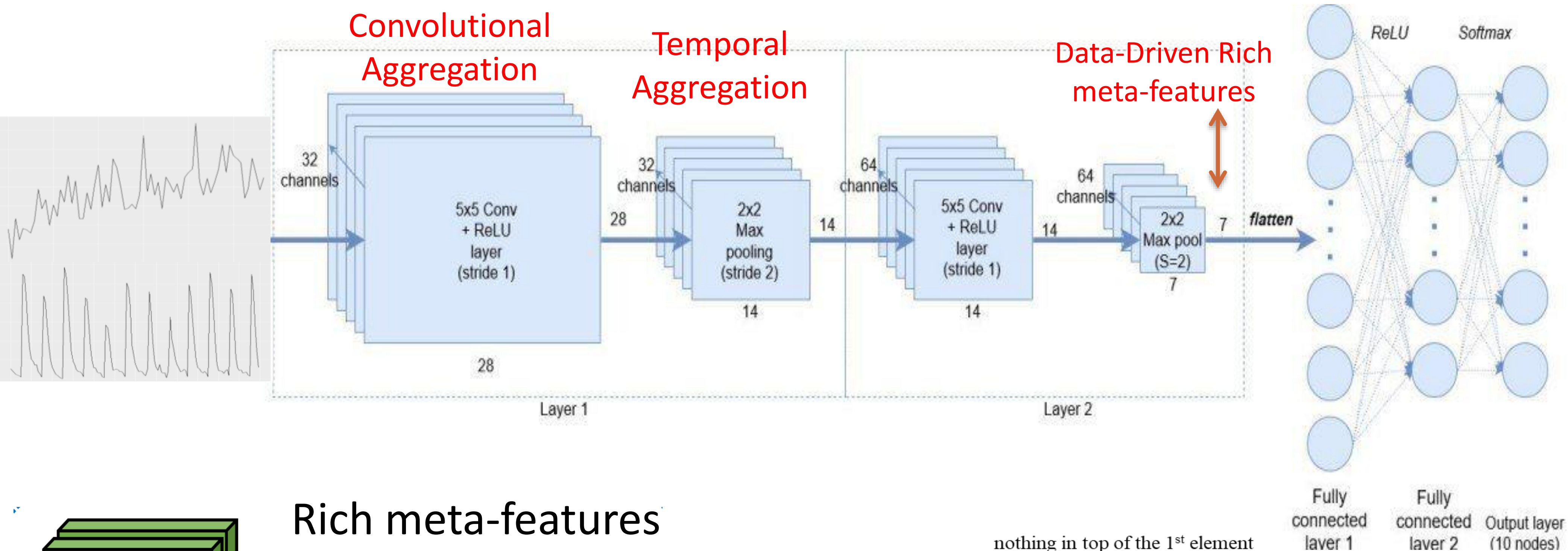
Deep-Learning (CNN)



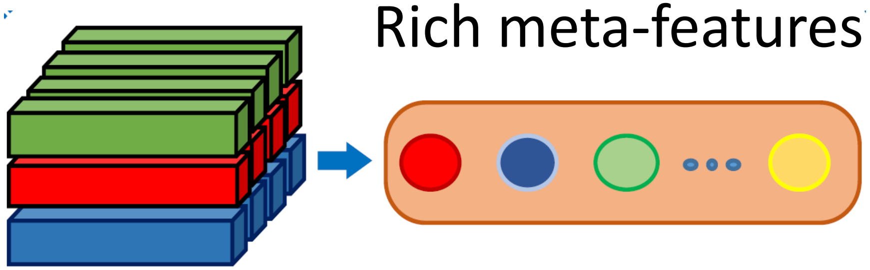
Two key principles:

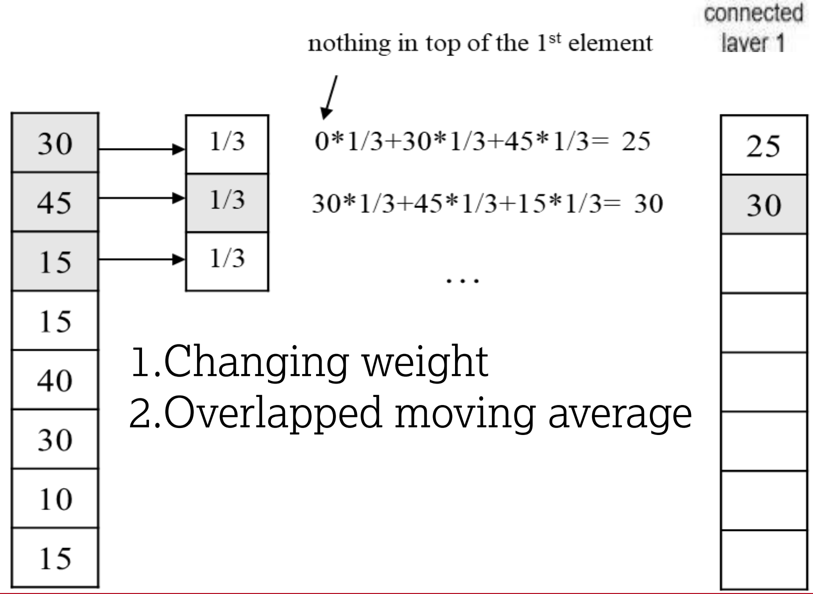
First, the base architecture should be simple and generic, yet deep. Second, the architecture should not rely on time-series-specific feature engineering or input scaling.

What CNN is doing in Forecasting ?



Rich meta-features



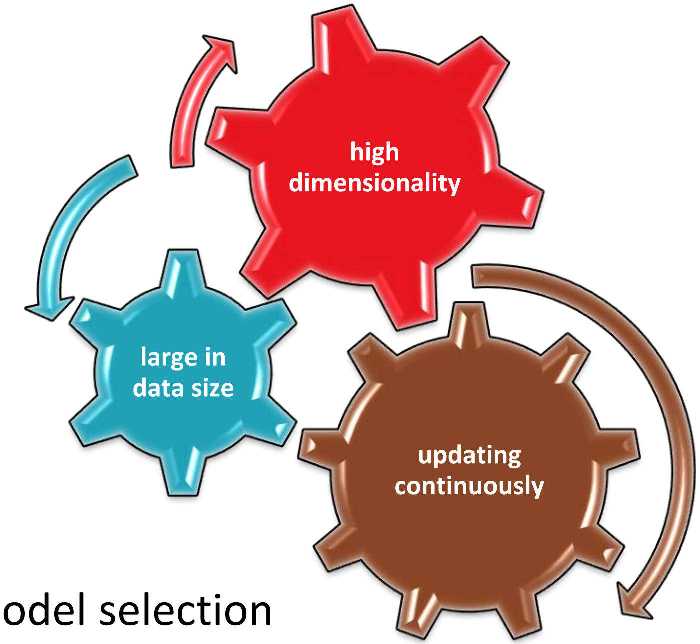


Types of layers:

- Convolution (aggregation)
- Pooling (aggregation)
- Fully connected (classifier-meta learner)

Contributions and Goals

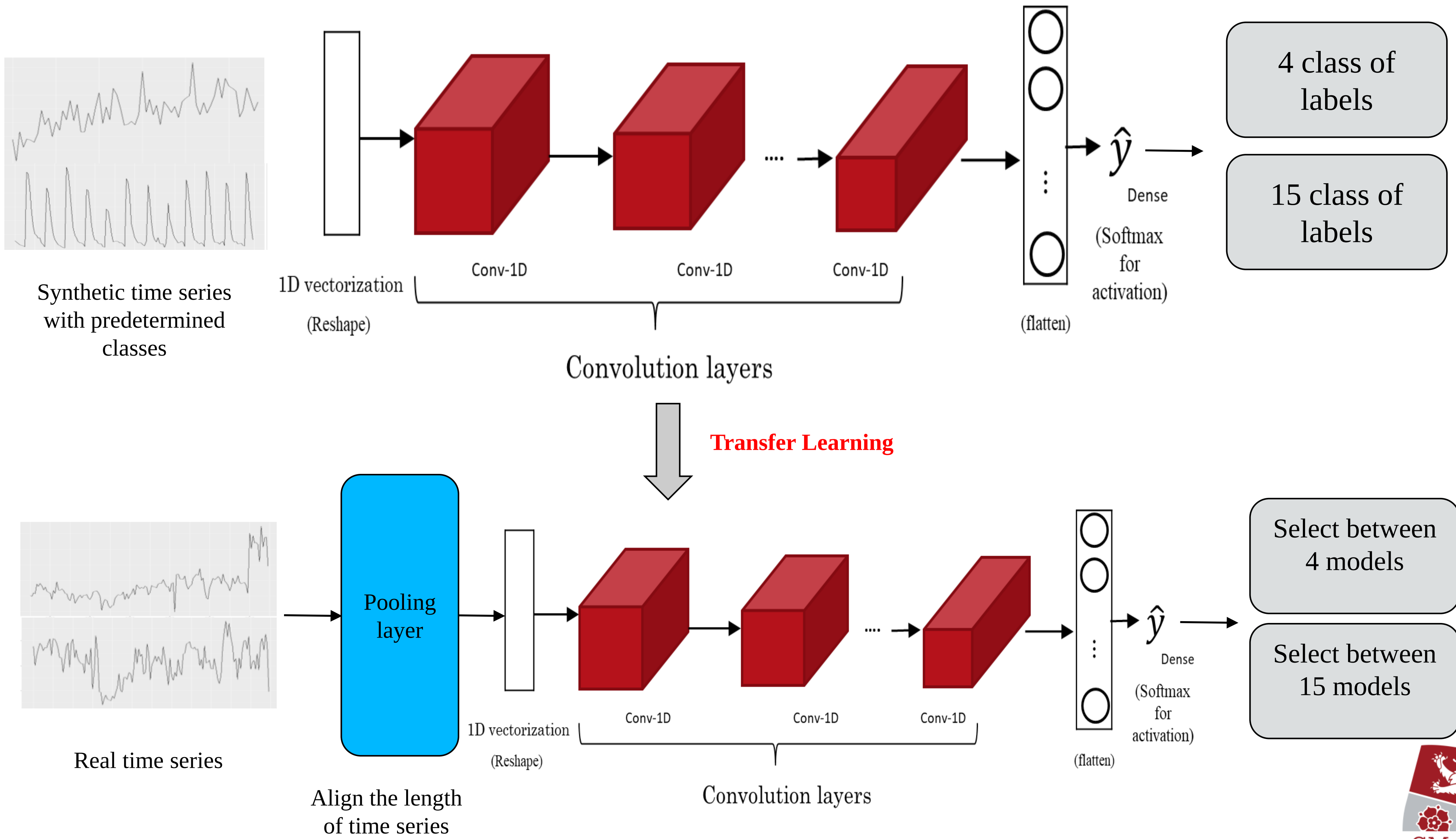
- Practitioners v.s. challenge of big data repository for time series components detection
- Due to the huge dataset we were able to implement complex nonlinear models without encountering much problem of over-fitting.
- Deep networks allowed us to add significant complexity to our model without specifying what forms the variation should take.



Our Goal

1. Implementing the convolutional neural network as a filter model selection
2. Time series components' identification with deep neural network
3. Comparison of the proposed model selection with individual and aggregate model forecasting
4. Evaluating the model robustness for short and long time series

Experimental Design



CMAF

Experimental Design details

Dataset:

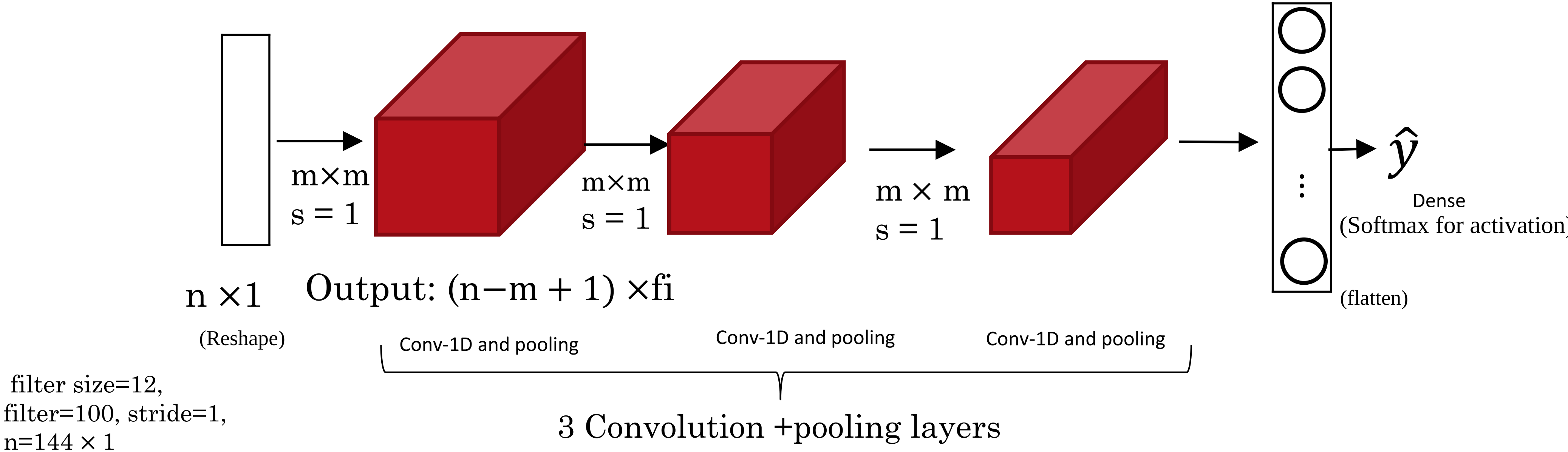
40000-time series with four approaches (ANN, ANA, AAN, ANA)

- Time series without trend and seasonality
- Time series with trend
- Time series with seasonality
- Time series with trend and seasonality

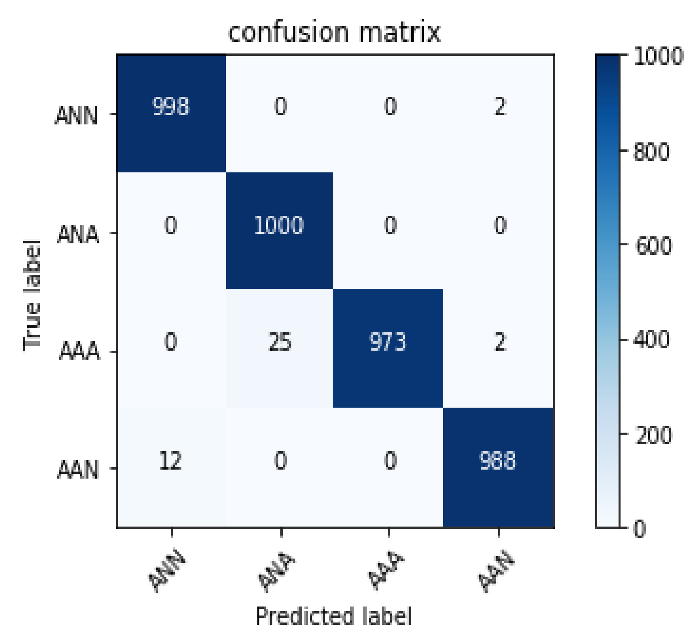
- Random number generation function for error term was randomly chosen between Normal, T-student, Uniform, and Beta distributions.
- Parameters for beta: sshape1=1.5, sshape2=1.5; for normal: mean=0, SD=100; for uniform: min=-0.5, max=0.5; and for T- distribution: mean=0, SD=100
- Frequency : 12
- Lengths: 48 and 144 Test, and Train, and Validation: 10 %, 80 %, 10 %
- Initial states of level, trend, and seasonality are generated randomly.

CMAF

Simulated data: Classification Accuracy



Classification accuracy %	4 class	
Length of time series	48	144
AICc	90.65	93.82
BICc	93.47	94.57
CNN	**98.97**	**98.87**
CS-Friedman	49.82	48.8
MK-Friedman	64.65	60.92



Simulated data: Comparison of forecast error

Two Error Measures: AvgRelAME and AvgRelMAE

CNN Outperforms in both measures for long and short series

AvgRelMAE	4 class	
	48	144
AICc	1	1
BICc	0.9828	1.015
CNN	**0.9460**	**0.9963**
CS-Friedman	1.0458	1.0408
MK-Friedman	1.0276	1.0298
AICc Comb	1	1
BICc Comb	0.9623	1.0262
CNN Comb	**0.9511**	**0.9899**

AvgRelAME	4 class	
	48	144
AICc	1	1
BICc	0.9321	0.9864
CNN	**0.8471**	**0.985**
CS-Fridman	1.1734	1.1873
MK-Friedman	1.1458	1.1851
AICc Comb	1	1
BICc Comb	0.9323	0.9998
CNN Comb	**0.8528**	**0.9921**

Benchmark is AICc

AvgRelAME: Geometric average of Relative Absolute Mean Error

AvgRelMAE: Geometric average of Relative Mean Absolute Error

CMAF

Real data: Comparison of forecast error on real data (M3)

	4 class	
	AvgRelMAE	AvgRelAME
AICc	1	1
BICc	**0.9679**	0.7679
Stat (CS-Friedman)	1.0453	1.0770
Stat(MK-Friedman)	1.1425	1.1250
CNN	**0.9682**	**0.6708**
AICc Comb	1	1
BICc Comb	**0.9742**	0.8262
CNN combined	**0.9722**	**0.7079**

Results are compared to benchmarks of statistical tests for seasonality and trend, wrapper-based model selection using information criteria, and forecast combination errors.

CMAF

Data condition on real data (M3 competition)

AvgRelMAE	Short			medium			long		
	AICc	BICc	CNN	AICc	BICc	Deep	AICc	BICc	CNN
level	1	1	**0.9849**	1	1	1.0319	1	1	**0.9802**
trend	1	0.8998	**0.8819**	1	0.8892	0.9422	1	0.9577	**0.8065**
season	1	0.9913	1.1008	1	0.9625	1.1748	1	1.0033	1.0395
trend-season	1	0.8401	0.9717	1	0.9112	1.2655	1	1.0150	1.1168
overall	1	0.9422	**0.9328**	1	0.9539	1.0968	1	0.9840	**0.9559**

Problem with Seasonal and trend-seasonal

Different Filter size?

CMAF

Future research

1. Generality of the problem for the real data

- a. Finding representative data
- b. Adding noise and data simulating with different Parameters (classes)

2. CNN on Seasonal plot of time series

CMAF

Thank you for your attention!

Q&A?!

Sasan Barak,

S.barak@Lancaster.ac.uk

Marketing Analytics & Forecasting | Lancaster University Management School