Student:	..
Supervisor/s:	Tobias Feigl
Client:	..
Thesis:	Bachelor or Masters Thesis

Pitfalls of the Recurrent Neural Network Family

Keywords: Time Series, (I)RNN, GRU, LSTM, Capacity, Error Carusel, HCNN, ECNN, Forecasting, Stacked-, Dense-, Skipped- Architecture.

Motivation. The analysis of the applicability and the strategic utilization has shown that Historical-Consistent Neural Networks (HCNNs) in particular offer great potential for industrial and macroeconomic applications, as they provide better forecast quality compared to state-of-the-art methods. However, in the scientific discourse and in a series of tests in practical projects, three major challenges have emerged that have hindered the widespread and successful use of HCNNs so far:

Challenge 1 - Benchmarking: Although there is theoretical evidence and first empirical evidence that HCNNs [1] are superior to other state-of-the-art methods such as LSTMs [5] and gradient boosting methods [10] this model class has received little attention in the scientific community so far. This is particularly due to a lack of an evaluation with the necessary scientific rigor. In addition, particularly in macroeconomic applications, it is questioned whether neural networks are suitable for making good long-term forecasts [9]. Although preliminary investigations showed that HCNNs can be superior to other methods with longer forecast horizons, it is still unclear in which (macroeconomic) applications HCNNs predict more precisely than other approaches.

Challenge 2 - Feature Selection: We expect that HCNNs are particularly efficient in economic applications. With these applications, the question arises: *"Which potentially conceivable influencing variables actually provide predictive added value?"* To address this question, either feature selection procedures or manual selection, paired with domain knowledge, are necessary. Conventional feature selection methods are not suitable, since with methods that also forecast the explanatory features (as with HCNNs), it is not only the influence of the feature that is decisive. In addition, the predictability of the features is important as the prediction of the target variable depends on the predictions of the explanatory features. In the case of the HCNN, this feature selection step can be circumvented by an architectural extension. However, as far as we know, this theoretical idea is new and has not been implemented in any software environment known to us, so that a wounding is not yet possible. Furthermore, an analysis is required to what extent this idea can be generalized to LSTMs, ECNNs, and CRCNNs. This could make the issue of feature selection easier for other use cases such as forecasting.

Challenge 3 - Uncertainty Quantification: In most forecast applications it makes sense not only to consider the forecast, but also to take into account the forecast's uncertainty. Only then can a well-founded decision be derived from the prognosis with sufficient certainty. For HCNNs, this has so far only been possible by creating a large ensemble of HCNNs and determining the uncertainty about the spread of the results. However, the high computational effort limits the number of possible models in the ensemble, which means that a sufficiently precise uncertainty cannot be determined. It is also unclear to what extent the errors in the models are also related to the errors in the forecasts. Hence, there is a need to analyze whether the spread of HCNN ensembles is suitable to clarify the uncertainty of the prognosis.

Related work. HCNNs in particular showed their advantage over "no-risk scenarios" in some price forecast applications (electricity price forecast, copper price forecast, steel price forecast, etc.). One of the reasons for this is that not only the target variable is forecast, but also the explanatory features. This allows the model to use the future values of explanatory features when forecasting future time steps, which improves the forecast. This approach is comparable to a vector autoregressive model [6], whereas an HCNN, in contrast, is non-linear and can therefore also explain non-linear relationships.

Student:	..
Supervisor/s:	Tobias Feigl
Client:	..
Thesis:	Bachelor or Masters Thesis

However, since many of these applications were developed within the framework of industrial projects and not within the framework of research activities, an evaluation with the necessary scientific rigor on publicly available datasets has so far been lacking. HCNNs can also quantify the uncertainty due to the spread of the prognosis of the submodels in the ensemble [7]. Alternatively, however, other methods of estimating the uncertainty in the forecast are also conceivable: For example, Salinas et al. [6] estimates the parameters of a Gaussian distribution for each time step. By repeatedly drawing from this distribution, various scenarios are created that show the uncertainty of the forecast. Gal et al. [8] describe Monte Carlo dropout using dropout layers to obtain uncertainties. These methods have not yet been applied to HCNNs.

Overall goal. In this qualification work, the limits of modern data-driven regression methods for analysis and prediction on time series data are to be examined. Specifically:

- (1) (I)RNN, ECN, Elman, GRU, LSTM, and HCNN cells will be implemented;
- (2) Suitable public datasets for evaluating different criteria, e.g., capacity, sequence length, sampling rate, data consistency, frequency, information complexity, short- and long-time dependencies, and input invariance) will be generated;
- (3) The methods are evaluated against the state-of-the-art, e.g., n-Beats, Prophet, Transformer, TCN, ResNet, more?;
- (4) An academic paper will be written.

Idea 1 - Benchmarking: To show the potential of HCNNs in (macroeconomic) applications, it is necessary to achieve significantly better solutions than current state-of-the-art methods (LSTMs, XGBoost) in a comprehensive benchmark with scientific rigor on several relevant datasets , Prophet, etc.) and share these findings with the research community. To carry out the benchmark, a time series framework such as GluonTS, supplemented by in-house developments, should be used, which enables a large number of different models to be compared on many datasets.

Idea 2 - Feature Selection: Instead of performing a feature selection step, all possibly relevant features are entered into the HCNN model. As a rule, this would lead to an input dimension that is too large. For this reason, the features in the model should first be converted into a low-dimensional representation that only contains the necessary information. In contrast to an autoencoder solution, in which the compressed representation contains all information of the input, with this solution only the information relevant to the target variable is retained. This enables an even smaller dimension. This solution significantly reduces the cost of feature selection in macroeconomic applications. So that the solution is publicly available, the architecture is to be implemented in the PyTorch framework in Python and added to the HCNN package. This means that industrial projects in this area can be carried out successfully and much faster.

Idea 3 - Uncertainty Quantification: The goal is to determine whether the required number of models in an ensemble can be reduced through additional scenarios (repeated sampling of the forecast values from a Gaussian distribution) or additional Monte Carlo dropouts. In particular, we want to examine the combination of these three methods. For the evaluation, we check whether the methods described provide quantiles for the prognosis that also agree with the quantiles of the actual values: Are actually 50% of the true values in the 50% quantile of the prognosis? This can be assessed using the so-called Weighted Scaled Pinball Loss [4]. In addition to the direct economic benefit, the quantification of the uncertainty contributes to the explainability of the models [3].

Student:	..
Supervisor/s:	Tobias Feigl
Client:	..
Thesis:	Bachelor or Masters Thesis

Timetable (6 months, in person weeks [PW]).

4 PW Literature and patent research; Familiarization with relevant work on the subject areas;

10 PW Methodological work: adaptation of the individual components to the state-of-the-art methods and advances to the state-of-the-art based on recent deep learning methods;

4 PW Evaluation and real-world demonstration;

6 PW Transcript.

Expected results and scientific contributions.

- ...;
- ...

References

[1] Zimmermann et al. 2012

[2] Ioannides R.T., Pany T., Gibbons G. (2016). Known Vulnerabilities of Global Navigation Satellite Systems, Status, and Potential Mitigation Techniques. Proc. IEEE. 104:1174–1194.

[3] Dominik Seuss 2021

[4] Spyros Makridakis et al. 2020

[5] Hochreiter and Schmidhuber 1997

[6] Salinas et al. 2017

[7] Lakshminarayanan et al. 2016

[8] Gal and Ghahramani et al. 2015

[9] Benidis et al. 2020

[10] Friedman 2001