Uncertainty Estimation for Sequence-to-Sequence Regression on Sparse Time Series

Sotirios Vavaroutas[1], Ting Dang[2], Emma Rocheteau[1], Cecilia Mascolo[1]

Abstract— **Machine Learning (ML) models for sequence-to-sequence tasks predicting one series from another expect continuous time series, but missing points and inconsistencies are common in mobile and wearable data. Additionally, models are often not integrated with uncertainty-aware solutions: uncertainty estimations are crucial, as they can discern confident vs nonconfident predictions. We propose uncertainty-aware sequence-to-sequence prediction on sparse time series. We enhance the state-of-the-art evidential regression with time series interpolation and modify its loss function for irregular series, tuning it to assign different weights to different types of points, as required by distinct uncertainty meanings varying per task and requirement. We also propose novel metrics for assessing the success of uncertainty estimations on sequence-to-sequence predictions, offering a robust way to assess uncertainty given by ML models, as opposed to accuracy-focused metrics.**

I. INTRODUCTION

Sequence-to-sequence regression tasks [1] are prevalent in real-world ubiquitous applications, like in using smartwatch PPG for heart rate monitoring. However, relying solely on accuracy as a measure of model goodness is insufficient, as the uncertainty can offer an orthogonal dimension to the value of the model prediction. Evidential Deep Learning [2] is widely accepted as the state-of-the-art for capturing uncertainty and has been validated for regression [3], but it has neither been verified nor tailored on sparse time series, widespread in cases of sensor data loss. Additionally, many commonly-used uncertainty evaluation metrics like the Brier score and the Expected Calibration Error are designed for classification and are rarely applicable to regression, while others like the Negative Log Likelihood [4] only give indirect insights about uncertainty. Also, many works focus on representing model uncertainty, but there are no metrics to quantify and compare the success and usefulness of these uncertainty estimates.

II. METHODS

Interpolation Sensors suffer from transmission errors or temporary malfunctions, leading to irregular time series with missing or incomplete data. To address this, we interpolate missing values, ideal for data with an irregular sampling rate.

Loss Function for Evidential Regression We adapt the loss function of Deep Evidential Regression [3] to accommodate for the irregularity in sparse time series by weighting the interpolated points and the actual values of the series on a point-by-point basis. This allows the model to focus more or less on the real points or the missing points for uncertainty estimation, depending on the specific application. As a result, this method provides the advantage of tuning the uncertainty in the model's predictions based on different requirements.

Uncertainty Metrics We introduce novel uncertainty metrics to assess uncertainty representations given by different solutions. The first is the Uncertainty Mean Absolute Error (Uncertainty MAE), and the second is the Uncertainty Mean Absolute Percentage Error (Uncertainty MAPE). These are calculated by adding the uncertainty to the error of the predicted values and can quantify the "worst case" prediction in critical tasks. The third is the ratio of samples outside the range of predicted values ± uncertainty, while the fourth is the MAE of samples outside that range. These allow for evaluating the "trade-off" between having a lower or a more accurate uncertainty, and help quantify its significance.

III. EVALUATION

Case Studies First we predict the hypoglycemia risk (output sequence) of type 1 diabetics given readings from Continuous Glucose Monitoring (CGM) wearables. This can reduce the chances of hypoglycemia that can occur suddenly and with symptoms often preventing the user from treating their condition. For training, we use UVA/PADOVA data [5].

Second, we focus on predicting weekly COVID-19 hospital admissions (output sequence), given the percentages testing positive in a population (input sequence) using weekly trends data [6]. This can help in planning hospital resources.

Baselines For ground truth, we start with regular datasets and turn them to irregular by randomly dropping a set of samples. This gives an upper case baseline in which the model is trained on the regular dataset ("best-case"), and a lower case baseline in which the model is trained on its irregular counterpart ("worst-case") using conventional EDL.

Model Structure We use an LSTM network [7], being the most well-known for sequence-to-sequence regression.

Results On hypoglycemia risk, the upper baseline shows a higher uncertainty than the lower baseline, which can also identify when the user might be suffering from previously-unseen symptoms and needs to take corrective action. Thus, we tune our proposal for more confident results, and achieve confidence intervals up to 3 times more accurate, just like the upper baseline. Using our novel metrics, we identify that this captures the true risk for 90% of the samples, compared to overconfident alternatives erroneously estimating uncertainty.

On COVID-19 data, the upper case shows a smaller uncertainty, which is expected as the data benefits from more accurate predictions and less variability. Here, we tune for moderately more samples outside the range of predicted values ± uncertainty, and we achieve a 20% lower MAE, and a 3 times smaller and more focused uncertainty estimation.

IV. CONCLUSION

Our work puts forward a novel approach and evaluation metrics for uncertainty-aware sequence-to-sequence predictions in the context of sparse time series, which are common in data harvested from mobile and wearable devices.

REFERENCES

[1] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," *CoRR abs/1409.3215*, 09 2014.
[2] M. Sensoy, M. Kandemir, and L. M. Kaplan, "Evidential deep learning to quantify classification uncertainty," *CoRR abs/1806.01768*, 06 2018.
[3] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," *CoRR abs/1910.02600*, 10 2019.
[4] P. A. Bosman and D. Thierens, "Negative log-likelihood and statistical hypothesis testing as the basis of model selection in ideas," 08 2000.
[5] C. D. Man, F. Micheletto, D. Lv, M. Breton, B. Kovatchev, and C. Cobelli, "The UVA/PADOVA type 1 diabetes simulator," 01 2014.
[6] UK Office for National Statistics, "Coronavirus latest insights," 01 2023.
[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," 11 1997.

[1]University of Cambridge
[2]Nokia Bell Labs, Cambridge