Gluformer: Transformer-Based Personalized Glucose Forecasting with Uncertainty Quantification

Renat Sergazinov Mohammadreza Armandpour Irina Gaynanova

Department of Statistics, Texas A&M University

Background

According to the International Diabetes Federation, **463 million adults worldwide have diabetes** [1]. **Glucose management** is a critical component of diabetes care. **Continuous glucose monitoring (CGM)** is a technology that enables patients to track their glucose levels at **short regular intervals**. Coupled with a **predictive model**, CGM can help people with diabetes manage their blood glucose levels either by providing **feedback** or as a part of **artificial pancreas (AP)** that can regulate blood glucose automatically.

Main limitation of the models: poor uncertainty quantification or poor distributional fit.

Data generation: consider a **toy example**, where the **data is generated** following:

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$$
$$\boldsymbol{\mu}_{1:n} \sim \mathcal{N}(\mathbf{0}, I_n) \qquad \boldsymbol{\mu}_{n:2n} \sim \frac{1}{2}\mathcal{N}(\mathbf{0}, I_n) + \frac{1}{2}\mathcal{N}(\mathbf{1}, I_n).$$

Why is it relevant to the CGM data?

1. Patients typically adhere to daily schedules (periodicity).
2. Fixed part corresponds to fixed portions of the schedule.
3. Mixture part corresponds to the variable portion of the schedule.

Setup: given past observations, $\boldsymbol{y}_{1:t}$, we want to form a predictive distribution $p(\boldsymbol{y}_{t:t_l}|\boldsymbol{y}_{1:t})$. Our training data consists of randomly subsampled chunks: $\{\boldsymbol{y}_{1:t}, \boldsymbol{y}_{t:t+l}\}$.

We consider 2 models:

1. Transformer (DL) trained with **MSE** $\Leftrightarrow$ additive Gaussian assumption on the noise,
2. Transformer (DL) trained with our infinite mixture model **(IMM)** objective.

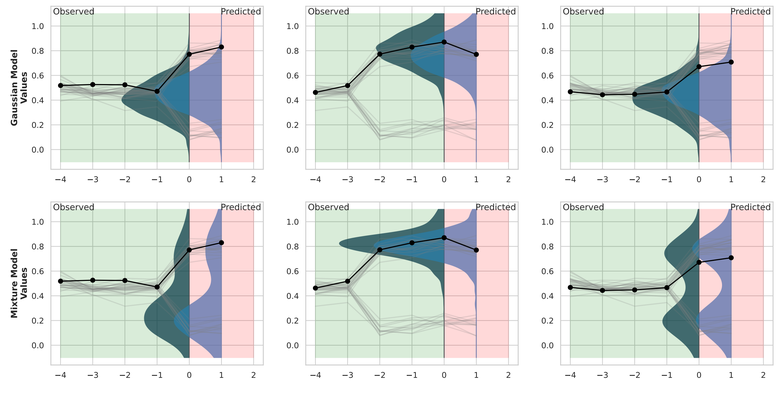

Figure 1. **Top:** model trained with **MSE**, **bottom:** our model trained with **IMM**.

We change model loss by conditioning on latent variables $\boldsymbol{z}$:

$$\boldsymbol{y}_{t:t+l}|\boldsymbol{y}_{1:t}, \boldsymbol{z} \sim \mathcal{N}(\hat{\boldsymbol{\mu}}, \hat{\sigma}^2 I_l) \qquad (\hat{\boldsymbol{\mu}}, \hat{\sigma}^2) = f_\theta(y_{1:t}|z) \qquad \boldsymbol{z} \sim p_{\boldsymbol{z}}.$$

We optimize model parameters θ to minimize negative log-likelihood:

$$\theta^\star = \arg\min \left(-\log \int p(\boldsymbol{y}_{t:t+l}|\boldsymbol{y}_{1:t}, \boldsymbol{z}) p(z) d\boldsymbol{z} \right).$$

During training, we approximate the integral using Monte Carlo, which amount to several **stochastic passes through the network**.

During inference, we form the predictive distribution as a finite mixture using the stochastic passes:

$$p(\boldsymbol{y}_{t:t+l}|\boldsymbol{y}_{1:t}) \approx \frac{1}{n}\sum_1^n \mathcal{N}(\hat{\boldsymbol{\mu}}_{\boldsymbol{i}}, \hat{\sigma}_i^2 I_l) \qquad (\hat{\boldsymbol{\mu}}_i, \hat{\sigma}_i^2) = f_\theta(\boldsymbol{y}_{1:t}|z_i) \qquad z_i \sim p_{\boldsymbol{z}}.$$

Overall, the **computational cost grows linearly** with the number of estimated components.

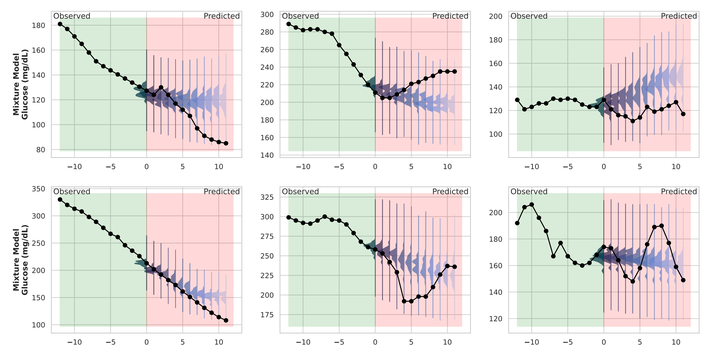

Figure 2. Predictive distribution for a sample of CGM curves based on **IMM**.

Key takeaway: The effect of unobserved variables on future glucose trajectories presents a challenge to traditional approaches for uncertainty quantification. We propose a novel approach that explicitly incorporates the impact of latent covariates on glucose dynamics, resulting in accurate uncertainty quantification.

We use a publicly available CGM data set [2], which contains information on glucose levels of **38 subjects** tracked continuously throughout multiple disjoint intervals with the measurement frequency of 5 minutes. Similar to the previous studies, we apply the following pre-processing steps:

1. Interpolation of small gaps;
2. Drop readings where measurements fluctuate by more than 40 mg/dl in 5 minutes;
3. Split into train / validation / test in chronological order in 20:1:1 proportion.

Total: 399,302 measurements with 30 uninterrupted sequences per subject.

Results

Model	Full	Event	Hypo	Hyper	Likelihood
ARIMA	9.85 / 17.65	8.91 / 19.86	19.94 / 14.53	8.51 / 22.17	-14.93
RF:Rec	9.04 / 17.15	8.97 / 20.36	18.84 / **12.43**	8.68 / 23.41	-14.58
RF:MO	10.22 / 18.27	8.61 / 19.90	21.64 / 17.36	7.99 / 21.58	-15.34
PolySeqMO	8.55 / 15.68	8.27 / 18.81	22.86 / 21.87	**6.77 / 18.30**	-15.61
RNN	8.17 / 15.67	8.29 / 19.37	18.72 / 16.26	6.99 / 19.22	-13.50
TFT	7.80 / 15.78	8.03 / 18.23	16.23 / 14.62	6.87 / 18.98	–
Our: IMM	**7.78** / 15.40	**7.89 / 17.85**	**15.75** / 14.03	7.08 / 19.58	**-2.67**
Our: Gaussian	7.82 / **14.73**	7.93 / 18.62	18.28 / 13.55	6.81 / 18.94	-12.82

Table 1. APE/RMSE for 60-minute prediction window (Full), hypoglycemia, hyperglycemia, and event (hypo-or hyperglycemia), and model log-likelihood on test data.

Further considerations that we have addressed to make the model and fitting more stable:

1. **Personalization.** The model is trained simultaneously on the data from all patients. We learn an embedding of patient-specific demographic information and add it to the model input.
2. **Long-term temporal dependence.** To be able to accommodate larger context window size, i.e., accommodate large t, we use an adaptation of the Transformer that down-samples the input in-between attention layers [4].
3. **Within-batch dependence.** During training, samples for each mini-batch of gradient descent are taken from different patients or different CGM curves for a patient.

Computational cost of training the model on the data set:

1. GPU: NVIDIA RTX 2080 with 12GB
2. Each epoch of SGD: $\approx$ 10 minutes
3. Total time for 100 epochs: 16 hours

Future work:

1. Learn latent distribution $p_{\boldsymbol{z}}$ to optimally explore different regimes;
2. Learn number of *active* mixture components;
3. Incorporate different data sources;
4. Conditional prediction based on future activity.

[1] International Diabetes Federation. *International Diabetes Federation Diabetes Atlas*. International Diabetes Federation, 2021.

[2] Ian Fox, Lynn Ang, Mamta Jaiswal, Rodica Pop-Busui, and Jenna Wiens. Deep Multi-Output Forecasting: Learning to Accurately Predict Blood Glucose Trajectories. *ACM SIGKDD*, pages 1387–1395, July 2018.

[3] Silvia Oviedo, Josep Vehí, Remei Calm, and Joaquim Armengol. A review of personalized blood glucose prediction strategies for T1DM patients. 33(6):e2833.

[4] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI*, volume 35, pages 11106–11115, 2021.