MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

Personalized Prediction of Parkinson's Disease Progression Based on Deep Gaussian Processes

Changrong Pan

Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China

Introduction

- Parkinson's disease (PD) is a chronic progressive neurodegenerative disorder, whose motor syndrome mainly includes bradykinesia, rigidity, rest tremor and postural instability.
- Inter-individual and intra-individual heterogeneity in PD leads to uncertainty of diagnosis, especially in the early stages, so developing highly personalized methods for accurate prediction of PD progression is necessary.

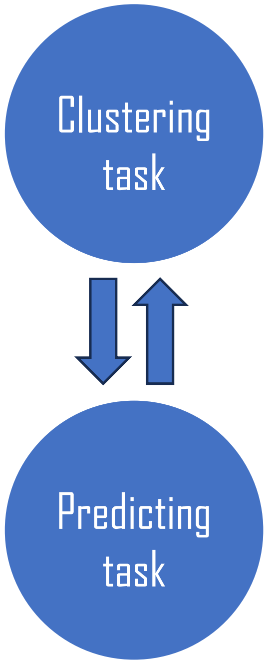
Introduction

- Some research about PD progression have been reported.
- The present research cannot achieve good predictive performance, model interpretability and data heterogeneity processing at the same time.

Author	Year	Research purpose	Method
Latourelle JC, et al	2017	Predicting UPDRS score progression and identifying novel predictors	GNS Healthcare's REFS platform
Nilashi M, et al	2018	Discovering the relationship between speech signal properities and UPDRS score	Incremental Support Vector Regression
Shahid AH, et al	2020	Predicting UPDRS score progression	Deep neural network
Severson KA, et al	2021	Defining PD disease states	Hidden Markov model

Introduction

- We propose a novel multi-task learning framework that divides PD progression modeling into **an unsupervised clustering task** and **a disease progression prediction task**.
  - For the clustering task, the method can discover new progression patterns which can help predict the future progression more accurately through parameter sharing.
  - For the prediction task, a combination of neural networks and GP regression is used for personalized prediction.



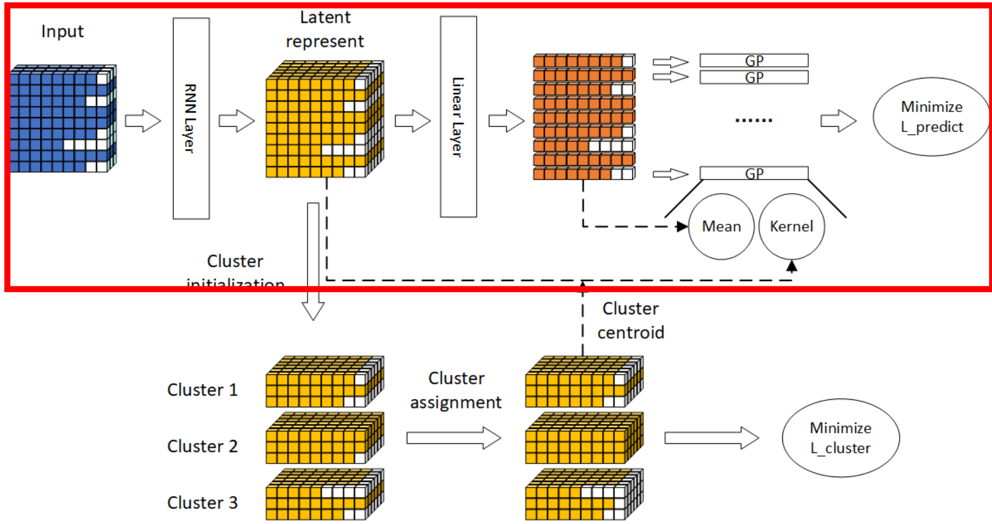
Methods - *GP prediciton*

A GP-based deep mixed-effect framework contains global function and personalized function for the i^{th} patient(Chung I et al, 2020) . We further assume that the i^{th} patient belong to the L_i cluster:

$$f^{(i)}(x_t) = g(x_t) + l^{(i)}(x_t)$$

$$\Downarrow$$

$$f^{(i)}(x_t) = g(x_t) + l^{(i)}(x_t) + c^{(L_i)}(x_t)$$



Flow chart of disease progression model based on multi-task learning framework

Methods - *GP prediciton*

The framework: $f^{(i)}(x_t) = g(x_t) + l^{(i)}(x_t) + c^{(L_i)}(x_t)$

- RNN layer: $g(x_t) = \mu(x_t)$
  - $h_t = \tanh(v_{xh}x_t + v_{hh}h_{t-1})$
  - $w = \{v_{xh}, v_{hh}\}$
- $l^{(i)}(x_t) \sim \mathcal{GP}\left(0, k^{(i)}(x_t, x_{t'})\right)$
- $c^{(L_i)}(x_t) \sim \mathcal{GP}\left(0, k^{(L_i)}(x_t, x_{t'})\right)$

- Matrix representation:

$$\begin{bmatrix} f^{(1)} \\ \vdots \\ f^{(P)} \end{bmatrix} \sim \mathcal{GP}\left(\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_P \end{bmatrix}, \begin{bmatrix} K^{(1)} + K^{*(L_1)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K^{(P)} + K^{*(L_P)} \end{bmatrix}\right)$$

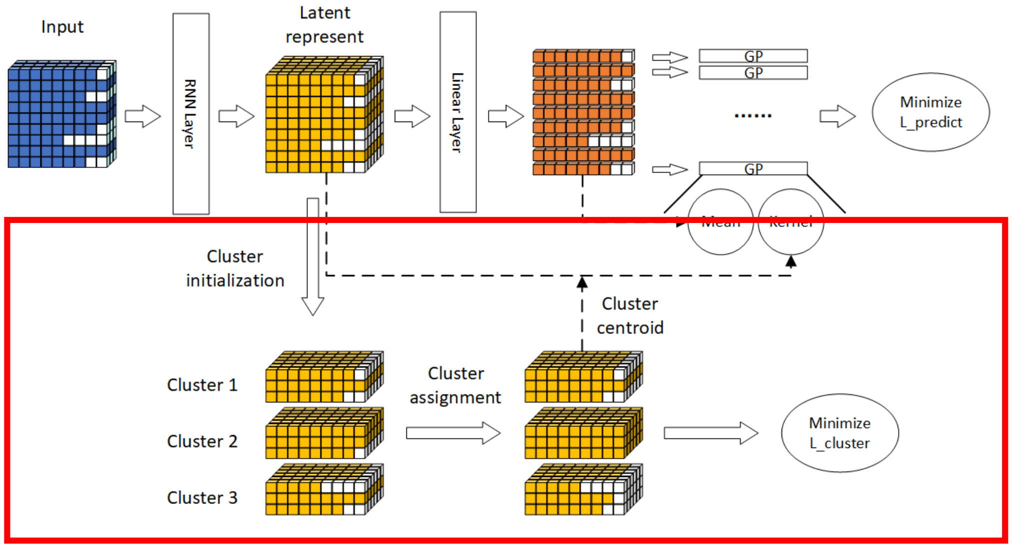
$$\Downarrow$$

$$f^{(i)}(x_t) \sim \mathcal{GP}\left(\mu(h_t|w), k'^{(i)}(h_t, h_{t'}, \theta_i, L_i)\right)$$

Methods - *Cluster layer*

Cluster initialization:

- Traditional clustering method is used to cluster the output signal of the initial hidden layer obtained by the RNN layer.
- Average the signals of each cluster to obtain the initial label and the initial estimated clustering centroids.



Methods - *Cluster layer*

Cluster assignment: the cluster label is assigned with the largest probability and the cluster centroids are re-estimated according to the new cluster results.

- The similarity $D(x, y)$ between the i^{th} patient and different centroids is computed based on Complexity Invariant Similarity(CID):

$$D(x,y) = ED(x,y)\,\frac{\max(CE(x), CE(y))}{\min(CE(x), CE(y))}$$

- The soft assignment probability distribution q_{ij} is computed based on the student t-distribution kernel.

$$q_{ij} = \frac{(1 + D(x,y))^{-1}}{\sum_{j=1}^{C}(1 + D(x,y))^{-1}}$$

Methods - *Cluster layer*

KL divergence is used to compute the loss $\mathcal{L}_{cluster}$ between the probability distribution q_{ij} and the target probability distribution p_{ij}.

$$\mathcal{L}_{cluster} = \sum_{i=1}^{N} \sum_{j=1}^{C} p_{ij} log \frac{p_{ij}}{q_{ij}}$$

- Auxiliary distribution is considered as the target distribution p_{ij}:

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{N} q_{ij}}{\sum_{j=1}^{C} \left(q_{ij}^2 / \sum_{i=1}^{N} q_{ij} \right)}$$

Methods - *Inference*

The goal of the learning is to maximize the marginal log-likelihoods $\mathcal{L}_{predict}$ under the assumption of the multi-task learning framework to optimize global parameters v_{xh}, v_{hh} and personalized parameters $\theta = \{\theta_i\}_{i=1}^{N}$ and meanwhile to minimize the KL divergence loss $\mathcal{L}_{cluster}$ to optimize clustering assignment.

$$\mathcal{L}_{predict} = \sum_{i=1}^{N} logp(Y_i|X_i|v_{xh}|v_{hh}|\theta_i)$$

$$\frac{\partial \mathcal{L}_{predict}}{\partial \theta_i} = \frac{\partial \mathcal{L}_{predict}}{\partial K^{(i)}} \cdot \frac{\partial K^{(i)}}{\partial \theta_i}$$

$$\frac{\partial \mathcal{L}_{predict}}{\partial v} = \frac{\partial \mathcal{L}_{predict}}{\partial K^{(i)}} \cdot \sum_{t=1}^{T_i} \frac{\partial \mathcal{L}_{predict}}{\partial h_t} \cdot \frac{\partial h_t}{\partial v} + \sum_{t=1}^{T_i} \frac{\partial \mathcal{L}_{predict}}{\partial g_t} \cdot \sum_{t'=1}^{T_i} \frac{\partial g_t}{\partial h_{t'}} \cdot \frac{\partial h_{t'}}{\partial v}$$

$$\mathcal{L}_{cluster} = \sum_{i=1}^{N} \sum_{j=1}^{C} p_{ij} log \frac{p_{ij}}{q_{ij}}$$

$$\frac{d\mathcal{L}_{cluster}}{dc_j} = \frac{1}{2} \sum_{j=1}^{C} \left(\left(1 + D(g_i, c_j)\right) * (p_{ij} - q_{ij}) \frac{d\left(D(g_i, c_j)\right)}{dc_j} \right)$$

c_j: the j^{th} centroids
g_i: the i^{th} representation

Results - *Dataset*

- The Parkinson's Progression Markers Initiative (PPMI) is an international, multicenter, prospective study designed to discover and validate biomarkers of disease progression in newly diagnosed PD participants.
- We obtained a PD dataset containing **361 patients**, of which had **at least 9 follow-up visits spanning 3 years** and **at most 16 follow-up visits spanning 8 years** with **95 variables** based on general information and classical clinical scales.

For up-to-date information on the study, visit www.ppmi-info.org.

Results - *Prediction Performance*

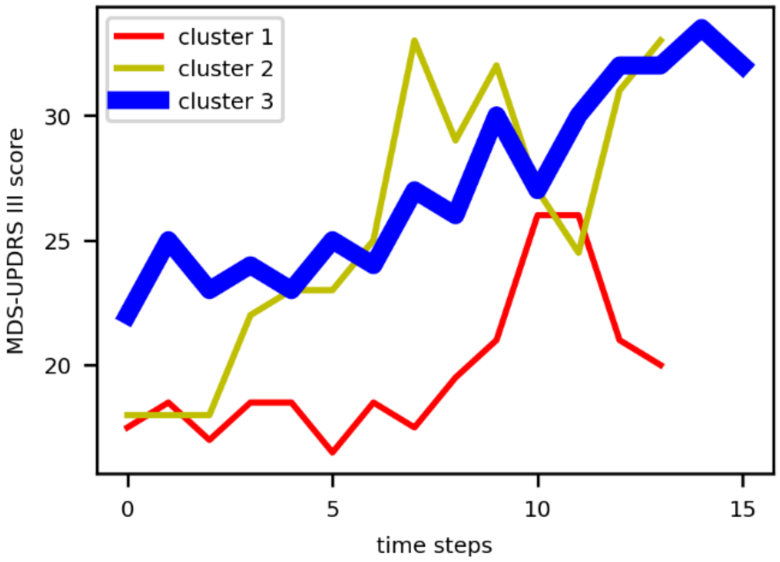
Ten-fold cross validation was used for evaluation. Our proposed model has the relatively best prediction performance, and the best performance is obtained when the cluster count is specified as 3.

Model	MAE	RMSE	R^2
GP	5.506	8.300	0.646
RNN	5.593	7.932	0.676
LSTM	5.534	7.887	0.680
GRU	5.397	7.648	0.699
DME-GP	5.222	7.392	0.719
Proposed Method(C=2)	5.179	7.331	0.724
Proposed Method(C=3)	**5.108**	**7.291**	**0.727**
Proposed Method(C=4)	5.194	7.339	0.723

The p-value were 0.006, 0.018 and 0.003 respectively for MAE, RMSE and R2.

Results - *Clustering Results*

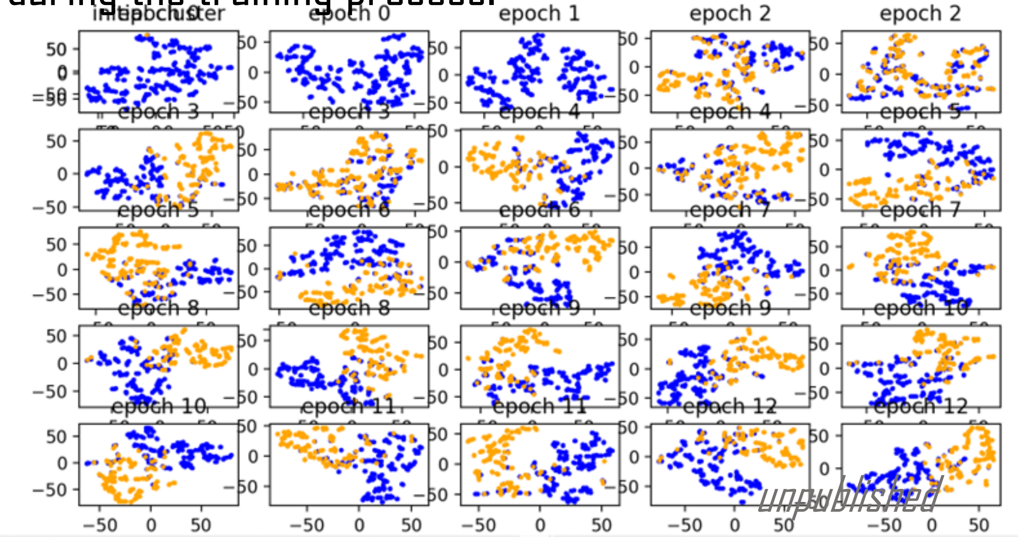
- The thickness of the lines represents the count of patients in the clusters.
- two progression patterns are identified among population:
  - the score remains at a relatively low level and even drops in the later stages in cluster 1.
  - the score fluctuates widely from low level to high level in cluster2.



Results - *Unpublished*

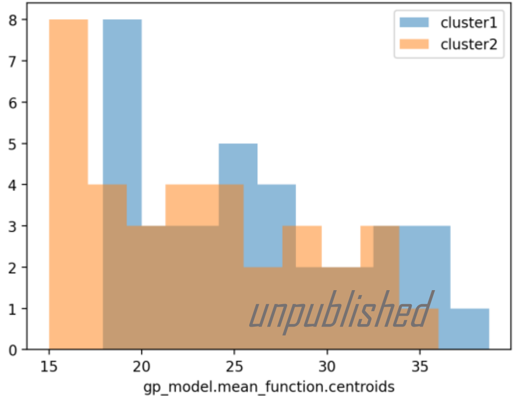
We further designed a training process, taking the clustering results as the sampling basis, and constantly updating the clustering results during the training process.

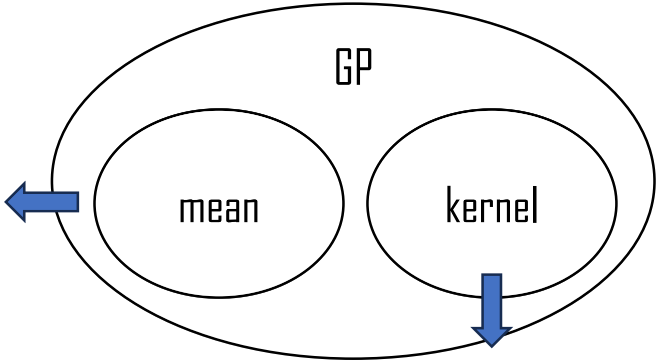
- No need to calculate $\mathcal{L}_{cluster}$.
- The situation of empty cluster in training is improved.



Results - *Unpublished*

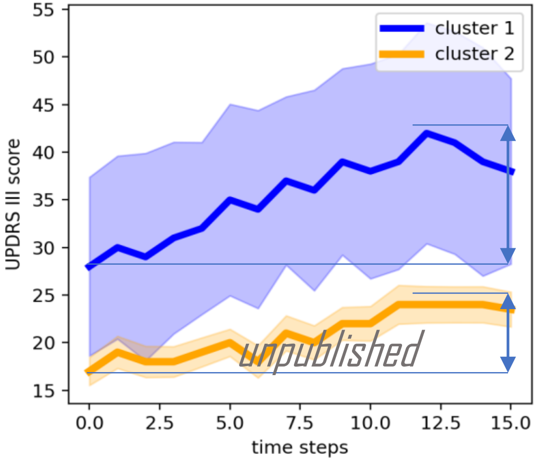
Interpretable analysis of disease progression:





$$k_{\mathrm{EQ}}\left(x, x'\right) = v^2 \exp\left(-\frac{\left(x - x'\right)^2}{2\ell^2}\right)$$

$$v_{cluster1} > v_{cluster2}$$



Solid line thickness: Number of samples within the cluster

Shadow area: Variance of the cluster

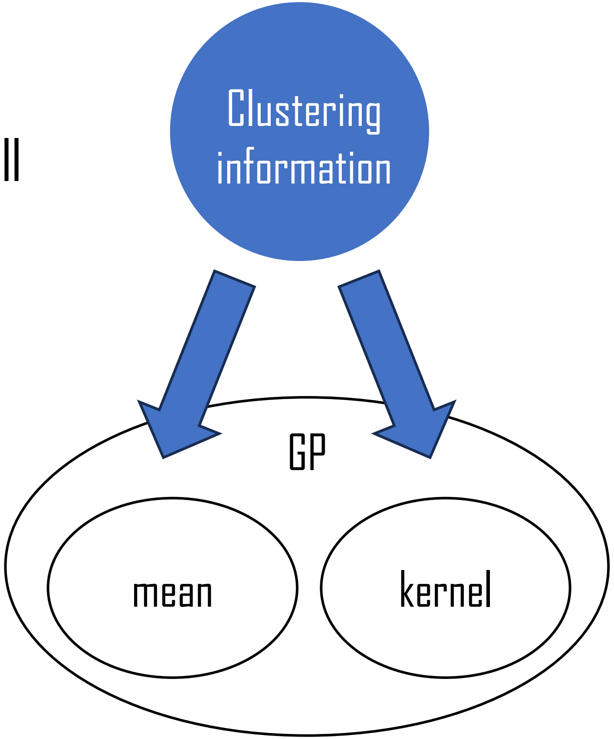
Discussion

- We confirms **that integrating heterogeneity among clusters is helpful to improve the prediction performance of the models**.
- Unlike the probability graph models, the clusters we obtained are **specific to patients rather than disease states**, which is clinically more meaningful.

Discussion

Different clusters may only have significant differences in the overall mean, but not in their position, trends, smoothness, etc.

In the future work, we can further integrate the cluster information into the kernel function and the mean function respectively or simultaneously.



Conclusions

In summary, we propose a novel approach based on a multi-task learning framework to divide Parkinson's disease progression modeling into an unsupervised clustering task and a disease progression prediction task.

- prediction
- heterogeneity
- interpretability

Inference

[1]Kalia LV, Kalia SK, and Lang AE, Disease-Modifying Strategies for Parkinson's Disease, Movement Disorders 30 (2015), 1442-1450.
[2]Ma L-Y, Tian Y, Pan C-R, et al., Motor Progression in Early-Stage Parkinson's Disease: A Clinical Prediction Model and the Role of Cerebrospinal Fluid Biomarkers, Frontiers in Aging Neuroscience 12 (2021).
[3]Latourelle JC, Beste MT, Hadzi TC, et al., Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation, Lancet Neurology 16 (2017), 908-916.
[4]Shahid AH and Singh MP, A deep learning approach for prediction of Parkinson's disease progression, Biomedical Engineering Letters 10 (2020), 227-239.
[5]Nguyen KP, Raval V, Treacher A, et al., Predicting Parkinson's disease trajectory using clinical and neuroimaging baseline measures, Parkinsonism & Related Disorders 85 (2021), 44-51.
[6]Severson KA, Chahine LM, Smolensky LA, et al., Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning, Lancet Digital Health 3 (2021), E555-E564.
[7]Ziegler G, Ridgway GR, Dahnke R, et al., Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects, Neuroimage 97 (2014), 333-348.
[8]Peterson K, Rudovic OO, Guerrero R, et al., Personalized Gaussian Processes for Future Prediction of Alzheimer's Disease Progression, 31st Conference on Neural Information Processing Systems (NIPS 2017) (2018).
[9]Chung I, Kim S, Lee J, et al., Deep Mixed Effect Model Using Gaussian Processes: A Personalized and Reliable Prediction for Healthcare, Proceedings of the AAAI Conference on Artificial Intelligence 34 (2020), 3649-3657.
[10]Batista GEAPA, Wang X, and Keogh EJ, A Complexity-Invariant Distance Measure for Time Series, Proceedings of the 2011 SIAM International Conference on Data Mining (SDM) (2011), 699-710.
[11]Hinton GE, Osindero S, and Osindero S, A Fast Learning Algorithm for Deep Belief Nets, Neural Computation 18 (2006), 1527–1554.
[12]Ibrahim ZM, Wu HH, Hamoud A, et al., On classifying sepsis heterogeneity in the ICU: insight using machine learning, Journal of the American Medical Informatics Association 27 (2020), 437-443.
[13]Beaulieu-Jones BK, Greene CS, and Open-Access PR, Semi-supervised learning of the electronic health record for phenotype stratification, Journal of Biomedical Informatics 64 (2016), 168-178.

Thank you for listening!

Changrong Pan
Zhejiang University
2023.7.12