MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231068

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

Changrong PAN[a], Yu TIAN[a], Tianshu ZHOU[b] and Jingsong LI[a,b,1]
[a] *Engineering Research Center of EMR and Intelligent Expert System, Ministry of Education, College of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, China*
[b] *Research Center for Healthcare Data Science, Zhejiang Laboratory, Hangzhou, China*

Abstract. Parkinson's disease is a chronic progressive neurodegenerative disease with highly heterogeneous symptoms and progression. It is helpful for patient management to establish a personalized model that integrates heterogeneous interpretation methods to predict disease progression. In the study, 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. On the one hand, the method can cluster patients with different progression trajectories and discover new progression patterns of Parkinson's disease. On the other hand, the discovery of new progression patterns helps to predict the future progression of Parkinson's disease markers more accurately through parameter sharing among multiple tasks. We discovered three different Parkinson's disease progression patterns and achieved better prediction performance (MAE=5.015, RMSE=7.284, r2=0.727) than previously proposed methods on Parkinson's Progression Markers Initiative datasets, which is a longitudinal cohort study with newly diagnosed Parkinson's disease.

Keywords. Gaussian process, Parkinson's disease, disease progress prediction, multi-task learning

1. Introduction

Parkinson's disease (PD) is a chronic progressive neurodegenerative disorder [1]. Inter-individual and intra-individual variabilities 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 [2].

Recent studies about PD progression based on statistical analysis, machine learning and classical deep learning have been reported [3-5] but it's difficult to balance the prediction accuracy and prediction interpretability due to the large heterogeneity of PD symptoms. Besides, probability graph models have been proposed to discover different PD states and depict the disease progression trajectory [6]. However, the disease states cannot be well evaluated on account of the lack of powerful assessment methods.

[1] Corresponding Author: Jingsong LI, ljs@zju.edu.cn.

Gaussian process (GP) is a popular non-parametric model which is suitable for personalized prediction of temporal data [7,8]. Ingyo et al. proposed a GP-based deep mixed-effect framework (DME-GP), which combined neural networks with multiple GPs [9]. Although these models have relatively good prediction performance, they cannot well explain the clinical heterogeneity of PD which has been proved [1,2].

To address the problems mentioned above, we propose a novel multi-task learning framework that divides PD progression modeling into a disease progression prediction task and an unsupervised clustering task.

2. Methods

2.1. Disease Progression Model Based on Multi-task Learning Framework

An overview of the flow chart of disease progression model based on multi-task learning framework is shown in Figure 1 (with the total clusters count C=3 as an example).

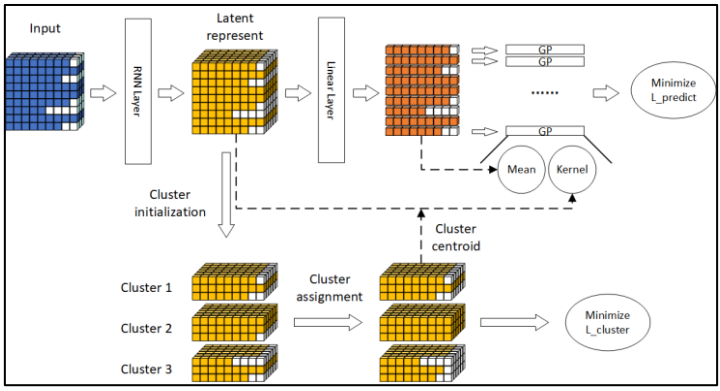

Figure 1. Flow chart of disease progression model based on multi-task learning framework.

As is mentioned obviously [9], a GP-based deep mixed-effect framework contains global function and personalized function for the i^{th} patient. We assume that the i^{th} patient belongs to the L_i cluster, so the multi-task learning framework can be represented as Eq. (1), where $g(\cdot)$, $l^{(i)}(\cdot)$ and $c^{(L_i)}(\cdot)$ respectively model the common trends of all patients, the i^{th} patient and each cluster of patients. We use GP to represent $l^{(i)}(\cdot)$ and $c^{(L_i)}(\cdot)$. We assume that the difference between different clusters can be reflected in the kernel function. So, the overall kernel function can be represented as Eq. (2), where $k^{(i)}(\cdot,\cdot)$ is the covariance function of GP. All the covariance functions in the paper use the common squared exponential kernel (RBF). δ_{ij} is the Kronecker delta function. $m_t^{L_i}$ is the cluster centroid of the ${L_i}^{th}$cluster which will be explained further in the section 2.2. We use RNN to represent $g(\cdot)$ to capture complex patterns in high-dimensional medical data in a relatively easy way to compute. Finally, the framework is represented as Eq. (3) and the loss $\mathcal{L}_{predict}$ can be computed.

$$f^{(i)}(x_t) = g(x_t) + l^{(i)}(x_t) + c^{(L_i)}(x_t) \tag{1}$$

$$\tilde{k}\left(x_t^{(i)}, x_{t'}^{(j)}\right) = \delta_{ij} \cdot k^{(i)}\left(x_t^{(i)} + m_t^{L_i}, x_{t'}^{(j)} + m_{t'}^{L_j}\right) \quad (2)$$

$$f^{(i)}(x_t) \sim \mathcal{GP}\left(g(h_t|w), k^{(i)}(h_t, h_{t'}|\theta_i, L_i)\right) \quad (3)$$

$$\mathcal{L}_{predict} = \sum_{i=1}^{N} logp(Y_i|X_i, w, v_{xh}, v_{hh}, \theta_i) \quad (4)$$

The goal of the learning is to maximize the marginal log-likelihoods $\mathcal{L}_{predict}$ and meanwhile minimize the clustering loss $\mathcal{L}_{cluster}$. In order to ensure that the initial clustering centroids can effectively represent data, we first carried out on the marginal log-likelihood function training. Then, the hidden layers outputs are fully connected to traditional clustering to obtain initial clustering labels and initial clustering centroids. Finally, the clustering and prediction parameters are optimized based on Adam algorithm.

2.2. Clustering Layer

A clustering layer is added after hidden layer of the deep model, to obtain the cluster label of each patient and integrate heterogenous information. Firstly, traditional clustering method is used to cluster the output signal of the initial hidden layer and average the signals of each cluster to obtain the initial label. Then, the similarity $D(x, y)$ between the i^{th} patient and different centroids and the soft assignment probability distribution q_{ij} based on the student t-distribution kernel are computed. In the paper, we use Complexity Invariant Similarity (CID) [10] as the similarity measurement. Finally, the cluster label is assigned with the largest probability and the cluster centroids are re-estimated. KL divergence is used to compute the loss $\mathcal{L}_{cluster}$ between the probability distribution q_{ij} and the target probability distribution p_{ij}. Since it's unsupervised learning and we cannot know the true probability distribution p_{ij}, auxiliary distribution [11] is considered as the target distribution.

$$q_{ij} = \frac{(1+D(x,y))^{-1}}{\sum_{j=1}^{C}(1+D(x,y))^{-1}} \quad (5)$$

$$p_{ij} = \frac{q_{ij}^2 / \sum_{i=1}^{N} q_{ij}}{\sum_{j=1}^{C}\left(q_{ij}^2 / \sum_{i=1}^{N} q_{ij}\right)} \quad (6)$$

$$\mathcal{L}_{cluster} = \sum_{i=1}^{N}\sum_{j=1}^{C} p_{ij} log \frac{p_{ij}}{q_{ij}} \quad (7)$$

3. Results

3.1. Dataset

Data used in the paper was obtained from the Parkinson's Progression Markers Initiative (PPMI) database. The PPMI database was accessed on June 23, 2022 and be integrated into a PD dataset containing 361 patients, of which had 9-16 follow-up visits among 3-8 years with 95 variables. Our prediction target is the MDS-UPDRS Part III score in final year, which is a relatively recognized marker of movement progress [3-5].

3.2. Prediction Performance

We compared the proposed method with general GP, classical RNN (LSTM, GRU) and DME-GP. Ten-fold cross validation was used for evaluation. The prediction results are

shown in Table 1. Our proposed model has the relatively best prediction performance, and the best performance is obtained when the cluster count is specified as 3. The t-test was used to assess the statistical significance of the difference between DME-GP and our proposed model (C=3). The p-value were 0.006, 0.018 and 0.003 respectively for MAE, RMSE and R2, indicating the model performance is significantly improved.

Table 1. Comparison of performance in predicting PD progression.

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)	5.184	7.326	0.724
Proposed Method(C=3)	**5.015**	**7.284**	**0.727**
Proposed Method(C=4)	5.252	7.403	0.718

3.3. Clustering Results

Progression patterns in different clusters when cluster count is specified as 3 are showed in Figure 2. The thickness of the lines represents the count of patients in the clusters. We can see that two progression patterns are identified among population, where the score remains at a relatively low level and even drops in the later stages in cluster 1, while the score fluctuates widely from low level to high level in cluster 2.

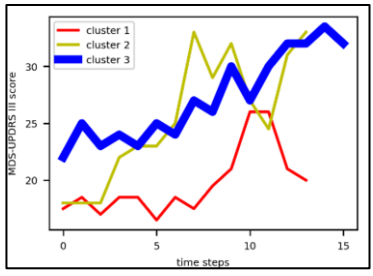

Figure 2. Progression patterns in different clusters when cluster count is 3.

4. Discussion

Our approach focuses on the heterogeneity in PD progression through integrating the different progression patterns information obtained by the clustering layer into the kernel function of GP so that the model gains a better interpretability than classical neural networks [4,5] and a better prediction performance than other previously proposed methods such as DME-GP [9]. This confirms that integrating heterogeneity interpretation is helpful to improve the prediction performance of the models which has been proved in some studies [12,13].

In addition, unlike the probability graph models [6], the clusters we obtained are specific to patients rather than disease states, which is clinically more meaningful. Furthermore, the approach can predict disease progression while clustering, thus

providing effective assessment indicators that can be replicated on other data sets.

5. Conclusions

In the paper, we propose a novel approach based on a multi-task learning framework. The results show that the method can cluster patients with different progression trajectories and discover new progression patterns which helps to predict the future progression more accurately than previously proposed methods.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 82172069) and the Key Research Project of Zhejiang Laboratory (No. 2022ND0AC01).

References

[1] Kalia LV, Kalia SK, Lang AE. Disease-modifying strategies for Parkinson's disease. Mov Disord. 2015 Sep;30(11):1442-50, doi: 10.1002/mds.26354.
[2] Ma LY, Tian Y, Pan CR, Chen ZL, Ling Y, Ren K, Li JS, Feng T. Motor progression in early-stage Parkinson's disease: a clinical prediction model and the role of cerebrospinal fluid biomarkers. Front Aging Neurosci. 2021 Jan;12:627199, doi: 10.3389/fnagi.2020.627199.
[3] Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, Wuest DM, Church BW, Khalil IG, Hayete B, Venuto CS. 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 Neurol. 2017 Nov;16(11):908-16, doi: 10.1016/S1474-4422(17)30328-9.
[4] Shahid AH, Singh MP. A deep learning approach for prediction of Parkinson's disease progression. Biomed Eng Lett. 2020 May;10:227-39, doi: 10.1007/s13534-020-00156-7.
[5] Nguyen KP, Raval V, Treacher A, Mellema C, Yu FF, Pinho MC, Subramaniam RM, Dewey Jr RB, Montillo AA. Predicting Parkinson's disease trajectory using clinical and neuroimaging baseline measures. Parkinsonism Relat Disord. 2021 Apr;85:44-51, doi: 10.1016/j.parkreldis.2021.02.026.
[6] Severson KA, Chahine LM, Smolensky LA, Dhuliawala M, Frasier M, Ng K, Ghosh S, Hu J. Discovery of Parkinson's disease states and disease progression modelling: a longitudinal data study using machine learning. Lancet Digit Health. 2021 Sep;3(9):e555-64, doi: 10.1016/S2589-7500(21)00101-1.
[7] Ziegler G, Ridgway GR, Dahnke R, Gaser C, Alzheimer's Disease Neuroimaging Initiative. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. Neuroimage. 2014 Aug;97:333-48, doi: 10.1016/j.neuroimage.2014.04.018.
[8] Peterson K, Rudovic O, Guerrero R, Picard RW. Personalized gaussian processes for future prediction of alzheimer's disease progression. arXiv preprint arXiv:1712.00181. 2017 Dec, doi: 10.48550/arXiv.1712.00181.
[9] Chung I, Kim S, Lee J, Kim KJ, Hwang SJ, Yang E. Deep mixed effect model using Gaussian processes: a personalized and reliable prediction for healthcare. InProceedings of the AAAI Conference on Artificial Intelligence 2020 Apr (Vol. 34, No. 04, pp. 3649-3657). doi: 10.1609/aaai.v34i04.5773.
[10] Batista GE, Wang X, Keogh EJ. A complexity-invariant distance measure for time series. InProceedings of the 2011 SIAM international conference on data mining 2011 Apr (pp. 699-710). SIAM, doi: 10.1137/1.9781611972818.60.
[11] Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Comput. 2006 Jul;18(7):1527-54, doi: 10.1162/neco.2006.18.7.1527.
[12] Ibrahim ZM, Wu H, Hamoud A, Stappen L, Dobson RJ, Agarossi A. On classifying sepsis heterogeneity in the ICU: insight using machine learning. J Am Med Inform. 2020 Mar;27(3):437-43, doi: 10.1093/jamia/ocz211.
[13] Beaulieu-Jones BK, Greene CS. Semi-supervised learning of the electronic health record for phenotype stratification. J Biomed Inform. 2016 Dec;64:168-78, doi: 10.1016/j.jbi.2016.10.007.