

Predictive Modeling of Hospitalization Occurrence Timing Using a Joint Model Based on Longitudinal Clinical Features

K. Kobayashi, H. Kondo, Y. Hasegawa, and S. Yui

Abstract— In order to provide efficient health services, it is necessary to predict individual health conditions. In this study, we constructed a method that predicts the probability of hospitalization and its timing by applying the Joint model, which models both past data transitions and future hospitalization occurrence timing. Experimental result shows that the discrimination accuracy of the proposed method is 0.81. From the result, it is suggested that an effective health service planning is possible using the proposed method.

Clinical Relevance— This achieves to predict not only the aggravation probability of lifestyle-related diseases, but also the time of aggravation.

I. INTRODUCTION

Japanese Health Insurance Societies provide health services to prevent diseases including lifestyle-related diseases with the aim of maintaining and promoting the health of their members. To conduct the services efficiently, it is important to extract target members who can significantly obtain the effects of the services by predicting their health conditions. Although we have already developed techniques for predicting the onset of lifestyle-related diseases and future hospitalization risks, if not only their occurrence, but also their timing is predicted, will help to plan more precise health business services. Therefore, in this study we developed a predictive model for hospitalization occurrence and its timing for diabetes, one of the lifestyle-related disease, using longitudinal data of medical checkup and past medical history.

II. METHODS

Generalized linear model (GLM) was used as predictive model for hospitalization occurrence in our past study, however, the timing prediction using GLM would contain accumulated error caused by sequential modeling. As a multivariate time series analysis method, Vector autoregressive model (VAR model) which is commonly used in the financial field and Long Short-Term Memory (LSTM) which is a neural network specialized for time series data are known recently. However, these methods are difficult to be applied to the longitudinal data of medical checkup and past medical history, because the observation interval of the data is once a year, and the number of samplings obtained are extremely small.

Therefore, we constructed a predictive model using the Joint model method [1, 2]. The Joint model we used performs survival time analysis using the Cox proportional hazard model while modeling mixed effect of repeatedly observed

longitudinal data with the Linear mixed effect model (LME model). The joint model is given as following,

$$y_{ij} = \alpha_L(x_{ij}) + \beta_{0i} + \beta_{1i}t_{ij} + \varepsilon_{ij}, \quad (1)$$

$$\lambda(t) = \lambda_0(t) \exp(\alpha_S(x_i) + \gamma(\beta_{0i} + \beta_{1i}t)), \quad (2)$$

where the y_{ij} are the longitudinal data observed per member i at sampling timing t_{ij} with $j = 1, \dots, n_i$, where n_i is the number of observations for member i . The ε_{ij} denotes mutually independent normally distributed error terms. The hazard function $\lambda(t)$ evaluated at time t is based on $\lambda_0(t)$ which represents the baseline hazard. The coefficients β_{0i} and β_{1i} are member-specific random intercept and slope while $\alpha_L(x_{ij})$ and $\alpha_S(x_i)$ are the longitudinal and the survival sub-predictor, respectively.

III. RESULTS

In order to verify the performance of the proposed method, we predicted diabetes hospitalization risk in the next 3 years using past 3 years longitudinal data which includes 99 case samples and 100 control samples. As a result of 5 cross-validation (CV) which is shown in TABLE I, the proposed method achieved average accuracy of 0.81.

TABLE I. ACCURACY OF GENERALIZATION PERFORMANCE

	1 year later	2 years later	3 years later	AVE.
CV1	0.92	0.85	0.78	0.85
CV2	0.90	0.70	0.88	0.83
CV3	0.73	0.73	0.88	0.78
CV4	0.80	0.80	0.85	0.82
CV5	0.82	0.69	0.79	0.77
AVE.	0.83	0.75	0.84	0.81

IV. DISCUSSION & CONCLUSION

In this paper, we proposed a joint model which predicts diabetes hospitalization occurrence and its timing. Although the model's accuracy was satisfactory, expansion of the sample size and experiments with other diseases remains as future works.

REFERENCES

- [1] Rizopoulos, D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. Chapman & Hall/CRC Biostatistics Series. Taylor & Francis, 2012.
- [2] E. Bergherr *et al.*, "Boosting Joint Models for Longitudinal and Time-to-Event Data", in *Biometrical Journal*. 59, 6, 2016, pp. 1104–1121.