The Workshops of the Thirty-Second
AAAI Conference on Artificial Intelligence

Feature Extraction from Electronic Health Records of Diabetic Nephropathy Patients with Convolutioinal Autoencoder

Takayuki Katsuki,[1] Masaki Ono,[1] Akira Koseki,[1] Michiharu Kudo,[1] Kyoichi Haida,[2] Jun Kuroda,[3] Masaki Makino,[4] Ryosuke Yanagiya,[5] Atsushi Suzuki[4]

[1]IBM Research AI, IBM Research - Tokyo, Japan.
[2]Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Japan.
[3]IT Business Process Planning Department, The Dai-ichi Life Insurance Company, Limited, Japan.
[4]Division of Endocrinology and Metabolism, Department of Internal Medicine, Fujita Health University, Japan.
[5]Division of Medical Information Systems, Fujita Health University, Japan.
{kats,moono,akoseki,kudo}@jp.ibm.com, {khaida,Kuroda394}@dl.dai-ichi-life.co.jp,
{makinom,yanagiya,aslapin}@fujita-hu.ac.jp

Abstract

This paper describes a feature extraction technology from event sequence of lab tests in electronic health record (EHR) for modeling diabetic nephropathy. We used a stacked convolutional autoencoder which can extract both local and global temporal information from the event sequence. The extracted features can be interpreted as similarities to a small number of typical sequences of lab tests. The extracted features in our prototyping experiment were promising for understanding of the long-term course of the disease.

Introduction

Diabetic nephropathy is a kidney disease which is commonly complicated with diabetes mellitus (International Diabetes Federation 2017). For its risk prediction and detailed health guidance, growing attention is being paid to analyzing the electronic health records (EHRs) (Shimizu et al. 2013; Perotte et al. 2015). Here, artificial intelligence (AI) technologies offer a solutions to the problem.

This study undertook a machine learning-based approach of feature extraction from a long-term EHR which consists of event sequence of lab tests. We developed a prototyping system trained with real-world EHRs of patients of diabetic nephropathy from a Japanese hospital. The major features of our system are twofold. First, the system can extract generally useful features for modeling diabetic nephropathy, such as for risk prediction and management, where the feature extractor is trained in an unsupervised way. Second, the extracted features are interpretable, which helps us to obtain knowledge of the characteristics and the long-term course of diabetic nephropathy.

To implement the above features, there were a number of technical hurdles to overcome. In particular, our challenge for the first feature was how to handle the event sequence of lab tests in an unsupervised way. As shown in Figure 1, since the events are recorded irregularly and sparsely, we need to design the feature extraction method in consideration of time shift invariance and correlations between lab tests both on local and global time scales. This requires a hierarchical convolution and pooling mechanism across time in the feature extraction process.

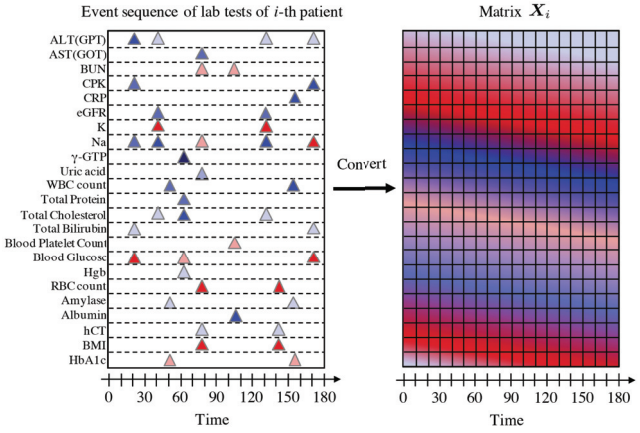

Figure 1: Event sequence of lab tests and its matrix representation. The horizontal and vertical axes respectively correspond to the time stamps and different lab tests. A triangle indicates the value of lab tests at the corresponding day, where red and blue mean a high and low values respectively.

The challenge for the second feature was how to make the extracted features interpretable while meeting the above requirement. It may be helpful if we can consider that the features are represented by affiliations to a small number of typical patterns of the lab tests sequences, i.e., some kind of filter for the sequence. This is because we can obtain these typical patterns through inverse analysis of the feature extractor in this case and the extracted features can be interpreted as the similarities with these few patterns.

To meet these challenges, we propose a solution based on the convolutional autoencoder (Masci et al. 2011), which is a variant of the autoencoder constructed by stacking convolutional layers and max-pooling layers. In the subsequent sections, we summarize our system and show the prototyping results of feature extraction which reveal the typi-

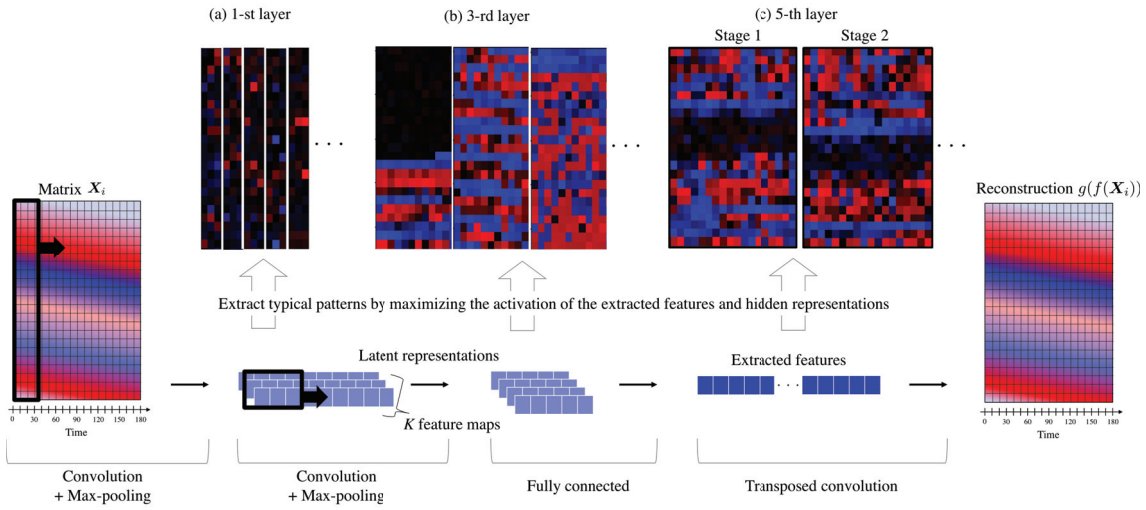

Figure 2: Framework for learning feature extractor from lab test sequences by using SCAE and extracting typical patterns of the sequence in each layer.

cal sequences that are related to the aggravation of diabetic nephropathy from stage 1 to stage 2.

Convolutional Autoencoder for Extracting Features from EHR

Our goal is to extract generally useful and interpretable features from the EHR of a diabetic nephropathy patient. As a preprocessing, we convert the event sequences of lab tests of every i-th patient into N-sets of matrices $\{\boldsymbol{X}_i\}_{i=1}^N$ as in (Wang et al. 2012), where $\boldsymbol{X}_i \in \mathbb{R}^{D\times T}$ whose horizontal dimension corresponds to the time stamp and vertical dimension corresponds to the lab tests as shown in Figure 1. The (d,t)-th entry of $\boldsymbol{X}_i$ is d-th real-valued lab test result at time stamp t.

For the matrices, we first learn a stacked convolutional autoencoder (SCAE) repeatedly applying one-dimensional convolution and pooling across time. It works as a hierarchical feature extractor from the sequence, which can compactly represent both local and global temporal information. Then, since the output of SCAE becomes affiliations to a small number of typical patterns, we generate typical sequences of lab tests as what maximally activate the SCAE outputs, which describe the features extracted with SCAE. Figure 2 summarizes the concept of the proposed framework.

For learning the SCAE, we minimize the following reconstruction error across the N-lab-test sequences $\{\boldsymbol{X}_i\}_{i=1}^N$:

$$\sum_{i=1}^{N} \|\boldsymbol{X}_i - g(f(\boldsymbol{X}_i))\|_2^2, \tag{1}$$

where the function $f(\boldsymbol{X}_i)$ is the feature extractor repeatedly mapping the input $\boldsymbol{X}_i$ to the latent representations. We define the latent representation in the l-th layer as $\boldsymbol{h}_l$. Then, we use $f(\boldsymbol{X}_i)$ to reconstruct the input $\boldsymbol{X}_i$ by a reverse mapping $g(f(\boldsymbol{X}_i))$. As shown in Figure 2, $f(\boldsymbol{X}_i)$ has five hidden layers: 1) a convolutional layer with 32 $D \times 3$ filters; 2) a max-pooling layer of 1×2 filters; 3) a convolutional layer with 32 1×3 filters per map; 4) a max-pooling layer of 1×2 filters; 5) a fully connected layer of 64 hidden neurons. The k-th feature map in the l-th convolutional layer is a deterministic function of the type:

$$\boldsymbol{h}_l^{(k)} = \sigma(\boldsymbol{h}_{l-1} * \boldsymbol{W}_l^{(k)} + \boldsymbol{b}_l^{(k)}), \tag{2}$$

where $\sigma(\bullet)$ is an activation function (we used rectified linear unit (ReLU) (Nair and Hinton 2010)) and the operator $*$ denotes a one-dimensional convolution across time with the parameter $\boldsymbol{W}_l^{(k)}$ and $\boldsymbol{b}_l^{(k)}$. The max-pooling layers downsample the latent representation by taking the maximum value over sub-temporal regions. The fully connected layer is also a deterministic function,

$$\boldsymbol{h}_5 = \sigma(\boldsymbol{W}_5\boldsymbol{h}_4 + \boldsymbol{b}_5), \tag{3}$$

where $\boldsymbol{W}_5$ and $\boldsymbol{b}_5$ are parameters. The first layer receives its input from $\boldsymbol{X}_i$ as $\boldsymbol{h}_{l-1}$ and each of the other layers receives its input from the latent representation of the layer below. Note that the size along the lab-test dimension for the convolutional filter in the first layer is D and that for the max-pooling filter is 1, since our convolutional and pooling filters are applied only along the temporal dimension. The reconstruction function $g(f(\boldsymbol{X}_i))$ is the transposed convolution (Long, Shelhamer, and Darrell 2015). Through the unsupervised learning, the SCAE, $f(\boldsymbol{X}_i)$, works as a feature extraction function from a complicated event sequence of lab tests without manually designing features. From the hierarchical convolutional structure, the SCAE can capture both local and global temporal information in early layers and later layers, respectively. As the feature extraction results, we used the 64-dimensional features $\boldsymbol{h}_5$, which were the output of the fully connected layer (5-th layer). They can be used for modeling with any other features extracted by other methods.

We output a typical sequence $\hat{\boldsymbol{X}}$ for the m-th feature, $[\boldsymbol{h}_5]_m$, through inverse analysis solving the following optimization problem:

$$\max_{\boldsymbol{X}} \ [\boldsymbol{h}_5]_m, \tag{4}$$

where we extract the pattern by maximizing the activation of the hidden representation $[\boldsymbol{h}_5]_m$. We can interpret the hidden representation by the pattern and the extracted feature can be interpreted as the similarity to the pattern.

We used ADAM with the recommended hyper-parameters in (Kingma and Ba 2015) and mini-batches of 128 examples for the above optimization problems.

Experimental Results

We trained our system by using real-world EHRs, where $\boldsymbol{X}_i$ was a 180-day history of lab-tests sequence in the i-th record. $\boldsymbol{X}_i$ had $D = 23$ attributes, each representing a real-valued result of the lab test, (the tests are listed in Figure 1). We made the time length T 18 by taking the mean of each 10-day result in the 180 days (which can be viewed as mean-pooling). We used EHRs obtained from a cumulative total of 30,810 patients for the training.

The matrices in Figure 2 (c) show examples of the results $\hat{\boldsymbol{X}}$ of Eq. (4), typical sequences which are related to the risk of diabetic nephropathy progressing from stage 1 to stage 2. The left matrix is a typical pattern of a 180-day sequence of patients who stayed in stage 1. The right matrix is that of patients who progressed to stage 2. Red and blue square mean a high test value and low test value, respectively. We determined $[\boldsymbol{h}_5]_m$ for the groups respectively as the latent representations having the highest correlations with the disease progression (stayed in stage 1 or progressed to stage 2) after 180 days from the latest EHR by using L1-regularized canonical correlation analysis. These results may provide knowledge on typical disease course and health guidance.

Figure 3 is a detailed comparison between the patients who stayed in stage 1 and those who progressed to stage 2, and it shows features having notable differences between these groups. The typical sequence of those progressing to stage 2 has higher C-reactive protein (CRP) values throughout the period than that of the patients staying in stage 1. The estimate glomerular filtration rate (eGFR) of patients who progressed to stage 2 is typically lower than that of patients staying in stage 1. The potassium (K) value fluctuates through time for stage 2 patients, but remains low for stage 1 patients. The value of amylase test is higher for progressing patients. These results are mostly consistent with knowledge about diabetic nephropathy. Additionally, if only the latest lab test values were analyzed, it would miss some temporal behaviors of the lab tests that we were discovered in this analysis.

We also showed an example of typical sequence of lab tests activating the latent representation in each middle convolutional layer in Figures 2 (a) and (b). Here, the patterns for the latent representations in the early layers represent local temporal information. The ones for those in the latter layers represent information with a longer time period.

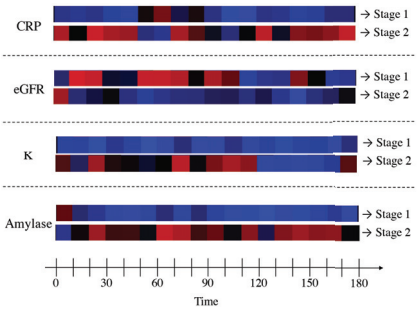

Figure 3: Comparison of the sequences between the patients who stayed in stage 1 and those who progressed to stage 2.

Related work

Healthcare is one of the major fields of application for AI. Risk prediction of disease from diverse data sources, such as Twitter (Paul and Dredze 2011; Sadilek, Kautz, and Silenzio 2012), is a promising research topic. Recently, many studies and applications focus on EHRs as data sources (Yadav and Simon 2015; Goldstein et al. 2017).

Kidney disease is a growing problem. (Echouffo-Tcheugui and Kengne 2012) reviewed risk prediction models for it and found that most of the work reported did not take into account the temporal behavior of the lab tests and focused on past medical history or lab tests at a certain point in time. None of the previous work studies tried to use SCAE to understand the temporal behavior of the lab tests. In addition, we prefer SCAE because of its application flexibility that originates from the unsupervised nature of its learning.

Neural network and deep learning models are being used in the analysis of EHR (Ravì et al. 2017); examples include recurrent neural networks (RNN) for prediction of the diagnosis and medication categories (Choi et al. 2016), an autoencoder for phenotyping (Lasko, Denny, and Levy 2013), a stack of denoising autoencoders for health states prediction (Miotto et al. 2016), and convolutional neural networks (CNN) for the risk prediction (Cheng et al. 2016). An advantage of deep learning is its prediction accuracy and a disadvantage is the difficulty understandability it results. Since one of the motivations of medical informatics is to understand a disease more deeply, we focus on the interpretability of the learned model.

SCAE has been used actively in the image processing and vision community. From its convolutional nature inspired from the biological structure of the visual cortex, many popular applications of SCAE have an aspect of being feature extraction or representation learning from images, such as image generation (van den Oord et al. 2016), extraction of per-pixel material information (Schwartz and Nishino 2013), super resolution (Wang et al. 2015), and medical imaging (Nishio et al. 2017). In this study, we used 1D-convolution across time-series. This type of convolution was proposed for time-delay neural networks (Waibel et al. 1989). Recently, convolutional autoencoder has been undertaken for feature extraction from sequential data, such as radio communication signals (O'Shea, Corgan, and Clancy

2016) and for videos (Bascol et al. 2016).

Future Directions

We have demonstrated that a stacked convolutional autoencoder can work as a feature extractor from event sequence in EHR and that the obtained typical sequences showed reasonable patterns. The next step is constructing a prediction model for the stages aggravation of diabetic nephropathy by combining the extracted features with other multi-modal information, such as medication and life-style information. Considering what new information can be obtained from the extracted typical patterns is another interesting topic.

Acknowledgments

Takayuki Katsuki was supported in part by JST CREST Grant Number JPMJCR1304, Japan.

References

Bascol, K.; Emonet, R.; Fromont, E.; and Odobez, J.-M. 2016. Unsupervised interpretable pattern discovery in time series using autoencoders. In *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*, 427–438. Springer.

Cheng, Y.; Wang, F.; Zhang, P.; and Hu, J. 2016. Risk prediction with electronic health records: A deep learning approach. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, 432–440. SIAM.

Choi, E.; Bahadori, M. T.; Schuetz, A.; Stewart, W. F.; and Sun, J. 2016. Doctor ai: Predicting clinical events via recurrent neural networks. In *Machine Learning for Healthcare Conference*, 301–318.

Echouffo-Tcheugui, J. B., and Kengne, A. P. 2012. Risk models to predict chronic kidney disease and its progression: A systematic review. *PLOS Medicine* 9(11):1–18.

Goldstein, B. A.; Navar, A. M.; Pencina, M. J.; and Ioannidis, J. P. A. 2017. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *Journal of the American Medical Informatics Association* 24(1):198.

International Diabetes Federation. 2017. IDF diabetes atlas eighth edition. http://www.diabetesatlas.org/resources/2017-atlas.html.

Kingma, D., and Ba, J. 2015. Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations (ICLR2015)*.

Lasko, T. A.; Denny, J. C.; and Levy, M. A. 2013. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. *PloS one* 8(6):e66341.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 3431–3440.

Masci, J.; Meier, U.; Cireşan, D.; and Schmidhuber, J. 2011. Stacked convolutional auto-encoders for hierarchical feature extraction. *Artificial Neural Networks and Machine Learning–ICANN 2011* 52–59.

Miotto, R.; Li, L.; Kidd, B. A.; and Dudley, J. T. 2016. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports* 6:26094.

Nair, V., and Hinton, G. E. 2010. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814.

Nishio, M.; Nagashima, C.; Hirabayashi, S.; Ohnishi, A.; Sasaki, K.; Sagawa, T.; Hamada, M.; and Yamashita, T. 2017. Convolutional auto-encoder for image denoising of ultra-low-dose ct. *Heliyon* 3(8):e00393.

O'Shea, T. J.; Corgan, J.; and Clancy, T. C. 2016. Unsupervised representation learning of structured radio communication signals. In *Sensing, Processing and Learning for Intelligent Machines (SPLINE), 2016 First International Workshop on*, 1–5. IEEE.

Paul, M. J., and Dredze, M. 2011. You are what you tweet: Analyzing twitter for public health. *Icwsm* 20:265–272.

Perotte, A.; Ranganath, R.; Hirsch, J. S.; Blei, D.; and Elhadad, N. 2015. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *Journal of the American Medical Informatics Association* 22(4):872–880.

Ravì, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; and Yang, G.-Z. 2017. Deep learning for health informatics. *IEEE journal of biomedical and health informatics* 21(1):4–21.

Sadilek, A.; Kautz, H. A.; and Silenzio, V. 2012. Predicting disease transmission from geo-tagged micro-blog data. In *AAAI*.

Schwartz, G., and Nishino, K. 2013. Visual material traits: Recognizing per-pixel material context. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 883–890.

Shimizu, M.; Furuichi, K.; Toyama, T.; Kitajima, S.; Hara, A.; Kitagawa, K.; Iwata, Y.; Sakai, N.; Takamura, T.; Yoshimura, M.; et al. 2013. Long-term outcomes of japanese type 2 diabetic patients with biopsy-proven diabetic nephropathy. *Diabetes Care* 36(11):3655–3662.

van den Oord, A.; Kalchbrenner, N.; Espeholt, L.; Vinyals, O.; Graves, A.; et al. 2016. Conditional image generation with pixelcnn decoders. In *Advances in Neural Information Processing Systems*, 4790–4798.

Waibel, A.; Hanazawa, T.; Hinton, G.; Shikano, K.; and Lang, K. J. 1989. Phoneme recognition using time-delay neural networks. *IEEE transactions on acoustics, speech, and signal processing* 37(3):328–339.

Wang, F.; Lee, N.; Hu, J.; Sun, J.; and Ebadollahi, S. 2012. Towards heterogeneous temporal clinical event pattern discovery: a convolutional approach. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, 453–461. ACM.

Wang, Z.; Yang, Y.; Wang, Z.; Chang, S.; Han, W.; Yang, J.; and Huang, T. 2015. Self-tuned deep super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 1–8.

Yadav, P., and Simon, G. 2015. Mining electronic health records (ehr): a survey. *Technical Report, Department of Computer Science and Engineering*.