Oral Presentation 11

日時：2020 年 10 月 30 日 14:45-16:15
チャンネル：２

口頭発表１１
『データサイエンス/ 構造相関解析』
Data Science/Structure Activity Relationship

座長：　植沢 芳広 （明治薬科大学）
Yoshihiro Uesawa (Meiji Pharmaceutical University),
湯田 浩太郎 （株式会社インシリコデータ）
Kotaro Yuta (In Silico Data, Ltd.)

O2-12
"Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites"
Haris Hasic
(Department of Computer Science, School of Computing, Tokyo Institute of Technology)

O2-13
"Deep neural network models to predict ion channel inhibitors with pre-learning of activity information of analogous proteins"
Tomohiro Sato
(Center for Biosystems Dynamics Research, RIKEN)

O2-14
"Statistically estimating life-long progression of chronic disease from the information of short clinical trials"
Ryota Jin
(Laboratory of Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University)

O2-15
"Prediction models of peroxisome proliferator-activated receptor-gamma through a deep learning-based QSAR analysis, DeepSnap"
Yasunari Matsuzaka
(Department of Medical Molecular Informatics, Meiji Pharmaceutical University)

O2-12

Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites

Haris Hasic[1] hasic@cb.cs.titech.ac.jp
Takashi Ishida[1] ishida@c.titech.ac.jp

[1] Department of Computer Science, School of Computing, Tokyo Institute of Technology, W8-85, 2-12-1 Ookayama, Meguro, 152-8550 Tokyo

Keywords: Machine Learning – AI Method Development, Descriptor (Parameter) Development

Retrosynthetic analysis or retrosynthesis is a technique used by chemists during the planning of chemical synthesis routes [1]. The manual application of retrosynthesis requires a lot of knowledge and experience to identify possible transformations for a given target compound. Furthermore, since the complexity scales with the complexity of the targets, efficient application on compounds with complex structures becomes an almost impossible task for human chemists. The idea of using computers in such situations has existed for a long time, but the accuracy was insufficient for practical applications. However, with the steady improvement of artificial intelligence in recent years, computer-assisted retrosynthesis has been gaining research attention again. Because of the overall lack of available chemical reaction data, the main challenge of recent computer-assisted retrosynthesis approaches is low exploration ability during the analysis of target and intermediate compounds [2]. This analysis is referred to as single-step retrosynthesis.

In this research, we developed a novel, template-free approach for the single-step retrosynthesis task for any given molecule. To ensure general applicability of the approach, we used only individual molecular substructures of the target compound to determine potential disconnection sites, without relying on additional information. The model for the identification of potential disconnection sites was constructed to use a novel molecular substructure fingerprint representation based on the extended connectivity fingerprints [3]. Then, a simple structural similarity-based reactant retrieval and scoring implementation was applied for each of the identified disconnection sites. As a result of the evaluation experiments, the proposed method outperforms the state-of-the-art method [4] for the single-step retrosynthesis task by achieving 48.9% Top-1 accuracy on the pre-cleaned U.S. Patent Office dataset [5]. If the reaction class predicted by the model was used to further narrow down reactant candidate search space, the performance is additionally improved to 61.4% Top-1 accuracy [6].

.

[1] Clayden, J., Greeves, N., and Warren, S., *Organic Chemistry*, Oxford University Press, 2012.
[2] Coley, C.W., Rogers, L., Green, W.H., and Jensen, K.F., Computer-Assisted Retrosynthesis based on Molecular Similarity, *ACS Central Science*, 3(12):1237-1245, 2017.
[3] Rogers, D., and Hahn, M., Extended-connectivity Fingerprints, *Journal of Chemical Information and Modelling*, 50(5):742-754, 2010.
[4] Zheng, S., Rao, J., Zhang, Z., Xu, J., and Yang, Y., Predicting retrosynthetic reactions using self-corrected transformer neural networks. *Journal of Chemical Information and Modelling*, 60:47-55, 2020.
[5] Schneider, N., Lowe, D.M., Sayle, R.A., Tarselli, M.A., and Landrum, G.A., Big Data from Pharmaceutical Patents: A Computational Analysis of Medicinal Chemists' Bread and Butter. *ACS Journal of Medicinal Chemistry*, 59(9):4385-4402, 2016.
[6] Hasic H., and Ishida T., Single-step Retrosynthesis Prediction based on the Identification of Potential Disconnection Sites using Molecular Substructure Fingerprints. Submitted to the *BMC Journal of Chemoinformatics*, 2020.

O2-13

Deep neural network models to predict ion channel inhibitors with pre-learning of activity information of analogous proteins

Tomohiro Sato **Teruki Honma**

tomohiro.sato@riken.jp honma.teruki@riken.jp

Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

Keywords: deep neural network, pre-learning, machine learning, ion channel, cardiotoxicity

The inhibitions of ion channels are closely related to various drug side effects such as torsade de pointes caused by the inhibition of hERG potassium channel. Recent studies revealed that the inhibition of various cardiac ion channels could result in diverse cardiotoxicities[1]. Comprehensive prediction of inhibitory activities for cardiac ion channels could greatly reduce the risk of cardiotoxicities in drug discovery. Though we reported the construction of a highly accurate hERG prediction model by integrating information dispersing in various databases[2,3], the limited number of inhibitory activity information could become a stabling block to construct accurate statistical models for most of the other cardiac ion channels. In this study, we demonstrated that the pre-learning about activity information of analogous proteins could effectively improve the prediction performances of deep neural network (DNN) models for ion channels with less activity information.

In the case of discrimination analysis of Nav1.5, Kv1.5, and Cav1.2 inhibitors, the pre-learning of other sodium channels, potassium channels, and calcium channels improved the kappa statistics from 0.68 by normal DNN model to 0.73 by pre-learned DNN model for Nav1.5, from 0.496 to 0.545 for Kv1.5, and from 0.387 to 0.548 for Cav1.2 (Figure 1), suggesting the usefulness of pre-learning technique when the activity information of the target protein is insufficient to construct an accurate statistical model. The prediction models were available at https://drugdesgin.riken.jp/hERG.

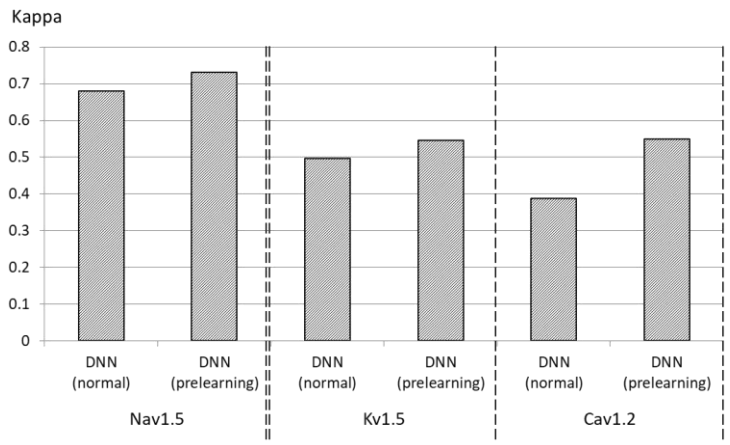

Figure 1. Kappa statistics of normal and pre-learned DNN models for Nav1.5, Kv1.5, and Cav1.2.

References

[1] Bartos DC, Grandi E, Ripplinger CM. Ion Channels in the Heart. *Compr Physiol*. **2015,** 5(3), 1423-1464

[2] Sato, T., Yuki, H., Ogura, K., Honma, T. Construction of an integrated database for hERG blocking small molecules. *PLOS ONE* **2018**, 13(7)

[3] Ogura, k., Sato, T., Yuki, H., Honma, T. Support Vector Machine model for hERG inhibitory activities based on the integrated hERG database using descriptor selection by NSGA-II. *Sci. Rep*. **2019**, 9, 12220

O2-14

Statistically estimating life-long progression of chronic disease from the information of short clinical trials

Ryota Jin[1]
ryota-j@chiba-u.jp

Shinya Kawamatsu[1, 2]
s.kawamatsu.gifu@gmail.com

Shogo Araki[1]
shogo.1202.tandf@gmail.com

Hiromi Sato[1]
hiromi-s@chiba-u.jp

Yasunori Sato[3]
Yasunori.sato@keio.jp

Akihiro Hisaka[1]
hisaka@chiba-u.jp

[1] Laboratory of Clinical Pharmacology and Pharmacometrics, Graduate School of Pharmaceutical Sciences, Chiba University, Chiba, 260-8675, Japan.
[2] Clinical Operations, Japan Development, GlaxoSmithKline, Tokyo, Japan.
[3] Department of Preventive Medicine and Public Health, Keio University School of Medicine, Tokyo, Japan

Keywords: Nonlinear mixed-effect model, Population pharmacokinetics analysis, Long-term pathological model, Chronic disease

Progressive central nervous system diseases including Alzheimer's disease and Parkinson's disease, and various chronic diseases including diabetes, cardiovascular disease, COPD and rheumatism, persistently progress throughout life. The change of biomarkers associated with the progression and factors that affect the progression are essential information as the basis of remedy, but they are usually difficult to elucidate due to the need for decades of cohort studies. However, if a large number of patients with different degrees of disease progression are enrolled, even with short-term studies, that is, in case of that the ergodic hypothesis (anzats) in mathematics is approximately satisfied for patients, it is possible to estimate the long-term progression. Based on this concept, we developed SReFT (Statistical Restoration of Fragmented Time course) and analyzed the progression of Alzheimer's disease [1]. In this study, the applications of SReFT to Parkinson's disease and COPD [2] will be described. In each case, we succeeded in analyzing the change of biomarkers for more than 30 years and factors influencing progressions such as gene mutation and smoking from individual information of several biomarkers of hundreds to thousands of patients. The advantages of SReFT have become apparent, such as detecting long-term changes that could not be according to conventional analyses and distinguishing early and late changes in the disease. In this presentation, we will outline the algorithm of SReFT, which can be said to be a new type of machine learning, along with the currently available analysis environments, and report the results of clinical data analysis.

[1] Ishida, T., *et al.*, A Novel Method to Estimate Long-Term Chronological Changes From Fragmented Observations in Disease Progression, *Clin Pharmacol Ther*, 105:436-447, 2019.
[2] Kawamatsu, S., et al., Scores of health-related quality of life questionnaire worsen consistently in patients of COPD: estimating disease progression over 30 years by SReFT with individual data collected in SUMMIT trial, *J Clin Med*, 9:2676, 2020

O2-15

Prediction models of peroxisome proliferator-activated receptor-gamma through a deep learning-based QSAR analysis, DeepSnap

Yasunari Matsuzaka matsuzys@my-pharm.ac.jp

Yoshihiro Uesawa uesawa@my-pharm.ac.jp

Department of Medical Molecular Informatics, Meiji Pharmaceutical University, Tokyo,

Keywords: DeepSnap, Deep learning, peroxisome proliferator-activated receptor-gamma, QSAR, Tox21,

We have previously reported a novel technique for capturing images of chemical compounds for deep learning (DL)-based quantitative structure-activity relationship (QSAR) analysis, namely, DeepSnap-DL [1,2]. This approach can automatically perform feature extraction using snapshot images derived from 3D chemical structures and create input image using data augmentation by angle-designated rotation, zoom specification, and RGB conversion, that expect to increase in the amount of information on chemical structure, not just data padding and to prevent over-fitting that occur when learning is performed only with to match training data [1-5]. Thus, it has been shown that the prediction performances of the DeepSnap-DL approach were higher than those of conventional machine learning, including random forest, XGBoost, LightGBM, and CatBoost [3-5]. In general, the perturbation of nuclear receptor (NR) signaling pathway due to unwanted interactions of chemical compounds with some NRs is associated with various adverse effects on human health. To develop new chemical risk assessment approaches, we established cell-based *in vitro* high-throughput screening assays for agonist/antagonist activities of some NRs using the reporter assay in Toxicology in the 21st Century (Tox21) program [6]. Among these NRs, peroxisome proliferator-activated receptor-gamma (PPARg) plays a paramount role as an antioxidant response in tissue injury. Hence, we constructed prediction models of PPARg agonist activity by the DeepSnap-DL approach using the Tox21 chemical library.

[1] Uesawa, Y. Quantitative structure-activity relationship analysis using deep learning based on a novel molecular image input technique, Bioorg Med Chem Lett, 28(20), 3400-3403, 2018.
[2] Matsuzaka, Y., Uesawa, Y. Optimization of a Deep-Learning Method Based on the Classification of Images Generated by Parameterized Deep Snap a Novel Molecular-Image-Input Technique for Quantitative Structure-Activity Relationship (QSAR) Analysis, Front Bioeng Biotechnol, 7:65, 2019.
[3] Matsuzaka, Y., Uesawa, Y. Prediction Model with High-Performance Constitutive Androstane Receptor (CAR) Using DeepSnap-Deep Learning Approach from the Tox21 10K Compound Library, Int J Mol Sci, 20(19):4855, 2019.
[4] Matsuzaka, Y., Uesawa, Y. DeepSnap-Deep Learning Approach Predicts Progesterone Receptor Antagonist Activity With High Performance, Front Bioeng Biotechnol, 7:485, 2020.
[5] Matsuzaka Y, Hosaka T, Ogaito A, Yoshinari K, Uesawa Y. Prediction Model of Aryl Hydrocarbon Receptor Activation by a Novel QSAR Approach, DeepSnap-Deep Learning, Molecules, 25(6):1317, 2020.
[6] Matsuzaka, Y., Uesawa, Y. Molecular Image-Based Prediction Models of Nuclear Receptor Agonists and Antagonists Using the DeepSnap-Deep Learning Approach with the Tox21 10K Library, Molecules, 25(12):2764, 2020.