KP202 Development of a Software Tool for Pharmacological Activity Prediction by using Collective SVM Models

Satoshi FUJISHIMA[*1], Kentaro KAWAI[*2], Yoshimasa TAKAHASHI[*1]

1) Department of Knowledge-based Information Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tempaku-cho, Toyohashi, Aichi 441-8580 JAPAN

2) Central Research Laboratories, Kaken Pharmaceutical Co., Ltd., 14, Shinomiya, Minamigawara-cho, Yamashina-ku, Kyoto 607-8042 JAPAN

fujisima@mis.tutkie.tut.ac.jp

1. Introduction

A serious side effect of a drug is part of the pleiotropic effects, but it is really undesirable. However, for the beneficial case, identifying some of unanticipated effects would often lead us to the discovery of new drug candidates. Therefore, it is quite useful for us to predict the pleiotropic effects of a drug.

We have investigated the usage of machine learning techniques in drug design and discovery. In a previous work, we reported that an artificial neural network approach combined with the topological fragment spectra (TFS) [1] as an input signal pattern allowed us to successfully classify dopamine antagonists that interact with four different types of dopamine receptors [2]. It was also shown that support vector machine (SVM) works for this type of problem [3, 4] much better. Moreover, we have investigated the multi-label classification of a large number of drugs [5] that have one or more of 100 different kinds of activity labels to explore a comprehensive model which can be used to predict individual biological activities and pleiotropic effects.

In this study, we have developed a software tool that makes us easy to use the SVM models obtained. The tool allows us to predict biological activities of unknown compounds and/or to give the alert of adverse effects of the drug candidates using only chemical structures.

2. System Overview and Features

The software tool consists of three major modules. They are the model input interface, the structure input module and the prediction module. The trained SVM models to be used for the prediction are entered through the model input interface into the system. A parameter file for the SVM model is also specified. The system loads the files of the SVM models and its parameters and then passes the data to the prediction module to construct the classifiers. The structure input module includes a structure editor. It enables us to draw the chemical structure of a candidate and submit to system. Various ring templates and user-defined templates are also available to make it easy to draw a chemical structure. The structure input module can import a connection table of the chemical structure drawn by the editor and displays it. The TFS method was used to describe structural features of the query chemical structure within the system. The TFS is derived by enumeration and

characterization of all possible substructures that have a specified number of bonds. The present system enumerates possible structure fragments that have five or less connected bonds. All of the fragments were characterized by the sum of atomic mass numbers of the constituent atoms. The imported structure is described as a multidimensional pattern vector and then the vector is passed to the prediction module. In the prediction module, the query received from the structure input module is tested whether it belongs to individual biological activities by using the constructed SVM classifiers. The prediction results for individual activities are displayed on the screen. The name of drug activities, the output value given by the individual classifiers and sign ("+" if output value is positive, otherwise "-") are shown. If sign is "+", the classifier indicates the query might belong to corresponding activity. The biological activity can be predicted continuously by the edit of the chemical structure.

3. Results and Discussion

In the present work, the collective SVM classifiers for 100 different types of activity to drug molecules which was developed in the previous work were used as the classifier. Figure 1 shows an example of the prediction by our system. Diazepam, which is known as anxiolytic, anticonvulsant, sedative, and skeletal muscle relaxant, was entered as a query molecule by using the structure editor. All the prediction results of 100 different types of activity are shown individually. In this case, it was predicted that the query drug (diazepam) would have the activities of anxiolytic, bronchodilator, and anticonvulsant. The system correctly identified the types of diazepam which have never used in training the models.

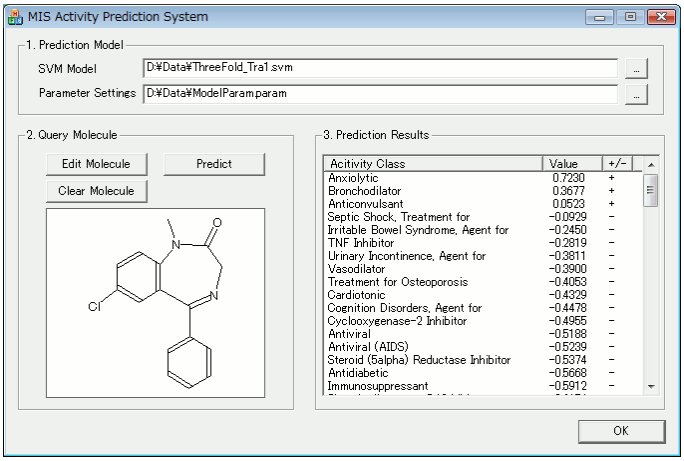

Fig.1 Prediction results of diazepam as query.

It is apparent that the present tool helps chemists to predict potential biological activities of unknown compounds or to alert adverse effects of known compounds.

References

1. Takahashi Y, Ohoka H, Ishiyama Y, In *Advances in Molecular Similarity*, 1998, Vol. 2, pp 93-104.
2. Fujishima S, Takahashi Y, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1006–1009.
3. Takahashi Y, Nishikoori K, Fujishima S, In *Active Mining, Second International Workshop, Lecture Notes in Computer Science*, 2005, Vol. 3430, pp 303-311.
4. Fujishima S, Takahashi Y, Nishikoori K, Kato H, Okada T, *New Gener. Comput.* **2007**, *25*, 203–212.
5. Kawai K, Fujishima S, Takahashi Y, *J.Chem. Inf. Model.* **2008**, *48*, 1152-1160.