Precision indices of neural networks for medicines: structure-activity correlation relationships

Hanxi ZHU, Tomoo AOYAMA*, Ikuo YOSHIHARA
Seungwoo LEE, Wookhyun KIM**

The Faculty of Engineering, Miyazaki University
Gakuen Kibanadai-nishi 1-1, Miyazaki 889-2192, Japan
*E-mail: t0b217u@cc.miyazaki-u.ac.jp
**College of Engineering Yeungnam University
Kyongsan, Kyoungbook, 712-749, Korea

Abstract

We investigated the structure-activity relationships on use of multi-layer neural networks. The relationships are techniques required in developments of medicines. Since many kinds of observations might be adopted on the techniques, we discussed some points between the observations and the properties of multi-layer neural networks. In the structure-activity relationships, an important property is not that standard deviations are nearly equal to zero for observed physiological activity, but prediction ability for unknown medicines. Since we adopted non-linear approximation, the function to represent the activity can be defined by observations; therefore, we believe that the standard deviations have not significance. The function was examined by "leave-one-out" method, which was originally introduced for the multi-regression analysis. In the linear approximation, the examination is significance, however, we believe that the method is inappropriate in case of non-linear fitting as neural networks; therefore, we derived a new index for the relationships from the differential of information propagation in the neural network. By using the index, we discussed physiological activity of an anti-cancer medicine, Mitomycine derivatives. The neuro-computing suggests that there is no direction to extend the anti-cancer activity of Mitomycine, which is close to the trend of anticancer developing.

Keyword: structure-activity relationships, neural network, drug design, medicine development, Mitomycine, anti-cancer

1. Introduction

The relationships between molecular structure and physiological activity have been studied from 1960's [1-5]. Since the relationships suggest the principle for developing of active-catalysts and useful medicines, the study is very important for chemical industries and pharmacy.

The relationships are represented by non-linear functions generally. Then, anyone imagine that the non-linear fitting ability of multi-layer neural networks would be effective tool for the relations [6]. However, there is a problem to be considered; any observation data, that are even random ones, can make a relation under the non-linear fitting. We must avoid the meaningless relation, and obtain true structure-activity relationships for unknown medicines. We select physical and chemical significance data, that are the volume of the substitutions, stereo specificity, attractive- or donating effects of electrons, and so on, at first. They have a possibility to be adopted, however, it is not a necessary and sufficient condition. We discussed the inevitability, and proposed "reconstruction-learning" that is a learning method with decreasing of connection amplitude among neurons [7].
Someone may be misunderstood the method with one kind of optimization for the number of neurons on hidden layer. The optimization is necessary, however, it cannot operate the selection of the descriptors. Then, the "reconstruction-learning" is quite different against the optimizations.

There is another approach, i.e. calculations of the differential coefficients between the output and input data of the neural network. The explicit expression is listed in section 5 in the paper. The expressions are already published [8], however, we cannot find papers to apply the structure-activity relationships on medicines. The calculation of the coefficients is very rapid, and the coefficients indicate the activity change for the substitutions directly; that is following. If the absolute values of the coefficients are negligible, or if the signals of the coefficients take alternative change, the descriptor would be omitted.

We are sure that this is a useful selection method for descriptors; therefore, we research the coefficients on the anticancer medicines.

2. Characters of the data for structure-activity relationships

There are particular data in the calculation of the structure-activity relationships, which are dummy- and grade-data [9]. The dummy-data are binary variables that are introduced for substitutions of the base compound. The grade-data are also discrete and take multi-values, for example {0,1/3,2/3,1}; if appropriate measurement were found, the data are not necessary. In spite of efforts of empirical scientists, because of the individual differences of objects, nowadays, it is hard to find a quantitative measure of the grade-data.

They are not direct observation data, however, in order to improvement the relationships, they have used empirically. Anyone, information scientists, hesitate handling them together with observations. Even if they interrelated with observed data, and if neurons were activated, the information on the neurons would not be significant. However, it would be undesirable, since chemists have adopted them, we use them carefully.

We are sure that calculations for the structure-activity relationships are separated into two types. One is