

Метод предсказания констант связывания стероидов на основе искусственных нейронных сетей

И.В. Федюшкина¹, И.В. Ромеро Рейес¹, В.С. Скворцов¹, И.С. Левина²

¹ФГБНУ «Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича», ²ФГБУН Институт органической химии им. Н.Д. Зелинского РАН

В настоящее время при создании новых лекарственных средств активно применяют различные методы компьютерного поиска, моделирования структур и конструирования фармакологически перспективных соединений *de novo* [1]. Их применение призвано существенно снизить стоимость, ускорить процессы их разработки и внедрения. Одной из важных задач является совмещение компьютерных оценок с традиционными величинами, используемыми экспериментаторами. Такой величиной является, например, константа связывания химических соединений с молекулой-мишенью, которая показывает эффективность их взаимодействия [2]. В задачи работы входило создание системы для предсказания величин констант связывания стероидов на основе параметров, вычисленных из структур лигандов и при молекулярном моделировании их комплексов с белком-мишенью. Для поиска корреляции применялась оригинальная реализация искусственных нейронных сетей.

Был осуществлен выбор нейронной сети и алгоритма её обучения. В работе применялась однонаправленная нейронная сеть с сигмоидальной функцией активации в скрытом слое и линейной функцией передачи в выходном слое, количество нейронов в скрытом слое было установлено равным 8. Обучение ИНС осуществлялось методом обратного распространения ошибки [3]. Минимизация ошибки между целевым и вычисленным значениями выполнялась по методу Левенберга-Марквардта [4]. С целью снижения ошибки переобучения в каждом случае начальная выборка была увеличена в 4 раза путем генерации новых точек в пределах погрешности (статистический бутстрап). Входные параметры представляли собой набор независимых дескрипторов лигандов (величин, описывающих свойства химических соединений) и энергетических параметров взаимодействия лигандов с белком (рассчитанных с помощью методов молекулярной динамики) [5, 6].

В качестве обучающей выборки в работе использовались данные о структуре 42 стероидов - селективных модуляторов рецептора прогестерона, синтезированных в Институте органической химии им. Н.Д. Зелинского РАН, для которых была экспериментально определена величина константы связывания с рецептором прогестерона. Для более адекватной оценки предсказательной способности полученной модели была использована независимая тестовая выборка, состоящая из 8 соединений, для которых были выполнены предсказания по созданной модели, а затем эти вещества были синтезированы и экспериментально определена константа связывания.

В результате работы была сконструирована нейросетевая модель для оценки констант связывания стероидных модуляторов рецептора прогестерона с хорошей предсказательной силой (коэффициент корреляции при скользящем контроле $Q^2 = 0,91$, среднеквадратичная ошибка = 0,29, для тестовой выборки $R^2 = 0,77$).

Данная нейросетевая модель может быть использована для оценки константы связывания с достаточно высокой точностью. На основе последовательного алгоритма была реализована параллельная версия этой модели с использованием графических процессоров. В результате скорость расчетов возросла в 70 раз (расчеты проводились на гибридной вычислительной системе на базе серверной платформы HP ProLiant G7 (AMD Opteron 6100) и вычислительной системы Tesla S2050 с использованием технологии NVIDIA CUDA 5.0). Характеристики отдельных функционалов модели представлены в таблице 1. Использование графических процессоров в соответствующей программной реализации позволит существенно увеличить размеры обрабатываемых данных при одновременном сокращении временных затрат.

Таблица 1. Характеристики выполнения последовательных и параллельных расчетов отдельных функционалов модели.

Реализация	MATLAB	C++CPU	C++CPU+CUDA GPU
Метод главных компонент PCA (сек.)	0,030	0,014	0,002
Неметрическое многомерное шкалирование MDS (сек.)	0,154	0,103	0,028
Обучение ИНС при PCA (эпох/ед. врем.)	54/1 сеть	78/1 сеть	5396/256 сетей
Обучение ИНС при MDS (эпох/ед.врем.)	61/1 сеть	82/1 сеть	5524/256 сетей

Литература:

1. Ivanov A.S., Veselovsky A.V., Dubanov A.V., Skvortsov V.S., Bioinformatics Platform Development: From Gene to Lead Compound, *Methods Mol Biol.* Vol.316, 2006, pp. 389-431.
2. Lawrenz M., Wereszczynski J., Ortiz-Sánchez J.-M., Nichols S.E., McCammon J.A., Thermodynamic integration to predict host-guest binding affinities, *Journal of Computer-Aided Molecular Design*, Vol.26, No.5, 2012, pp. 569-576.
3. D.E. Rumelhart, G.E. Hinton, R.J. Williams Learning internal representations by error propagation // Vol. 1 of *Computational models of cognition and perception*, chap. 8. Cambridge, MA: MIT Press, 1986. Pp. 319–362
4. M. Hagan Training Feedforward Networks with the Marquardt Algorithm” / M. Hagan, M. Menhaj // *IEEE Transactions on Neural Networks*. – November 1994. –Vol. 5, No. 6. – P. 989-993
5. I.D. Kuntz et al. // *J. Mol. Biol.* 1982, V.161, P. 269
6. Amber 9, <http://www.ambermd.org>

Method of predicting the binding constants of steroids based on artificial neural networks

Irina Fedyushkina, Ilacai Romero Reyes, Vladlen Skvortsov and Inna Levina

Keywords: artificial neural networks, progesterone receptor, steroids, parallel computation, GPU

The system for prediction the values of binding constants of steroids with progesterone receptor on the basis of parameters, calculated from the ligand structures and from molecular modeling of protein/ligand complexes was developed. The original implementation of the artificial neural networks was used for correlated these values.