

Т.И. Маджидов: В работе вы показали возможность применения подхода SEA. Имеются ли результаты по его валидации на новых данных? Например, по определению возможной мишени. Протокол валидации выглядит достаточно просто реализуемым.

А.А. Никитина: На данном этапе работы, мы не проводили валидацию метода на внешней выборке. Однако вопросы предсказательной способности алгоритма были рассмотрены в работах Keiser et.al. [1] и Wang et.al. [2], посвященных методу SEA, в том числе на тестовой выборке и экспериментально. Кроме того, было показано, что использование различных дескрипторов, варьирование порога активности для экспериментов и выбор типа распределения (нормальное, экстремальные значения) не оказывает существенного влияния на результаты прогноза. В дальнейшем, мы также планируем проанализировать влияние этих параметров на качество прогноза для выборки противовирусных соединений и при наличии достаточного количества данных провести time-split валидацию. Наиболее очевидный подход к валидации - введение граничного значения p-value (например 0.05) и использование стандартных метрик для задач классификации.

Т.И. Маджидов: По какой причине использовалось распределение экстремальных значений? Почему проводили оптимизацию подхода SEA, а не использовали оригинальную технику и подобранные в работе Kaiser et al параметры?

А.А. Никитина: распределения экстремальных значений в качестве модельного было предложено в оригинальной работе Keiser et.al.[1] и использовано нами в данной работе. На слайде 9 в формуле $f(z)$ была допущена опечатка, корректная формула, использованная в расчетах р-значения: $-\exp(-z * \pi / \sqrt{6}) - 0.577215665$. Сравнение моделей с различным видом распределения (нормальным и экстремальных значений) также проводилось в работе [2]. В настоящей работе мы повторили процедуру определения оптимальных параметров с помощью статистической модели, так как их значения зависят от исследуемой выборки.

1. Keiser MJ, Roth BL, Armbruster BN, Ernsberger P, Irwin JJ, Shoichet BK. Relating protein pharmacology by ligand chemistry. Nat Biotech 25 (2), 197-206 (2007)
2. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. J Cheminform. 2016;8:20 (2016)