

УДК 577.29

# НЕЙРОННЫЕ СЕТИ ГЛУБОКОГО ОБУЧЕНИЯ В МОДЕЛИРОВАНИИ ОТНОШЕНИЯ «СТРУКТУРА-АКТИВНОСТЬ»

*Аддылова Ф.Т.*

Институт математики им. В.И. Романовского АН РУз, Ташкент

Искусственные нейронные сети стали популярны в молекулярной и химической информатике (при разработке лекарств) примерно два десятилетия назад, но за последнее десятилетие нейронные сети глубокого обучения достигли значительных успехов в различных областях применения искусственного интеллекта. Эта технология, выросшая из исследований искусственных нейронных сетей, показала превосходную производительность по сравнению с другими алгоритмами машинного обучения в таких областях, как распознавание изображений и голоса, обработка естественного языка и другие. Первая волна приложений глубокого обучения в фармацевтических исследованиях появилась в последние годы, сегодня она продемонстрировала свою перспективу в решении разнообразных проблем в области открытия лекарств. В работе обсуждаются работы, касающиеся прогнозирования биоактивности по отношению «структура-активность» (QSAR) и решения этих задач в случае небольших объёмов исходных данных.

**Ключевые слова:** нейронные сети глубокого обучения, модели отношения «структура-активность», небольшие объёмы исходных данных.

**Цитирование:** Аддылова Ф.Т. Нейронные сети глубокого обучения в моделировании отношения «структура-активность» // Проблемы вычислительной и прикладной математики. — 2019. — № 3(21). — С. 5–18.

*Светлой памяти моих коллег, однокурсников-бауманцев:  
З.Т. Аддыловой, Ш.Е. Туляганова, Д.Х. Гулямова*

## 1 Введение

Цифровые данные во всех формах и размерах растут экспоненциально. Согласно информации Агентства Национальной Безопасности США (National Security Agency), Интернет обрабатывает 1826 петабайт данных в день [1]. В 2011 году отметили, что объём цифровой информации только за пять лет вырос в девять раз [2]; к 2020 году, как ожидается, он достигнет 35 триллионов гигабайт [3]. Высокий спрос на изучение и анализ больших данных (Big Data) стимулировал появление и использование алгоритмов глубокого обучения (Deep Learning, DL). DL добилось огромных успехов в широком спектре приложений: компьютерные игры, распознавание речи, компьютерное зрение, обработка естественного языка и др. [4]. По Gartner, в первой десятке трендов технологий 2017 года, технологии искусственного интеллекта расположились на верхней строчке [4].

За последнее десятилетие наблюдалось значительное увеличение количества соединений с доступной активностью и биомедицинских данных [5, 6] из-за появления новых экспериментальных методов, таких как HTS, параллельный синтез [7]. Эффективная обработка огромных массивов химических данных становится важной

проблемой для открытия лекарств. Большие объёмы данных в сочетании с усовершенствованной технологией автоматизации мотивировали дальнейшее использование машинного обучения. Помимо известных методов, таких как векторные машины (SVM) [8], нейронные сети (NN) [9] и случайный лес (RF) [10], которые применялись для разработки моделей «структура-активность» (QSAR моделей) в течение длительного времени, начали использоваться такие методы, как матричная факторизация [11] и глубокое обучение.

Архитектурой нейронных сетей DL, которые будут обсуждаться в этом обзоре, являются глубокие нейронные сети (DNN). Область применения DL в открытии лекарств быстро прогрессирует – новые статьи публикуются почти каждую неделю. Недавно вышло несколько обзоров приложений DL в области вычислительной химии и наук о жизни [12–17]. В данной работе внимание уделяется QSAR приложениям DL в разработке лекарств, выделяя используемые сегодня архитектуры DNN.

## 2 Используемые в вычислительной химии архитектуры глубокого обучения

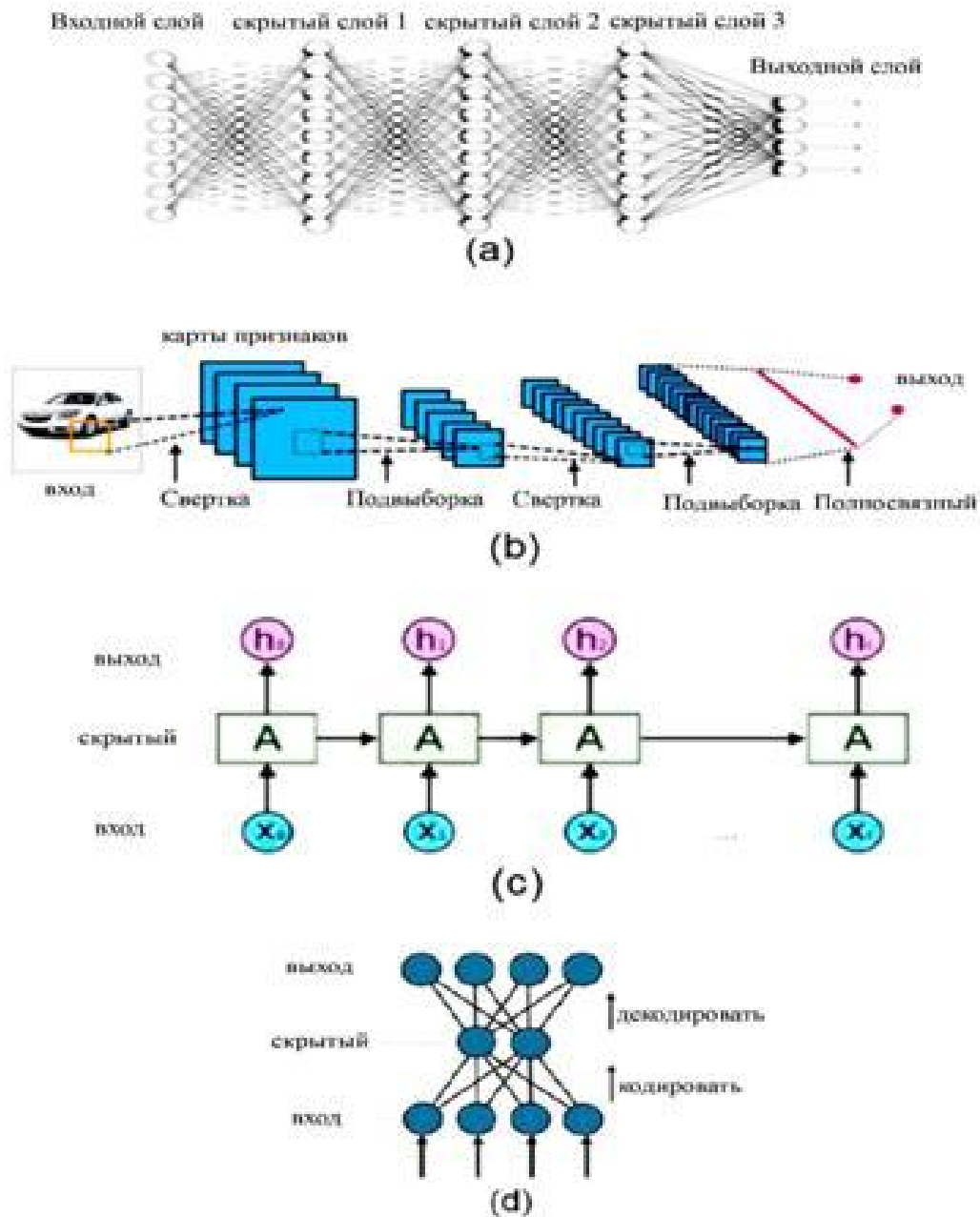
Мы не останавливаемся на принципах и приложениях глубоких сетей, отсылая читателя к замечательной книге [18]. Отметим только, что основное различие между DL и традиционными нейронными сетями (Artificial Neural Networks, ANN) – масштаб и сложность. DL использует большее количество скрытых слоев, тогда как ANN обычно имеют только один или два скрытых слоя. В сетях глубокого обучения используются много узлов и скрытых слоев благодаря появлению более мощной техники (GPU). Есть также много алгоритмических улучшений в DL, например, методы dropout [19] и DropConnect [20], которые решают проблему переобучения с применением выпрямленной линейной единицы (Rectified Linear Unit, ReLU) [21], чтобы избежать исчезновения градиентов, введение сверточных и объединенных слоев в качестве новых сетевых архитектур, позволяющих использовать большое количество входных переменных. Большинство программных пакетов DL являются доступными: TensorFlow [22], Caffe [23], PyTorch [24], Keras [25] и Theano [26]; они являются одними из самых популярных пакетов DL, используемых сегодня.

Вначале представим несколько популярных архитектур, используемых в DL (рис. 1). Полносвязная глубокая нейронная сеть (DNN), которая содержит несколько скрытых слоев, и каждый слой содержит сотни нелинейных нейронов, показана на рисунке 1a. DNN могут принимать большое количество признаков ввода, а нейроны в разных слоях DNN могут автоматически извлекать признаки на разных иерархических уровнях [27].

На рисунке 1b представлен пример очень популярного типа сети – сверточные или сверточные сети (CNN). Сеть обычно содержит несколько слоев свертки и слой подвыборки. Сверточный слой состоит из набора фильтров (или ядер), которые имеют небольшое восприимчивое поле и обучаемые параметры. Благодаря тому, что для каждого фильтра используются одни и те же параметры, CNN в значительной степени уменьшает количество свободных параметров, тем самым снижая потребляемую память и увеличивая скорость обучения.

Рекуррентные сети (RNN) позволяют в одном и том же скрытом слое формировать направленный цикл (рис. 1c). RNN могут принимать последовательные данные в качестве входных признаков, что очень подходит для задач, зависящих от времени. Использование технологии Long Short-Term Memory (LSTM) [28] в RNN решают проблему исчезновения градиента.

Четвертая архитектура ANN, показанная на рисунке 1d, называется autoencoder (AE) [29]. AE – это сеть, используемая для неконтролируемого обучения. Она содержит часть кодера, которая является нейронной сетью, чтобы преобразовать информацию, полученную с входного уровня, в ограниченное количество скрытых блоков, а затем соединяет декодер NN с уровнем вывода, имеющим такое же количество узлов, что и входной уровень. Недавно концепция AE стала более широко использоваться для обучения генеративных моделей [30].



**Рис. 1** Архитектура популярных нейронных сетей: (а) полносвязная глубокая нейронная сеть (DNN), (б) сверточная нейронная сеть (CNN), (с) рекуррентная нейронная сеть (RNN) и (d) автокодер (AE) [54]

### 3 Применение сетей глубокого обучения в прогнозировании свойства и активности соединения

Количественные отношения структура-активность (QSAR)- очень распространенный метод в фармацевтической промышленности для прогнозирования целевых и нецелевых активностей. Такие прогнозы помогают расставить приоритеты в экспериментах во время процесса разработки лекарств и существенно сокращают необходимую экспериментальную работу.

Идеальный метод QSAR должен быть способен построить прогнозную модель из 300 000 молекул с 10 000 дескрипторами в течение 24 часов без ручного вмешательства. Из-за этих ограничений только небольшое количество предложенных алгоритмов машинного обучения подходят для общих приложений QSAR. В настоящее время самыми успешно прогнозирующими методами являются случайный лес (RF) [31] и SVM [32].

Эталонным исследованием, доказывающим эффективность применения глубоких нейронных сетей в разработке лекарств, считается конкурс Kaggle ([www.kaggle.com](http://www.kaggle.com)), проведенный в 2012 году компанией Merck. Цель конкурса – выяснить, насколько хорошо современные методы машинного обучения могут справиться с проблемами QSAR. Выбрали 15 наборов данных QSAR различных размеров (2000-50 000 молекул), используя общий тип дескриптора. Каждый набор данных был разделен на обучающий и тестовый наборы. Участникам конкурса были даны дескрипторы и активности для обучающего набора, а также дескрипторы только для тестового набора. Было разрешено создавать модели, используя любой метод машинного обучения или их комбинации, и прогнозировать активность молекул на тестовом наборе.

Победитель (команда George Dahl), соавтора статьи [33], улучшила значение квадрата коэффициента корреляции Пирсона ( $R^2$ ), усредненное по 15 наборам данных, с 0,42 (для RF) до 0,49. Улучшение может показаться незначительным, но за последние 10 лет впервые нашли подход, который стабильно опережал RF с таким запасом. Победитель использовал множество разных методов, включая глубокую нейронную сеть (DNN), машину повышения градиента (Gradient Boosting Machine, GBM) [34], регрессию по гауссовскому процессу (GaussianProcess, GP) [35].

Остановимся на некоторых важных моментах для понимания сути применения глубокого обучения в QSAR на материалах конкурса Kaggle.

**Данные.** Исследовали 15 разнообразных наборов данных QSAR и доказали, что DNN в большинстве случаев могут давать лучшие прогнозы, чем RF. Более того, можно иметь один набор настраиваемых параметров, хорошо работающий для большинства наборов данных, и нет необходимости оптимизировать параметры для каждого набора данных в отдельности. Это делает DNN практическим методом для QSAR в промышленной среде разработки лекарств.

Каждый из наборов данных был разделен на два непересекающихся подмножества: обучающий и тестовый. Хотя обычным способом деления является случайный отбор, то есть «случайное разделение», в реальной фармацевтической практике модели QSAR применяются «проспективно», т.е. для каждого набора данных первые 75% молекул, проанализированных на конкретную активность, образуют обучающий набор, в то время как оставшиеся 25% соединений, проанализированных позже, образуют тестовый набор. Оказалось, что  $R^2$  на данных с разделением по времени лучше, чем  $R^2$  из обычной схемы «случайного разделения» [36].

**Дескрипторы.** Каждая молекула представлена списком признаков, то есть «дескрипторами» в терминологии QSAR. Здесь использовали набор дескрипторов, кото-

рый является объединением оригинального дескриптора «пары атомов» (AP) из [37] и дескрипторов «пара донор-акцептор» [38]. Оба дескриптора имеют следующую форму:

$$\text{тип атома } i - (\text{расстояние в связях}) - \text{тип атома } j.$$

Для AP тип атома включает в себя атом, число неводородных соседей и число рi-электронов. Для DP тип атома - один из семи (катион, анион, нейтральный донор, нейтральный акцептор, полярный, гидрофобный и другие).

## 4 Модели прогноза

*Случайный Лес или RF* – это ансамблевый метод рекурсивного разделения, в котором каждое «дерево» генерируется из загруженной выборки соединений, а случайное подмножество дескрипторов используется в каждом ветвлении каждого узла.

Хотя есть несколько настраиваемых параметров, качество прогнозов, как правило, нечувствительно к их изменениям. Следовательно, один и тот же набор параметров может эффективно использоваться в различных задачах. Для обработки больших объёмов данных в разумный срок версия RF была распараллелена для запуска одного дерева на один процессор в кластере. Для всех моделей RF сгенерировали 100 деревьев с  $m/3$  дескрипторами, используемыми в каждой точке ветвления, где  $m$  – количество дескрипторов в обучающем наборе. Узлы дерева с 5 или менее молекулами не расщеплялись дальше.

*Глубокие Нейронные Сети.* Структура DNN, включая количество слоев и нейронов в каждом слое, должна быть предварительно задана, наряду с выбором функции активации в каждом нейроне. Следовательно, обучение DNN означает максимизацию целевой функции путем оптимизации весов  $w_{i,j}$  и смещений  $b_j$  каждого нейрона.

$$\emptyset = (\{w_{i,j}\}, \{b_j\}, i = 1, \dots, N_j, j = 1, \dots, L + 1) \quad (1)$$

где  $N_j$  – количество нейронов в  $j$ -м слое, а  $L$  – количество скрытых слоев. Дополнительный слой  $j$  предназначен для выходного слоя. Процедура обучения – хорошо известный алгоритм обратного распространения (BackPropagation, BP) с использованием мини-пакетного стохастического градиентного спуска (Stochastic Gradient Descent, SGD) с моментом [39]. Отдельные значения для  $\emptyset$  сначала назначаются случайным образом; молекулы в обучающем наборе случайным образом перемешиваются, а затем равномерно делятся на небольшие группы, называемые «мини-пакетами». Каждый мини-пакет используется для обновления значений  $\emptyset$  один раз с использованием алгоритма BP. Когда просчитаны все мини-пакеты из обучающей выборки, говорят, что процедура обучения заканчивается одной «эпохой». Понятно, что обычно требуется много эпох, а их количество является настраиваемым параметром.

Как правило, количество элементов в  $\emptyset$  для задачи QSAR может быть очень большим. Например, если набор обучения имеет 8000 дескрипторов, а в DNN три скрытых слоя, каждый из которых имеет 2000 нейронов, тогда в  $\emptyset$  будет более 24 миллионов настраиваемых значений. Поэтому сети DNN, обученные с использованием алгоритма BP, склонны к переобучению. Известно несколько методов, позволяющих это избежать. Два наиболее популярных из этих методов: предварительное обучение, предложенное в [40], и процедура «drop-out» или отсева», предложенная в [41]. Первый подход может смягчить переобучение, потому что действует как зависимый от данных регуляризатор  $\emptyset$ , то есть ограничивает его значения. Вместо использования случайных значений для инициализации  $\emptyset$  в DNN, он генерирует

значения  $\emptyset$ , используя процедуру обучения без учителя, проводимую только для входных дескрипторов, без учета активности соединений, которые они представляют. Последующий контролируемый алгоритм обучения BP просто точно настраивает  $\emptyset$ , исходя из значений, полученных в результате обучения без учителя.

Второй подход привносит нестабильность в архитектуру DNN путем случайного “drop-out” некоторых нейронов в каждом мини-пакете обучения. Эти два подхода могут использоваться по отдельности или совместно в процессе обучения DNN. DNN может иметь несколько нейронов в выходном слое, причем каждый выходной нейрон соответствует одной модели QSAR. Этот случай объединения DNN называют совместными или многозадачными (multitask DNN) [33]. Такие сети могут одновременно моделировать несколько задач QSAR, все встроенные модели QSAR имеют одинаковые веса и смещения в скрытых слоях, но имеют свой уникальный вес и смещение в выходном слое. Совместные DNN особенно полезны для задач QSAR с небольшим обучающим набором, т.к. в этом случае обучающий набор совместного DNN формируется путем объединения обучающих наборов всех вовлеченных задач QSAR. Ниже мы покажем разницу в производительности между совместными DNN и отдельными DNN, обученными с помощью одного набора данных.

Подчеркнем, что обучение DNN является сложным в вычислительном отношении из-за большого количества молекул в обучающем наборе и большого количества необходимых нейронов. Но поскольку при обучении DNN используются в первую очередь матричные операции, то вычисления на графических процессорах могут быть очень эффективными в смысле времени обучения. Таким образом, можно настроить следующие алгоритмические параметры (также называемые метопараметрами или гиперпараметрами DNN):

а) по данным:

– опции для преобразования дескриптора:

- 1) без преобразования;
- 2) логарифмическое преобразование;
- 3) двоичное преобразование;

б) по архитектуре сети:

- количество скрытых слоев;
- количество нейронов в каждом скрытом слое;
- выбор функций активации скрытых слоев:

- 1) сигмоидная функция;
- 2) ReLU;

в) по стратегии обучения DNN:

- обучение DNN из одного тренировочного набора или совместного DNN из нескольких тренировочных наборов;
- процент dropout нейронов в каждом слое;
- использование (или нет) предварительного обучения для инициализации  $\emptyset$ ;

г) по процедуре мини-пакетного стохастического градиентного спуска в алгоритме BP:

- количество молекул в каждом мини-пакете, то есть его размер;
- количество эпох, то есть сколько раз использовался тренировочный набор;
- параметры для управления процедурой оптимизации градиентного спуска, включая:

- 1) скорость обучения;

- 2) силу импульса (momentum strength);
- 3) увеличение веса (weight cost strength).

Ключевой вопрос в применении глубоких сетей, заключается в том, что можно ли найти такой набор значений для алгоритмических параметров DNN, чтобы для разнообразных наборов данных QSAR DNN могли делать более точные прогнозы, чем RF? Из-за большого количества настраиваемых параметров трудно оценить все комбинации возможных значений. Поэтому рекомендовалось тщательно выбирать разумное количество настроек параметров, корректируя значения одного или двух параметров за один раз, а затем вычислять  $R^2$  DNN, обученных с выбранными настройками параметров.

## 5 Результаты

Более 50 DNN прошли обучение с использованием различных настроек параметров. Эти настройки параметров были выбраны произвольно, но они охватывают достаточный диапазон значений каждого настраиваемого параметра. При произвольно выбранных настройках параметров DNN в среднем лучше RF в 11 из 15 наборов данных Kaggle. Более того, в пяти наборах данных DNN работают лучше, чем RF для всех настроек параметров. Среднее значение  $R^2$ , усредненное по всем DNN и всем 15 наборам данных, на 0.043 выше, чем у RF (улучшение на 10%). Это показывает, что, даже если для каждой задачи QSAR использовался наихудший параметр DNN, среднее значение  $R^2$  будет ухудшаться только с 0.423 до 0.412, то есть всего на 2,6%. Следовательно, результаты убеждают в том, что DNN, в целом, превосходят RF (таблица 1).

Чтобы понять, какие именно параметры обеспечивают превосходство DNN, исследовались параметры архитектуры сети и функции активации. Ключевые причины улучшения производительности DNN: (1) входной слой, который принимает тысячи дескрипторов без необходимости их сокращения, и (2) метод dropout, который успешно исключает переобучение. Второй метод снижения риска переобучения (предварительное обучение) признан нежелательным в задачах QSAR, т.к. эта процедура ухудшает прогнозирующую способность. Но почему так происходит – вопрос дальнейшего изучения.

Чтобы устранить сложность использования DNN из-за большого количества настраиваемых алгоритмических параметров, авторы приводят рекомендуемый набор значений для всех параметров DNN, которые подходят для больших наборов данных QSAR. Однако следует отметить, что, хотя рекомендуемые параметры дают хорошие прогнозы для большинства наборов данных, это не обязательно означает, что эти параметры являются наиболее эффективными в вычислительном отношении для любого конкретного набора данных.

Функция активации ReLU в QSAR предпочтительнее, чем Sigmoid. Одна из проблем, связанных с использованием ReLU, заключается в том, что она иногда может вызывать числовую нестабильность, поскольку не имеет верхней границы. Решением этой проблемы является предварительная обработка входных данных с помощью логарифмического или двоичного преобразования.

Ещё на один вопрос, постоянно возникающий у моделировщиков QSAR, ответ получен на конкурсе. Речь идет о небольших объёмах исходных данных для построения моделей. Поэтому сравнивалось обучение объединенных (совместных) DNN с использованием нескольких обучающих наборов, с обучением отдельного DNN с использованием одного набора данных. При усреднении по всем наборам данных, ка-

Таблица 1 Сравнение тестовых значений  $R^2$  на разных моделях

DNN, обученная с произвольными параметрами					
Наборы данных	Случайный лес	DNN с рекомендованными значениями параметров	median $R^2$	Наихудший случай	Наилучший случай
3A4	0.469	<b>0.521</b>	0.515	0.476	0.550
CB1	0.283	<b>0.345</b>	0.331	0.294	0.337
DPP4	<b>0.232</b>	0.194	0.224	0.186	0.259
HIVINT	0.353	<b>0.403</b>	0.316	0.263	0.393
HIVPROT	0.530	<b>0.543</b>	0.524	0.467	0.581
LOGD	0.685	<b>0.822</b>	0.808	0.769	0.836
METAB	0.631	<b>0.687</b>	0.661	0.603	0.690
NK1	0.403	<b>0.448</b>	0.414	0.373	0.449
OX1	0.501	<b>0.657</b>	0.584	0.484	0.628
OX2	0.580	<b>0.707</b>	0.621	0.540	0.695
PGP	0.550	<b>0.616</b>	0.592	0.521	0.622
PPB	0.420	<b>0.598</b>	0.561	0.508	0.615
RAT_F	0.098	<b>0.168</b>	0.138	0.065	0.187
TDI	<b>0.381</b>	0.350	0.339	0.319	0.377
THROMBIN	0.222	<b>0.375</b>	0.351	0.315	0.380
<b>mean</b>	<b>0.423</b>	<b>0.496</b>	<b>0.465</b>	<b>0.412</b>	<b>0.509</b>

жется, есть разница в пользу совместного DNN. Однако размер обучающих наборов играет решающую роль в том, выгоден ли совместный DNN. Неоднозначным является утверждение о том, что преимущество объединенного DNN состоит в том, что он использует молекулы, совместно используемые в разных наборах данных. Определить провал такого, казалось бы, удачного решения дефицита входных данных пока не удается.

И, наконец, важная и до сих пор еще не решенная задача – автоматические методы настройки параметров DNN, которые зависят от эффективности валидации, являющейся точным предиктором производительности теста. Следовательно, прежде чем мы сможем максимально использовать преимущества DNN, необходимо разработать новые подходы, которые могут лучше указывать прогнозирующую способность DNN в наборе тестов с разделением по времени.

## 6 Применение сетей глубокого обучения в прогнозировании свойства и активности соединения на небольших выборках данных

Особенностью всех выборок, которые использовались на конкурсе, является то, что они большие (десятки тысяч). После этой работы многие группы продемонстрировали, что массовые многозадачные сети могут значительно повысить предсказательную силу моделей глубокого обучения для прогнозирования свойств [42–44]. Параллельно с этим, другие группы представили сложные глубокие архитектуры для обработки и извлечения элементов из молекулярных структур [45]. Граф-сверточные архитектуры [46, 47], в частности, обрабатывают малые молекулы, представленные



в виде неориентированных графов и создают элементы с использованием обучаемых слоев свертки. В отличие от более ранних алгоритмов, позволяющих использовать небольшие молекулы, такие как круговые отпечатки пальцев [48], эти новые граф-сверточные архитектуры извлечения признаков являются удобными для обучения. Практический эффект этих инноваций был ограничен, т.к. большинство вышеупомянутых систем глубокого обучения требуют больших объемов данных. Например, массовые многозадачные сети до сих пор обучались с миллионами точек данных [44]. Это резко контрастирует с состоянием существующих технологий обнаружения лекарств, которые часто пытаются работать с несколькими десятками соединений, и даже сотня соединений часто слишком ресурсоемка для стандартных компаний, занимающихся разработкой лекарств.

Недавние достижения в области машинного обучения продемонстрировали, что в некоторых обстоятельствах нетривиальные предикторы могут быть обучены только по нескольким точкам данных [49–51]. Для сравнения новых точек данных с небольшим количеством доступных данных и последующего прогнозирования свойств этих новых точек данных используется сложная метрика. В более широком смысле, эти методы известны как «однократное обучение». Здесь задача состоит в том, чтобы предсказать поведение молекулы в новой архитектуре глубокого обучения [52], итеративное уточнение долговременной памяти (long short-term memory, LSTM) с модификацией архитектуры сопоставимых сетей (matching-networks, MN) и остаточной сверточной сети.

На рис. 2 представлена схема архитектуры однократного обучения.

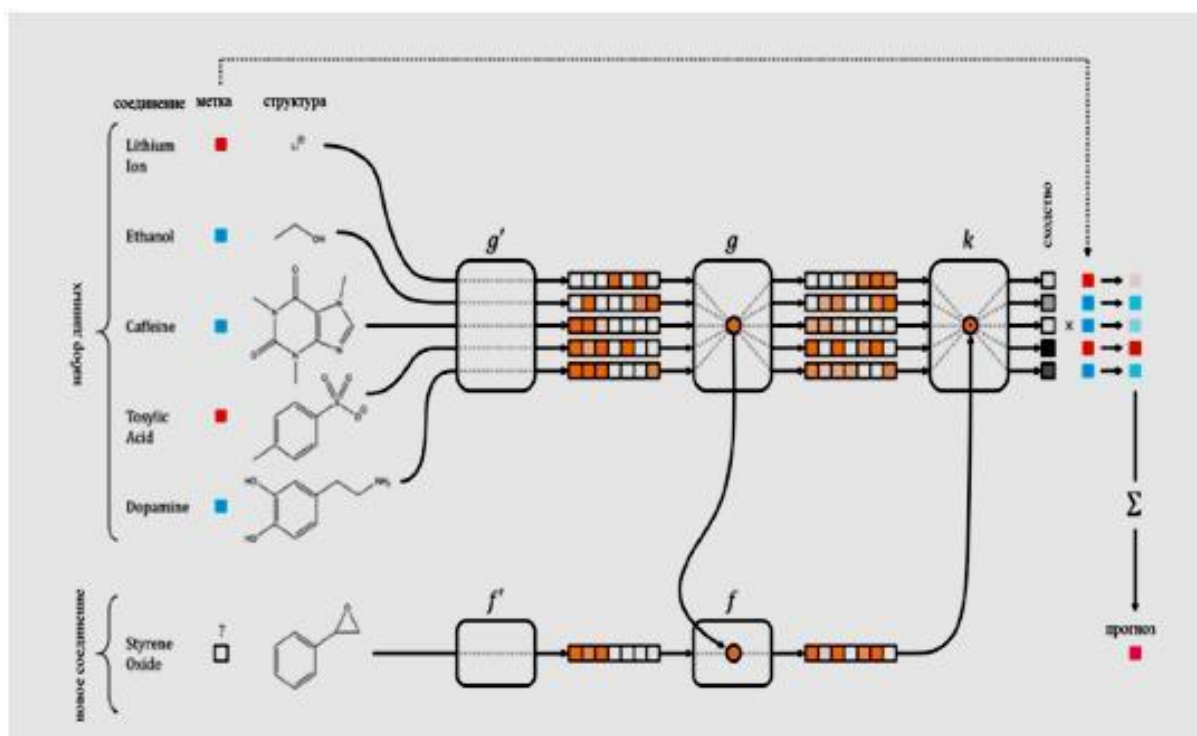


Рис. 2 Схема архитектуры однократного обучения [55]

Задача обучения, выполняемая ранее алгоритмами однократного обучения, состояла в том, чтобы выполнить распознавание объектов для новых классов изображений, используя только несколько примеров из каждого класса. Для молекул ана-

логичной задачей обучения было бы изучение поведения соединений в новом молекулярном каркасе для фиксированного экспериментального анализа, учитывая только несколько точек данных из нового каркаса. Но результаты [52] идут дальше и демонстрируют, что итеративные уточнения LSTM могут обобщаться для новых экспериментальных анализов, связанных, но не идентичных тестам в обучающей выборке. Следовательно, делается концептуальный шаг вперед, показывая, что однократные методы способны к более сильному обобщению, чем продемонстрировано ранее. В то же время ясно, что существуют сильные ограничения в обобщающей способности современных однократных методов обучения, и это есть предмет дальнейших исследований.

Не останавливаясь на подробном описании математического формализма и проведенных экспериментах (наша будущая работа с другими данными), отметим, что все однократные методы обучения демонстрируют сильное повышение по сравнению с исходным уровнем случайного леса, а модели LSTM с итеративным уточнением показывают более устойчивое повышение при наличии меньшего количества данных.

В глубоких сетях есть понятие «перенос обучения», которое было протестировано и в однократном обучении. До сих пор эксперименты проверяли способность однократных методов обучения обобщать задачи обучения на тесно связанные задачи тестирования. Чтобы проверить, способны ли однократные методы обучения к более широкому обобщению (переносу обучения), были обучены модели на наборе данных Tox21 и оценена точность прогнозирования на данных SIDER. Обратите внимание, что эти два набора данных в целом различны: Tox21 содержит результаты измерений из анализа ядерных рецепторов, а в SIDER измеряют побочные реакции реальных пациентов. Однако, ни одна из моделей однократного обучения не достигла значимой прогнозирующей силы в этой задаче, доказывая тем самым, что модели имеют ограниченные возможности обобщения для несвязанных систем.

## 7 Заключение

В то время как в соответствии с стандартными протоколами, были созданы доступные программные средства для интеллектуального анализа данных и извлечения знаний, в медицинской химии такие инструменты и инфраструктуры ещё недоступны. Это препятствует анализу данных в разработке лекарств, но и предоставляет большие возможности, и потому есть большие надежды на DL в разработке лекарств.

Частично они были подкреплены результатами конкурса прогнозирования активности в 2012 году, инициированного Merck, которые здесь цитировались, и Challenge-NINTox21 2014 года. В обоих случаях архитектуры DNN добились большей точности прогнозирования, чем другие методы машинного обучения, но ожидания до сих пор не были достаточно обоснованы. В этих случаях точность прогнозирования DL была в основном сопоставима с другими методами машинного обучения, и улучшения с помощью DL, если они были обнаружены, были незначительными.

Таким образом, потенциальное воздействие DL на медицинскую химию, следует рассматривать более критически. Однако, хотя предсказание свойств соединения представляет собой непосредственную область применимости DL, опыт показывает, что эта методология не будет «революционизировать» текущие исследования. Задача поиска новых активных соединений будет решающей. В будущем исследователи должны учитывать ограничения машинного обучения, чтобы последовательно идентифицировать новые химические объекты, независимо от используемых методов. Возможно, следует рассмотреть концептуально новые молекулярные представ-

ления и, возможно, многозадачное обучение на необработанных данных скрининга было бы перспективным для изучения по методологии DL [53].

## Литература

- [1] National Security Agency statement. URL: <https://www.nsa.gov/news-features/pressroom/statements/2013-08-09-the-nsa-story.shtml>.
- [2] *Gantz J., Reinsel D.* Extracting value from chaos. 2011. URL: <https://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>.
- [3] *Gantz J., Reinsel D.* The digital universe decade — are you ready? 2010. URL: <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-are-you-ready.pdf>.
- [4] *Howard J.* The business impact of deep learning // Proc. of the 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. — 2013. P. 1135. URL: <http://www.gartner.com/technology/research/top-10-technology-trends/>.
- [5] *Papadatos G. et al.* Activity, assay and target data curation and quality in the ChEMBL database // J. Comput., 2015. Vol. 29. P. 885–896. doi: <http://dx.doi.org/10.1007/s10822-015-9860-5>.
- [6] *Kim S. et. al.* PubChem substance and compound databases // Nucleic Acids Res, 2016. V. 44. P. 1202–1213. doi: <http://dx.doi.org/10.1093/nar/gkv951>.
- [7] *Gilson M.K. et. al.* BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology // Nucleic Acids Res, 2016. V. 44. P. 1045–1053. doi: <http://dx.doi.org/10.1093/nar/gkv1072>.
- [8] *Cortes C. and Vapnik V.* Support-vector networks // Mach. Learn, 1995. V. 20. №. P. 273–297. doi: <http://dx.doi.org/10.1007/BF00994018>.
- [9] *Salt D W. et. all.* The use of artificial neural networks in QSAR // Pesticide Science, 1992. V. 36. №2. P. 161–170. doi: <http://dx.doi.org/10.1002/ps.2780360212>.
- [10] *Ho T. K.* The random subspace method for constructing decision forests // Pattern Anal. Mach. Intell., 1998. V. 20. №8. P. 832–844. doi: <http://dx.doi.org/10.1109/34.709601>.
- [11] *Ammad-Ud-Din M. et. all.* Drug response prediction by inferring pathway-response associations with kernelized Bayesian matrix factorization // Bioinformatics, 2016. V. 32. №9. P. 455–463. doi: <http://dx.doi.org/10.1093/bioinformatics/btw433>.
- [12] *Ching T. et. all.* and obstacles for deep learning in biology and medicine // J. R. Soc. Interface, 2017. V. 15. №4. P. 69–77. doi: <http://dx.doi.org/10.1098/rsif.2017.0387>.
- [13] *Goh G. B. et. all.* Deep learning for computational chemistry // J. Comput. Chem., 2017. V. 38. №3. P. 1291–1307. doi: <http://dx.doi.org/10.1002/jcc.24764>.
- [14] *Gawehn E. et. all.* Deep learning in drug discovery // Mol. Inform, 2016. V. 35. P. 3–14. doi: <http://dx.doi.org/10.1016/j.drudis.2018.01.039>.
- [15] *Mamoshina P. et. all.* Applications of deep learning in biomedicine // Mol. Pharm, 2016. V. 13. №5. P. 1445–1454. doi: <http://dx.doi.org/10.1021/acs.molpharmaceut.5b00982>.
- [16] *Ekins S.* The next era: deep learning in pharmaceutical research // Pharm. Res, 2016. V. 33. №11. P. 2594–2603. doi: <http://dx.doi.org/10.1007/s11095-016-2029-7>.
- [17] *Baskin I. I. et. all.* A renaissance of neural networks in drug discovery // Expert Opin. Drug Discov., 2016. V. 11. №8. P. 785–795. doi: <http://dx.doi.org/10.1080/17460441.2016.1201262>.
- [18] *Goodfellow I., Bengio Y. and Courville A.* Deep Learning. — MIT Press, 2016. URL: <http://www.deeplearningbook.org>.
- [19] *Srivastava N. et. all.* Dropout: a simple way to prevent neural networks from overfitting // Machine Learning Res, 2014. V. 15. P. 1929–1958.

- [20] Wan L. et. all. Regularization of neural networks using DropConnect. // In Proc. of the 30th Int. Conf. on Machine Learning, 2013. V.28. P. 1058–1066.
- [21] Nair V. and Hinton G. E. Rectified linear units improve restricted boltzmann machines // In Proc. of the 27th Int. Conf. on Machine Learning, 2010. P. 807–814.
- [22] TensorFlow. URL: <https://www.tensorflow.org/>
- [23] Caffe. URL: <http://caffe.berkeleyvision.org/>
- [24] PYTORCH. URL: <http://pytorch.org/>
- [25] Keras. URL: <https://keras.io/>
- [26] Theano. URL: <http://deeplearning.net/software/theano/>
- [27] Lee H. et. all. Unsupervised learning of hierarchical representations with convolutional deep belief networks // Commun. ACM, 2011. V.54. №10. P. 95–103. doi: <http://dx.doi.org/10.1145/2001269.2001295>.
- [28] Hochreiter S. and Schmidhuber J. Long short-term memory // Neural Comput., 1997. V. 9. №8. P. 1735–1780.
- [29] Bengio Y. Learning deep architectures for AI // Found. Trends Mach. Learn, 2009. V. 2. №1. P. 1–127. doi: <http://dx.doi.org/10.1561/2200000006>.
- [30] Kingma D. P. and Welling M. Auto-encoding variational bayes // 2013. doi: <http://dx.doi.org/ArXivdoi:1312.6114>.
- [31] Breiman L. Random forests // Machine Learning, 2001. V.45. №1. P. 5–32. doi: <http://dx.doi.org/10.1023/A:1010933404324>.
- [32] Cortes C., Vapnik V. N. Support-vector networks // Machine Learning, 1995. V. 20. P. 273–297. doi: <http://dx.doi.org/10.1007/BF00994018>.
- [33] Dahl G. E., Jaitly N., Salakhutdinov R. Multi-task neural networks for QSAR predictions // 2014. doi: <http://dx.doi.org/http://arxiv.org/abs/1406.1231>.
- [34] Svetnik V. et. all. Boosting: an ensemble learning tool for compound classification and QSAR modeling // J. Chem. Inf. Comput. Sci., 2005. V.45. №3. P. 786–799 doi: <http://dx.doi.org/10.1021/ci0500379>.
- [35] Burden F. R. Quantitative structure-activity relationship studies using Gaussian Processes // J. Chem. Inf. Comput. Sci., 2001. V.41. №3. P. 830–835 doi: <http://dx.doi.org/10.1021/ci000459c>.
- [36] Sheridan R. P. Time-split cross-validation as a method for estimating the goodness of prospective prediction // J. Chem. Inf. Model., 2003. V.53. №4. P. 783–790. doi: <http://dx.doi.org/10.1021/ci400084k>.
- [37] Carhart R. E., Smith D. H., Ventkataraghavan R. Atom pairs as molecular features in structure-activity studies: definition and application // J. Chem. Inf. Comput. Sci., 1985. V. 25. №2. P. 64–73.
- [38] Kearsley S. K. et. all. Chemical similarity using physiochemical property descriptors // J. Chem. Inform. Comp. Sci., 1996. V. 36. №1. P. 118–127. doi: <http://dx.doi.org/10.1021/ci950274j>.
- [39] Rumelhart D. E., Hinton G. E., Williams R. J. Learning representations by back-propagating errors // Nature, 1986. V.323. P. 533–536. doi: <http://dx.doi.org/10.1038/323533a0>.
- [40] Hinton G. E., Osindero S., Teh Y. W. A fast learning algorithm for deep belief nets // Neural Computation, 2006. V.18. №7. P. 1527–1554. doi: <http://dx.doi.org/10.1162/neco.2006.18.7.1527>.
- [41] Srivastava N. et. all. Dropout: A simple way to prevent neural networks from overfitting // J. Mach. Learn. Res, 2014. V.15. №6. P. 1929–1958.

- [42] *Ma J. et. all.* Deep neural nets as a method for quantitative structure-activity relationships // *J. Chem. Inf. Model.*, 2015. V. 55. №2. P. 263–274. doi: <http://dx.doi.org/10.1021/ci500747n>.
- [43] *Ramsundar B. et. all.* Massively multitask networks for drug discovery. 2015 // doi: <http://dx.doi.org/arXivpreprintarXiv:1502.02072>.
- [44] *Unterthiner T. et. all.* Deep Learning as an Opportunity in Virtual Screening // *Neural Inf. Proc. Sys. DL Workshop*, 2014. V. 27.
- [45] *Lusci A., Pollastri G., Baldi P.* Deep architectures and deep learning in chemoinformatics: the prediction of aqueous solubility for drug-like molecules // *J. Chem. Inf. Model.*, 2013. V. 53. №7. P. 1563–1575. doi: <http://dx.doi.org/10.1021/ci400187y>.
- [46] *Duvenaud D. K. et. all.* Convolutional networks on graphs for learning molecular fingerprints // *Neural Inf. Proc. Sys.*, 2015. P. 2224–2232.
- [47] *Kearnes S. et. all.* Molecular Graph Convolutions: Moving Beyond Fingerprints // *J. Comput.-Aided Mol. Des.*, 2016. V. 30. P. 595–608. doi: <http://dx.doi.org/10.1007/s10822-016-9938-8>.
- [48] *Rogers D., Hahn M.* Extended-connectivity fingerprints // *J. Chem. Inf. Model.*, 2010. V. 50. №5. P. 742–754. doi: <http://dx.doi.org/10.1021/ci100050t>.
- [49] *Santoro A. et. all.* One-shot Learning with Memory-Augmented Neural Networks // 2016. doi: <http://dx.doi.org/arXivpreprintarXiv:1605.06065>.
- [50] *Vinyals O. et. all.* Matching Networks for One Shot Learning // *Advances in Neural Information Processing Systems*, 2016. P. 3630–3638.
- [51] *Han Altae-Tran, BharathRamsundar, Aneesh S. Pappu, Vijay Pande Low.* Data Drug Discovery with One-Shot Learning // *ACS Cent. Sci.*, 2017. V. 3. №4. P. 283–293. doi: <http://dx.doi.org/10.1021/acscentsci.6b00367>.
- [52] *He K. et. all.* Identity Mappings in Deep Residual Networks // *European Conference on Computer Vision*, 2016. P. 630–645.
- [53] *Jurgen Bajorath* Data analytics and deep learning in medicinal chemistry // *Future Med. Chem.*, 2018. doi: <http://dx.doi.org/10.4155/fmc-2018-0129>.
- [54] *Chen H., Engkvist O., Wang Y., Olivecrona M., Blaschke T.* The rise of deep learning in drug discovery // *Drug Discovery Today*, 2018. Vol. 23. Issue 6. P. 1241–1250. doi: <http://dx.doi.org/10.1016/j.drudis.2018.01.039>.
- [55] *Altae-Tran H., Ramsundar B., Pappu A.S., Pande V.* Low Data Drug Discovery with One-Shot Learning // *ACS Central Science*, 2017. Vol. 3. Issue 4 P. 283–293. doi: <http://dx.doi.org/10.1021/acscentsci.6b00367>.

Поступила в редакцию 04.01.2019

UDC 577.29

## DEEP LEARNING NEURAL NETWORKS IN MODELING THE STRUCTURE-ACTIVITY RELATIONSHIP

*Adilova F.T.*

Institute of Mathematics named after V.I. Romanovskiy NAS RUz, Tashkent

Artificial neural networks became popular in molecular and chemical informatics (in drug design) about two decades ago, but over the past decade, deep learning neural networks have achieved remarkable success in various areas of artificial intelligence. This

technology, which grew out of research on artificial neural networks, has shown excellent performance compared to other machine learning algorithms in areas such as image and voice recognition, natural language processing, and others. The first wave of applications of deep learning in pharmaceutical research has appeared in recent years; today it has demonstrated its perspective in solving various problems in the field of drug discovery. The paper discusses the works related to the prediction of bioactivity based on the quantitative structure-activity relation (QSAR) and the solution of these problems in the case of small amounts of input data.

**Keywords:** deep learning neural networks, models of structure-activity relationship, small amounts of input data.

**Citation:** Adilova F.T. 2019. Deep learning neural networks in modeling the structure-activity relationship. *Problems of Computational and Applied Mathematics*. 3(21):5–18.