

## НЕЙРОСЕТЕВЫЕ МЕТОДЫ ПРЕДСКАЗАНИЯ ЛИПОФИЛЬНОСТИ МАЛЫХ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ

Б. И. Пякилля

Научный руководитель – д.т.н., профессор ОАР ИШИТР ТПУ В. И. Гончаров

*Национальный исследовательский Томский политехнический университет  
Россия, г. Томск, пр. Ленина, д. 30, morphism@tpu.ru*

Предсказание молекулярных свойств веществ является одной из важнейших задач хемоинформатики – научной дисциплины, которая занимается разработкой и применением методов информатики для решения химических задач. Некоторые молекулярные свойства играют основополагающую роль в задачах биохимии и конкретно – в проектировании лекарственных средств (drug design). Примером такого молекулярного свойства является физико-химическое свойство липофильности, т. е. способности вещества растворяться в липидах – жирах и маслах. Липофильность является ключевым физическим свойством лекарственного средства т. к. определяет способность вещества проникать через клеточную мембрану, которая состоит из билипидного слоя фосфолипидов [1]. Кроме того, липофильность является важным дескриптором (признаком) для моделей quantitative structure-activity relationship. Данный тип моделей может использовать липофильность как входной признак, описывающий изучаемое вещество, либо как выходной, т. е. предсказываемый. Конкретный же вид моделей выбирается исследователем и может включать в себя как простые линейные регрессионные модели, так и сложные нелинейные модели типа глубоких нейронных сетей. Поиск же архитектуры последних является непростой задачей и часто занимает значительное время, что усложняет процесс исследования возможных биохимических свойств соединений.

Таким образом, целью данной работы является исследование влияния сложности архитектуры нейронной сети на качество её работы.

В основе вычислительного эксперимента будут лежать данные о липофильности, собранные из открытой базы данных ChEMBL, а точнее из MoleculeNet. Количество соединений в выбранных данных равняется 4200 молекулам. Разделение на обучающую и тестовую выборки проводилось в соотношении 80/20. Каждая молекула описывается строкой SMILES, для которой были посчитаны дескрипторы Моргана в размере 2048 битов (Morgan Fingerprints).

В ходе работы архитектуру нейросетевой модели определили следующим образом:

- 1) Два полносвязных скрытых слоя с  $N$  и  $N/2$  нейронами соответственно;
- 2) Один выходной нейрон, выходом которого является предсказанное значения липофильности.

Определенная выше неизвестная величина  $N$  отражает количество нейронов и именно это значение нам нужно найти. Поиск будем осуществлять обычным перебором (поиск по сетке), начиная с 50-и нейронов и заканчивая 500-ми нейронами с шагом, равном 50, а результаты в виде RMSE представим в виде таблицы 1.

**Таблица 1.** Результаты обучения модели

Количество нейронов $N$	Обучающее RMSE	Тестовое RMSE
50	2,081	2,693
100	2,005	2,651
150	1,971	2,510
200	1,747	2,201
250	1,516	2,003
300	1,884	2,192
350	1,857	2,112
400	1,914	2,214
450	1,759	2,321
500	1,997	2,307

Исходя из таблицы 1 видно, что оптимальным значением количества скрытых нейронов является 250. В случае увеличения сложности сети происходит переобучение, в случае меньшем значении чем 250 – недообучение.

Таким образом показано, что поиск структуры нейронной сети – это процесс итеративный и занимающий определенное время, которое сильно зависит от условий поиска (в данном случае это шаг изменения количества нейронов в скрытых слоях).

## Список литературы

1. Кольман Я., Рем К. Г. *Наглядная биохимия*. – мир, 2000. – Т. 469. – С. 23.

## ФАЗОВОЕ СОСТОЯНИЕ СИСТЕМЫ ДИОКСИД УГЛЕРОДА – n-БУТАНОЛ ПРИ ПОВЫШЕННЫХ ТЕМПЕРАТУРАХ

В. И. Репина

Научный руководитель – к.х.н., доцент Т. Э. Скребец

Северный (Арктический) федеральный университет имени М. В. Ломоносова  
Архангельск, набережная Северной Двины, 17, public@narfu.ru

В настоящее время сверхкритические (СК) флюиды широко применяются в процессах экстракции в качестве альтернативы классическим растворителям. Наиболее часто используют диоксид углерода в около и сверхкритическом состоянии. В обоих случаях он является неполярным растворителем, поэтому нередко требует применения соразтворителя, позволяющего регулировать его растворяющую способность. В качестве таких соразтворителей применяют n-спирты, ацетон, ксилолы и другие органические вещества, способные образовывать гомогенные системы с  $\text{CO}_2$ . Возникает необходимость определения условий, в которых такие бинарные системы будут однофазны. Современные экспериментальные методы, позволяющие решать эту задачу для СК флюидов, разделяют на две группы – статические (реализуются в замкнутой ячейке постоянного или переменного объема) и динамические (реализуются в проточной системе). При использовании статических методов жидкость и флюид находятся в контакте при перемешивании, пока не достигнут

состояния равновесия. Эти методы позволяют визуализировать процесс [1]. Изменяя температуру и давление в системе заданного состава, можно наблюдать исчезновение или появление второй фазы. Статический метод может быть реализован в трех вариантах: аналитическом, синтетическом и гравиметрическом. Наиболее прост в исполнении синтетический метод, не требующий отбора пробы и ее анализа. В ячейку, снабженную сапфировыми окнами, помещают известные количества компонентов. Изменяя давление при заданной температуре, наблюдают исчезновение границы раздела фаз, после чего снижают давление до точки помутнения, которая и свидетельствует о растворимости вещества во флюиде.

Систему  $\text{CO}_2$ -этанол наиболее часто используют для СК экстракции, поэтому она наиболее изучена с точки зрения фазового состояния [2]. Существуют опубликованные данные о системах с изо-пропанолом, ацетоном в качестве соразтворителей.

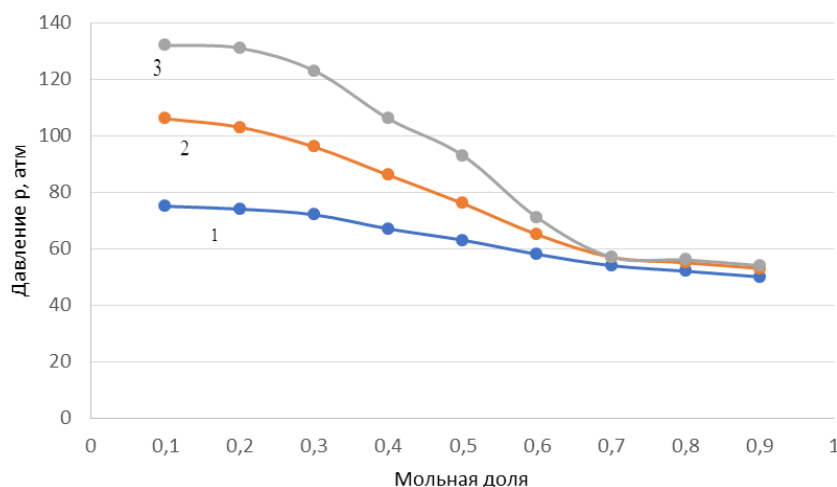


Рис. 1. Зависимости давления гомогенизации бинарной системы диоксид углерода – n-бутанол от состава при различных температурах: 1 – 313 К, 2 – 333 К, 3 – 358 К