

Применение интеллектуальной ДСМ-системы для предсказания токсичности

Блинова В.Г.

Федеральный исследовательский центр

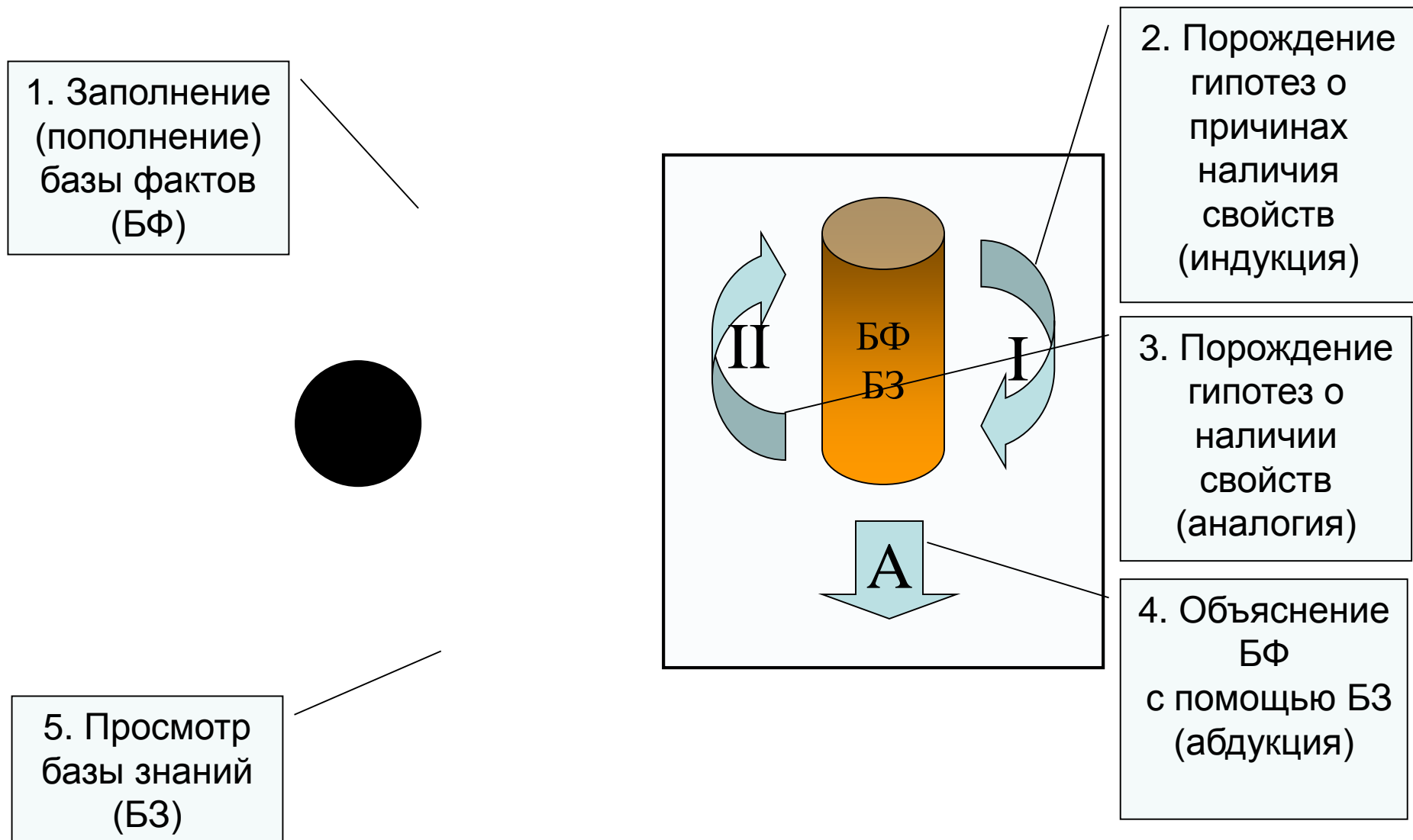
«Информатика и управление» РАН

Харчевникова Н.В., Жолдакова З.И.

НИИ экологии человека и гигиены окружающей среды им.

А.Н. Сысина Минздрава РФ

Схема работы интеллектуальной системы типа ДСМ

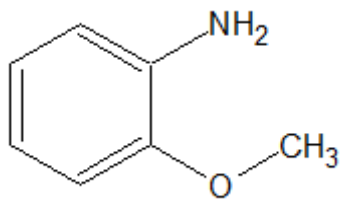


Интеллектуальная ДСМ система для прогнозирования биологической активности химических соединений

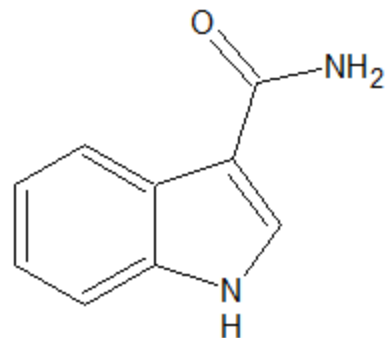
- ДСМ решатель, который на основе анализа сходств порождает гипотезы о причинах и гипотезы о предсказаниях;
- автоматический кодировщик, который в пакетном режиме преобразует химические структуры в код ФКСП;
- квантовый вычислитель, используемый для расчета устойчивых метаболитов;
- средства определения сходства, которые реализуют операции пересечения, объединения и проверки включения над множествами кодов ФКСП;
- средства графической визуализации исходных данных и результатов эксперимента;
- интерфейсная оболочка, которая взаимодействует с пользователем и управляет процессом проведения эксперимента.

Пример кодирования ФКСП

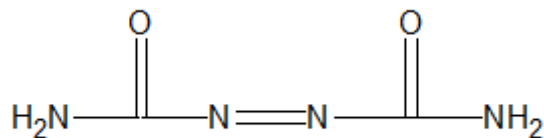
(проблемно ориентированный язык)



6,06 0200331
0262121
1200331
1201411



56,10N1
0201131
0201331



0201051 0200051
0201131 0201131
0500051 0500051
0501131 0501131

ДСМ эксперименты

Основные исследованные активности:

1. Антибактериальная
2. Антилепрозная
3. Астмагенная
4. Ингибиторы холинэстеразы
5. Канцерогенность
6. Кумулятивная токсичность
7. Мутагенная
8. Противоопухолевая
9. Хроническая токсичность
10. Цитотоксичность

Преимущества ДСМ-подхода:

- работа с малыми выборками (36-150 соединений);
- возможность работы с выборками размера 500-700 соединений;
- автоматическое выявление активных фрагментов (фармакофоров и антифармакофоров);

**Текущие эксперименты:
соединения, обладающие сенсibiliзирующим
эффектом (астмагенная активность)**

**Малеиновый ангидрид
Метилбензоат
Ампициллин
Этаноламин
Фталевый ангидрид
Диизоцианатметилбензол
Пиперазин
Бензилпенициллин
Диметилэтанолламин
Этилендиламин
Дифенилметандиизоцианат
Цефалексин**

Прогноз наличия сенсibiliзирующего эффекта веществ при ингаляционном воздействии

- Обучающая выборка содержала 80 соединений (40 активных и 40 неактивных). Проведен компьютерный эксперимент ДСМ-методом. Результаты сопоставлены с полученными на той же выборке с использованием программ Multicase¹, cat-SAR² и методом read-across³.
- Были выделены структурные фрагменты(гипотезы), определяющие аллергенную активность соединений и ее отсутствие. Положительными гипотезами были изоцианатные группы $-N=C=O$, структуры с пятичленным циклом, содержащим в цикле атом кислорода и две соседние карбонильные группы (ангидриды), молекулы с четырехчленными циклами (антибиотики пенициллинового ряда). Выделены также в качестве положительных гипотез алкилдиамины. Наличие в молекулах сложных эфиров длинных углеводородных цепей определяет отсутствие активности соединений.
- Применение скользящего контроля в ДСМ экспериментах позволило осуществить прогноз, точность которого составила 95%, чувствительность 94%, специфичность 97%, что не хуже результатов других авторов на том же массиве.

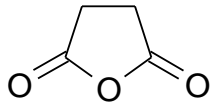
¹ Graham C., Rozencranz H.S. et al. Regul. Toxicol. Pharmacol., 1997, v.26, 266.

² Cunningham A.R. et al. SAR and QSAR in Environ. Res., 2005, v.16, 273-285.

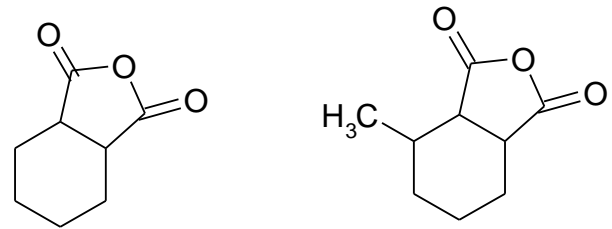
³ Enoch S.J., Roberts D.W., Cronin M. T. D.Chem. Res. Toxicol., 2010, v.23, 1547-1555.

Примеры гипотез

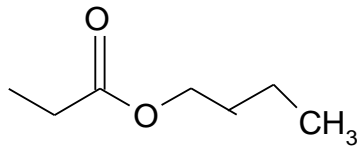
- Положительная гипотеза



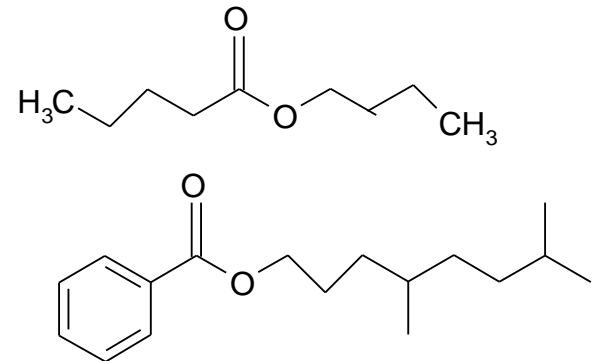
«Родители» гипотезы



Отрицательная гипотеза



«Родители» гипотезы



Сопоставление результатов различных методов на одной и той же выборке из 80 соединений. Оценка по скользящему контролю

Метод	Точность	Специфичность	Чувствительность
Cat-SAR	92%	89%	95%
Multicase	83-87%	95%	72-80%
Cat-SAR с физико-химическими параметрами	79%	74%	85%
ДСМ-метод	95%	97%	94%

Точность прогноза определяется как процентное отношение общего числа правильно предсказанных активных и неактивных соединений к числу тестовых соединений, для которых был сделан прогноз, чувствительность - как процент правильно предсказанных активных соединений от числа активных тестовых соединений и специфичность - это процент правильно предсказанных неактивных от числа неактивных тестовых соединений.

Выводы

Решаемые задачи:

- Задачи прогнозирования антипродуктивных свойств (токсичности, канцерогенности, мутагенности) химических соединений;
- Задачи мониторинга опасных соединений в рамках экологической экспертизы;
- Возможность создания лекарственных соединений на основании работы ДСМ-системы.

Получены авторские свидетельства на программные продукты, входящие в систему. Создаются новые версии ДСМ-системы с последующей защитой авторских прав.

ДСМ-система продемонстрировала преимущество над остальными системами машинного обучения в рамках международного соревнования (предсказательной токсикологии, Фрайбург, 2002 г.)