

Вопросы по курсу «Анализ молекулярных графов»

Проф. Кумсков М.И. kumskov@mail.ru

1. Постановка задачи анализа и классификации структурных объектов (молекулярных графов – граф с маркированными вершинами – М-графов).
2. Линейные нотации представления молекулярных графов в ЭВМ: WLN нотация Виссвесера, SMILE-код М-графа, волновой алгоритм формирования матрицы топологических расстояний.
3. Представление молекул химических соединений в ЭВМ. Формат *.mol. Представление выборки – формат *.sdf. Представление молекулярных графов в компьютере, графические редакторы
4. Особые точки (ОТ) – элементы описания структурных объектов, представленных графами. Вершины и ребра М-графа как варианты ОТ. Маркирование особых точек. Использование маркеров для учета локальных физ.-химических свойств химических соединений.
5. Матрица объект-признак (признак – фрагмент). Особенности построения ОП-матриц. Формирование ОП-матриц на основе молекулярных дескрипторов. Топологические индексы М-графов.
6. Описание объектов в виде (фрагмент, число повторений). Эквивалентность фрагментов. Формирование имен k-фрагментов (k = 1, 2, 3, 4). Индуктивное построение. Структурные символьные спектры М-графов
7. Теорема о расщеплении столбцов ОП-матрицы. Адаптивный поиск описания М-графа под заданное свойство химического соединения
8. Линейная множественная регрессия. Скользящий контроль. Коэффициент множественной корреляции. Построение линейных моделей на очень широких матрицах – использование эволюционных алгоритмов (МГУА).
9. Эволюционные алгоритмы - метод группового учета аргументов(МГУА), генетические алгоритмы.
10. Метод Группового Учета Аргументов (МГУА – линейная версия). Схема и параметры настройки алгоритма. Сокращение комбинаторного взрыва за счёт итерации «МГУА – формирования ОП-матрицы» по k-фрагментам (k = 2, 3, 4)
11. Зависимость и отсев признаков в ОП-матрице перед ее анализом. Корреляция признаков. Отбор значимых признаков (фрагментов) – факторный анализ (SVD-разложение ОП-матрицы).

12. Геометрический подход к анализу данных. Близость M-графов в евклидовом пространстве дескрипторов. Метрики M-графов. Алгоритмы кластерного анализа
13. Выбор оптимальной метрики при построении дерева решений. Генетический алгоритм. Факторный анализ (метод главных компонент).
14. Использование кластеров при формировании обобщённых деревьев решений. Отказ от прогнозов (классификатор). Двухфазная схема построения моделей классификаторов.
15. Алгоритмы кластер-анализа: минимальное покрывающее дерево, k-средних (классический и с ядрами), иерархический, формальный элемент (ФОРЭХО).

Литература

1. Прогнозирование свойств химических соединений. Унифицированный Репозиторий моделей «структура-свойство» - Москва, 2012, Изд-во МАКС Пресс, ISSN 978-5-317-04335-3,
2. Система прогнозирования свойств химических соединений. Алгоритмы и модели. – Москва, 2008, Издательство МАКС Пресс, ISBN 978-5-317-02704-9,
3. Химические приложения топологии и теории графов, под ред. Р. Кинга. — М.: Мир, 1987. — 560 с
4. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003 - 649 с.
5. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger - Weinheim: Wiley-VCH, 2003 - 1870 с
6. Хельтье Х.-Д. и др. Молекулярное моделирование: теория и практика - пер. с англ. - М: Бином. Лаборатория знаний, 2009.-318 с.
7. В.В. Корнеев, А.Ф. Гареев, С.В. Васютин, В.В. Райх. Базы данных. Интеллектуальная обработка информации. Нолидж, 2000.

Дополнительная литература

1. Chemoinformatics: a textbook. Под ред. J. Gasteiger, T. Engel. - Weinheim: Wiley-VCH, 2003 - 649 с.
2. Leach A.R., Gillett V.J. An introduction to Chemoinformatics. - Dordrecht: Kluwer Academic Publishers, 2003 - 259 с.
3. Handbook of Chemoinformatics: From Data to Knowledge. В 4 т. Под ред. J. Gasteiger - Weinheim: Wiley-VCH, 2003 - 1870 с.
4. M. Mohri, A. Rostamizadeh, A. Talwalkar. Foundations of Machine Learning, MIT Press, 2012.
5. C. M. Bishop, Pattern Recognition and Machine Learning (Information Science and Statistics), Springer, 2006