

Цели исследования

Цель работы

Предложить метод отбора признаков, учитывающий взаимное расположение признаков и целевого вектора.

Проблема

Методы отбора признаков дают избыточное подмножество мультикоррелирующих признаков.

Метод решения

Использование постановки задачи квадратичного программирования для получение оптимального подмножества признаков.

- Тематические модели *неполны и неустойчивы*.
- Получение хорошей тематической модели, как правило, требует больших затрат времени.
- Не существует идеального автоматического способа оценивания качества тематических моделей.

Решение

Банк тем — инструмент для сохранения интерпретируемых тем, построенных при многократных запусках, с целью последующего их использования для оценки качества моделей.

Цели

Реализовать метод построения банка тем и оценивания качества тематических моделей с помощью банка тем.

Цель: предложить алгоритм поиска характерных квазипериодических сегментов внутри временного ряда, полученных при помощи мобильного акселерометра.

Задачи

- 1 Предложить признаковое описание точек временного ряда.
- 2 Предложить функцию расстояния между точками временного ряда в новом признаковом описании, для их дальнейшей кластеризации.

Исследуемая проблема

- 1 Понижение размерности пространства признаков. Построение признакового описания точек временного ряда.

Метод решения

Алгоритм поиска характерных сегментов основывается на методе главных компонент для локального снижения размерности сегмента фазовой траектории в окрестности каждой точки временного ряда. Главные компоненты рассматриваются как признаковое описание точек временного ряда.

Требуется

Построить модель предсказания молекулярного графа основного продукта химической реакции по графам исходных веществ.

На модель накладываются ограничения:

- применима к данным в виде несвязанного молекулярного графа;
- допускает использования экспертных знаний о локальной структуре молекулярного графа;

Проблема

Пространство молекулярных структур высоко-размерное. Количество механизмов реакций растет с ростом числа известных структур.

Метод

Графовая нейронная сеть, допускающая использование экспертных знаний о структуре молекулярного графа.

Значимость

Предлагаемый подход предназначен для улучшения систем информационного поиска, основанных на экспертных оценках релевантности документа запросам.

Коллекции документов

Следуя традициям сообщества ИП, мы ставим своей целью построение ранжирующих функций, дающих высокий MAP на коллекциях TREC.

Актуальность

Постоянное развитие TREC-сообщества, программных пакетов, связанных в т.ч. с ранжирующими функциями (напр. Terrier) демонстрирует актуальность поставленной задачи.

Цель исследования: создать метод выбора мультимodelей при построении моделей банковского кредитного скоринга.

Мотивация: Логистическая модель является де-факто стандартом в банковском скоринге, мультимodelи являются интерпретируемым обобщением, позволяющим учитывать неоднородности в данных.

Проблема: мультимodelь может содержать большое число похожих modelей, что ведет к ее неинтерпретируемости и низкому качеству прогноза. Признаковые пространства modelей могут не совпадать, в частности иметь разную размерность.

Метод решения задачи: анализ пространства параметров мультимodelи с помощью введенной функции сравнения modelей.

Задача

Построить прогнозы семейства временных рядов, связанных в иерархическую многоуровневую структуру и описывающих объемы погрузки ряда грузов в заданных узлах РЖД с разным уровнем детализации.

Требования к модели

- прогнозы должны быть точны — обеспечивать минимально возможное значение заданной функции потерь;
- прогнозы должны удовлетворять физическим ограничениям — лежать в заданном интервале для каждого временного ряда;
- прогнозы должны удовлетворять условию согласованности (структуре иерархии).

Проблема согласования прогнозов

Прогнозы, полученные для каждого временного ряда независимо, могут не удовлетворять структуре иерархии, т. е. не быть *согласованными*.

Снижение размерности траекторного пространства

Задача

Решается задача поиска связей между временными рядами.

Проблема

Размерность траекторного пространства временного ряда может быть избыточна. Это усложняет описание ряда и приводит к неустойчивости прогностических моделей.

Требуется

Понизить размерность траекторного пространства временного ряда. В полученном пространстве меньшей размерности построить аппроксимацию исходного временного ряда.

Предлагается

Использовать метод сферической регрессии для снижения размерности траекторного пространства.

Цель: Предложить метод оценки объема выборки на основе близости между эмпирическими распределениями побвыборок для получения оптимального качества классификации при выборе между порождающим и разделяющим подходами.

- 1 Определение достаточного объема выборки. Оценка объема выборки на основе расстояния Кульбака-Лейблера
- 2 Свойства расстояния Кульбака-Лейблера
- 3 Задача классификации: разделяющий и порождающий подходы
- 4 Оценка объема выборки при выборе между подходами
- 5 Основные результаты

Классификация временных рядов

Цель

Предложить способ построения ансамбля моделей локальной аппроксимации для классификации сигналов носимых устройств.

Гипотеза

Ансамбль моделей локальной аппроксимации предпочтительнее в парето-оптимальном смысле универсальной модели (нейросети): точнее, устойчивее, проще.

Задача

Требуется построить признаковое описание временных рядов на используя параметры моделей локальной аппроксимации.

Метод

Предложить критерий сложности ансамбля для выбора оптимального признакового описания.

Задача декодирования временного ряда

Цель

Исследовать зависимости в пространствах объектов и ответов и построить устойчивую модель декодирования временных рядов в случае коррелированного описания данных.

Проблема

Целевая переменная – вектор, компоненты которого являются зависимыми.

Требуется построить модель, адекватно описывающую как пространство объектов так и пространство ответов при наблюдаемой мультикорреляции в обоих пространствах высокой размерности.

Решение

Для учёта зависимостей в пространствах объектов и ответов предлагается снизить размерность с использованием скрытого пространства.

Задача молекулярного докинга (CASF)

Ранжирование синтезированных молекулярных комплексов (конформаций) по энергетической устойчивости.

Проблема

Существующие подходы моделируют физический потенциал взаимодействия с привлечением данных из множества источников и используют избыточно сложные модели.

Предлагается

- 1 разбить молекулярные комплексы на элементарные взаимодействующие пары аминокислота — лиганд,
- 2 построить модели вероятностных распределений взаимного расположения элементарных пар в \mathbb{R}^3 ,
- 3 использовать метрические методы в пространстве полученных распределений.