

КЛАСТЕРИЗАЦИЯ ПУЧКОВ ВРЕМЕННЫХ РЯДОВ ПУТЕМ ДИНАМИЧЕСКОЙ СВЕРТКИ СВОБОДНОЙ ПЕРЕМЕННОЙ

Стрижов В.В.

Вычислительный центр РАН, 119333, Москва, Вавилова, 40, тел.: (495) 1354163,
strijov@ccas.ru

Задано множество временных рядов. Временные ряды имеют, в общем случае, различное число элементов. Каждый из временных рядов относится к одному из двух кластеров. Требуется найти такие инвариантные преобразования свободной переменной — времени, при которых сумма ошибок первого и второго рода, нормированных числом элементов классов, минимальна. Инвариантные преобразования отыскивается на индуктивно заданном множестве параметрических функций.

Существует, по крайней мере, два пути найти инвариант, разделяющий временные ряды на кластеры. Первый путь — поиск таких параметрических моделей временных рядов, которые бы достаточно хорошо аппроксимировали эквивалентные временные ряды [1]. При этом классификация выполняется в пространстве параметров полученных моделей.

Второй, предлагаемый, путь — поиск расстояний между временными рядами на декартовом произведении их элементов. При этом отыскивается не модель множества эквивалентных рядов, а инвариант, отображающий классы эквивалентности из множества временных рядов во множество оценок принадлежности кластеру. Для выбора инвариантов используется техника динамической свертки свободных переменных, DTW (dynamic time warping). Путь минимальной стоимости, полученный в результате построения матрицы DTW, аппроксимируется монотонной функцией, после чего вычисляется расстояние невязки между путем и его аппроксимацией. Функция расстояния является классификатором временных рядов.

Предложенный метод был использован для кластеризации временных рядов давления в камере внутреннего сгорания дизельного двигателя. Непосредственное вычисление значения пути оптимальной стоимости DTW не позволило решить задачу кластеризации исследуемых временных рядов, так как стоимость двух несовпадающих путей некоторых временных рядов из разных кластеров оказывалась одинаковой. Создание моделей, аппроксимирующих эквивалентные временные ряды также не позволило решить данную задачу, так как кластеризацию при этом приходилось выполнять в пространстве параметров, которое имело большую размерность. Предложенный метод позволил разделить данные временные ряды на кластеры, так как кластеризация выполнялась в пространстве параметров небольшой размерности. Работа поддержана грантом РФФИ 07-07-00181.

Литература.

1. *Стрижов В.В.* Поиск параметрической регрессионной модели в индуктивно заданном множестве. Вычислительные технологии. 2007. Том 12, N 1. С. 93-102.