

Локальная аппроксимация временных рядов для построения признакового описания в задачах классификации и прогнозирования

Цель: создание методов анализа влияния параметров локальных моделей на качество классификации.

Мотивация: при создании метапрогностических моделей (моделей прогнозирования прогностических моделей) остается открытой проблема использования значений параметров локальных моделей при создании метамodelей. Цель нижеприведенного проекта - создание инструмента для анализа этой проблемы.

Объект исследования – сегмент временного ряда. Требуется спрогнозировать класс сегмента. (Вариант: спрогнозировать окончание сегмента, последующий сегмент, его класс. При этом класс последующего сегмента может отличаться от класса предыдущего).

Задача: в задаче классификации, а затем в задаче прогнозирования временных рядов, получить компактное признаковое описание объекта.

План решения задачи классификации:

- 1) написать формальную постановку задачи по этому плану,
- 2) написать раздел «аппроксимация сплайнами», предложив и обосновав нужный вид сплайна,
- 3) взять за основу выборку Santa Fe или WISDM (выборки состоят из сегментов со многими элементарными движениями и соответствующими сегментам метками классов), вариант OPPORTUNITY Activity Recognition Challenge
- 4) создать выборку, состоящую из сегментов с не более чем двумя элементарными движениями,
- 5) вариант: создать выборку из сегментов с единственным элементарным движением путем сегментации временных рядов,
- 6) аппроксимировать сегменты сплайнами с небольшим фиксированным числом узлов,
- 7) представить каждый объект выборки в виде соединенного вектора из значений узлов и параметров, назовем этот вектор локальным описанием,
- 8) возможно потребуются преобразовать значения узлов в абсолютные для указания на отдельные части сегментов,
- 9) отказаться от временного ряда, оставить его локальное описание и построить по нему модель классификации, метамодель,
- 10) проинтерпретировать информативность локального описания – узлов и параметров сплайна, например, класс не зависит от точки начала или конца сплайна,
- 11) для этого построить график дисперсии и ковариации параметров метамодели, график стандартного отклонения функции ошибки (логарифма функции правдоподобия логистической регрессии),
- 12) исключить маловажные признаки, посмотреть, как это отразится на качестве,
- 13) для контрольной части выборки по сегментам временных рядов построить локальное описание, по нему восстановить метки классов, вычислить качество классификации,
- 14) сравнить результаты классификации с результатами, полученными а) Михаилом Кузнецовым – Никитой Ивкиным и б) Михаилом Карасиковым, должно получиться похожее качество при похожей небольшой сложности метамодели.

План решения задачи прогнозирования:

- 1) построить выборку типа: начало элементарного движения-конец элементарного движения, восстановить регрессию по их локальным описаниям.
- 2) вариант: построить выборку локальная история-прогноз как два последовательно идущих элементарных движения, при этом по возможности, собрав выборку из сегментов с одинаковыми и различными метками классов,

- 3) из пар история-прогноз,
- 4) восстановить регрессию истории на прогноз по локальным описаниям,
- 5) вариант: восстановить классификацию истории на метку класса следующего сегмента, решив задачу опережающей классификации.

По сплайнам в матлабе (документация в матлабе ко всему этому есть, примерчики тоже)

Approximation:

least-squares spline approximation [spap2/csaps](#), данные многомерные

Smoothing splines: [spaps](#) или с помощью [функции fit](#): fitobject = fit(x,y, 'smoothingspline')

Interpolation:

Cubic spline interpolation [csapi](#), cubic spline interpolation with end conditions [csape](#), cubic spline data interpolation [spline](#), spline interpolation [spapi](#)