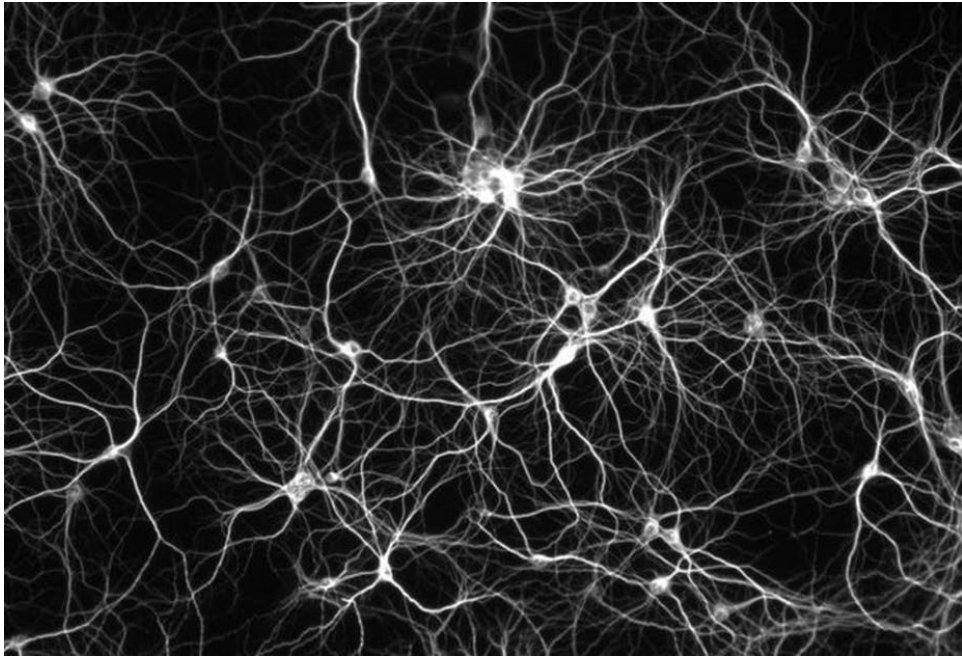


Суперпозиции



Задача

классификация временных рядов

Данные

одномерные короткие временные ряды

$$D = \{t_i, y_i\}_{i=1}^N, \text{ где}$$

$$t_i \in \mathbb{R}^z$$

- независимая переменная, а y_i

- зависимая переменная, где $z = 1$

для одномерных.



- размер грудной клетки спящего пациента
- сердечный ритм
- концентрация кислорода
- цены на акции компании
- курс золота
- количество солнечных пятен
- ...

множество примитивов для $z = 1$

$$G = \{g_1, \dots, g_m, x\}$$

, где g_i

- математические функции, а x

- переменная (терминал).

типы примитивов

- Описательные
- Вспомогательные
- Связующие

F

- множество всех суперпозиций

T

- синтаксическое дерево суперпозиции $f \in F$

Решение

аппроксимируем выборку D

с помощью символьной регрессии, которая порождается генетическим алгоритмом

$$f^* = \operatorname{argmin}_{f \in F} \operatorname{MSE}(f) = \operatorname{argmin}_{f \in F} \sum_{i=1}^N (f(t_i) - y_i)^2.$$

$C(D)$

- классификатор временного ряда

- логистическая регрессия
- AdaBoost
- метод опорных векторов
- нейронные сети
- метод k ближайших соседей
- ...

$F(T)$

- признаки синтаксического дерева T

$$D \rightarrow f \rightarrow T \rightarrow F \rightarrow C(F)$$

Имеется набор M

временных рядов $\{D_j, z_j\}_{j=1}^M$

, где $D_j = \{t_i^j, y_i^j\}_{i=1}^{N_j}$

, а z_j

- метка класса временного ряда

$$C^*(D) = \operatorname{argmin}_C \sum_{j=1}^M [C(D) \neq z_j]$$

$$F^*(f) = \operatorname{argmin}_{F(f)} \sum_{j=1}^M [C^*(D) \neq z_j]$$

Извлечение структурных признаков из T

- удаление связующих вершин
 - удаление вспомогательных вершин
-

$$U = \cup_{j=1}^M F(D_j)$$

- множество всех признаков временных рядов

наши признаки

- частота встречаемости признака U_i среди $F(D)_j$
 - структурная сложность суперпозиций
 - высота их синтаксических деревьев
 - точность аппроксимации
 - гистограмма частот примитивов в функции
 - ...
-
-