

RELATORIO

Aluno: Alexander Ylner Choquenaira Florez

Nro. USP: 11360389

O meu projeto faz classificação de séries temporais utilizando a distância Dynamic Time Warping (DTW), escolhendo 1-vizinho mais próximo.

É implementado em Python (v 3.7). Faz uso de **numpy** (para trabalhar com arrays).

1. MÉTODO DE EXECUÇÃO

1.1. INPUT

O método de execução é:

`python DTW.py treino.txt teste.txt`

- **DTW.py** : nome de programa
- **treino.txt** : arquivo (txt) onde o programa vai ler os dados de training
- **teste.txt** : arquivo (txt) onde o programa vai ler os dados de testing

Os arquivos tem que ficar na mesma pasta que os programas.

1.2. OUTPUT

É gerado um arquivo (Classification.txt) com a classificação de cada caso de teste e a distância mínima. Também, é mostrado a precisão (accuracy) do classificador.

2. IMPLEMENTAÇÃO

2.1. DTW

Para fazer uma implementação eficiente do DTW, é necessário fazer uso de programação dinâmica. O processo é feito salvando os dados calculados numa matriz, da mesma forma que em [\[1\]](#).

2.2. VIZINHO MAIS PRÓXIMO

A classificação é feita escolhendo o vizinho mais próximo. Para isso é necessário calcular a distância de cada elemento do arquivo de teste com todas as séries do arquivo de training e escolher a classe da série mais próxima.

3. RESULTADOS

Ao final tem uma precisão de 0.84583333, têm também uma duração de execução de quasi 645.2804 segundos (10 min e 45 segundos).

O arquivo gerado têm o formato:

Number of series: 957	Classification: 12	True Class: 12
Number of series: 958	Classification: 12	True Class: 12
Number of series: 959	Classification: 12	True Class: 12
Number of series: 960	Classification: 12	True Class: 12
Accuracy: 0.8458333333333333		
Global Time(seg): 645.2804		

Onde se pode olhar o número de série, a classificação feita e a verdadeira classe.

Os experimentos foram feitos num computador que têm um processador IntelCore i7 (7th generation) com 8GB RAM.

4. CONCLUSÕES

A classificação feita com o vizinho mais próximo têm um bom resultado. Más acho que a duração é um problema. Isso é porque o custo computacional de DTW é de $O(n*m)$ onde n e m são as dimensões das inputs series. Quando $n > m$, o custo pode ser de $O(n^2)$ [1]. Também, já que se precisa de calcular a distancia con todos as series de testing para encontrar a série mais próxima.

Uma alternativa para fazer más rápida a execução é a implementação de TRKNN como em [2], mas isso é algo fora do alcance do projeto.

5. REFERENCIAS

[1] Dynamic Time Warping. https://en.wikipedia.org/wiki/Dynamic_time_warping

[2] Eduardo T Bogue, Edson T. Matsubara, Anderson C.Bessa. Uma Abordagem em Reconhecimento de Movimentos Utilizando TRKNN e Dynamic Time Warping.