
K-NN SAMPLING FOR VISUALIZATION OF DYNAMIC DATA USING LION-TSNE - ANALYSIS

Gędlek Paweł

Wydział Informatyki, Elektroniki i Telekomunikacji
Akademia Górniczo-Hutnicza
Kraków
gedlek@student.agh.edu.pl

Wójtowicz Patryk

Wydział Informatyki, Elektroniki i Telekomunikacji
Akademia Górniczo-Hutnicza
Kraków
wojtowicz@student.agh.edu.pl

ABSTRAKT

TODO

1 Struktura raportu

Contents

1	Struktura raportu	1
2	Metoda tSNE	1
2.1	Czym właściwie jest tSNE?	1
2.2	Algorytm tSNE - podstawy matematyczne	2
3	Metoda LION tSNE	2
4	kNN sampling	2
5	Wnioski	2

2 Metoda tSNE

2.1 Czym właściwie jest tSNE?

Algorytm tSNE(t-Distributed Stochastic Neighbor Embedding) którego autorami są Laurens van der Maaten oraz Geoffrey Hinton bazuje na metodzie SNE, której głównym założeniem jest reprezentacja wielowymiarowych danych w możliwy do zobrazowania dla człowieka dwu- lub trzy-wymiarowej przestrzeni. Osiąga się to poprzez modelowanie wysoko wymiarowych obiektów poprzez dwu- lub trzy-wymiarowe punkty w taki sposób, że zbliżone obiekty modelowane są poprzez bliskie sobie punkty, a oddalone obiekty modelowane są poprzez oddalone od siebie punkty z dużym prawdopodobieństwem.

2.2 Algorytm tSNE - podstawy matematyczne

- Algorytm tSNE konwertuje odległości między parami punktów w funkcje rozkładu prawdopodobieństwa określając podobieństwo pomiędzy parami punktów.
- Rozbieżność między podobieństwem wysoko wymiarowych danych z nisko-wymiarowymi danymi jest mierzona poprzez dywergencję Kullbacka-Leiblera i minimalizowana metoda gradientowa poszukiwania minimum lokalnego

Mamy dany zbiór wejściowy $X = x_1, x_2 \dots x_n$ gdzie dla każdego $x_i \in R^D$ jest D-wymiarowym wektorem. Zbiór ten zostanie przekształcony do postaci $Y = y_1, y_2 \dots y_n$ gdzie każde $y_i \in R^d$ jest d-wymiarowym wektorem oraz $d \ll D$ (zazwyczaj $d = 2$ lub 3). Podobieństwo pomiędzy parami punktów wejściowych x_i oraz x_j oznaczamy poprzez $p_{j/i}$, które jest prawdopodobieństwem wybrania x_j jako sąsiada x_i według funkcji gęstości prawdopodobieństwa na rozkładzie normalnym gdzie x_i stanowi centrum. $p_{j/i}$ definiujemy jako:

...

3 Metoda LION tSNE

4 kNN sampling

5 Wnioski

TODO

Źródła

- [1] Bheekya Dharamsotu ; K. Swarupa Rani ; Salman Abdul Moiz ; C. Raghavendra Rao Paper: k-NN Sampling for Visualization of Dynamic Data Using LION-tSNE. <https://ieeexplore.ieee.org/abstract/document/8990391>