* **Εισαγωγή**

Χρονική αποτίμηση αλγορίθμων εύρεσης προσεγγιστικά πλησιέστερου γείτονα και αλγορίθμων ομαδοποίησης δεδομένων, με τη χρήση κωδικοποιημένων χρονοσειρών

* **Χρήση μετρικών για τη σύγκριση αποτελεσμάτων**

Για τη σύγκριση των αποτελεσμάτων πριν και μετά τη κωδικοποίηση των χρονοσειρών χρησιμοποιήθηκαν οι μετρικές :

1. Approximate Average
2. Approximate True
3. Maximum Approximation Factor
4. Clustering Time
5. Δέλτα
6. Δέλτα %

όπου **Δέλτα** ορίζεται ως η διαφορά μεταξύ δύο μετρικών και **Δέλτα %** ως η κανονικοποιημένη, ποσοστιαία διαφορά μεταξύ δύο μετρικών, πριν και μετά τη κωδικοποίηση των χρονοσειρών

* **Ακρωνύμια**

**Appr. Avg. = Approximate Average**

**Appr. True = Approximate True**

**M.A.F. = Maximum Approximation Factor**

* **Αποτελέσματα αλγορίθμων εύρεσης προσεγγιστικά πλησιέστερου γείτονα**

Χαρακτηριστικά 1ης μεθόδου :

1. Αναπαράσταση χρονοσειράς **Ti** = **(Y1, Y2, …, Ym)**
2. Μετρική απόστασης = **Ευκλείδεια (L2)**
3. Dataset = **350 χρονοσειρές**
4. Queryset = **9 χρονοσειρές**

|  |  |
| --- | --- |
| Παράμετροι | |
| L |  |
| K |  |
| W | 6 |
| Threshold | 117 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Appr. Avg. |  |  |  |  |
| Appr. True |  |  |  |  |
| M.A.F. |  |  |  |  |

Χαρακτηριστικά 2ης μεθόδου :

1. Αναπαράσταση χρονοσειράς **Ti** = **(Y1, Y2, …, Ym)**
2. Μετρική απόστασης = **Ευκλείδεια (L2)**
3. Dataset = **350 χρονοσειρές**
4. Queryset = **9 χρονοσειρές**

|  |  |
| --- | --- |
| Παράμετροι | |
| K |  |
| M |  |
| Probes | 2 |
| W | 6 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Appr. Avg. |  |  |  |  |
| Appr. True |  |  |  |  |
| M.A.F. |  |  |  |  |

Χαρακτηριστικά 3ης μεθόδου:

1. Αναπαράσταση χρονοσειράς **Ti** = **((X1, Y1), (X2, Y2), …, (Xm,Ym))**
2. Μετρική απόστασης = **Discrete Fréchet**
3. Dataset = **10 χρονοσειρές** (αναφορά)
4. Queryset = **9 χρονοσειρές**

|  |  |  |
| --- | --- | --- |
| Παράμετροι | Αρχικές | Κωδικοποιημένες |
| L |  |  |
| K |  |  |
| W |  |  |
| Delta |  |  |
| Threshold |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Appr. Avg. |  |  | 13963.53 |  |
| Appr. True |  |  | 61435.46 |  |
| M.A.F. |  |  |  |  |

Χαρακτηριστικά 4ης μεθόδου:

1. Αναπαράσταση χρονοσειράς **Ti** = **((X1, Y1), (X2, Y2), …, (Xm,Ym))**
2. Μετρική απόστασης = **Continuous Fréchet**
3. Dataset = **10 χρονοσειρές** (αναφορά)
4. Queryset = **9 χρονοσειρές**

|  |  |  |
| --- | --- | --- |
| Παράμετροι | Αρχικές | Κωδικοποιημένες |
| L |  |  |
| K |  |  |
| W |  |  |
| Delta |  |  |
| Threshold |  |  |
| Epsilon |  |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Appr. Avg. |  |  | 451320.39 |  |
| Appr. True |  |  |  |  |
| M.A.F. |  |  |  |  |

Σχολιασμός αποτελεσμάτων :

Παρατηρούμε ότι για μία κωδικοποίηση στα δεδομένα της τάξης **~3.3**, επιτυγχάνεται όμοια χρονική επιτάχυνση όταν οι μέθοδοι **LSH** και **Hypercube** χρησιμοποιούν την **Ευκλείδεια** μετρική απόστασης, ενώ 3 φορές μεγαλύτερη επιτάχυνση όταν χρησιμοποιείται η μετρική **Discrete** ή **Continuous Fréchet.**

* **Αποτελέσματα αλγορίθμων ομαδοποίησης δεδομένων**

Χαρακτηριστικά 1ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Lloyd’s**
2. Αναπαράσταση χρονοσειράς **Ti** = **(Y1, Y2, …, Ym)**
3. Μετρική απόστασης = **Ευκλείδεια (L2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **NaN**

|  |  |
| --- | --- |
| Παράμετροι | |
| Clusters |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Clustering Time |  |  |  |  |

Χαρακτηριστικά 2ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **LSH Range Search**
2. Αναπαράσταση χρονοσειράς **Ti** = **(Y1, Y2, …, Ym)**
3. Μετρική απόστασης = **Ευκλείδεια (L2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **NaN**

|  |  |
| --- | --- |
| Παράμετροι | |
| Clusters |  |
| k |  |
| L |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Clustering Time |  |  |  |  |

Χαρακτηριστικά 3ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Hypercube Range Search**
2. Αναπαράσταση χρονοσειράς **Ti** = **(Y1, Y2, …, Ym)**
3. Μετρική απόστασης = **Ευκλείδεια (L2)**
4. Dataset = **350** **χρονοσειρές**
5. Queryset = **NaN**

|  |  |
| --- | --- |
| Παράμετροι | |
| Clusters |  |
| k |  |
| M |  |
| Probes |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Clustering Time |  |  |  |  |

Χαρακτηριστικά 4ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Lloyd’s**
2. Αναπαράσταση χρονοσειράς **Ti** = **((X1, Y1), (X2, Y2), …, (Xm,Ym))**
3. Μετρική απόστασης = **Discrete Fréchet**
4. Dataset = **10** (αναφορά)
5. Queryset = **NaN**

|  |  |
| --- | --- |
| Παράμετροι | |
| Clusters |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Clustering Time |  |  |  |  |

Χαρακτηριστικά 5ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης : **LSH Range Search**
2. Αναπαράσταση χρονοσειράς **Ti** = **((X1, Y1), (X2, Y2), …, (Xm,Ym))**
3. Μετρική απόστασης = **Discrete Fréchet**
4. Dataset = **10** (αναφορά)
5. Queryset = **NaN**

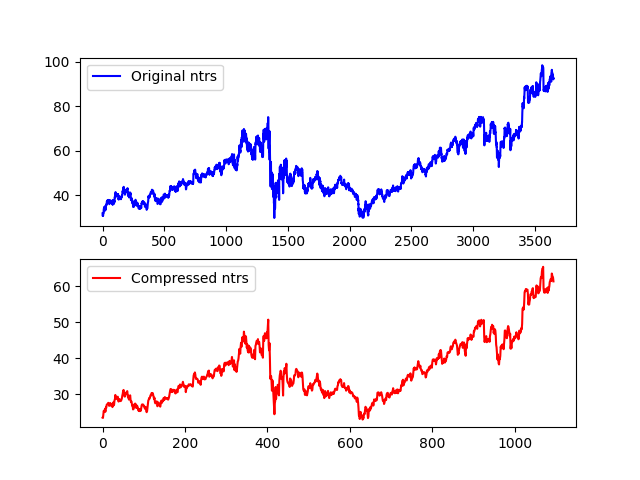
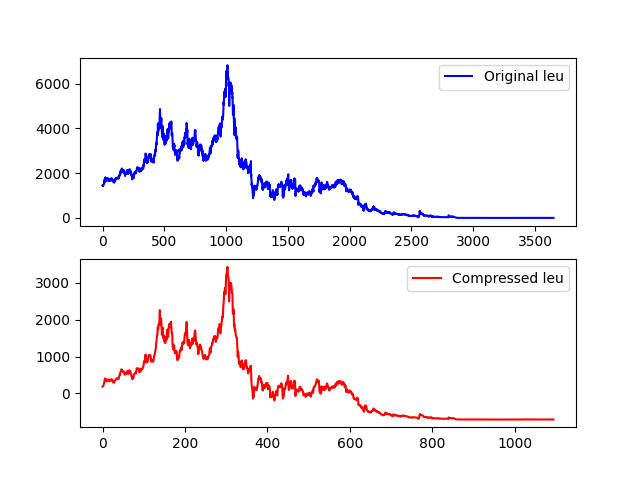
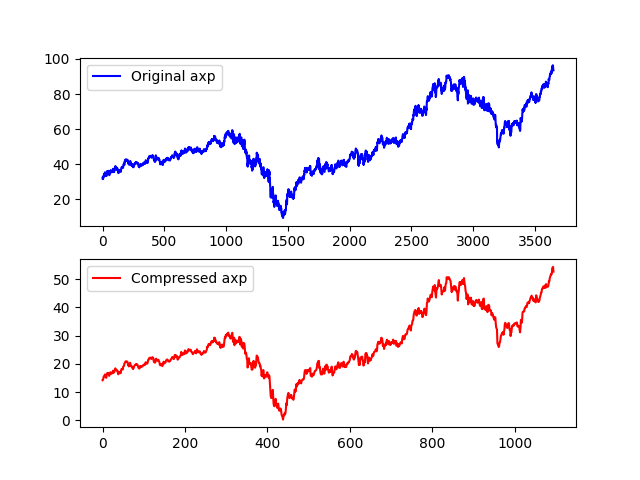
|  |  |
| --- | --- |
| Παράμετροι | |
| Clusters |  |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Αρχικές | Κωδικοποιημένες | Δέλτα | Δέλτα % |
| Clustering Time |  |  |  |  |

Σχολιασμός αποτελεσμάτων :

Ομοίως κι εδώ, βλέπουμε ότι για μία κωδικοποίηση στα δεδομένα της τάξης **~3.3**, επιτυγχάνεται παρόμοια χρονική επιτάχυνση όταν οι μέθοδοι **Lloyd’s**, **LSH Range Search** και **Hypercube** **Range Search** χρησιμοποιούν την **Ευκλείδεια** μετρική απόστασης, ενώ περίπου 3 φορές υψηλότερη επιτάχυνση στη μετρική απόστασης **Discrete Fréchet .**

* **Σύγκριση καμπυλών πριν και μετά τη κωδικοποίηση :**



* **Αναφορές**

1: <https://eclass.uoa.gr/modules/forum/viewtopic.php?course=DI352&topic=33210&forum=54779>