

- **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 1^{ης} μεθόδου :

1. Αλγόριθμος αναζήτησης = **LSH**
2. Αναπαράσταση χρονοσειράς $T_i = (Y_1, Y_2, \dots, Y_m)$
3. Μετρική απόστασης = **Ευκλείδεια (L^2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **9 χρονοσειρές**

Παράμετροι	
L	5
K	4
W	6
Threshold	117

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Appr. Avg.	13.41 ms	3.42 ms	9.99 ms	74.49
Appr. True	110.44 ms	33.51 ms	76.93 ms	69.65
M.A.F.	2.31	2.45	-0.14	-6.06

Χαρακτηριστικά 2^{ης} μεθόδου :

1. Αλγόριθμος αναζήτησης = **Hypercube**
2. Αναπαράσταση χρονοσειράς $T_i = (Y_1, Y_2, \dots, Y_m)$
3. Μετρική απόστασης = **Ευκλείδεια (L^2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **9 χρονοσειρές**

Παράμετροι	
K	14
M	10
Probes	2
W	6

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Appr. Avg.	2.78 ms	1.05 ms	1.73 ms	62.23
Appr. True	109.79 ms	33.53 ms	76.26 ms	62.45
M.A.F.	5.41	3.90	1.51	27.91

Χαρακτηριστικά 3ης μεθόδου:

1. Αλγόριθμος αναζήτησης = **LSH**
2. Αναπαράσταση χρονοσειράς $T_i = ((X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m))$
3. Μετρική απόστασης = **Discrete Fréchet**
4. Dataset = **10 χρονοσειρές** (αναφορά)
5. Queryset = **9 χρονοσειρές**

Παράμετροι	Αρχικές	Κωδικοποιημένες
L	5	5
K	4	4
W	6	6
Delta	0.29	0.08
Threshold	4	4

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Appr. Avg.	16494.96 ms	2531.43 ms	13963.53 ms	84.65
Appr. True	70083.87 ms	8648.41 ms	61435.46 ms	87.65
M.A.F.	1.23	1.36	-0.13	-10.56

Χαρακτηριστικά 4^{ης} μεθόδου:

1. Αλγόριθμος αναζήτησης = **LSH**
2. Αναπαράσταση χρονοσειράς $T_i = ((X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m))$
3. Μετρική απόστασης = **Continuous Fréchet**
4. Dataset = **10 χρονοσειρές** (αναφορά)
5. Queryset = **9 χρονοσειρές**

Παράμετροι	Αρχικές	Κωδικοποιημένες
L	1	1
K	4	4
W	6	6
Delta	0.14	0.04
Threshold	4	4
Epsilon	1.00	1.00

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Appr. Avg.	496875.01 ms	45554.62 ms	451320.39 ms	90.83
Appr. True	1146955.51 ms	115447.53 ms	1031507.98 ms	89.93
M.A.F.	2.22	1.81	0.41	18.46

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 1^{ης} μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Lloyd's**
2. Αναπαράσταση χρονοσειράς $T_i = (Y_1, Y_2, \dots, Y_m)$
3. Μετρική απόστασης = **Ευκλείδεια (L^2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **NaN**

Παράμετροι	
Clusters	10

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Clustering Time	6.50 s	1.93 s	4.57 s	70.30

Χαρακτηριστικά 2^{ης} μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **LSH Range Search**
2. Αναπαράσταση χρονοσειράς $T_i = (Y_1, Y_2, \dots, Y_m)$
3. Μετρική απόστασης = **Ευκλείδεια (L^2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **NaN**

Παράμετροι	
Clusters	10
k	4
L	3

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Clustering Time	1.25 s	0.38 s	0.87 s	69.60

Χαρακτηριστικά 3^{ης} μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Hypercube Range Search**
2. Αναπαράσταση χρονοσειράς $T_i = (Y_1, Y_2, \dots, Y_m)$
3. Μετρική απόστασης = **Ευκλείδεια (L^2)**
4. Dataset = **350 χρονοσειρές**
5. Queryset = **NaN**

Παράμετροι	
Clusters	10
k	14
M	10
Probes	2

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Clustering Time	1.28 s	0.40 s	0.88 s	68.75

Χαρακτηριστικά 4^{ης} μεθόδου :

1. Αλγόριθμος ομαδοποίησης = **Lloyd's**
2. Αναπαράσταση χρονοσειράς $T_i = ((X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m))$
3. Μετρική απόστασης = **Discrete Fréchet**
4. Dataset = **10** (αναφορά)
5. Queryset = **NaN**

Παράμετροι	
Clusters	10

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Clustering Time	7057.39 s	794.63 s	6262.76 s	88.74

Χαρακτηριστικά 5ης μεθόδου :

1. Αλγόριθμος ομαδοποίησης : **LSH Range Search**
2. Αναπαράσταση χρονοσειράς $T_i = ((X_1, Y_1), (X_2, Y_2), \dots, (X_m, Y_m))$
3. Μετρική απόστασης = **Discrete Fréchet**
4. Dataset = **10** (αναφορά)
5. Queryset = **NaN**

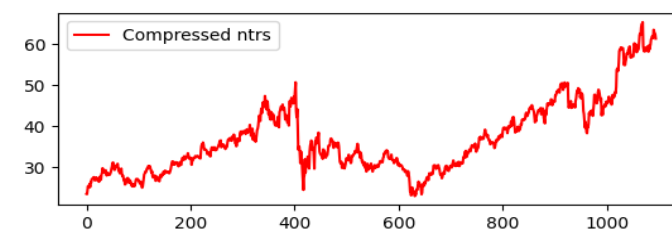
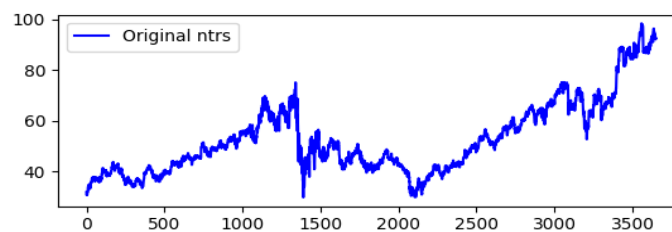
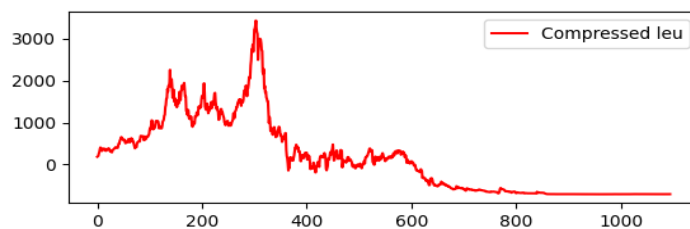
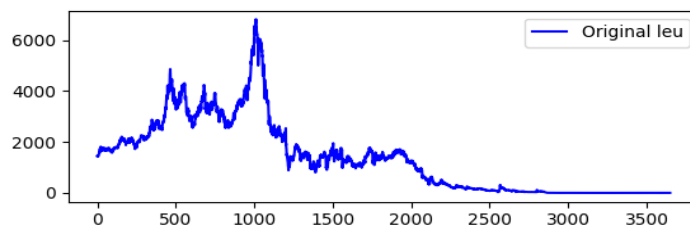
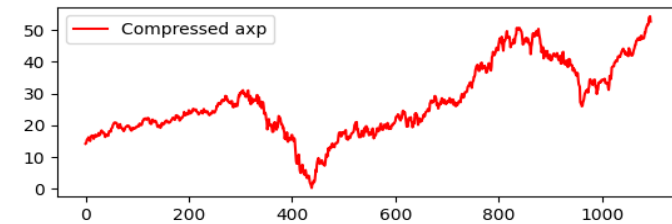
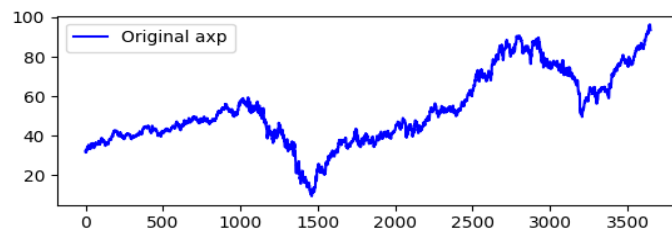
Παράμετροι	
Clusters	10

	Αρχικές	Κωδικοποιημένες	Δέλτα	Δέλτα %
Clustering Time	2215.52 s	290.52 s	1925 s	86.93

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>