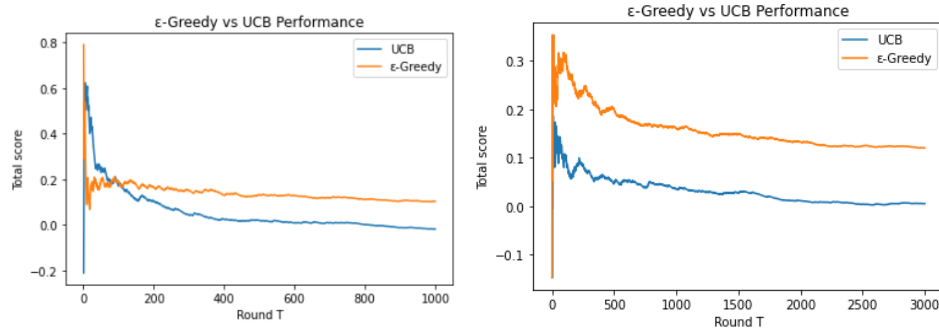


Large and Social Networks

ΚΩΝΣΤΑΝΤΙΝΟΣ ΝΙΚΟΛΟΣ

AM: 2019030096

Γραφήματα Regret-T και των δύο αλγορίθμων.



Μπορούμε εύκολα να διακρίνουμε ότι και στις δύο περιπτώσεις έχουμε πετύχει sublinear regret καθώς το regret μειώνεται σχεδόν σε όλες τις στιγμές και όχι με γραμμικό ρυθμό.

Παρατηρούμε ότι παρότι και στους δύο αλγόριθμους το regret μειώνεται δραστικά, ο αλγόριθμος UCB τελικά δίνει καλύτερο αποτέλεσμα καθώς το cumulative regret πλησιάζει το 10%, σε αντίθεση με τον ε-Greedy όπου το cumulative regret συγκλίνει στο 20%.

Διαπιστώνεται ότι στον αλγόριθμο ε-Greedy παρότι έχουμε μείωση του cumulative regret υπάρχει μια διαρκής ταλάντωση σε μικρό βαθμό σε σύγκριση με τον αλγόριθμο UCB ο οποίος παρουσιάζει αύξηση του regret σπανιότερα. Γεγονός που οφείλεται στην πιθανότητα να χρησιμοποιηθεί κάποιο 'bad arm' από τον αλγόριθμο ε-Greedy. Ο αλγόριθμος UCB εμφανίζει σημαντική βελτίωση ελέγχοντας καλύτερα πότε γίνεται μία εξερεύνηση ενός 'bad arm'

UCB

```
round= 48.000000 arm played 8.000000, ucb=1.850760, Q affecting ucb=0.819894
round= 49.000000 arm played 1.000000, ucb=1.814130, Q affecting ucb=0.724386
round= 50.000000 arm played 3.000000, ucb=1.564130, Q affecting ucb=0.848167
round= 51.000000 arm played 6.000000, ucb=1.564130, Q affecting ucb=0.848167
round= 52.000000 arm played 8.000000, ucb=1.564130, Q affecting ucb=0.848167
round= 53.000000 arm played 9.000000, ucb=1.814130, Q affecting ucb=0.724386
round= 54.000000 arm played 2.000000, ucb=1.804231, Q affecting ucb=0.515029
round= 55.000000 arm played 4.000000, ucb=1.792450, Q affecting ucb=0.442105
```

```
round= 409.000000 arm played 9.000000, ucb=1.146519, Q affecting ucb=0.418530
round= 410.000000 arm played 7.000000, ucb=1.165122, Q affecting ucb=0.158716
round= 411.000000 arm played 7.000000, ucb=1.164763, Q affecting ucb=0.158374
round= 412.000000 arm played 7.000000, ucb=1.164407, Q affecting ucb=0.158033
round= 413.000000 arm played 7.000000, ucb=1.164054, Q affecting ucb=0.157695
round= 414.000000 arm played 7.000000, ucb=1.163702, Q affecting ucb=0.157360
round= 415.000000 arm played 7.000000, ucb=1.163353, Q affecting ucb=0.157026
round= 416.000000 arm played 4.000000, ucb=1.163140, Q affecting ucb=0.227101
```

Παρατηρούμε ότι ο αλγόριθμος λειτουργεί σωστά, αποτελεσματικότερα από τον ε-Greedy

καθώς αρχικά τα ποσοστά επιρροής $p = \frac{\sqrt{\ln T}}{\sqrt{Q_i(t) Ucbi(t)}}$ είναι μεγαλύτερο του 50% στα πρώτα 50 rounds και διαλέγονται διαφορετικά arms συνεχώς πετυχαίνοντας το αντίστοιχο exploration κάθε arm.

Μετά από ένα διάστημα rounds= 400 το ποσοστό p μειώνεται σημαντικά με αποτέλεσμα την επιλογή των καλύτερων arms βασίζοντας την σύγκριση περισσότερο στον δείκτη $\mu(t)$ κάθε arm.