



Πολυτεχνείο Κρήτης - Τμήμα ΗΜΜΥ

Reinforcement Learning  
and Dynamic Optimization

Εαρινό Εξάμηνο 2022-2023  
2<sup>η</sup> Εργασία

Τασιόπουλος Ιωάννης 2015030002

### **Εισαγωγή**

Σκοπός της άσκησης ήταν η μελέτη του αλγορίθμου Multiplicative Weights ή Hedge και η σύγκρισή του με τον Upper Confidence Bound(UCB) αλγόριθμο.

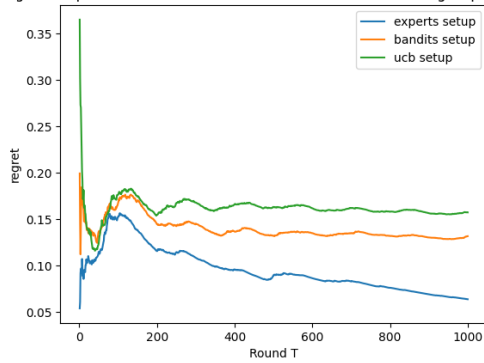
Αυτό πραγματοποιήθηκε χρησιμοποιώντας δεδομένα από φορτία κάποιων servers. Αυτά δεν ήταν Independent and Identically Distributed (IID) γεγονός που επέβαλε την εφαρμογή μη ντετερμινιστικών αλγορίθμων. Ο στόχος ήταν η εκμάθηση του αλγορίθμου ώστε να επιλέγει με την πάροδο του χρόνου τον server που έχει το λιγότερο φορτίο.

Ο αλγόριθμος δοκιμάστηκε σε περιβάλλον Experts(Γίνονται γνωστά τα φορτία όλων των servers για κάθε χρονική στιγμή), και σε Adversarial Bandits(Γίνεται γνωστό το φορτίο μόνο του server που επέλεξε ο αλγόριθμος). Τέλος, δοκιμάστηκε ο αλγόριθμος UCB, ο οποίος είχε πολύ καλά αποτελέσματα.

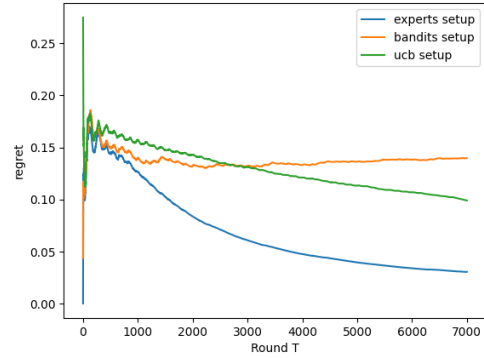
### **Αποτελέσματα**

Παρακάτω παρατίθενται τα διαγράμματα που παρήχθησαν από τα πειράματα για χρονικό ορίζοντα 1000 και 7000 timeslots.

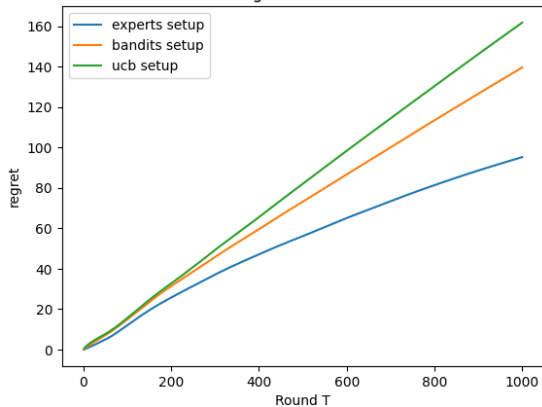
Algorithm performance in various enviroment (normalized regret per round)



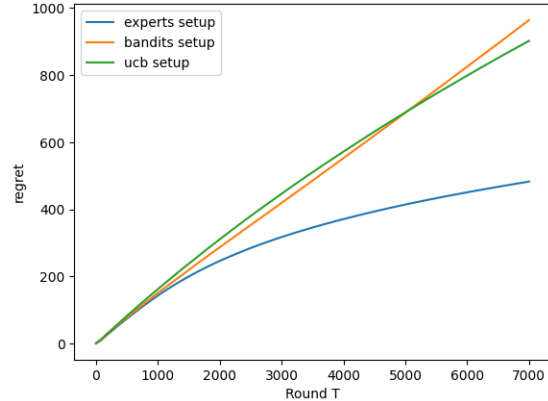
Algorithm performance in various enviroment (normalized regret per round)



Cumulative regret in various enviroment



Cumulative regret in various enviroment



## Συμπεράσματα

Είναι φανερό ότι ο αλγόριθμος Hedge αποδίδει καλύτερα σε περιβάλλον experts, όπως ήταν αναμενόμενο, καθώς έχει στη διάθεσή του όλα τα δεδομένα. Αυτό που είναι αξιοσημείωτο είναι ότι ο αλγόριθμος ucb μετά από πολλές επαναλήψεις και για το συγκεκριμένο σετ δεδομένων καταφέρνει να αποδώσει καλύτερα από τον Hedge σε περιβάλλον adversarial bandits.

**Σημείωση:** Το παραδοτέο αρχείο κώδικα δημιουργήθηκε στο περιβάλλον του Google Colab και για την εκτέλεσή του απαιτεί τη μεταφόρτωση του αρχείου με τα δεδομένα. (Κατά την εκτέλεση του προγράμματος ανοίγει το αντίστοιχο πλαίσιο διαλόγου.)