

NATIONAL TECHNICAL UNIVERSITY OF ATHENS



PROGRAMMING TOOLS AND TECHNOLOGIES FOR DATA SCIENCE

THE TIME-TRAVEL PROJECT
2020B EDITION

Christos Katsakioris
ckatsak@cslab.ece.ntua.gr
03002964

PhD Candidate
Computing Systems Laboratory
School of Electrical and Computer Engineering

Athens, Greece

Tuesday 26th January, 2021

Πίνακας Περιεχομένων

Πίνακας Σχημάτων	i
1 Εισαγωγή	1
2 Περιβάλλον και Εκτέλεση	2
3 Μικρή Ακολουθία	3
4 Μεγάλη Ακολουθία	5

Πίνακας Σχημάτων

1	Διάγραμμα αποτίμησης κατά τον υπολογισμό της μικρής ακολουθίας.	4
2	Διάγραμμα αποτίμησης κατά τον υπολογισμό της μεγάλης ακολουθίας.	6

1 Εισαγωγή

Στην παρούσα εργασία επεξεργαζόμαστε ιστορικά δεδομένα από το χρηματιστήριο της Νέας Υόρκης για τα έτη 1962 – 2017.

Το σύνολο δεδομένων που μας διατίθεται στην αρχική του μορφή αποτελείται από ένα σύνολο αρχείων CSV, καθένα από τα οποία αναφέρεται σε μία εταιρεία του χρηματιστηρίου. Οι στήλες κάθε τέτοιου αρχείου είναι:

- η ημερομηνία,
- η τιμή ανοίγματος της μετοχής την δεδομένη ημερομηνία,
- η τιμή κλεισίματος της μετοχής την δεδομένη ημερομηνία,
- η ελάχιστη τιμή της μετοχής που καταγράφηκε την δεδομένη ημερομηνία,
- η μέγιστη τιμή της μετοχής που καταγράφηκε την δεδομένη ημερομηνία, και
- ο συνολικός όγκος των μετοχών που διαπραγματεύτηκαν την δεδομένη ημερομηνία.

Στόχος της εργασίας είναι να καταφέρουμε να παράγουμε δύο ακολουθίες αγοραπωλησιών μετοχών εταιρειών, μία μικρή και μία μεγάλη, με τους δεδομένους περιορισμούς που περιγράφονται στην εκφώνηση της άσκησης.

Λόγω της εγγενούς δυσκολίας και της αυξημένης πολυπλοκότητας (NP-hard) του προβλήματος, καθώς και του ικανού μεγέθους του συνόλου δεδομένων εισόδου, δεν επιχειρήθηκε η εύρεση βέλτιστης λύσης, παρά μόνο μίας ικανοποιητικής (ανά υποπρόβλημα, δηλαδή ανά μέγεθος ακολουθίας) λύσης σύμφωνα με τον παρεχόμενο online επικυρωτή.

Στις επόμενες ενότητες εξηγούμε εν συντομία τη συλλογιστική του κάθε αλγορίθμου που σχεδιάστηκε και υλοποιήθηκε, αφού εξηγήσουμε τον τρόπο εκτέλεσης του παραδοτέου script για την αναπαραγωγή των αποτελεσμάτων.

Ο κώδικας της υλοποίησης, των σχημάτων, καθώς και του παρόντος εγγράφου (για το οποίο χρησιμοποιήθηκε περιβάλλον L^AT_EX), είναι δημοσιοποιημένα σε αποθετήριο του GitHub και προσπελάσιμο μέσω του υπερσυνδέσμου: <https://github.com/ckatsak/progds2020-timetravel>.

2 Περιβάλλον και Εκτέλεση

Η υλοποίηση της άσκησης περιλήφθηκε στο σύνολό της σε ένα Python script, το οποίο αναπτύχθηκε σε περιβάλλον Linux και Python 3.8.5 x86, και το οποίο, εκτελούμενο με τα κατάλληλα ορίσματα από την γραμμή εντολών, παράγει τα επιθυμητά κάθε φορά αποτελέσματα.

Οι διαθέσιμες επιλογές ορισμάτων κλήσης του παραδοτέου script από την γραμμή εντολών του Linux συνοψίζονται από το ίδιο το script με την χρήση της παραμέτρου `-h` ή `--help`, όπως φαίνεται παρακάτω:

```
$ python3 timetravel.py --help
usage: timetravel.py [-h] [-v] [-i STOCKS] [-o RESULT] [-p PLOT] (--small | --large)
```

ProgDS @ ECE NTUA, Fall 2020; by Christos Katsakioris

optional arguments:

<code>-h, --help</code>	show this help message and exit
<code>-v, --verbosity</code>	set the verbosity of the log (max: <code>`-vv`</code>)
<code>-i STOCKS, --stocks STOCKS</code>	path of the input directory (<code>`Stocks/`</code>)
<code>-o RESULT, --result RESULT</code>	path to flush the result sequence into
<code>-p PLOT, --plot PLOT</code>	path to flush the result plot into
<code>--small</code>	produce a small sequence
<code>--large</code>	produce a large sequence

Βάσει και των παραπάνω, μπορούμε να εκκινήσουμε την εκτέλεση του script για την παραγωγή της μικρής ακολουθίας χωρίς εκτενή μηνύματα εξόδου, με την παρακάτω εντολή:

```
$ python3 timetravel.py --small -i $PATH_TO_STOCKS_DIR
```

ενώ αντίστοιχα για την παραγωγή της μεγάλης ακολουθίας:

```
$ python3 timetravel.py --large -i $PATH_TO_STOCKS_DIR
```

Με αυτόν τον τρόπο, οι παραγόμενες ακολουθίες πρόκειται να αποθηκευτούν αντιστοίχως στα αρχεία `small.txt` και `large.txt` του τρέχοντος καταλόγου εργασίας, ενώ τα διαγράμματα αποτίμησης στα αρχεία `small.png` και `large.png` του τρέχοντος καταλόγου εργασίας, αντιστοίχως. Ωστόσο, παρέχεται η δυνατότητα ορισμού εναλλακτικών διαδρομών του συστήματος αρχείου τόσο για την περίπτωση του αρχείου εξόδου (όρισμα γραμμής εντολών `-o` ή `--result`), όσο και του αρχείου του διαγράμματος αποτίμησης (όρισμα γραμμής εντολών `-p` ή `--plot`), όπως είναι εμφανές και στην παραπάνω παρατετημένη έξοδο του script.

3 Μικρή Ακολουθία

Η παραγωγή των δύο ακολουθιών βασίζεται σε διαφορετικό αλγόριθμο διάσχισης του συνόλου δεδομένων, ανάλογα με το μέγεθος της ακολουθίας.

Ο αλγόριθμος για την παραγωγή της πρώτης ακολουθίας βασίζεται στη χρήση ενός χρονικού παραθύρου το οποίο μετακινείται εντός του συνόλου δεδομένων, αποφασίζοντας άπληστα την αγοραπωλησία μετοχών των διαθέσιμων εταιρειών. Ο καθορισμός του χρονικού παραθύρου βασίζεται στις δεδομένες σταθερές του προβλήματος, συνεπώς είναι υπολογίσιμος στατικά (εν προκειμένω είναι 40 ημέρες).

Η βασική λογική είναι ότι λαμβάνει χώρα πάντα μία αγοραπωλησία μετοχών μίας εταιρείας ανά χρονικό παράθυρο. Η αγορά των μετοχών της εκάστοτε εταιρείας γίνεται κάποια ημέρα του χρονικού παραθύρου στην ελάχιστη καταγραφείσα τιμή μετοχής της εταιρείας, και η αντίστοιχη πώλησή τους λαμβάνει χώρα σε κάποια από τις επόμενες ημέρες στην μέγιστη καταγραφείσα τιμή μετοχής της εταιρείας. Η επιλογή της εταιρείας μεταξύ όλων όσων είναι διαθέσιμες στο εκάστοτε χρονικό παράθυρο πραγματοποιείται υπολογίζοντας τη μέγιστη διαφορά τιμών μετοχής ανά εταιρεία για το χρονικό παράθυρο, και μετέπειτα σταχυολογώντας τις ευρεθείσες διαφορές (με πολλαπλά κριτήρια, όπως κατά πόσον είναι εφικτή η αγορά τους, κατά πόσον είναι επικερδής η αγοραπωλησία, λαμβάνοντας υπ'όψιν και την επιβεβλημένη προμήθεια, κλπ) ώστε να προκύψει μία άπληστα βέλτιστη ευκαιρία αγοραπωλησίας μεταξύ των εναπομεινασών.

Σημειώνεται ότι, μολονότι δεν επιχειρείται εύρεση βέλτιστης λύσης, λόγω της Άπληστης φύσης του ανωτέρω περιγραφέντος αλγορίθμου ενδέχεται και να μην εντοπιστεί καμία επικερδής αγοραπωλησία σε κάποια εκ των υπό εξέταση χρονικών παραθύρων, εξασφαλίζοντας τοιουτοτρόπως ότι το υπόλοιπο του λογαριασμού ουδέποτε παρουσιάζει ζημιά, παρότι μπορεί ενίοτε να παραμένει σταθερό.

Επιπροσθέτως, προκειμένου να ελαττωθεί ο χρόνος εκτέλεσης της υλοποίησής μας, μειώσαμε περαιτέρω το σύνολο δεδομένων εισόδου εφαρμόζοντας τον παραπάνω αλγόριθμο μόνο σε ένα υποσύνολο 130 εταιρειών. Η επιλογή των εταιρειών εισόδου έγινε κατά κύριο λόγο με τυχαίο τρόπο (λόγου χάριν, όλες οι εταιρείες των οποίων η συντετμημένη γραφή αρχίζει από “x”, “y”, “mo” και “oi”, χωρίς κάποιον ιδιαίτερο λόγο). Παρ'όλ'αυτά, κάποιες επιλέχθηκαν λιγότερο τυχαία λόγω της περιέργειας του γράφοντος να συμπεριλάβει γνωστές διεθνούς βεληνεχούς εταιρείες ποικίλων τομέων δραστηριότητας, αλλά κυρίως του κλάδου της Τεχνολογίας (παραδείγματα είναι οι Alphabet, Amazon, Microsoft, IBM, Oracle, Intel, Cisco, αλλά και οι Coca-Cola, J.P. Morgan, BP, Procter & Gamble, Disney, και άλλες).

Η περιγραφείσα προσέγγιση εξασφάλισε κέρδος 1870282663.4389617 ($\approx 1.87 \times 10^9$) δολλαρίων (σύμφωνα με τον παρεχόμενο online επικυρωτή) κατόπιν 922 συναλλαγών.

Στο σχήμα 1 που ακολουθεί, βλέπουμε το διάγραμμα αποτίμησης για το συνολικό χρονικό διάστημα για το οποίο εργαστήκαμε, σύμφωνα με τον ορισμό της αποτίμησης που περιλαμβάνεται στην εκφώνηση της εργασίας.

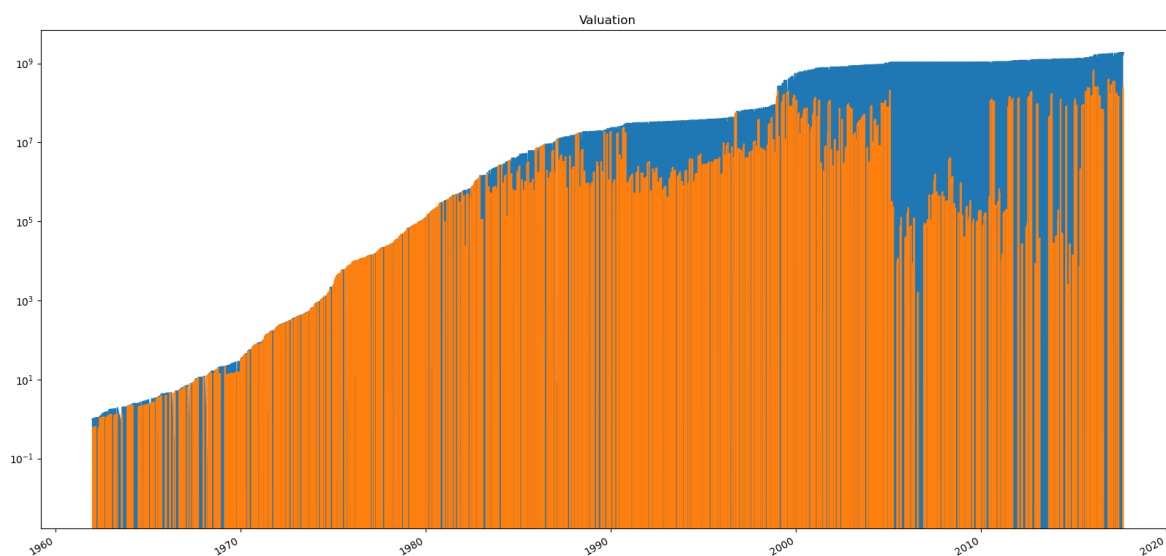


Figure 1: Διάγραμμα αποτίμησης κατά τον υπολογισμό της μικρής ακολουθίας.

4 Μεγάλη Ακολουθία

Ο αλγόριθμος που υλοποιήθηκε για την παραγωγή της μεγάλης ακολουθίας βασίζεται σε μία απλούστερη συλλογιστική. Αυτή τη φορά ο αλγόριθμος εφαρμόστηκε σε ολόκληρο το σύνολο δεδομένων και δεν πραγματοποιήθηκε καμία επιλογή εταιρειών με οποιονδήποτε τρόπο.

Γνωρίζοντας πλέον ότι το μέγιστο επιτρεπτό πλήθος συναλλαγών είναι πολύ μεγαλύτερο από το πλήθος των ημερών του συνόλου δεδομένων μας, μπορούμε να περιορίσουμε το χρονικό παράθυρο σε μία μόνο ημέρα, και μάλιστα να πραγματοποιήσουμε πολλαπλές συναλλαγές ανά ημέρα.

Εκμεταλλευόμαστε λοιπόν το “intra-day trading”, όπως αυτό περιγράφεται από την εκφώνηση της άσκησης, και εφαρμόζουμε μία αντίστοιχα άπληστη λογική με αυτήν που περιγράφηκε στην προηγούμενη ενότητα. Επικεντρωνόμαστε σε δύο είδη ομάδων συναλλαγών:

- Αγορά στην τιμή ανοίγματος της μετοχής και πώληση στη μέγιστη καταγραφείσα τιμή της, και
- Αγορά στην χαμηλότερη καταγραφείσα τιμή της μετοχής και πώληση στην τιμή κλεισίματός της.

Κάθε ημέρα λαμβάνουν χώρα συναλλαγές μόνο της μίας εκ των δύο αυτών ομάδων. Ωστόσο, για την επιλογή μίας εκ των δύο, υπολογίζεται το κέρδος και των δύο αυτών περιπτώσεων. Για την ακρίβεια, κάθε ημέρα χωρίζουμε τις εταιρείες σε δύο ομάδες με κριτήριο την περίπτωση (μεταξύ των προαναφερθέντων δύο) στην οποία μεγιστοποιείται το κέρδος από την αγοραπωλησία τους, εφόσον η αγοραπωλησία τους μπορεί να είναι επικερδής. Στην συνέχεια, μετά την κατάλληλη σταχυολόγησή τους με τρόπο παρόμοιο με αυτόν που περιγράφηκε για τον αλγόριθμο παραγωγής της μικρής ακολουθίας, υπολογίζουμε ποιας ομάδας εναπομεινασών εταιρειών η αγοραπωλησία μπορεί να αποφέρει μεγαλύτερο κέρδος, και επιλέγουμε άπληστα την βέλτιστη εκ των δύο. Μία εύκολη βελτιστοποίηση του αλγορίθμου που θα μπορούσε να γίνει στο σημείο αυτό είναι η επιπρόσθετη αγορά μετοχών και των υπόλοιπων εταιρειών (της άλλης ομάδας), βέβαια μόνο εφόσον παραμένουν επικερδείς και στην περίπτωση του έτερου intra-day trading παραθύρου.

Παρ’όλ’αυτά, δεν χρειάστηκε να υλοποιηθούν και να εφαρμοστούν περαιτέρω βελτιστοποιήσεις στον αλγόριθμο, διότι το κέρδος που επετεύχθη αποδεικνύεται αρκούντως ικανοποιητικό για τον παρεχόμενο online επικυρωτή, σύμφωνα με τον οποίον η ακολουθία που παράχθηκε αποφέρει κέρδος 67974244180.00747 ($\approx 6.8 \times 10^{10}$) δολλαρίων κατόπιν 314164 συναλλαγών.

Παρατηρούμε ότι παρότι το πλήθος των συναλλαγών είναι αρκετών τάξεων μεγέθους μεγαλύτερο από το αντίστοιχο της περίπτωσης της παραγωγής της μικρής ακολουθίας της προηγούμενης ενότητας, εν τέλει το κέρδος δεν είναι αντίστοιχα αυξημένο κατά πολλές τάξεις μεγέθους. Η βασική αιτία που εξηγεί αυτή την συμπεριφορά είναι βέβαια η φύση του αλγορίθμου που υλοποιήθηκε.

Επιβάλλοντας την άπληστη αγοραπωλησία μετοχών των εταιρειών σε ημερήσια βάση, χάνονται πολλές ενδεχόμενες μεγάλες αυξήσεις τιμών μετοχών εταιρειών μεταξύ πολλαπλών ημερών – περιπτώσεις τις οποίες ο αλγόριθμος που υλοποιήθηκε για τον υπολογισμό της μικρής ακολουθίας θα εντόπιζε επιτυχώς και θα εκμεταλλευόταν. Συν τοις άλλοις, για την περίπτωση του υπολογισμού της μεγάλης ακολουθίας καταβλήθηκε συνολικά σημαντικά λιγότερη προσπάθεια στην βελτιστοποίηση του ίδιου του αλγορίθμου, ακριβώς διότι υπήρχε η γνώση ότι το αρκετά μεγαλύτερο επιτρεπτό πλήθος συναλλαγών θα μπορούσε να οδηγήσει σε αποδεκτές λύσεις οι οποίες θα μπορούσαν να βασίζονται σε λιγότερο σύνθετη συλλογιστική.

Το αντίστοιχο διάγραμμα αποτίμησης της περίπτωσης της μεγάλης ακολουθίας παρατίθεται στο παρακάτω σχήμα. Παρατηρούμε ότι αυτή την φορά στο κλείσιμο κάθε ημέρας δεν υπάρχει καμία μετοχή εταιρείας στην κατοχή μας, αφού ο αλγόριθμος ορίζει ότι πρέπει να έχουν όλες πωληθεί μέχρι

το τέλος κάθε ημέρας. Επομένως, το διάγραμμα αποτίμησης έχει πρακτικά εκφυλιστεί σε διάγραμμα του υπολοίπου του λογαριασμού.

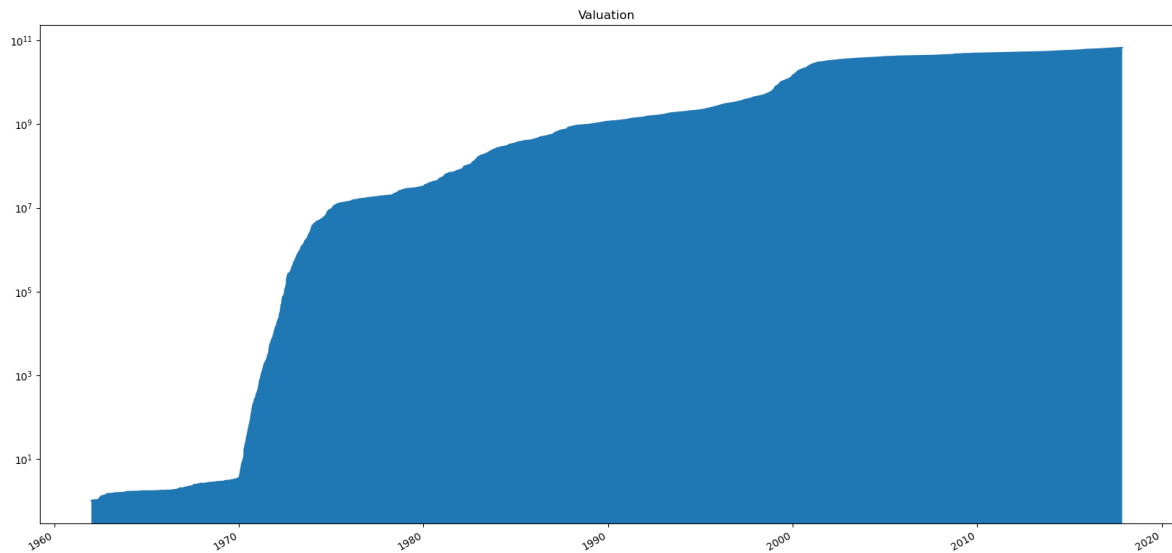


Figure 2: Διάγραμμα αποτίμησης κατά τον υπολογισμό της μεγάλης ακολουθίας.

Μία άλλη παρατήρηση που αξίζει να σημειωθεί είναι ότι, και πάλι λόγω της φύσης του παρόντος αλγορίθμου σε σχέση με αυτού της προηγούμενης ενότητας, το προκείμενο διάγραμμα αποτίμησης “ακολουθεί” καλύτερα την ευρύτερη τάση του χρηματιστηρίου (εφόσον η πώληση κάθε μετοχής πραγματοποιείται την ημέρα της αγοράς της), σε αντίθεση με το αντίστοιχο διάγραμμα της προηγούμενης ενότητας, το οποίο χρησιμοποιούσε το χρονικό παράθυρο “αλλοιώνοντας” έτσι την αντίστοιχη απεικόνιση της τάσης του χρηματιστηρίου σε επίπεδο ευαισθησίας 40 ημερών.