

Πολυπρακτορικά Συστήματα

(ΠΛΗ 512)

Αναφορά Προγραμματιστικής Εργασίας



Κωνσταντίνος Νικολός

2019030096

Γιώργος Βαλαβάνης

2019030065

Κυριότερες Προκλήσεις:

Επιλογή Hardliner ή Conceding Agent.

Η σημαντική πρόκληση ήταν η απόφαση σχετικά με την ιδέα που θα ακολουθηθεί για την υλοποίηση του Agent. Χρησιμοποιήθηκαν οι Agents (CSE3210) ως σημείο αναφοράς με κύριο ζητούμενο την κατάληξη στην υλοποίηση του πράκτορα ως hardliner ή περισσότερο conceding. Αφού έγιναν ορισμένα tests μεταξύ των ήδη υπάρχων υλοποιήσεων που δόθηκαν προέκυψε ότι οι πράκτορες με πιο αυστηρή στρατηγική κερδίζουν περισσότερο αν και καταλήγουν σε λιγότερα agreements. Με βάση αυτά τα αποτελέσματα επιλέχθηκε ένα μοντέλο κοντινό σε έναν hardliner agent το οποίο θα ζητάει κυρίως υψηλά utilities αλλά με την προσπάθεια να βρεθεί κάποιο καλό bid κοντά στο σημείο Nash που να ικανοποιεί και τους δύο παίκτες.

Επιλογή Βάσης του Agent.

Όσον αφορά την εκκίνηση της εργασίας αποφασίστηκε να χρησιμοποιηθεί ένας υπάρχων υλοποιημένος πράκτορας από εκείνους που δόθηκαν και να λειτουργήσει ως βάση στην οποία θα μπορούσαν να προστεθούν οι δικές μας ιδέες. Για την επιλογή αυτή χρειαζόταν ένας απλός κώδικας εύκολα κατανοητός που χρησιμοποιούσε time dependent strategy για να μπορεί να προσαρμοστεί στις νέες ιδέες. Επιλέχθηκε ο Agent 26 καθώς χρησιμοποιούσε ήδη ένα σύστημα βαρών και μία time dependent στρατηγική, τα οποία θα μπορούσαν να αναπτυχθούν σύμφωνα με τις ιδέες που ξεχωρίσαμε από τα raref που αναφέρθηκαν προηγουμένως.

Εύρεση Θεωρητικού Υπόβαθρου.

Ένας βασικός προβληματισμός αποτελούσε το γεγονός ότι δεν υπάρχει τεράστιο υλικό στο διαδίκτυο σχετικά με το διαγωνισμό ANAC ή την διαπραγμάτευση μεταξύ πρακτόρων γενικότερα. Βρέθηκαν κάποια papers που αναφέρονταν σε προηγούμενους νικητές του διαγωνισμού με μια πολύ βιαστική εξήγηση της βασικής ιδέας, αλλά ήταν δύσκολο να κατανοηθούν και να μεταφερθούν χωρίς την πρόσβαση στον πηγαίο κώδικά τους. Η επιθυμία μας ήταν να φτιάξουμε κάποιου είδους opponent modeling το οποίο θα ήταν επιστημονικά σωστό, ήδη ελεγμένο και κατ' επέκταση να συνάδει με την θεωρία του μαθήματος. Χωρίς όμως οι ίδιοι να γνωρίσουμε τα μαθηματικά για την δημιουργία δικών μας συναρτήσεων ήταν δύσκολη η εύρεση ενός τέτοιου paper. Τελικά, ξεχώρισαν δύο paper με κάποια υλοποίηση opponent modeling που μπορούσαν να μεταφερθούν στον υπάρχον project.

- Το 1ο paper [1] αφορά τον διαγωνισμό ANAC 2017 και παρουσιάζει τα χαρακτηριστικά και τις στρατηγικές των διαφόρων agents που συμμετείχαν σε εκείνο τον διαγωνισμό. Στόχος ήταν η εύρεση ενός time dependent threshold, το οποίο θα αντιστοιχούσε σε κάποιον agent που βρισκόταν στην πρώτη τετράδα. Ο Agent που ταίριαζε στις ανάγκες του threshold, ήταν ο AgentKN από τον Keita Nakamura. Χρησιμοποιεί τη μέση τιμή και τη διασπορά των bids που έχει προτείνει ο αντίπαλος για να προβλέψει το μέγιστο utility του αντιπάλου τη δεδομένη στιγμή. Επιπλέον, το threshold του είναι time dependent, επομένως ταιριάζει στον πράκτορά μας. Παρακάτω, υπάρχει η αναλυτική περιγραφή των τύπων που χρησιμοποιήθηκαν για τη δημιουργία του Acceptance Threshold.
- Το 2ο paper [2] εξηγεί τις στρατηγικές του νικητή πράκτορα του ANAC 2020, agentKT. Ο agentKT δημιουργεί ένα frequency map των bids του αντιπάλου και χρησιμοποιεί αυτά τα frequency maps για να δημιουργήσει ένα μετρικό αξιολόγησης της σημασίας των διάφορων issues (importance). Εκμεταλλεύοντας το importance και τα frequency maps υλοποιήθηκε η Find Most Similar Bid, η οποία προσπαθεί να βρει ένα bid το οποίο θα έχει όσο το δυνατόν περισσότερα κοινά values για κάθε issue με το συχνότερο bid του αντιπάλου, με σκοπό να προσεγγίσει κάποιο Nash Equilibrium και να διασφαλίσει agreement. Παρακάτω περιγράφεται αναλυτικά η Find Most Similar Bid (Finding Most Similar Bid).

Στρατηγική του Πράκτορα:

Η στρατηγική του συγκεκριμένου πράκτορα αποτελείται από τέσσερα βασικά στοιχεία:

- **Random Walk** ($0 \leq t < 0.4$)

Για το χρονικό διάστημα $0 \leq t < 0.4$ εκτελείται Random Walk στο οποίο ο πράκτορας επιλέγει τυχαία από μια συλλογή με τα 40 καλύτερα Bids με βάση το utility που μας προσφέρθηκε.

- **Accept Good Offers** ($0 \leq t < 1$)

Καθόλη τη διάρκεια του simulation υπολογίζεται ένα acceptance threshold του πράκτορα σύμφωνα με το maximum utility που θα μπορούσε να προσφέρει ο αντίπαλος. Το acceptance threshold, εκτός από φυσικά να αποτελεί το πάσο για τις προσφορές που δέχεται ο πράκτορας, επηρεάζει σημαντικά και τα Bids που ενδέχεται να σταλούν στον αντίπαλο.

- **Find Most Similar Bid** ($0.4 \leq t < 1$)

Ο πράκτορας ξεκινάει την αποστολή bids τα οποία έχουν ομοιότητες με προηγούμενες προσφορές του αντιπάλου με στόχο την εύρεση κάποιας συμφωνίας, ιδανικά το σημείο Nash. Τα bids που προτείνονται εκτός από κοινά με τον αντίπαλο οφείλουν να υπερβαίνουν σε utility το acceptance threshold της δεδομένης στιγμής.

- **Accept More** ($0.9 \leq t < 1$)

Για την επίτευξη περισσότερων συμφωνιών αλλά και την δημιουργία καλύτερων προτάσεων μετά το πέρας του συγκεκριμένου χρονικού διαστήματος το acceptance threshold μειώνεται δραστικά. Επιπλέον δημιουργήθηκαν συναρτήσεις, ώστε ο πράκτορας να προτείνει αυτούσια bids που έχει προσφέρει ο αντίπαλος ακόμη και με χαμηλότερο utility.

Ο σκοπός της συγκεκριμένης στρατηγικής είναι αρχικά για το στάδιο του Random Walk η αποθήκευση και η εκμετάλλευση της πληροφορίας που προσφέρεται από τα bids του αντιπάλου, ενώ παράλληλα δεν δίνεται κάποιο στοιχείο για την συμπεριφορά του πράκτορα μας. Αφού υπάρχει ένα ικανοποιητικό δείγμα, προσπαθούν να βρεθούν προσφορές που θα ικανοποιούν και τους δύο αντιπάλους, οδηγώντας ιδανικά σε pareto optimality. Τέλος, η σταδιακή μείωση του acceptance threshold στοχεύει στην επίτευξη συμφωνιών σε ενδεχόμενα domains, όπου το maximum utility που μπορεί να προσφερθεί είναι μικρό ή οι αντίπαλοι agents είναι αρκετά “αυστηροί” στις προσφορές τους.

Finding Most Similar Bid:

Για την εύρεση ενός κοινού bid, επί της ουσίας καταγράφεται η συχνότητα πρότασης κάθε value ενός issue από τον αντίπαλο και δημιουργείται ένας πίνακας συχνοτήτων.

- Υπολογίζεται η συχνότητα προσφοράς ενός value (j) στο συγκεκριμένο issue (k). Ορίζεται η τιμή $fk(j) = 1$ εάν πρόκειται για το περισσότερο προτεινόμενο value και $fk(j) = \frac{hk(j)}{Z}$ για τα υπόλοιπα. Όπου Z το άθροισμα όλων των ψήφων σε ένα issue εξαιρώντας εκείνο με τις περισσότερες ψήφους και $hk(j)$ οι ψήφοι του value j στο issue k.
- Υπολογίζεται το ποσοστό σημασίας κάθε issue για τον αντίπαλο πράκτορα σύμφωνα με τα αθροίσματα συνολικών προσφορών και συχνοτήτων.
$$imp(k) = \frac{\sum_{i=1}^{Jk} fk(i) * hk(i)}{\sum_{i=1}^{Jk} hk(i)}$$
 Όπου Jk το άθροισμα των πιθανών values σε ένα issue k.
- Δίνονται πόντοι σε ένα value με βάση το γινόμενο συχνότητας του value και σημαντικότητας του issue $fk(j) * imp(k)$.
- Τέλος προστίθενται σε ένα bid το άθροισμα πόντων του κάθε value που περιέχει.

Η συγκεκριμένη υλοποίηση χρησιμοποιείται αφού βρεθούν τα καλύτερα 40 bids με βάση το utility threshold εκείνης της στιγμής. Υπολογίζεται το most similar bid από τα 40 αυτά πιθανά bids και στέλνεται στον αντίπαλο.

Acceptance Threshold:

Για τον υπολογισμό του acceptance threshold χρησιμοποιείται πάλι εν μέρει opponent modeling καθώς υπολογίζεται με βάση το maximum expected utility που θα προσφέρει ο αντίπαλος.

- $threshold(t) = 1 - (1 - emax(t)) * t^a$
- Η τιμή emax αποτελεί το expected maximum utility του αντιπάλου agent και υπολογίζεται από: $emax(t) = \mu(t) + (1 - \mu(t)) * d(t)$ όπου
- Ορίζεται $\mu(t)$ ως τη μέση τιμή των utilities που προσφέρει ο αντίπαλος
- Ορίζεται $d(t)$ ως το bid range του αντιπάλου που δίνεται από τη σχέση:
$$d(t) = \frac{\sigma(t)}{\sqrt{\mu(t) * (1 - \mu(t))}}$$
- Ορίζουμε $\sigma(t)$ ως το standard deviation of utility που προσφέρθηκε από τον αντίπαλο
- Η τιμή threshold ανανεώνεται με κάθε προσφορά του αντιπάλου και μετά από χρονικό διάστημα $t > 9$ μειώνεται κατά 0.07 ώστε να επιτευχθούν καλύτερες προσφορές.

Heuristics:

Προφανώς μετά το simulation με άλλους αντιπάλους παρατηρήθηκαν κάποιες μη αποδεκτές συμπεριφορές για τις οποίες έγινε προσπάθεια να λυθούν κυρίως με βάση την εμπειρία που πρόσφεραν τα αποτελέσματα. Το βασικότερο ζητούμενο ήταν πως ο πράκτορας δεν έστελνε κάποιο αυτούσιο bid του αντιπάλου ξανά, που είναι πιθανό να δεχθεί αλλά κοινά σε αυτό.

- ShouldAccept
Παρατηρήθηκε ότι ο πράκτορας ορισμένες φορές αρνούνταν προτάσεις που προσέφεραν καλύτερο utility από εκείνο που έστειλε νωρίτερα στον αντίπαλο για αυτό δημιουργήθηκε ο έλεγχος ώστε να δέχεται το συγκεκριμένο bid.
- SendRecieved
Συνάρτηση που παίρνει τα καλύτερα 20 bids που έστειλε ο αντίπαλος και τα κατατάσει με βάρος ανάλογο του utility τους. Τέλος επιστρέφει ένα από τα 20 bids με πιθανότητα ίση με το weight τους. Το bid αυτό στέλνεται στον αντίπαλο εάν δίνει καλύτερο utility συγκριτικά με εκείνο που θα διάλεγε ο Agent (best_bid). Μετά το πέρας του χρονικού διαστήματος $0.9 \leq t$, το ελάχιστο utility του bid που διαλέγει η sendRecieved αρκεί να υπερβαίνει ένα threshold που ξεκινά από 0.88 και μειώνεται σταδιακά στο 0.71 όσο πλησιάζει το τέλος της διαπραγμάτευσης, για να επιστραφεί. Η συγκεκριμένη υλοποίηση υπήρχε ήδη στον κώδικα του agent 26 και χρησιμοποιήθηκε για την βελτίωση του πράκτορα όσον αφορά την αποστολή αυτούσιων bids του αντιπάλου.

Εφαρμογή Θεωρίας που διδάχθηκε στο μάθημα:

Κατά τη διάρκεια της σχεδίασης του πράκτορα μπορούσαν να υιοθετηθούν αρκετές θεωρητικές έννοιες όπως το pareto optimality ή η εύρεση του σημείου Nash, στην πράξη όμως τα πράγματα ήταν κάπως πιο περίπλοκα αποτρέποντας από την κατάλληλη επιστημονική εφαρμογή τους.

Nash Point:

Ένας από του στόχους ήταν να βρεθεί το σημείο Nash που θα ικανοποιούσε και τους δύο παίκτες με τον μέγιστο τρόπο. Η στρατηγική find similar bid θα έδινε περισσότερο βάρος σε αυτό το σενάριο αν δεν παρέμβαινε η ανταγωνιστική φύση του διαγωνισμού. Η find similar bid μπορεί να καταλήξει στην εύρεση ενός καλού bid, κοινό και για τους δύο παίκτες, πιθανά κοντά στο σημείο Nash εάν το επιτρέψει το αντίστοιχο domain που θα δίνει και στους δύο παίκτες καλές αμοιβές. Επιπλέον ένας λόγος που δεν μπορεί να λειτουργήσει καταλλήλα η εύρεση του σημείου Nash, παρότι θα θέλαμε είναι ότι τα utilities του αντιπάλου δεν γίνονται γνωστά και ακόμη τελικός στόχος του διαγωνισμού ήταν η "καταπάτηση" του αντιπάλου γεγονός που οδηγεί στην αποστολή bids μεγαλύτερου reward.

Pareto Optimallity:

Μία ακόμη θεωρητική έννοια που ακούγεται αρκετά ελκυστική αλλά δύσκολη στην εφαρμογή της είναι η αποστολή pareto optimall bids. Πάλι έγινε μια προσπάθεια εκτέλεσης της συγκεκριμένης θεωρίας με την εύρεση ενός κοινού bid με τον αντίπαλο, αλλά προέκυψαν τα ίδια εμπόδια, καθώς το pareto σημείο εξασφαλίζει το καλύτερο δυνατό σενάριο, δεδομένου ότι κανένας αντίπαλος δεν επωφελείται του άλλου, που αντικρούεται με τις ανάγκες του διαγωνισμού. Ίσως θα μπορούσε μια άλλη προσέγγιση, μειώνοντας το expected utility του πράκτορα, να οδηγήσει σε καλύτερα αποτελέσματα διεκδικώντας περισσότερες συμφωνίες, αλλά δεν εφαρμόστηκε στην προκειμένη περίπτωση.

Fictitious Play:

Τέλος ένας ακόμη στόχος της δημιουργίας του find similar bid αποτελεί την εφαρμογή εν μέρει του fictitious play. Σύμφωνα με το fictitious play, και οι δύο παίκτες δημιουργούν και ανανεώνουν πεποιθήσεις για την στρατηγική του αντιπάλου μέσω της καταγραφής των προτιμήσεών του, κατά τη διάρκεια των προηγούμενων γύρων. Μία τέτοια εφαρμογή είναι περίπου η καταγραφή της συχνότητας των προτιμήσεων του αντιπάλου σε ορισμένα values για κάθε διαθέσιμο issue στο συγκεκριμένο domain και η προσαρμογή των bets ανάλογα με την στρατηγική που ακολουθεί ο αντίπαλος. Ένα πρόβλημα με την εφαρμογή της συγκεκριμένης θεωρίας είναι πως προϋποθέτει ότι και οι δύο παίκτες χρησιμοποιούν την ίδια στρατηγική, γεγονός το οποίο δεν μπορούσαμε να γνωρίζουμε, καθώς και ότι δεν οδηγεί πάντα στα καλύτερα αποτελέσματα ή στην εύρεση ενός σημείου Nash.

Αποτελέσματα Διαγωνισμού:

Η πρώτη και πιο αναμενόμενη παρατήρηση είναι ο αριθμός των disagreements. Εφόσον ο Agent δεν ρίχνει το utility του κάτω από 0.71 είναι αναμενόμενο να έχει αρκετά disagreements. Αυτό είναι και ένας βασικός παράγοντας που δεν πέτυχε ο Agent υψηλό average utility. Είχε παρατηρηθεί και στα πειράματα, όμως αυτό είχε ως αποτέλεσμα να καταλήγει σε agreements με πολύ υψηλό utility για εκείνον και αρκετά χαμηλότερο για τον αντίπαλο. Για αυτό, παρόλο τον μικρό αριθμό Agreements συγκριτικά με τους αντιπάλους παρατηρούμε average utility 0.61.

Επιπλέον, ο Agent έχει ένα από τα χαμηλότερα Nash products, χωρίς όμως να βγαίνει από τους τελευταίους στην κατάταξη. Αυτό σημαίνει πως στα περισσότερα agreements είχε αρκετά παραπάνω utility συγκριτικά με τον αντίπαλο, κάτι το οποίο

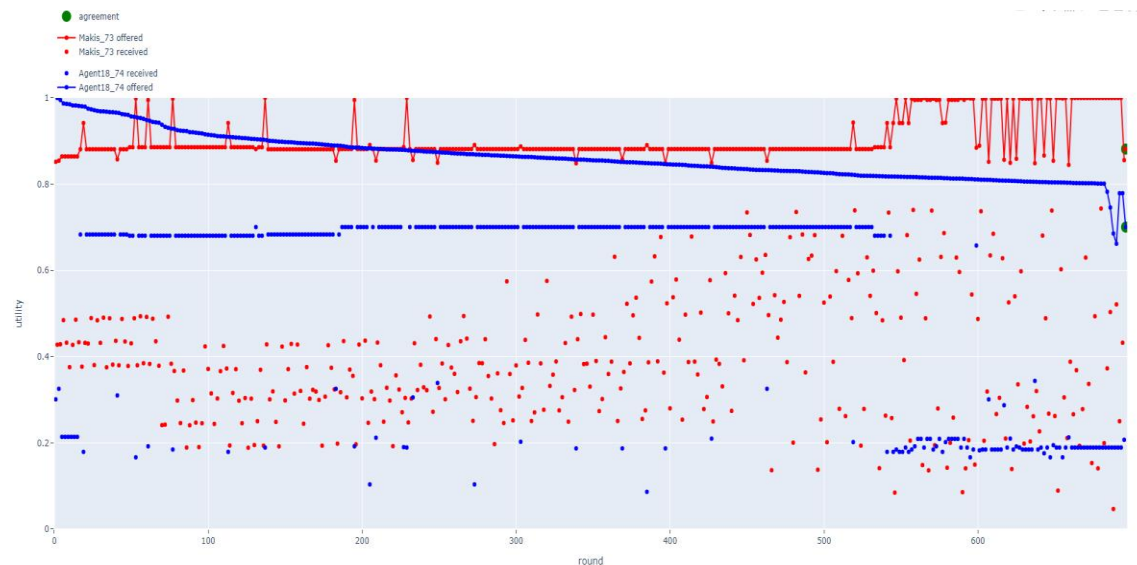
ήταν αναμενόμενο για έναν hardliner. Παρακάτω παρατίθενται πειραματικά αποτελέσματα ενάντια του CSE3210 Agent 11 και του CSE3210 Agent 18:

Mean Own Score: 0.7917624796722221	Mean Own Score: 0.8472879610588236
Mean Opponent Score: 0.6951205126166666	Mean Opponent Score: 0.6535657993117647
Total Agreements Made: 33	Total Agreements Made: 32
Total Disagreements : 7	Total Disagreements : 8

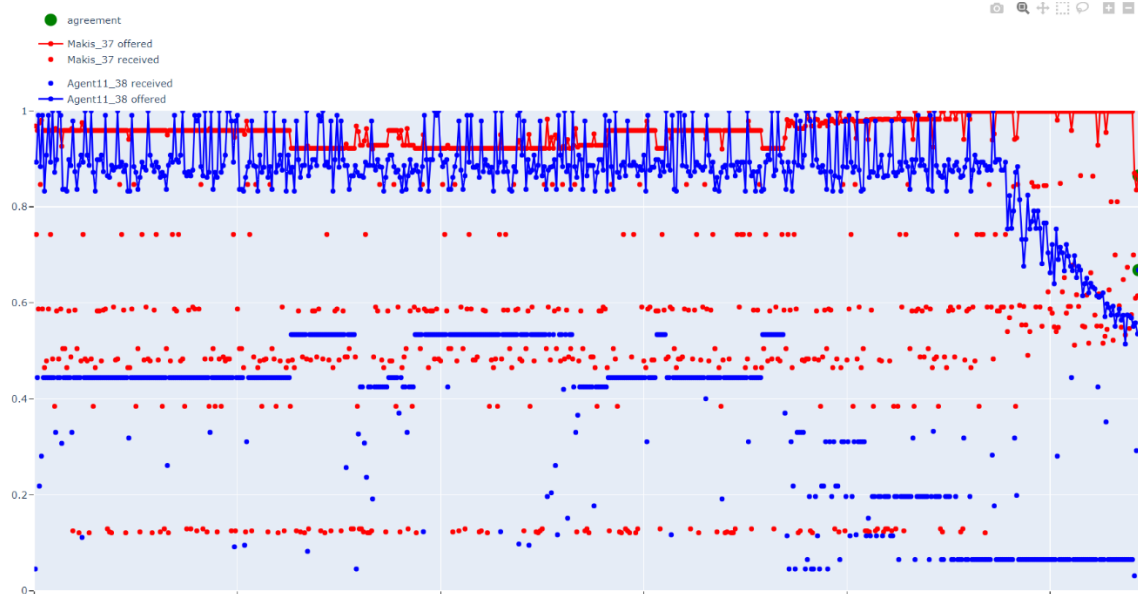
**Εικόνα 1: Makis vs Agent 11 (hardliner)
σε 40 γύρους**

**Εικόνα 2: Makis vs Agent 18 (conceder)
σε 40 γύρους**

Παρατηρείται ότι τα πειραματικά αποτελέσματα είναι λίγο καλύτερα από τα αποτελέσματα του διαγωνισμού γεγονός που λογικά οφείλεται στον πειραματισμό με λιγοστούς πράκτορες κατά την διάρκεια των δοκιμών σε αντίθεση με τον διαγωνισμό όπου υπήρχαν περισσότερες στρατηγικές παικτών.



**Εικόνα 3: Makis vs Agent 18 (conceder) αποτελέσματα ενός γύρου
Εδώ γίνεται αποδοχή ενός bid που δίνει μεγαλύτερο utility από το
bid με το ελάχιστο utility που πρότεινε ο Makis.**



Εικόνα 4: Makis vs Agent 11 (hardliner) αποτελέσματα ενός γύρου.
Εδώ γίνεται χρήση ενός προηγούμενου bid που προτάθηκε από τον αντίπαλο, η διαπραγμάτευση τελειώνει ($\text{progress} > 0.9$) και το enemy bid πληρεί το time dependent threshold και στέλνεται ξανά από τον πράκτορα.

Σχετικά με τον αριθμό των offers, είναι μικρός συγκριτικά με των άλλων πρακτόρων, διότι όταν το progress γίνει μεγαλύτερο του 0.4, ο Agent προτείνει το most similar bid με πιθανότητα 55%. Από ένα σημείο και έπειτα αυτό το bid δεν αλλάζει επομένως κάθε φορά που επιλέγει να προτείνει το most similar bid καταλήγει να προτείνει το ίδιο bid. Επιπλέον, επειδή το threshold μειώνεται πολύ αργά μέχρι να πλησιάσει το τέλος της διαπραγμάτευσης ακόμα και η τυχαία επιλογή ενός bid που βρίσκεται μέσα στο threshold καταλήγει να μην είναι και τόσο τυχαία τελικά, καθώς τα bids που βρίσκονται μέσα στο threshold είναι περιορισμένα.

Τέλος παρατηρείται ότι σε αρκετές περιπτώσεις το most similar bid που προτείνεται διαρκώς μετά το πέρας μια χρονικής στιγμής, συχνά κρίνεται άστοχο (πολύ αυστηρό) καθώς έχουν νωρίτερα προταθεί καλύτερα bids που μειώνουν λίγο το utility του πράκτορα αλλά αυξάνουν αρκετά του αντιπάλου τα οποία δεν επαναλαμβάνονται.

Ενδεχόμενες βελτιώσεις:

Όπως προέκυψε και από τα δεδομένα που προαναφέρθηκαν, το κύριο ζήτημα του πράκτορα είναι ότι δεν πετύχαινε αρκετές συμφωνίες και σε συνδιαμό με το ικανοποιητικό utility που δέχθηκε στο διαγωνισμό καταλήγουμε στο συμπέρασμα πως θα ήταν καλύτερο να μειωθούν τα απαιτούμενα rewards με στόχο την επίτευξη περισσότερων συμφωνιών.

Μείωση Expected Utility:

Μία τέτοια ενδιαφέρουσα ιδέα ακούστηκε στην παρουσίαση των projects από μία ομάδα (Faker) κατά την οποία δινόταν περισσότερη έμφαση στις προτιμήσεις του αντιπάλου μέσω βαρών από τις προτιμήσεις του ίδιου του πράκτορα. Μια τέτοια προσθήκη θα μπορούσε να γίνει στην εύρεση κοινών bids με των αντίπαλων μειώνοντας το acceptance threshold σε ένα ενδεχόμενο δύσκολο domain τα οποία πρέπει να πληρούν για να σταλούν στον αντίπαλο.

Προσθήκη βαρών στην εύρεση του Most Similar Bid:

Παρατηρήθηκε ότι το most similar bid δεν άλλαζε από ένα σημείο της διαπραγμάτευσης και μετά. Επομένως, μια ενδεχόμενη βελτίωση θα ήταν η προσθήκη χρονικά εξαρτωμένων βαρών στο σύστημα των πόντων κάθε bid. Τα βάρη, όσο θα περνούσε ο χρόνος της διαπραγμάτευσης θα αυξανόντουσαν, με αποτέλεσμα τα μεταγενέστερα bids να έχουν μεγαλύτερη σημασία από τα προηγούμενα. Δηλαδή πέρα από το importance και το frequency, ο Agent θα κοιτάει και τον χρόνο. Συνεπώς, το most similar bid θα μεταβάλλεται πιο συχνά μέσα στο χρόνο, αυξάνοντας τη πιθανότητα για agreement.

Ολοκλήρωση του Opponent Modeling:

Μία ακόμη ιδέα που δεν υλοποιήθηκε σε επιθυμητό επίπεδο είναι η διάκριση των αντιπάλων σε κατηγορίες με βάση την συμπεριφορά τους και η προσαρμογή αντίστοιχα της στρατηγικής. Αυτή η ιδέα ακούστηκε και πάλι την παρουσίαση και από εμάς ως ενδεχόμενο βελτίωσης αλλά και από την ομάδα που κέρδισε τον διαγωνισμό η οποία το είχε υλοποιήσει. Η βασική ιδέα που θα μπορούσε να εφαρμοστεί είναι η παρατήρηση των utilities που προσφέρει ο αντίπαλος (εάν τα bids προσφέρουν στον δικό μας agent χαμηλό utility και εάν μειώνονται ή αυξάνονται με το πέρασμα του χρόνου οδηγώντας στην διάκριση του σε hardliner ή conceding. Στην περίπτωση ενός hardliner αντιπάλου το expected utility στις προσφορές του πράκτορα θα μειωνόταν ενώ στην περίπτωση ενός conceding τα bids θα γίνονταν πιο αυστηρά επιζητώντας μεγαλύτερο κέρδος.

Βιβλιογραφία

[1] : [ANAC_2017_Repeated_Multilateral_Negotiation_League](#)

[2] : [Baseline Strategies for the ANAC Automated Negotiation League](#)