Μπιμπίρης Πέτρος 2019030135

Χριστοφιλογιάννης Ιωάννης 2019030140

Περίληψη Research Paper:

**Intro:** Το κύριο πρόβλημα που εξετάζεται είναι ότι δεν υπάρχει ιδανική στρατηγική διαπραγμάτευσης για όλα τα domains.

Οι συγγραφείς αντιμετωπίζουν αυτό το πρόβλημα επιλογής αλγορίθμου (στρατηγικής) με τη δημιουργία meta-agent που εξετάζει ποσοτικά την αποτελεσματικότητα των αλγορίθμων. Χωρίζουν τα προβλήματα στο on-line και off-line, με κριτήριο το αν μπορεί να αλλάξει η στρατηγική μετά την αρχή της διαπραγμάτευσης \*. Το off-line το αντιμετωπίζουν με supervised machine learning βασισμένο στο performance και τα χαρακτηριστικά του domain, ενώ το on-line με τεχνικές multi-armed bandit (MAB)\*. Αφού εστιάζουν στις διμερείς (bilateral) διαπραγματεύσεις με κρυφές στρατηγικές η αξιολόγηση των αποτελεσμάτων έγινε με βάση το GENIUS testbed διαπραγματεύσεων του διαγωνισμού ANAC και αποδεικνύουν ότι «λίγη μάθηση» βοηθάει σε αυτό και πιθανώς σε άλλα προβλήματα απόφασης πολυπρακτορικών συστημάτων.

3: Ο σχεδιασμός του μετα-πράκτορα περιλαμβάνει τη δυναμική πρόβλεψη της απόδοσης των στρατηγικών διαπραγμάτευσης.

Third, it empirically shows that the techniques significantly improve the state-of-the-art agents in an international competition setting. These results provide insights for agent-designers in negotiation, demonstrating that “a little learning goes a long way”, and suggesting that the algorithm selection approach may also be feasible for other multi-agent optimization problems such as planning and decision-making

Επιλογή Αλγορίθμου:

s: Which algorithm out of a set of possible candidates is likely to perform best for a well-defined problem

that characterize problem instances that can be computed off-line and that correlate with the

computational hardness/complexity of the problem.

s run-time on a representative set of problem instances, and then to

use only the algorithm that offered the best

**Ml black boxes**

**Gia to 4:**

Ο αλγόριθμος μάθησης αντιμετωπίζει τους αλγορίθμους πρακτόρων σαν «μαύρο κουτί» μετρώντας μόνο την απόδοση τους.

**Our setting differs because negotiation is a multi-agent optimization problem, and the performance of a negotiation strategy highly depends on the negotiation strategy used by otheρ agents**

2.2 Reaching agreement over joint resources or task assignments,

Το άρθρο μελετά 3 περιπτώσεις. Πρώτα τα άγνωστα domains με γνωστούς agents, ύστερα τα άγνωστα domains με γνωστούς agents και τέλος άγνωστοι agents και domains όπως στον διαγωνισμό ANAC.

the fact that an agent is the overall winner of the tournament does not imply that the agent is optimal for every domain.

Ο πράκτορας που κερδίζει τον διαγωνισμό ANAC δεν είναι ο βέλτιστος για κάθε domain.

The aggregate behaviour of PDAs has been shown to closely follow that of humans in some negotiation settings [34–38] and can thus be used as a proxy for measuring aggregate human behaviour.

**4 Off-line algorithm selection**

Στην off-line παραλλαγή του προβλήματος η στρατηγική/πράκτορας επιλέγεται στην αρχή και δεν αλλάζει και όπως αναφέρουν οι συγγραφείς δεν μπορούμε να “αλλάξουμε άλογο” στην μέση του αγώνα.

Μετράνε τη διαφορά απόδοσης μεταξύ πρακτόρων και τη μέση απόδοση (avg) για standardization αφού η απόδοση μπορεί να μην αποτελεί ιδανικό metric από μόνη της.

Ένας agent είναι “νικητής” αν είναι εντός μίας τυπικής απόκλισης με τον βέλτιστο/ιδανικό. Αυτό είναι σημαντικό αφού μπορεί πρακτικά κάποιος που είναι σχεδόν όσο καλός όσο ο optimal να είναι αρκετά καλός στην πράξη (+no free lunch?)

Χρησιμοποίησαν μεθόδους μάθησης με σκοπό την πρόβλεψη της απόδοσης όπως: γραμμική παλινδρόμηση στα χαρακτηριστικά του domain, CART (για αντιμετώπιση μη-γραμμικότητας) και νευρωνικό δίκτυο με ένα κρυφό layer. Επίσης με σκοπό την πρόβλεψη του νικητή χρησιμοποίησαν: Λογιστική παλινδρόμηση έναντι της γραμμικής και τις άλλες 2 μεθόδους χωρίς αλλαγές.

Ο αλγόριθμος meta πράκτορα που δημιούργησαν, μαθαίνει δίνοντας του τα training domains D και τους πράκτορες A και μετά το πέρας της μάθησης αποφασίζει την καλύτερη στρατηγική (πράκτορα) με βάση την πρόβλεψη της απόδοσης της, επιλέγοντας με δικά του κριτήρια νικητή σε περίπτωση ισοπαλίας.

Αποτελέσματα:

Οι 3 παραλλαγές του περιβάλλοντος που ερευνήθηκαν είναι: Νέα domains γνωστοί agents, γνωστά domains νέοι agents, νέα domains νέοι agents όπως και ο διαγωνισμός ANAC.

6-fold cross validation. All methods better than

Extra:

This strategy is equal to choosing the winning agent in A according to its average performance when negotiating with all agents in the training set (using the same rules as in the ANAC competition). (2) the “Oracle” strategy: selecting the agent in A associated with the maximum score on test domain d (s d,A′ ∗ of Eq. 2) assuming retrospect knowledge of the test agents

**5.On-line algorithm selection:**

Χρησιμοποιούν agent βασισμένο στο Multi-Armed Bandit (MAB), που αποτελεί τρόπο επίλυσης προβλημάτων ακολουθιακής απόφασης, με δίλημμα εξερεύνησης-εκμετάλλευσης. Χρησιμοποιούν δύο διαφορετικές παραλαγές. Ο απλός MAB δεν γνωρίζει τις κατανομές και δεν έχει ξεχωριστή φάση μάθησης. Χρησιμοποιείται ο αλγόριθμος Upper Confidence Bound (UCB) επειδή επιλέγει πιο συχνά τον βέλτιστο πράκτορα αλλά συνεχίζει να εξερευνεί. Ο prior-MAB βελτιώνει τον UCB συνδυάζοντας τον με τον αλγόριθμο μηχανικής μάθησης της ενότητας 4 για να δώσει μια εκτίμηση της απόδοσης κάθε πράκτορα .

**Αποτελέσματα:**

Εισάγοντας τον prior-MAB στον ANAC 2013 κατέλαβε τη δεύτερη θέση, λόγω μεγαλύτερης τυπικής απόκλισης από τον TheFawkes (1η θέση), αλλά παρατηρήθηκαν κάποιες μη βέλτιστες αποφάσεις λόγω της εξερεύνησης διαφορετικών πρακτόρων κάτι που είναι λογικό για τον MAB.

Και οι δύο μέθοδοι είχαν καλύτερη απόδοση από τον meta-agent κάτι που είναι λογικό αφού λόγω της δυναμικότητας του περιβάλλοντος δεν είναι περιορισμένοι στην επιλογή μίας μεθόδου μέχρι το πέρας της διαπραγμάτευσης, χωρίς προσαρμογές ακόμα και αν στην αρχή η απόδοση τους ήταν κοντινή.

Ο απλός MAB επιτυγχάνει αντίστοιχα αποτελέσματα με τον prior-MAB μόνο στην περίπτωση που υπάρχει περίοδος μάθησης ίσης με το 60% των γύρων δοκιμής.

**6. Συζήτηση – Συμπέρασμα:**

Η εργασία ορίζει διάφορες παραλλαγές επιλογής αλγορίθμων, προτείνει μια προσέγγιση μηχανικής μάθησης και καταδεικνύει την αποτελεσματικότητά της. Ξεπερνά τους φιναλίστ του διαγωνισμού, επιδεικνύοντας τη στιβαρότητα της προσέγγισης σε διαφορετικά περιβάλλοντα διαπραγμάτευσης. Αναγνωρίζονται οι περιορισμοί και προτείνεται μελλοντική εργασία, συμπεριλαμβανομένης της βελτίωσης της εξερεύνησης στις διαδικτυακές διαπραγματεύσεις και της επέκτασης της προσέγγισης σε άλλες αλληλεπιδράσεις πολλών πρακτόρων.

Offline Ο meta agent συμφωνεί συχνα με το oracle choice και νικα τον average-best agent. Online O MAB βελτίωνε την απόδοση του με την πάροδο του χρόνου. Μια συνδυαστική μέθοδος, η prior-MAB αποδείχτηκε ιδιαίτερα αποτελεσματική στο να νικήσει τον meta-agent. Ο meta-agent δεν συμφωνούσε πάντα με το oracle choice λόγω της επανεκκίνησης μεταξύ γύρων διαπραγμάτευσης. Επίσης ο αλγόριθμος UCB δεν είναι απαραίτητα κατάλληλος για αυτού του είδους διαπραγματεύσεων αφού απαιτεί μεγάλο αριθμό γύρων για να συγκλίνει

Μελλοντική δουλειά:

Βελτίωση της εξερεύνησης, εύρεση και εκμετάλλευση μιας προσφοράς ισορροπίας Nash, πιο κατάλληλη επιλογή μοντέλου μάθησης με βάση την προκαθορισμένη αποδοτικότητα. Τέλος προτείνουν την αλλαγή του διαγωνισμού ANAC για αποθήκευση πληροφοριών μεταξύ γύρων διαπραγματεύσεων μεταξύ διαφορετικών πρακτόρων, κάτι που έγινε και ευελπιστούν τα ευρήματα τους στην επιλογή αλγορίθμου να εφαρμοστούν σε προβλήματα ψηφοφορίας και συντονισμού.