



[ΠΛΗ 512]- Πολυπρακτορικά Συστήματα

Παρουσίαση επιστημονικού άρθρου:

Approximating Game-Theoretic Optimal Strategies for Full-scale Poker

D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Schauenberg, and D. Szafron

Ιωάννης Περίδης

A.M. 2018030069

Εισαγωγή:

Το πόκερ, είναι ένα στοχαστικό, ακολουθιακό παιχνίδι με ατελής πληροφόρηση. Επομένως, ακόμα και γνωρίζοντας τις πτυχές του game theoretic value του παιχνιδιού (και αγνοώντας τις διαφορετικές στρατηγικές κάθε αντιπάλου που προφανώς δεν ακολουθάει game theory optimal strategy) είναι αρκετά δύσκολο να βρεθεί μια βέλτιστη λύση για όλα τα πιθανά σενάρια. Αυτό λόγω του τεράστιου υπολογιστικού κόστους που θα χρειαζόταν. Το άρθρο αυτό λοιπόν, έχει σκοπό να βρει νέους τρόπους για να μειώσει την πολυπλοκότητα της εύρεσης των λύσεων αυτών. Το επιτυγχάνει, εξερευνώντας την χρήση αφηρημένων μαθηματικών μοντέλων που καταγράφουν τις βασικές ιδιότητες του πραγματικού χώρου καταστάσεων, έτσι ώστε μια ακριβής λύση στο μικρότερο πρόβλημα, να παρέχει μια χρήσιμη προσέγγιση μιας βέλτιστης στρατηγικής στον πραγματικό τομέα.

Μοντελοποίηση Προβλήματος & Game Theory:

Αφού το πόκερ, είναι όπως προαναφέραμε ένα παιχνίδι ατελούς πληροφόρησης, το οποίο εξαρτάται από αποτελέσματα τύχης, μπορεί να αναπαρασταθεί σαν ένα ατελούς πληροφόρησης game tree, με κόμβους τύχης και απόφασης, που χωρίζονται σε information sets. Στο δέντρο θα υπάρχουν 4 ακολουθίες πονταρισμάτων μεταξύ των παικτών, μια αρχική και μια μετά από κάθε έναν από τους 3 γύρους, οι οποίοι αποτελούν κόμβους τύχης. Παρόλα αυτά, αφού οι κόμβοι του δέντρου δεν είναι ανεξάρτητοι, οι κλασικές μέθοδοι διαίρει και κυρίευε που κάνουν υπολογισμούς σε υπο-δέντρα, δεν γίνεται να εφαρμοστούν. Η πιο συμβατική μέθοδος για την επίλυση τέτοιων προβλημάτων, είναι η αναπαράσταση του προβλήματος σαν ένα σύστημα γραμμικών εξισώσεων, το οποίο στην συνέχεια θα λυθεί μέσω γραμμικού προγραμματισμού. Στην περίπτωση μας όμως, εφόσον το game tree, μπορεί να είναι μέχρι πολύ μεγάλου μεγέθους, πάλι η χρήση της μεθόδου αυτής, έχει μεγάλους περιορισμούς ή είναι απαγορευτική. Άρα καταλήγουμε στο συμπέρασμα, πως πρέπει να βρούμε έναν εξυπνότερο τρόπο να προσεγγίσουμε το πρόβλημα.

Μέθοδοι επίλυσης-Abstractions:

Ο σκοπός των abstractions, είναι να μειώσουν την κλίμακα του προβλήματος χωρίς να αλλοιώσουν τα θεμελιώδη στοιχεία του παιχνιδιού ή τις βέλτιστες στρατηγικές. Υπάρχουν, πολλοί διαφορετικοί τρόποι απλοποίησης του παιχνιδιού, οι οποίοι ποικίλουν στο επίπεδο αφαίρεσης ή στην ακρίβεια προσέγγισής τους. Αρχικά, διερευνήθηκαν κάποιες αφαιρέσεις εξαιρετικά υψηλού βαθμού, όπως για παράδειγμα η μείωση του πλήθους των καρτών της τράπουλας, η μείωση των καρτών στο χέρι των παικτών από 2 σε 1 ή η μείωση των καρτών που μοιράζονται στο board. Παρόλο που στις μεθόδους αυτές, βρέθηκαν σχεδόν βέλτιστα αποτελέσματα, πρακτικά παράγουν μικρότερα sub-games του πόκερ, τα οποία δεν είναι συμβατά για να προσεγγίσουν το Texas Hold'em, διότι αλλάζουν την βασική δομή του ίδιου του παιχνιδιού.

Betting round reduction: Στους κανονισμούς του limit Hold'em, επιτρέπονται ανα γύρο στους παίκτες έως 4 πονταρίσματα/αυξήσεις πονταρίσματος. Σε παιχνίδι 2 παικτών, μας δίνει 17 πιθανούς συνδυασμούς πονταρισμάτων από τους οποίους 8 καταλήγουν σε τερματικό κόμβο(κάποιος παίκτης κάνει fold), ενώ μόνο 9 καταλήγουν σε call, άρα συνεχίζουν στον επόμενο κόμβο τύχης. Μειώνοντας τον αριθμό αυτόν σε 3, μειώνουμε το branching factor του δέντρου από 9 σε 7, χωρίς να επηρεάζουμε το παιχνίδι κάθε παίκτη (σε αντίθεση με το να το μειώναμε σε 2, όπου προκαλεί ραγδαίες αλλαγές στην στρατηγική τους).

Elimination of betting rounds: Στην μέθοδο αυτή, γίνεται περικοπή του 4^{ου} γύρου πονταρισμάτων, πρακτικά δεν μοιράζεται τελευταία κάρτα του board και έχουμε ένα 3 round game tree. Ύστερα, γίνεται Estimated Value roll out, για όλους τους πιθανούς συνδυασμούς καρτών που απομένουν και κάθε παίκτης παίρνει ένα ποσοστό του pot όσο η αξία του χεριού του. Υπάρχουν και μοντέλα ακραίων περικοπών (π.χ. psOpt1, 1 round game, βρέθηκε ο optimal Preflop solver), τα οποία παρόλο που δεν μπορούν να σταθούν σαν στρατηγικές από μόνα τους, αρκετές φορές μας είναι χρήσιμα (σύνθεση μοντέλων). Παρόλα αυτά, δεν θα μπορέσουμε ποτέ να εξετάσουμε τους γύρους πονταρισμάτων ξεχωριστά, εφόσον είναι ισχυρά εξαρτώμενοι και έτσι θα χάναμε πληροφορία για τις επόμενες αποφάσεις μας.

Composition of preflop and postflop models: Στην μέθοδο αυτή, γίνεται μια εφικτή αποσύνθεση του game tree σε preflop και postflop κομμάτια που θα διαχειριστούμε ξεχωριστά και ύστερα θα τα συνθέσουμε. Με αυτόν τον τρόπο,

επιτυγχάνεται η αποφυγή της πολυπλοκότητας του preflop γύρου πονταρισμάτων. Το υποδέντρο του postflop, παίρνει σαν όρισμα την λύση του preflop και δημιουργεί μέσω των υπο συνθήκη πιθανοτήτων αυτών, λύσεις για διαφορετικές ακολουθίες πονταρισμάτων, που αντιστοιχούν στον αριθμό των raises που έγιναν preflop. Για την preflop λύση, μπορεί να χρησιμοποιηθεί είτε αποκοπή 3 γύρων (όπως προαναφέρθηκε), είτε ένα 3 round betting model. Τα μοντέλα αυτά, υπολογίζουν την αρχική κατανομή πιθανότητας των χειρών των παικτών και έτσι δίνονται σωστές παράμετροι εισόδου στις ακολουθίες πονταρισμάτων.

Bucketing: Αυτή, είναι η πιο σημαντική μέθοδος και αυτή που επιφέρει τα καλύτερα αποτελέσματα. Στην μέθοδο αυτή, όλοι οι πιθανοί συνδυασμοί χειρών, χωρίζονται σε κλάσεις ή αλλιώς Buckets, ανάλογα με την στρατηγική τους ομοιότητα (δηλαδή παίζονται με παρόμοιο τρόπο). Μια mapping function, αποφασίζει πως θα ομαδοποιηθούν τα χέρια. Υπάρχουν απλές μονοδιάστατες συναρτήσεις bucketing, που υπολογίζουν την δύναμη ενός χεριού, αλλά οι κυρίαρχες συναρτήσεις, είναι οι διδιάστατες οι οποίες συνυπολογίζουν την δύναμη του χεριού, δεδομένου του roll out των μελλοντικών καρτών και το hand potential, δηλαδή την ικανότητα ενός χεριού, να εξελιχθεί σε δυνατό συνδυασμό ή σε κάποια ισχυρή στρατηγική ακόμα και αν έχει μικρή αρχική δύναμη. Καταλήγουμε πως με n διαθέσιμους buckets, χρησιμοποιούμε τους $n-1$ σε roll-out hand strength, με κάθε Bucket, να έχει διαφορετικό πλήθος χειρών (τα δυνατότερα έχουν λιγότερους σε σχέση με τα πιο μέτρια), και ανάλογα με το κάθε διαχωρισμό εκτελούνται οι αντίστοιχες στρατηγικές, περισσότερο ή λιγότερο επιθετικές/παθητικές. Το 1 bucket που απομένει, χρησιμοποιείται για χέρια που έχουν υψηλό potential (δυνατά draws) και χρησιμεύει στο να μπλοφάρεις (ή κυρίως να κάνεις semi-bluffs). Για να ολοκληρωθεί η αποτύπωση του παιχνιδιού, από την μέθοδο, θα πρέπει να συνυπολογιστούν και οι κόμβοι τύχης του δέντρου. Οι κόμβοι αυτοί, αναπαρίστανται σαν πιθανότητες μετάβασης μεταξύ γύρων, από το ένα bucket σε κάποιο άλλο, αυτό διότι μετά από κάθε γύρω προφανώς τα χέρια κερδίζουν ή χάνουν δύναμη σε σχέση με το board. Οι μεταβάσεις αυτές, υλοποιούνται από ένα δίκτυο μεταβάσεων (nxn) προς (nxn). Τέλος, με γραμμικό προγραμματισμό, παράγεται ένας πίνακας μικτών στρατηγικών για όλα τα πιθανά σενάρια το παιχνιδιού και το πρόγραμμα, αναζητά την αντίστοιχη κατάσταση βασισμένο σε ίδιο hand strength και potential και διαλέγει μια ενέργεια από τις μικτές στρατηγικές.

Πειραματικά αποτελέσματα & Σχολιασμός:

Εξάχθηκαν πειράματα μεταξύ των διαφορετικών πρακτόρων, αλλά και μεταξύ πρακτόρων και πραγματικών παικτών, με διαφορετικά επίπεδα εξοικείωσης με το πόκερ. Τα ποσοστά επιτυχίας, μετρώνται σε small blinds/hour. Μεταξύ πρακτόρων η μέθοδος Bucketing ήταν η καλύτερη, παρόλα αυτά, δεν παρατηρήθηκαν μεγάλες διακυμάνσεις στα ποσοστά. Αυτό διότι ο τρόπος παιξίματος που ακολουθούν οι πράκτορες GTO, δεν εγγυάται την νίκη, αλλά ότι δεν θα έχεις μακροχρόνια ήττα, αφού δεν θα παρεκκλίνεις από τις θεωρητικά σωστές αποφάσεις. Αντιθέτως, ενάντια των αληθινών παικτών, υπήρχαν ξεκάθαρα winning και losing παίκτες, είτε λιγότερο είτε περισσότερο και τα αποτελέσματα ήταν αρκετά πιο απρόβλεπτα. Γεγονός που είναι αναμμένο, εφόσον δεν υπάρχει μοντελοποίηση αντιπάλων στα προγράμματα και ο τρόπος αντιμετώπισης σε συμπεριφορές ενός πολύ κακού ή ενός πολύ καλού παίκτη μπορεί να παρεκκλίνει απόλυτα από τις ορθολογικές στρατηγικές. Μεγάλο ενδιαφέρον είχε πως πολλοί παίκτες ακόμα και μεσαίου-καλού επιπέδου, μην γνωρίζοντας πως δουλεύουν τα προγράμματα, ήταν πεπεισμένοι πως οι πράκτορες είχαν μοντελοποιήσει την συμπεριφορά τους και είχαν βρει την κατάλληλη στρατηγική για να παίξουν συγκεκριμένα εναντίον τους, γεγονός που δεν ισχύει. Παρόλα αυτά, ένας world class παίκτης, ο οποίος μετά από πολλές παρτίδες με τον καλύτερο πράκτορα (ακόμα και αν έχανε στην αρχή), κατάλαβε πως ο πράκτορας ακολουθούσε την ίδια τακτική χωρίς να έχει αλλαγές στρατηγικής. Έτσι, αλλάζοντας τον τρόπο παιξίματος του στα περισσότερα σενάρια σε πολύ λιγότερο επιθετικό (αφού ο τρόπος που εκφόβιζε τους άλλους παίκτες, δεν έπιανε στον πράκτορα που δεν έχει το αίσθημα του φόβου), και περιμένοντας απλά να εκμεταλλευτεί έντονα τις καταστάσεις που έχει ο ίδιος ξεκάθαρο πλεονέκτημα, κατάφερε να νικήσει με διαφορά τον πράκτορα. Τέλος, ένας καταλυτικός παράγοντας που παρατηρήθηκε να δημιουργεί μεγάλες αλλαγές στα ποσοστά νίκης, είναι ο αυτός της ανθρώπινης ψυχολογίας. Οι παίκτες, και ιδικά οι λιγότερο έμπειροι, σε σενάρια κακοτυχίας (badbeat, lucky escape), έσπερναν απερίσκεπτες αποφάσεις και αμελούσαν την λογική, λόγω της έντονης συναισθηματικής τους αστάθειας (tilt). Επομένως στις περιπτώσεις αυτές έχαναν EV σε εξωτερικούς παράγοντες, το οποίο δεν μπορούν να ισοσταθμίσουν κάπως, αφού σε αντίθεση με αυτούς με τα ρομπότ, δεν έχουν συναισθήματα.

Συμπεράσματα & Βελτιώσεις:

Συμπεραίνουμε πως μια GTO στρατηγική, δεν είναι αρκετή για να νικήσει κάποιον κορυφαίο παίκτη. Παρόλα αυτά, όταν εισάγονται στον πράκτορα μοντελοποίηση αντιπάλων και διαρκείς αλλαγές στρατηγικών, τότε μπορεί το μακροχρόνιο αποτέλεσμα να είναι πολύ κοντά στο ανίκητο. Μια ιδέα για κάποιον τέτοιου είδους βελτίωση, θα ήταν ο πράκτορας στην αρχή του παιχνιδιού με έναν αντίπαλο, να ξεκινάει παίζοντας GTO, και προχωρώντας κατά την διάρκεια, να ανανεώνει με κάποιους δείκτες (3bet%, fold%, vrip, bluff%) τα δεδομένα του και να γίνεται αντίστοιχα πιο παθητικός ή επιθετικός, να αλλάζει τις κατανομές των χειρών που παίζει σε πιο σφιχτές ή πιο ανοιχτές και να κάνει μεταβάσεις των χειρών μεταξύ των buckets. Τέλος, θα μπορούσε ακόμα και να προσπαθήσει να εκμεταλλευτεί την ανθρώπινη αδυναμία που αναφέρθηκε πριν, με το να γίνεται πιο exploitative-επιθετικός, όταν δέχεται σαν είσοδο κάποια ακολουθία γεγονότων, που έχει μεγάλη πιθανότητα να επηρεάσει συναισθηματικά τον αντίπαλο.