

NEŠA LĪDZSVARA UN Oponentu EKSPLUATĀCIJAS REALIZĀCIJA AUTOMATISKĀ POKERA SPĒLĒTĀJĀ

MAĢISTRA KURSA DARBS

Autors: Romāns Ledeņovs
Studenta apl. nr. r11034
Darba vadītājs: Dr. Dat. Dmitrijs Kravčenko
Rīgā 2017

Pokera boti un algoritmi

Kursā darbā tiek aprakstītas visi maģistra darba nepieciešami teorētiskie jautājumi pokerā un spēļu teorijā, kā arī citu autoru darbu izpēte, kas attiecas uz izvēlēto tēmu.

Pokers ir nulles summas kāršu spēle ar nepilnu informāciju. Tam tiek izmantota 52 kāršu kāva. Spēles pamatā ir dažas kāršu kombinācijas, likmes un solīšana.

Vispopulārākie pokera paveidi ir: Tehasa bezlimita holdems (*NLHE*), Tehasa limita holdems (*LHE*) un banka limita Omaha (*PLO*) un citi. Pokera variācijas atšķiras ar izdalīto kāršu skaitu, kombinācijām, likmes lieluma un spēlētāju skaitu. Bezlimita spēlēs ir iespēja bez ierobežojuma pacelt likmi.

GTO (game theory optimal), jeb Neša līdzsvara stratēģija ir tāda stratēģija, kuras pielietošana garantē lielāku iespējamu matemātisku cerību. Par vienkāršu piemēru no šīm varētu kalpot Neša līdzsvara atrašana spēlē "Akmens, šķēres, papīrit's", kur ir 3 iespējamās tīras stratēģijas: uzradīt "akmens", "šķēres" vai "papīrit's". Pie tam "papīrit's" > "akmens", "šķēres" > "papīrit's" un "akmens" > "šķēres". Neša līdzsvars šai spēlei ir pielietot katru no šīm trim tīrām stratēģijām ar varbūtību 1/3. Mūsu oponentam nebūs iespējas pielāgoties mūsu jauktai stratēģijai lai vinnētu vairāk par 0. Saka ka šādu stratēģiju nevar ekspluatēt. Piemēram, ja mēs visu laiku izmantotu "šķēres" tad acīmredzami ka mūsu oponentam labāka stratēģija būtu izmantot stratēģiju akmens visu laiku. Tomēr, pēc neliela laika mēs varētu pamanām, ka zaudējam visu laiku un tad sākt mainīt stratēģiju, izmantojot stratēģiju "papīritis" biežāk. Pēc šīs izmaiņas oponentam būs lietderīgi biežāk izmantot "šķēres" un tā tālāk. Tādā veidā mēs pakāpeniski nonākam cik vien tuvu Neša līdzsvaram.

Neša līdzsvaru relatīvi nelielam spēlēm (līdz 10^8 stāvokļiem) ar nepilnu informāciju var atrast, izmantojot lineāras programmēšanas metodes. Spēlēm ar lielāko stāvokļu skaitu parasti izmanto iteratīvas metodes, kas konverģē uz līdzsvaru. Populārāks no tiem ir CFR (counterfactual regret) minimizācija. Tajā pamatā ir zuduma minimizācijas koncepts. Katrā iterācijā mēs izrēķinām vērtību, kuru mēs zaudējam attiecība pret mūsu izvēli.

Pierādīts, ka pēc pietiekoša iterāciju skaita mēs pakāpeniski nonākam tuvāk pie Neša līdzsvarā.

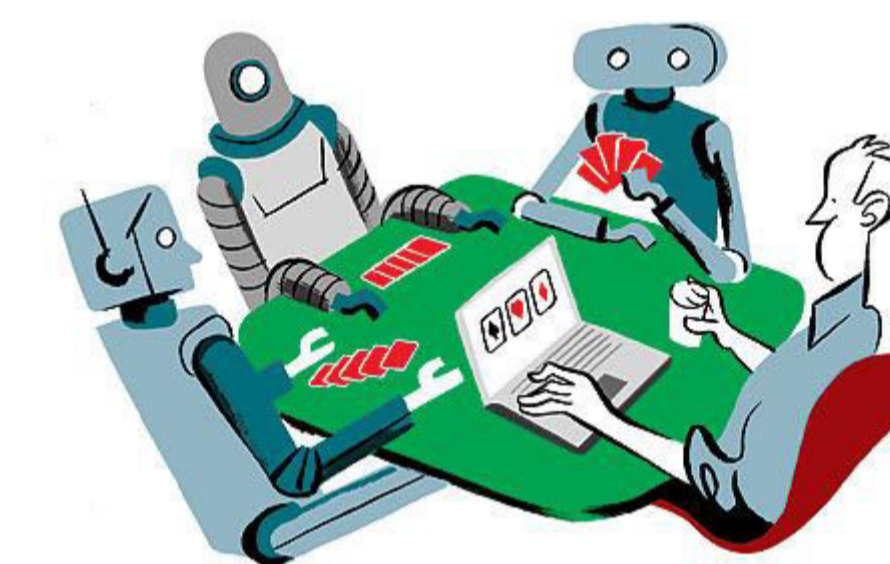
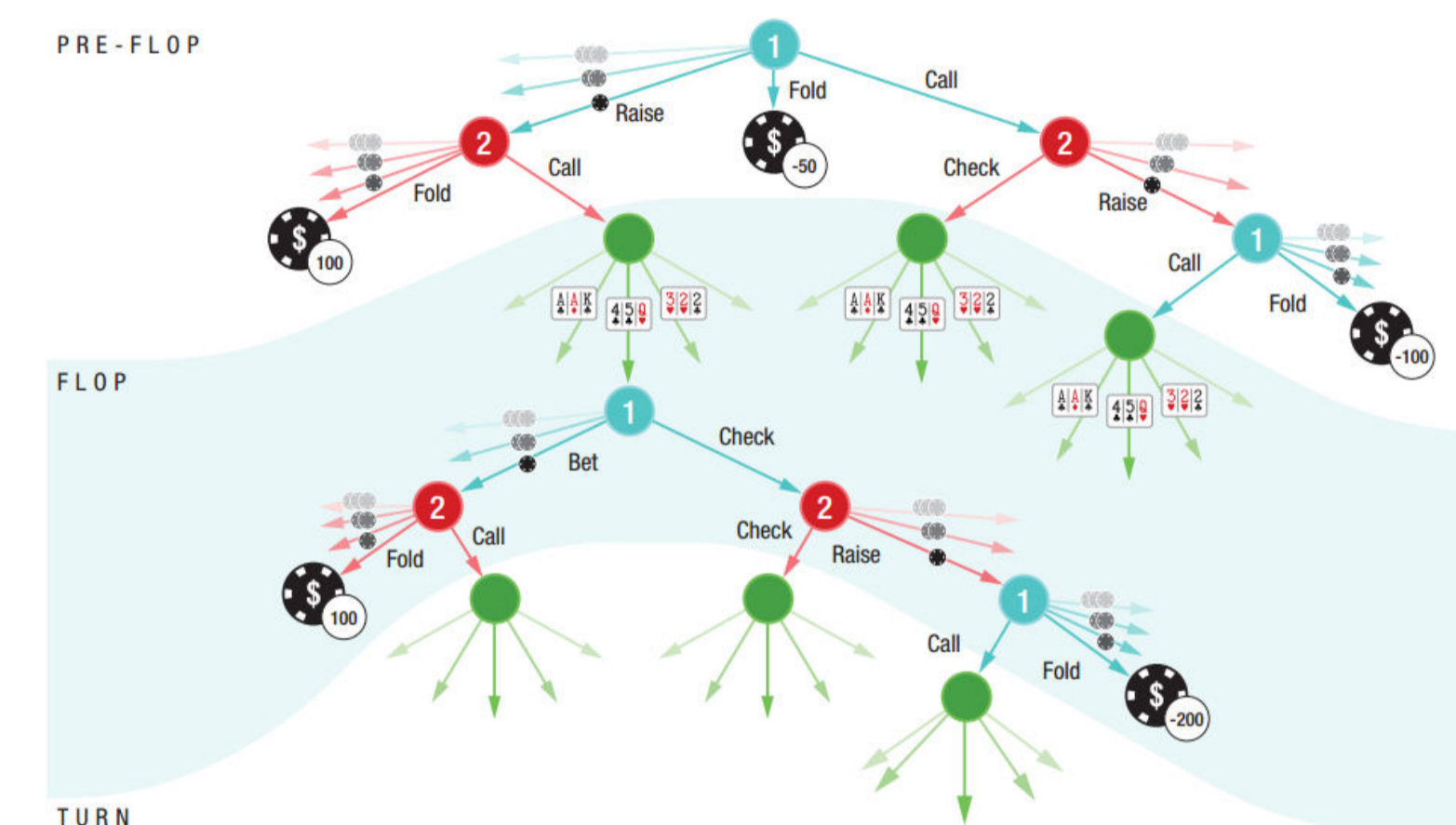
CEPHEUS

2015. gadā janvārī Albertas Universitātes pētnieciska grupa pasludināja limita Holdemu (10^{12} spēles stāvokļu) "vāji atrisinātu". Tas nozīmē, ka ir atrasta stratēģijā, kura ir tik tuvu Neša līdzsvaram, ka maksimāli iespējama šīs stratēģijas ekspluatācija varētu vinnēt 0.000986 BB/spēle. Šī vērtība praktiski nozīme, ka vajadzēs ļoti liela roku izlase (vairāk par miljonu), lai ar diezgan lielo statistisku nozīmību apgalvot, ka šī kontrstratēģija tiešam ir labāka.

CLAUDICO un LIBRATUS

2015. gadā aprīlī notika pasaules pirmās nopietnas (80000 rokas nospēlētas) bezlimita holdema (10^{12} spēles stāvokļu) pokerbotu sacensības pret pokera profesionāļiem. *Carnegie Mellon* universitātes grupas bots *Claudico* nospēlēja 20000 rokas pret katru no 4 profesionāļiem. Pēc 13 dienām cilvēki kopā vinnēja 731,000\$, kas tomēr neļauj ar 95% (tikai ar 90%) statistisku nozīmību secināt, ka cilvēki ir tiešam nospēlējuši labāk par datoru.

Pašlaik notiek sacensību jauna kārta, uzlabots pokerbots *Libratus* vinnē par cilvēkiem ar diezgan pārliecinošo priekšrocību, tomēr nav skaidrs, ka 120000 roku izlase ir pietiekoša nopietnam secinājumam [2].



Secinājumi:

Apkopot literatūru, varam secināt, ka ir paveikts liels darbs optimālo stratēģiju meklēšanai, un tuvāka laikā labākie boti pārliecinoši apspēlēs pokera profesionāļus.

Tomēr viens pokera aspekts, kur cilvēka smadzenes ir joprojām piemērotākas par datoru, ir vāju oponentu ekspluatācija.

Darbā mērķis ir mēģināt uzlabot pastāvošas metodes, lai tās veiksmīgāk imitētu labāku spēlētāju uzvedību, meklējot bilanci starp ekspluatējošo un GTO stratēģijām, lai maksimizētu vinnestu pret vājākiem spēlētājiem.

Avoti:

1. Michael Bowling, Neil Burch, Michael Johanson, Oskari Tammelin, "Heads-up Limit Hold'em Poker is Solved", Department of Computing Science, University of Alberta.
2. Noam Brown, Christian Kroer, Tuomas Sandholm, "Dynamic Thresholding and Pruning for Regret Minimization", Carnegie Mellon University, Computer Science department.