Czy uczenie się powoduje chaos komunikacyjny?*

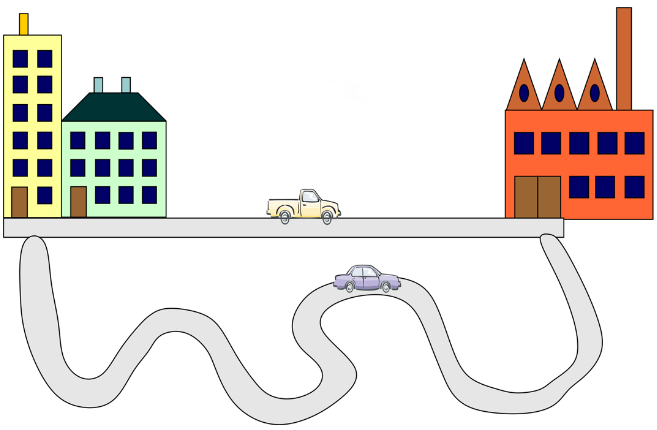
*Grzegorz KOSIOROWSKI***, Kraków*

Pglądowo i jego kłopoty

Jest to tekst związany z odczytem wygłoszonym na LXV Szkole Matematyki Poglądowej, *Uogólnienia*, Siedlce, sierpień 2022.

Redakcja

Nie tak dawno temu, za górami, za lasami stało sobie miasteczko zwane Poglądowem. Ze względu na szczególną rzeźbę terenu, miasteczko dzieliło się na dwie dzielnice: w jednej znajdowały się wszelkie lokalne przedsiębiorstwa, fabryki i większość punktów usługowych oraz urzędów (dzielnica przemysłowa), a w drugiej mieszkania większości poglądowian (dzielnica mieszkalna). Każdego poranka znacząca liczba poglądowian dojeżdżała samochodami (niestety transport publiczny w Poglądowie jest bardzo nieefektywny) do pracy ze swoich domów. Oczywiście, między dzielnicą mieszkalną a przemysłową wybudowano drogę dobrej jakości, która pozwala dojechać na miejsce bardzo szybko... pod warunkiem, że ta droga jest pusta. Oczywiście, w godzinach dojazdu do pracy trudno na to liczyć, dlatego na tej drodze tworzą się korki, które marnują czas i budzą irytację kierowców. Dlatego kuszącą alternatywą jest objazd – trasa jest dłuższa, nienajlepszej jakości i poza godzinami szczytu jedzie się nią znacznie dłużej niż główną drogą. Ale jeśli przygniatająca większość poglądowian będzie stać w korku na głównej drodze, objazdem można na miejsce dotrzeć szybciej.



Rys. 1. Poglądowo i jego drogi dojazdowe.

Dlatego co rano każdy z kierowców z Poglądowa musi podjąć decyzję: którą drogą dzisiaj wyruszy do pracy, by dotrzeć do dzielnicy przemysłowej jak najszybciej. W swoich decyzjach oczywiście kieruje się jakością drogi, ale też oczekiwaniami dotyczącymi liczby pozostałych mieszkańców podróżujących jedną, bądź drugą drogą. Gdyby jakiś poglądowianin zignorował prognozowanie zachowań pozostałych, to na pewno wyruszyłby drogą krótszą i wygodniejszą.

Ale jeśli spodziewa się, że wszyscy pozostali podejmą taką decyzję, to może spróbować ich „przechytrzyć” i pojechać wolnym od samochodów objazdem zamiast tkwić w gigantycznym korku. Nie może jednak wykluczyć, że wszyscy inni o tym pomyślą i zator tak naprawdę uformuje się na trasie objazdowej! Czy potrafimy jakoś wymodelować proces decyzyjny poglądowian i przewidzieć ich długoterminowe wybory? Czy uformuje się jakaś „równowaga” polegająca na mniej więcej stałej liczbie pojazdów poruszających się każdą z tras w kolejne dni? A może zatłoczenie dróg będzie zmieniać się okresowo lub w jakimś innym, przewidywalnym schemacie? Czy też może będziemy mieć do czynienia z kompletnie nieprzewidywalnymi, niestabilnymi zachowaniami?

Sytuacjami podobnymi do dylematu mieszkańców Poglądowa zajmuje się teoria gier. Bez wchodzenia w techniczne detale, przez grę rozumiemy matematyczny model sytuacji, w której wielu jej uczestników (zwanych graczami) wykonuje swoje działania, które z kolei mają wpływ na końcowy efekt wysiłków każdego z nich. Kłopoty mieszkańców Poglądowa są dla mnie pretekstem przedstawienia stosunkowo nowatorskiego podejścia do teorii gier, znanego jako algorytmiczna (lub ewolucyjna) teoria gier. Polega ono na założeniu, że taką grę rozgrywa się wielokrotnie i dopuszczamy „uczenie się” graczy, czyli aktualizowanie strategii podejmowania wyborów na podstawie tego, co wydarzyło się do tej pory.
Na przykład, kierowca z Poglądowa może zdecydować się na wybór takiej drogi, która oznaczała krótszy czas dojazdu poprzedniego dnia – albo wręcz przeciwnie, spodziewając się, że na tę właśnie drogę „przeskoczy” wystarczająco wielu współdojeżdżających. Jak zobaczymy, nawet przy bardzo prostych założeniach, rezultaty takiego modelowania mogą się okazać bardzo skomplikowane.

**Uniwersytet Ekonomiczny w Krakowie, grzegorz.kosiorowski@uek.krakow.pl

*https://doi.org/10.34739/mp.2024.09.02

Gry zagęszczeniowe

Spróbujmy sformalizować, o jakiej grze w sensie matematycznym mówimy, analizując dylemat kierowców z Poglądowa. Najpierw musimy ustalić bazowe elementy takiej gry:

- Zbiór graczy: w naszej sytuacji zakładamy, że graczy jest tak wielu, że praktycznie ich liczbę można przybliżać zbiorem ciągłym mocy continuum (czyli $[0, 1]$). Ponadto, nie mamy żadnych informacji wyróżniających któregokolwiek z graczy, więc zakładamy, że są oni symetryczni: dysponują tymi samymi zasobami, takim samym dostępem do informacji i mają takie same preferencje.
- Zbiór strategii, czyli zasad wyborów możliwych działań dokonywanych przez graczy. Dla Poglądowa są to dwie *czyste* strategie: wybór pierwszej lub drugiej drogi dojazdu.
- Funkcję wypłaty (kosztu), która przedstawi wielkość kosztu poniesionego przez każdego z graczy w postaci czasu poświęconego na dojazd w zależności od tego jaką strategię wybierze każdy z nich. Dokładne postaci funkcji wypłat omówimy w dalszej części artykułu. Oczywiście, zakładamy, że uczestnicy gry chcą ponieść jak najmniejszy koszt swoich działań, czyli zużyć na dojazd jak najmniej czasu.

W teorii gier często dopuszcza się strategie mieszane, czyli takie, w których każdej opcji przypisuje się pewne prawdopodobieństwo jej wykonania. Jednakże, w przypadku Poglądowa rozważanie strategii mieszanych nie wnosi do gry nic nowego. Po pierwsze, modelujemy sytuację, w której obserwujemy tylko działania graczy, a więc realizacje strategii: w rzeczywistości, nigdy nie zaobserwujemy, że strategią danego mieszkańca był wybór pierwszej drogi z prawdopodobieństwem p, a drugiej z prawdopodobieństwem $1 - p$, tylko to, którą drogą faktycznie pojechał. Drugi, bardziej matematyczny powód, to liczebność zbioru graczy. Jeśli jakiś jego podzbiór o mierze niezerowej w danym momencie preferowałby strategię mieszaną, to można tę preferencję bez utraty informacji o modelu zastąpić wartością oczekiwaną realizacji takiej strategii mieszanej. Na przykład, jeśli pewien podzbiór zbioru graczy wybierałaby pierwszą drogę z prawdopodobieństwem 75%, to równie dobrze można założyć, że po prostu 3/4 uczestników gry z tego zbioru wybiera pierwszą drogę, a reszta drugą.

Analizowana przez nas sytuacja jest przykładem gry zagęszczeniowej (*congestion game*). Takie gry charakteryzują się funkcją kosztu użycia danej strategii rosnącą ze względu na ilość jej użytkowników. W naturalny sposób takie gry modelują zagadnienia związane z transportem (takie jak dylemat poglądowian): im więcej pojazdów użytkuje daną drogę (czy jest to ulica, czy tor kolejowy, czy też ścieżka rowerowa), tym wolniej średnio mogą się przemieszczać i tym większy jest czasowy koszt dotarcia do celu. Ale nie jest to jedyny scenariusz, w którym gry zagęszczeniowe mają zastosowania. Podobnymi efektami charakteryzują się gry rynkowe ze względu na mechanizm popytu: im więcej osób chce kupić dane dobro na rynku (np. akcje pewnej firmy na giełdzie), tym większy dają bodziec sprzedającym do podniesienia ceny tego dobra, co oczywiście zwiększa koszt jego zakupu. Podobne zasady występują również w biologicznych grach populacyjnych w postaci konkurencji o zasoby: na przykład, im więcej organizmów wykorzystuje to samo źródło żywności, tym trudniej jest pozyskać pożywienie z tego źródła. Generalnie, wszędzie, gdzie różni uczestnicy gry próbują skorzystać ze skończonych zasobów, gry zagęszczeniowe są idealnym narzędziem modelowania zachowań graczy lub też pozwalają nimi sterować (przykładem tego drugiego podejścia są algorytmy inwestowania na giełdach lub przyporządkowywania przekaźników do obsługi telefonów komórkowych bez przeciążania sieci).

Jak zatem sformułować funkcje kosztów dla mieszkańców Poglądowa, by otrzymać grę zagęszczeniową o interesujących rozwiązaniach? Okazuje się, że wystarczy najprostsze możliwe podejście: rozważenie kosztów proporcjonalnych do liczby użytkowników obydwu dróg podczas danej rozgrywki. Przez $x \in [0, 1]$ będziemy oznaczać ułamek populacji wybierający pierwszą drogę podczas dojazdu do pracy (oczywiście, wtedy $1 - x$ wybiera drogę drugą). Wtedy czas (czyli koszt) przejazdu pierwszą drogą (czyli wyboru pierwszej strategii) wynosi:

$$c_1(x) = \alpha x,$$

a czas (czyli koszt) przejazdu drugą drogą (czyli wyboru drugiej strategii):

$$c_2(x) = \beta(1 - x),$$

dla pewnych $\alpha, \beta \in \mathbb{R}^+$. Dodatkowo zakładamy, że $\alpha \neq \beta$, gdyż w innym przypadku drogi stają się identyczne pod względem funkcji kosztu i całe zagadnienie przestaje być szczególnie interesujące. Oczywiście, sytuację, w której pierwsza droga jest generalnie szybsza przy tej samej liczbie użytkowników modeluje założenie $\beta > \alpha$.

Przedstawiliśmy już wszystkie elementy gry: czas ją rozwiązać. Ale co rozumiemy przez rozwiązanie gry? W klasycznej teorii gier, za rozwiązanie uważa się stan równowagi (w skrócie: równowagę), czyli zestaw strategii wybieranych przez wszystkich graczy, taki, że żadnemu z nich, a raczej żadnej grupie graczy (gdyż przy założeniu istnienia continuum graczy, zmiany strategii pojedynczych uczestników gry nie mają znaczenia) nie opłaca się zmienić swojej strategii przy założeniu, że inni pozostaną przy swoich wyborach.

Definicja 1. Równowagą *dla rozważanej gry nazywamy taką proporcję $x_0 \in [0,1]$ graczy wybierających pierwszą strategię, że dla każdego $\varepsilon > 0$:*

$$c_1(x_0+\varepsilon) \geqslant c_2(x_0), \text{ gdy } x_0+\varepsilon \leqslant 1; c_2(x_0-\varepsilon) \geqslant c_1(x_0) \text{ gdy } x_0-\varepsilon \geqslant 0.$$

Założenie z opisu sytuacji Poglądowa, że wybór pierwszej drogi oznacza mniejszy koszt przy tej samej liczbie użytkowników obydwu dróg przekłada się na warunek $b > \frac{1}{2}$.

Łatwo sprawdzić, że dla naszej gry jedyną równowagą będzie $b = \frac{\beta}{\alpha+\beta}$ i w tej równowadze czas podróży każdą z dróg jest taki sam:

$$c_1(b) = c_2(b) = \frac{\alpha\beta}{\alpha+\beta}.$$

Faktycznie zatem, jeśli właśnie część mieszkańców Poglądowa równa b będzie podróżować pierwszą drogą i będą oni świadomi, że taki właśnie jest stan rzeczywistości, to równowaga będzie się utrzymywać z niewymuszonej woli graczy.

Czy to już koniec historii? Zdecydowanie nie! Dla tego typu zagadnień, klasyczna teoria gier, a zwłaszcza rozwiązanie w formie równowagi niekoniecznie jest najlepszym podejściem. O ile faktycznie stan równowagi, jeśli już raz zajdzie, zapewne zostanie utrzymany, to warto się zastanowić, dlaczego i w jaki sposób mieszkańcy Poglądowa mogą do tego stanu dojść.

W najbardziej klasycznym ujęciu, powodem jest założenie o racjonalności, a właściwie superracjonalności graczy. Zakłada się, że gracze znają wszystkie zasady gry (w szczególności, dokładne sformułowanie funkcji kosztu), zgadzają się z jej celami (więc celem każdego w każdej sytuacji jest jedynie minimalizacja kosztu), potrafią idealnie realizować zaplanowane przez siebie strategie (np. nigdy nie skręcą w złą drogę ani nie zepsuje się im samochód) i dodatkowo są pewni, że pozostali gracze mają takie samo podejście do sytuacji i te same umiejętności. W szczególności, potrafią obliczyć równowagę, zrealizować ją w praktyce i mieć świadomość, że jest ona realizowana przez innych. O ile takie założenie jest dogodne przy rozpatrywaniu prostych, abstrakcyjnych gier, to jest bardzo słusznie krytykowane jako podstawa rozwiązywania rzeczywistych problemów. Mieszkańcy rzeczywistych miast z całą pewnością nie są w stanie spełnić tych wszystkich założeń. Warto też wspomnieć, że równowaga może być trudna do obliczenia w przypadku skomplikowanych gier lub też prowadzić do powstawania układów stabilnych, lecz nieoptymalnych dla wszystkich zainteresowanych stron.

Takim przypadkiem jest słynny dylemat więźnia lub jego populacyjne ujęcie znane jako tragedia wspólnego pastwiska, gdzie rozwiązanie równowagowe przewiduje kompletne zużycie ograniczonego zasobu przez jego użytkowników zamiast współużytkowania go w sposób zrównoważony.

Jednakże klasyczna teoria gier sprawdza się w wyjaśnieniu powstawania wielu społecznych konwencji, mimo braku superracjonalnych społeczeństw. Powodem jest druga możliwość osiągnięcia równowagi przez populację. Jeśli sytuacja modelowana przez grę powtarza się wielokrotnie, nawet kompletnie nieracjonalni gracze mogą dowiedzieć się o tej równowadze z doświadczenia: albo ucząc się bezpośrednio, dzięki rozegraniu takiej interakcji wielokrotnie, albo dowiadując się o takim rozwiązaniu od innych, którzy w takiej sytuacji byli.

Powstaje pytanie: czy na pewno wielokrotne interakcje i samodzielne nabieranie doświadczenia przez graczy wystarczy do osiągnięcia równowagi? Czy w tak prostej grze o liniowej funkcji kosztu może zdarzyć się coś bardziej skomplikowanego? By na to odpowiedzieć, musimy się odwołać do ewolucyjnej (algorytmicznej) teorii gier, której matematyczne podstawy oparte są na teorii układów dynamicznych.

Gry ewolucyjne i proste algorytmy uczenia się

W grach ewolucyjnych odchodzimy od analizy statycznego stanu pojedynczej gry na rzecz badania sekwencji kolejno rozgrywanych (iterowanych) gier. Wynik każdej z takich gier wpływa na zachowanie graczy w kolejnych rozgrywkach. Podejście ewolucyjne porzuca założenie o superracjonalności graczy. Uczestnicy nie są już wszechwiedzący co do preferencji pozostałych graczy, ani co do dokładnych funkcji kosztów. Swoją wiedzę o stanie świata i w konsekwencji swój proces decyzyjny wyprowadzają z historii dotychchczas rozegranych gier, wiedzy o efektach dotychchczasowych działań i porównaniu ich z działaniami współgraczy i tym, jakie koszty oni ponieśli. Innymi słowy, gracze *się uczą* i za pomocą pewnego algorytmu starają się przełożyć efekty tego uczenia na lepsze wyniki w przyszłości.

W wypadku sytuacji panującej w Poglądowie, możemy założyć, że początkowy wybór dróg do pracy był mniej więcej losowy, a następnie, każdego dnia nasza gra w wybór trasy jest powtarzana, a kierowcy podejmują decyzję w oparciu o to, co się wydarzyło dawniej, w szczególności, jak długo im zajmował dojazd w porównaniu z sąsiadami. W rezultacie otrzymujemy ciąg $(x_n)_{n=0}^{\infty} \in [0,1]^{\mathbb{N}}$ którego wyrazy opisują, jaka część poglądowian wybrała pierwszą drogę każdego dnia.

W matematycznym ujęciu, gry ewolucyjne badają pewien układ dynamiczny. Dla naszej gry z Poglądowa stan początkowy $x_0 \in (0,1)$ jest ułamkiem liczby mieszkańców, którzy podczas pierwszej „rozgrywki" (czyli „zerowego" dnia dojazdu do pracy) wybrali pierwszą z dróg. Najważniejszym składnikiem naszego układu jest funkcja $f : [0,1] \to [0,1]$ (odpowiednio regularna) definiująca zmianę strategii populacji poglądowian z dnia na dzień (dokładniej: jak zmienia się udział użytkowników pierwszej drogi z dnia na dzień). Ciąg decyzji poglądowian $(x_n)_{n=0}^{\infty}$ otrzymujemy indukcyjnie z równania dynamiki:

Technicznie możliwe jest rozpoczęcie ze stanu $x_0 = 1$ lub $x_0 = 0$, jednak wtedy wszyscy uczestnicy gry zastosowali tę samą strategię i nie mają się od kogo nauczyć, jakie efekty dawało inne podejście – dlatego takie stany początkowe nie są dla nas interesujące.

$$x_{n+1} = f(x_n).$$

Takie ujęcie gry przedstawia ją jako dyskretny układ dynamiczny na zbiorze $[0,1]$ i pozwala nam użyć wszelkich narzędzi, które teoria układów dynamicznych wytworzyła. Jednak zanim z nich skorzystamy, musimy sprecyzować naszą dynamikę, czyli funkcję f. W algorytmicznej teorii gier ta funkcja nazywa się algorytmem uczenia się i może być wybrana na wiele sposobów. Najprostszym jest algorytm najlepszej odpowiedzi (*best response*): każdy wybiera strategię, która dałaby mu najlepszy wynik w ostatniej rozgrywce, czyli najkrótszy czas dojazdu poprzedniego dnia. Niestety, prostota w tym wypadku nie wystarcza: algorytm powinien dawać rezultaty, które modelują realistycznie zachowania graczy (wtedy może być użyty do wyjaśniania i prognozowania różnych zjawisk ze świata rzeczywistego) lub generować działania, które są efektywne dla społeczności (wtedy można go zastosować jako procedurę sterującą danym systemem we właściwym kierunku). Ewentualnie moglibyśmy się zainteresować algorytmem nie spełniającym tych kryteriów, gdyby chociaż był podstawą matematycznie interesującej dynamiki. Ale dynamika algorytmu najlepszej odpowiedzi jest bardzo prosta i bardzo nieefektywna: wszyscy mieszkańcy Poglądowa od pierwszego dnia wybierają za każdym razem tę samą drogę, a konkretnie taką, na której czas przejazdu poprzedniego dnia był mniejszy:

$$f_{BR}(x) = \begin{cases} 1, \text{ gdy } \alpha x < \beta(1-x) \\ b, \text{ gdy } \alpha x = \beta(1-x) \\ 0, \text{ gdy } \alpha x > \beta(1-x), \end{cases}$$

co łatwo przeliczyć na:

$$f_{BR}(x) = \begin{cases} 1, \text{ gdy } x < b \\ b, \text{ gdy } x = b \\ 0, \text{ gdy } x > b. \end{cases}$$

W rezultacie, jedynymi stabilnymi długoterminowymi trajektoriami są $(\ldots, 0, 1, 0, 1, 0, \ldots)$, opisujące ciągi wydarzeń w których na przemian wszyscy mieszkańcy Poglądowa jadą pierwszą drogą i wszyscy mieszkańcy Poglądowa jadą drugą drogą. Oczywiście, jest to zachowanie kompletnie nieefektywne, ale też nierealistyczne: można przypuszczać, że w rzeczywistości co najmniej część z nich zauważyłaby taki schemat i próbowała się z niego wyłamać.

Nieco bardziej skomplikowanym algorytmem uczenia się jest Follow-the-Leader (FTL). Bierze on pod uwagę nie tylko to, co się działo w ostatniej iteracji gry (czyli poprzedniego dnia), ale też w całej historii rozgrywania gry. Każdy gracz $(n+1)$-wszego dnia wybiera swoją drogę pierwszą z prawdopodobieństwem x_{n+1}, a drugą z prawdopodobieństwem $(1 - x_{n+1})$, gdzie x_{n+1} jest wybrany w taki sposób, że wartość oczekiwana czasu podróży z tak wybranej strategii mieszanej stosowanej do dni $0, 1, \ldots, n$ jest jak najmniejsza. Oczywiście, w rezultacie x_{n+1} populacji Poglądowa podróżuje $(n+1)$-wszego dnia pierwszą drogą.

Mówimy, że taka dynamika minimalizuje „żal" (*regret*) z wyboru danej strategii w czasie całej historii gry, czyli wyrażenie $\sum_{j \leqslant n}[c_1(x_j) \cdot x_{n+1} + c_2(x_j) \cdot (1 - x_{n+1})]$. Dla naszych funkcji kosztu mamy:

$$x_{n+1} = argmin_{x \in [0,1]} \sum_{j \leqslant n} \left[\alpha x_j \cdot x + \beta(1 - x_j) \cdot (1 - x)\right]$$

Co ciekawe faktycznie funkcja $f_{FTL}(x_n) = x_{n+1}$ jest dobrze zdefiniowana, mimo, że formalnie wzór na x_{n+1} zawiera wcześniejsze wyrazy ciągu (mechanizm za tym stojący prześledzimy na przykładzie kolejnego uogólnienia algorytmu uczenia się). Jednak niestety dynamika generowana przez FTL jest niemal równie nieciekawa jak w wypadku dynamiki najlepszej odpowiedzi. Jedyne stabilne „przyszłości" prognozowane przez dynamikę, to nadal takie, w których wszyscy kierowcy codziennie jeżdżą tą samą drogą – tym razem nie tą, która była korzystniejsza poprzedniego dnia, ale taką, która była średnio korzystniejsza w ciągu wszystkich poprzednich dni. Zmienia się jedynie regularność oscylacji pomiędzy drogami: niekoniecznie do „przeskoków" z jednej drogi na drugą dochodzi każdego dnia.

Na szczęście, te najprostsze algorytmy uczenia się da się uogólnić do bardziej realistycznej i interesującej postaci.

Algorytm FTRL i dynamika dla Poglądowa

Algorytm uczenia się znany jako Follow-the-Regularized-Leader (FTRL) opiera się na znanym z psychologii (np.[Yan]) spostrzeżeniu, że ludzie nie zmieniają swojego zachowania w procesie uczenia się gwałtownie, pod wpływem pojedynczej nowej informacji, lecz stopniowo: poprzez kompromis między pierwotnym podejściem, a lepszym rozwiązaniem, o którym się dowiadują. Takie podejście sprawdza się również, gdy algorytm uczenia się chcemy zastosować do wyznaczenia optymalnego rozwiązania jakiegoś problemu: brak nagłych przeskoków w kierunku lokalnych ekstremów potencjału często ułatwia osiągnięcie globalnych ekstremów efektywności.

Poprawę konstrukcji w algorytmie uzyskujemy przez dodanie do minimalizowanego wyrażenia „żalu" dodatkowego składnika – tak zwanego *regularyzatora*, który wyraża preferencje przyjmowane przez grupę graczy „a priori", gdyby nie zachodził proces uczenia się i jednocześnie zniechęca do szybkiego przyjmowania rozwiązań skrajnych. Formalnie, będziemy się opierać na następującej optymalizacji:

$$x_{n+1} = argmin_{x \in (0,1)} \left[\varepsilon \cdot \sum_{j \leqslant n} \big(c_1(x_j) \cdot x + c_2(x_j) \cdot (1 - x)\big) + r(x)\right], \tag{1}$$

gdzie $r : (0, 1) \to \mathbb{R}$ pełni funkcję wspomnianego regularyzatora. Bardzo istotny jest też współczynnik $\varepsilon \in [0, \infty)$, który mierzy prędkość uczenia się: dla $\varepsilon = 0$ gracze w ogóle się nie uczą, a o wybranej strategii decyduje jedynie minimalizacja regularyzatora. Z kolei im większa wartość ε, tym większe znaczenie dla wyboru strategii ma jej aktualizacja o informacje z historii gry. Można powiedzieć, że pierwsza część optymalizowanego wyrażenia odpowiada za adaptację graczy do warunków świata, druga – zachęca do eksploracji różnych możliwości.

Jakie założenia musi spełniać regularyzator? Okazuje się, że wcale nie aż takie silne. Wystarczy nam, żeby był symetryczny (tj. $r(1-x) = r(x)$), wypukły ($r''(x) > 0$) oraz „stromy" ($\lim_{x \to 1^-} r'(x) = \infty$ lub równoważnie $\lim_{x \to 0^+} r'(x) = -\infty$). Pierwsze założenie wynika z faktu, że nie chcemy by „a priori" któreś rozwiązanie było preferowane przez regularyzator, drugie gwarantuje, że sam regularyzator można zminimalizować, a trzecie pozwala nam zagwarantować, że maksimum będzie się znajdować odpowiednio daleko od końców przedziału, co pozwoli uniknąć trywializowania dynamiki, z którym spotkaliśmy się w przypadku dwóch poprzednio analizowanych algorytmów.

Algorytm FTRL używający tego regularyzatora znany jest też pod nazwą *Multiplicative Weight Update* (MWU).

Najbardziej powszechnie wykorzystywanym regularyzatorem spełniającym te warunki jest entropia Shannona rozkładu prawdopodobieństwa $(x, 1-x)$, czyli:

$$r_{Sh}(x) = x \log x + (1-x)\log(1-x).$$

Założenia regularyzatora spełniają generalnie entropie Arimoto, czyli funkcje postaci $r(x) = -\eta(x) - \eta(1-x)$, gdzie η jest funkcją wklęsłą taką, że $\lim_{x\to 0^+} \eta'(x) = +\infty$. Innym przykładem regularyzatora jest pochodząca od entropii Renyi'ego funkcja

$$r_{Re} = \frac{1}{q-1}\log(x^q + (1-x)^q); q \in (0,1).$$

Fakt, że algorytm FTRL w ogóle generuje dynamikę może być zaskakujący, gdyż wzór (1) wskazuje, że x_{n+1} zależy nie tylko od x_n, ale też od wcześniejszych wyrazów ciągu $(x_n)_{n=0}^{\infty}$. Jednakże, gdy zapiszemy wzór (1) z podstawieniem funkcji kosztu:

$$x_{n+1} = argmin_{x\in(0,1)}\big[\varepsilon \cdot \sum_{j\leqslant n}\big(\alpha x_j \cdot x + \beta(1-x_j)\cdot(1-x)\big) + r(x)\big],$$

to możemy zauważyć, że prawdziwe też jest

$$x_n = argmin_{x\in(0,1)}\big[\varepsilon \cdot \sum_{j\leqslant n-1}\big(\alpha x_j \cdot x + \beta(1-x_j)\cdot(1-x)\big) + r(x)\big].$$

Są to warunki równoważne na minimum ze względu na postać regularyzatora i funkcji kosztu.

Możemy obliczyć i przyrównać do zera pochodne wyrażeń minimalizowanych po prawych stronach tych równości, uzyskując odpowiednio:

$$r'(x_{n+1}) = -\varepsilon \cdot \sum_{j\leqslant n}\big(\alpha x_j - \beta(1-x_j)\big)$$

i

$$r'(x_n) = -\varepsilon \cdot \sum_{j\leqslant n-1}\big(\alpha x_j - \beta(1-x_j)\big).$$

Z powyższych równań otrzymujemy natychmiast:

$$r'(x_{n+1}) = r'(x_n) - \varepsilon\ \big((\alpha+\beta)x_n - \beta\big).$$

W ten sposób uzyskaliśmy zależność, w której nie występują już żadne wyrazy ciągu decyzji mieszkańców Poglądowa poza x_{n+1} i x_n. Dla wygody, podstawiamy $\Psi = -r'(x)$, $b = \frac{\beta}{\alpha+\beta}$ i $a = \varepsilon(\alpha+\beta)$, otrzymując funkcję generującą interesujący nas układ dynamiczny:

(2) $$f_{a,b}(x) = \Psi^{-1}(\Psi(x) + a(x-b)),$$

Wynika to z faktu, że a jest proporcjonalne do ε.

gdzie b jest punktem równowagi, a a – parametrem prędkości uczenia się. Warto zauważyć, że choć funkcja $f_{a,b}$ jest formalnie określona tylko na przedziale $(0,1)$ (co wynika z definicji (1)), to można ją przedłużyć w sposób ciągły na przedział $[0,1]$ definiując $f_{a,b}(0) = 0$ i $f_{a,b}(1) = 1$. Dzięki temu możemy wykorzystać szeroki zakres twierdzeń o układach dynamicznych, których przestrzenią fazową jest domknięty przedział.

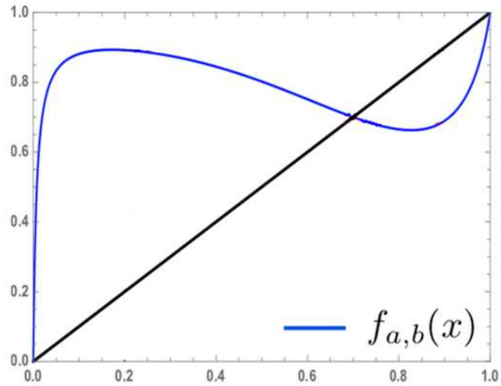
Jak się uczą poglądowianie?

Skoro zdefiniowaliśmy już dynamikę uczenia się dla Poglądowa, czas zbadać, do czego ona prowadzi. Przede wszystkim powstaje pytanie, czy na dłuższą metę poglądowianie osiągną stan równowagi b?

Pierwszym krokiem w badaniu układu dynamicznego jest znalezienie jego punktów stacjonarnych, czyli rozwiązanie równania $f_{a,b}(x) = x$. Oczywiście, 0 i 1 spełniają to równanie, jednak, ze względu na interpretację naszego układu, interesujące są tylko punkty stałe wewnątrz przedziału $(0, 1)$. Ze względu na monotoniczność funkcji $\Psi(x)$ (wynikającą z wypukłości r), łatwo obliczamy:

$$x = \Psi^{-1}(\Psi(x) + a(x - b)) \Rightarrow a(x - b) = 0 \Rightarrow x = b.$$

Zatem jedynym interesującym punktem stacjonarnym dynamiki jest równowaga „statycznej” wersji naszej gry, czyli b. Latwo zauważyć, że $f(x) > x$ dla $x < b$ i $f(x) < x$ dla $x > b$.



Rys. 2. Typowy wykres funkcji $f_{a,b}$. $f_{a,b}$ jest zawsze odwzorowaniem bimodalnym, co oznacza, że posiada dwa ekstrema w przedziale $(0, 1)$. Dynamika takich odwzorowań jest zdecydowanie słabiej zbadana niż dynamika odwzorowań unimodalnych (czyli o jednym ekstremum, takich jak np. $f(x) = 4x(1 - x)$).

Możemy stąd wnioskować, że jeśli ruch na drogach w Poglądowie ma się długoterminowo ustabilizować, to jedyną możliwością na trwałą proporcję między użytkowaniem obu tras jest równowaga b. Gdyby tak faktycznie było, nasza gra transportowa mogłaby osiągnąć równowagę bez założenia o superracjonalności poglądowian – opierając się wyłącznie na ich zdolności do uczenia się. I to właśnie jest podstawowe zastosowanie algorytmu FTRL: doprowadzanie skomplikowanych procesów do stanu równowagi bez zewnętrznej ingerencji, dzięki regularyzatorowi zapewniającemu ściąganie w kierunku równowagi. Algorytm FTRL cieszy się już zasłużoną sławą dzięki efektownym zastosowaniom w zagadnieniach informatycznych opartych o *machine learning*. Największy rozgłos uzyskał, gdy firma Google ujawniła, że stosuje algorytmy oparte o FTRL w przewidywaniu efektywności reklam na stronach internetowych ([McM]). Innym, mniej spektakularnym finansowo, ale nie mniej imponującym przykładem jego skuteczności było użycie go do stworzenia programu grającego w *Stratego* (niezwykle złożoną – znacznie bardziej niż szachy, czy go – klasyczną grę planszową z niepełną informacją) na poziomie porównywalnym do ludzkich mistrzów gry ([Per]). Jednakże, z matematycznego punktu widzenia, na reputację algorytmu FTRL cień rzucał fakt, że nie został dobrze zbadany zakres jego stosowalności. W szczególności, nie istniał dowód, że FTRL zawsze powoduje zbieżność do zachowania równowagi lub przynajmniej innych prostych zachowań (np. okresowych). Z drugiej strony, nie był znany przykład gry, dla której algorytm FTRL prowadziłby do długoterminowo skomplikowanej, niezbieżnej dynamiki. Tę sytuację zmieniły prace [FTRL1] i [FTRL2], których główne wyniki zaprezentuję poniżej.

Aby punkt równowagi mógł „przyciągnąć” do siebie całą dynamikę, niezależnie od stanu początkowego, musi być lokalnie asymptotycznie stabilny, czyli musi przyciągać przynajmniej najbliższe swoje otoczenie. Z twierdzenia o linearyzacji dla układów dynamicznych, warunkiem takiego przyciągania przez punkt b byłoby $\left|f'_{a,b}(b)\right| < 1$. Tymczasem obliczenia wskazują, że

$$f'_{a,b}(x) = \left(\Psi^{-1}\right)' (\Psi(x) + a(x - b)) \cdot (\Psi'(x) + a)\,,$$

czyli

$$f'_{a,b}(b) = \left(\Psi^{-1}\right)' (\Psi(b)) \cdot (\Psi'(b) + a) = \frac{\Psi'(b) + a}{\Psi'(b)}. \tag{3}$$

Zatem warunek asymptotycznej stabilności stanu równowagi jest równoważny nierówności:

$$|\Psi'(b) + a| < -\Psi'(b).$$

Ostatecznie, b jest przyciągającym punktem stałym naszej dynamiki tylko dla odpowiednio małych a: konkretnie, gdy $a \in (0, -2\Psi'(b))$. Dla $a > -2\Psi'(b)$, b staje się punktem odpychającym dynamiki i równowaga nigdy nie zaistnieje na drogach Poglądowa. Można powiedzieć, że zbieżność do równowagi możemy uzyskać tylko jeśli poglądowianie uczą się (a precyzyjniej: aktualizują nowymi informacjami swój proces podejmowania decyzji) wystarczająco wolno.

Okazuje się, że jeśli $\Psi'''(x) < 0$ dla $x \in (0, 1)$ (co jest spełnione np. przez regularyzator pochodzący od entropii Shannona) to warunek $a \in \big(0, -2\Psi'(b)\big)$ jest również warunkiem globalnej zbieżności do stanu równowagi.

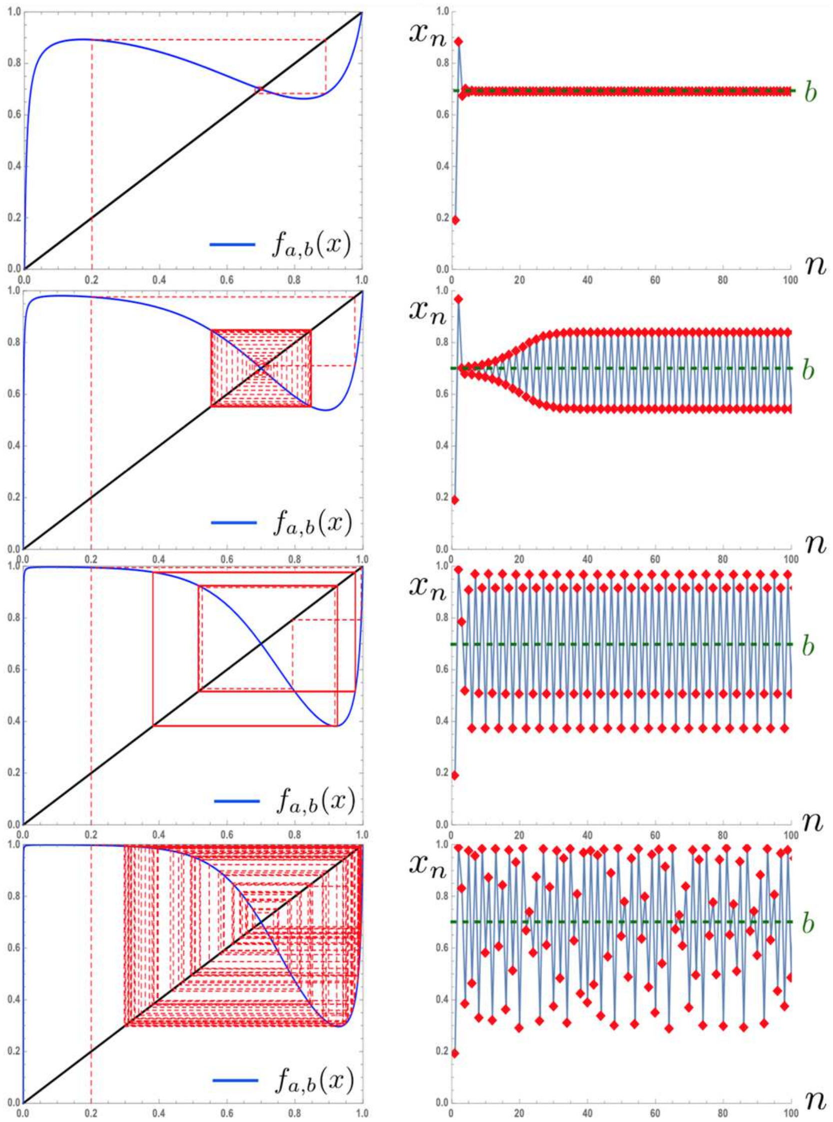
Z drugiej strony, pewna własność „statystyczna” naszej dynamiki wydaje się wskazywać, że nawet jeśli nie zawsze uzyskamy formalną zbieżność do równowagi, możemy liczyć na zbieżność „średnią” czyli na to, że średnia wartość każdej trajektorii dynamiki FTRL zbiega do b, co z kolei sugeruje, że dynamika nie może być zanadto skomplikowana.

Definicja 2. (Przyciąganie w sensie Cesaro) *Dla odwzorowania* $f : [0,1] \to [0,1]$ *punkt* p *jest* przyciągający w sensie Cesaro, *jeśli istnieje otoczenie* U *tego punktu takie, że dla każdego* $x \in U$ *zachodzi*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f^n(x) = p.$$

Twierdzenie 1. *(Przyciąganie Cesaro) Dla* $a > 0$, $b \in (0,1)$ *i dla każdego* $x_0 \in (0,1)$ *zachodzi:*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n} f^n_{a,b}(x_0) = b.$$



Rys. 3. Wykresy obok prezentują zmiany zachowania typowych trajektorii dynamiki (2) przy regularyzatorze

$$r_{Sh}(x) = x\log x + (1-x)\log(1-x)$$

i równowadze $b = 0.7$ oraz parametrach uczenia się a równych odpowiednio (od góry) 7, 10.2, 13.2 i 15.
W szczególności, (2) przyjmuje wtedy postać

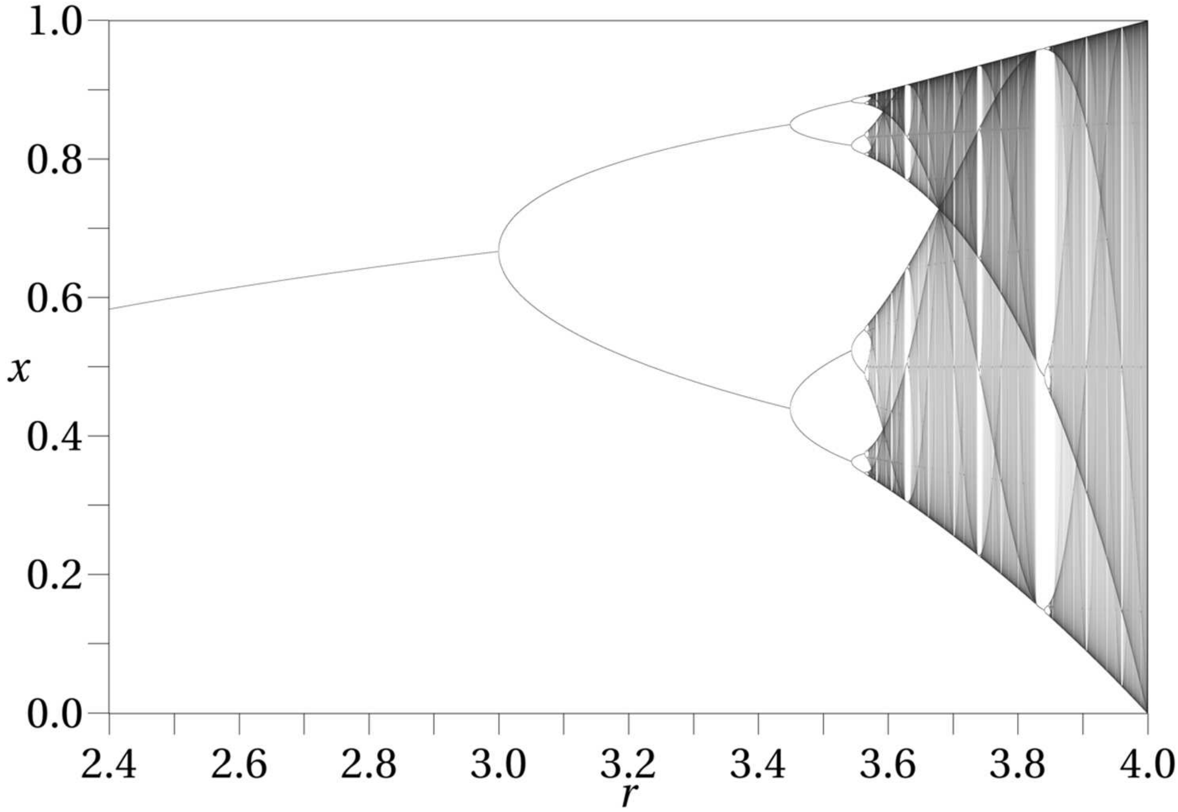
$$f_{a,b}(x) = \frac{x}{x + (1-x)\exp[a(x-0,7)]}.$$

W lewej kolumnie znajdują się odpowiednie wykresy funkcji $f_{a,b}(x)$ wraz z wykresem pajęczynowym, ilustrującym obliczanie kolejnych wyrazów ciągu decyzji poglądowian, a w prawej wykresy typowych trajektorii tej dynamiki. Dla $a = 7$ mamy zbieżność do równowagi b. Wzrost parametru a powoduje zbieżność do trajektorii o okresie 2 ($a = 10.2$), trajektorii o okresie 4 ($a = 13.2$) i wreszcie przejście do skomplikowanego, potencjalnie chaotycznego atraktora dla $a = 15$. Jednakże, średnia wartość tych trajektorii wynosi dokładnie b (zaznaczone przerywaną linią w prawej kolumnie). Jako punkt startowy wybrane jest $x_0 = 0.2$, jednak długoterminowe zachowanie typowych trajektorii nie zależy od tego wyboru.

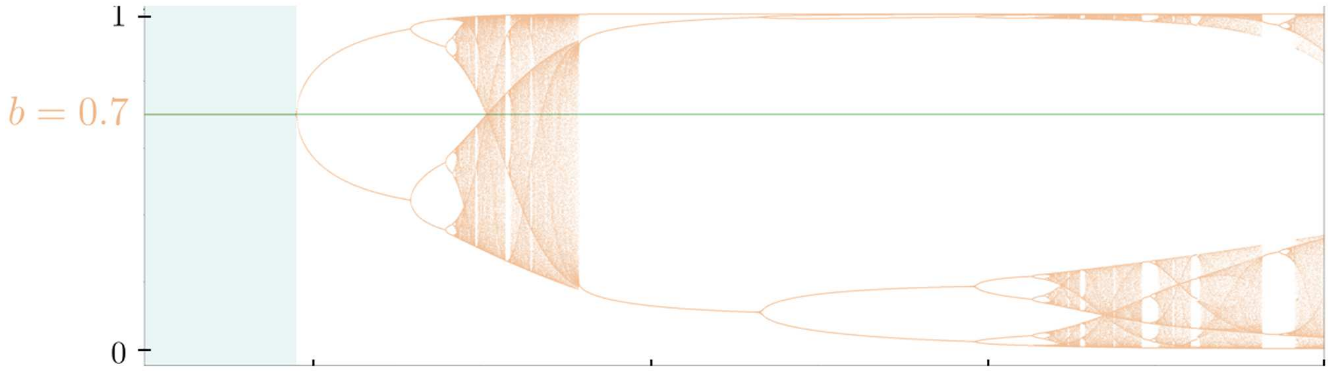
Jednakże, badanie numeryczne konkretnych przypadków (przedstawione na Rysunku 3) wskazuje na całkiem inną sytuację: dla stosunkowo małych wartości parametru uczenia się a, typowe trajektorie zbiegają do równowagi, dla nieco większych – do rozwiązań 2-okresowych, a następnie 4-okresowych, aż dla dużych wartości zaczynają pozornie losowo błądzić po odcinku $(0,1)$, nie zbiegając do żadnych rozwiązań okresowych (a przynajmniej – nie zbiegając szybko).

Chaos w Poglądowie?

Zmiana zbioru przyciągającego dynamikę wraz ze zmianą parametru (w tym przypadku a) polegająca na ciągłym podwajaniu okresu trajektorii przyciągającej (tak zwana „kaskada podwojenia okresu”) nie jest w teorii układów dynamicznych odcinka niczym nowym. Najbardziej kanonicznym przykładem takego efektu jest ewolucja atraktora przy zwiększeniu parametru r w dynamice zadanej równaniem logistycznym $f(x) = rx(1-x)$ (np. [Fei]). W tym klasycznym przykładzie, kaskada podwojenia okresu prowadzi do skomplikowanej własności dynamiki znanej pod nazwą chaosu deterministycznego.



Rys. 4. Tak zwany diagram bifurkacyjny przedstawiający zmiany zbioru przyciągającego dynamiki dla $f(x) = rx(1-x)$ (źródło wykresu: Wikipedia). Dla każdej wartości parametru r na wykresie zaznaczony jest globalny atraktor, czyli zbiór do którego zmierzają typowe trajektorie układu. Widać, że dla stosunkowo małych r (np. mniejszych od 3, w zbiorze $(0,1)$ istnieje jeden punkt przyciągający trajektorie. Dla kolejnego zakresu parametrów (nieco powyżej 3) mamy dwa takie punkty (a dokładniej: przyciągającą trajektorię 2-okresową). Nadal zwiększając parametry, obserwujemy „kaskadę podwojenia okresu” – moc zbioru przyciągającego się podwaja w odstępach, o których decyduje tzw. stała Feigenbauma. Wreszcie dla wystarczająco dużych parametrów (w okolicy 3.6 i większych) zbieżność do orbit okresowych przechodzi w niestabilność typowych trajektorii, które zaczynają „błądzić” po coraz większych częściach przestrzeni fazowej.



Rys. 5. Diagram bifurkacyjny dla przykładu dynamiki Poglądowa, z danymi jak na Rysunku 3, sugeruje powstawanie chaosu jako następstwo kaskady podwojeń okresów.

Współcześni specjaliści od teorii chaosu przedstawiają wiele jego silnie rozłącznych definicji, ale są zgodni co do tego, że chaotyczna dynamika powinna łączyć ze sobą istnienie odpowiednio nieregularnych trajektorii oraz pewnej struktury, która mimo wszystko wyłania się z tej nieregularności. W dalszej części artykułu, mówiąc o chaosie będę się odnosić do definicji chaosu Li-Yorke'a:

Definicja 3 (Chaos w sensie Li-Yorke'a)**.** *Niech $f : X \to X$ generuje układ dynamiczny, a $(x, y) \in X \times X$. Mówimy, że (x, y) jest parą Li-Yorke'a, jeżeli*

$$\liminf_{n\to\infty} dist(f^n(x), f^n(y)) = 0, \quad i \quad \limsup_{n\to\infty} dist(f^n(x), f^n(y)) > 0.$$

Układ dynamiczny zadany przez f jest chaotyczny w sensie Li-Yorke'a, jeśli istnieje nieprzeliczalny zbiór $S \subset X$ (scrambled set) taki, że każda para $(x, y) \in S^2$ taka, że $x \neq y$ jest parą Li-Yorke'a.

Definicja Li-Yorke'a wychwytuje podstawową własność skomplikowanych dynamik zawartą we wszystkich istotnych definicjach chaosu: tak zwaną tranzytywność, czyli „mieszanie" punktów całego zbioru, oznaczające, że dalekie punkty mogą się do siebie dowolnie zbliżać i jednocześnie dowolnie bliskie punkty mogą się od siebie istotnie oddalić. W szczególności, mamy zagwarantowaną silną zależność od warunków początkowych (najbardziej popularny chyba atrybut chaosu znany powszechnie jako *efekt motyla*), która oznacza, że dowolnie małe zaburzenie punktu startowego trajektorii może spowodować dużą różnicę w jej przebiegu po przedziale $[0, 1[$.

Dla dynamiki (2), Rysunek 3 przedstawia pojawianie się nieregularności w długoterminowych trajektoriach okresowych, a Twierdzenie 1 (o przyciąganiu w sensie Cesaro) świadczy o tym, że mimo to istnieje w tej dynamice pewna struktura. Pokażemy, że jeśli poglądowianie uczą się wystarczająco szybko (a dokładniej: wystarczająco gwałtownie dostosowują swoje strategie do zdobywanych informacji), to dynamika zmian ich decyzji (2) staje się chaotyczna w sensie Li-Yorke'a. Przy okazji, pokażemy ciekawy element struktury ukrytej w chaosie związany z trajektoriami okresowymi.

Twierdzenie 2. *Dla każdego $b \neq \frac{1}{2}$ istnieje a_0 takie, że dla każdego $a > a_0$ i każdego regularyzatora r spełniającego wcześniej przedstawione założenia, układ dynamiczny zadany na $[0, 1]$ przez (2) jest chaotyczny w sensie Li-Yorke'a i posiada trajektorie okresowe o dowolnym naturalnym okresie.*

Dowód chaotyczności oparty jest o piękną klasyczną ideę, której sednem jest zdumiewająca struktura liczbowa związaną z trajektoriami okresowymi i związek pomiędzy tą strukturą a chaotycznością dynamiki.

Definicja 4. Porządek Szarkowskiego *to porządek liniowy zdefiniowany na liczbach naturalnych w taki sposób, że jako najmniejsze są traktowane wszystkie liczby niepatrzyste (ustawione rosnąco tak jak w „zwyczajnym" uporządkowaniu liczb naturalnych), następnie wszystkie liczby nieparzste pomnożone przez kolejne potęgi dwójki (również w tej samej kolejności), a jako największe traktujemy same potęgi dwójki, ustawione w kolejności odwrotnej do „zwyczajnej" :*
$3 < 5 < 7 < \ldots < 2 \cdot 3 < 2 \cdot 5 < 2 \cdot 7 < \ldots <$
$< \ldots < 4 \cdot 3 < 4 \cdot 5 < 4 \cdot 7 < \ldots < 2^n < 2^{n-1} < \ldots < 4 < 2 < 1.$

Twierdzenie 3 (Twierdzenie Szarkowskiego)**.** ([Shar]) *Jeśli odwzorowanie ciągłe $f : [0,1] \to [0,1]$ generuje układ dynamiczny, który ma trajektorię o okresie m i $m < n$ w porządku Szarkowskiego, to ten układ dynamiczny ma trajektorię o okresie n.*

W szczególności, z twierdzenia Szarkowskiego wynika, że jeśli istnieje trajektoria o okresie 3, to istnieją też trajektorie o wszystkich innych okresach. Okazuje się, że to wystarcza do zaistnienia chaosu w sensie Li-Yorke'a.

Twierdzenie 4 (Twierdzenie Li-Yorke'a)**.** ([LY]) *Istnienie trajektorii o okresie 3 dla odwzorowania ciągłego $f : [0,1] \to [0,1]$ oznacza chaotyczność dynamiki w sensie Li-Yorke'a.*

By udowodnić, że dla dużych a dynamika (2) jest chaotyczna, potrzebujemy jeszcze następującego lematu:

Lemat 5. ([LMGY]) *Jeśli dla odwzorowania ciągłego $f : [0,1] \to [0,1]$ istnieje punkt x taki, że dla pewnego nieparzystego $n > 1$, $f^n(x) < x < f(x)$, to f ma orbitę o okresie 3.*

Dowód Twierdzenia 2. Na podstawie powyższego lematu, wystarczy wykazać, że dla odpowiednio dużych a istnieje x_0 taki, że $f^3_{a,b}(x_0) < x_0 < f_{a,b}(x_0)$. Ustalmy $a > 0$, $b, x_0 \in (0,1)$ i zdefiniujmy $x_n = f^n_{a,b}(x_0)$. Prostą indukcją możemy udowodnić, że:

$$x_n = f_{a,b}(x_{n-1}) = \Psi^{-1}\big(\Psi(x_0) + a\big(\sum_{k=0}^{n-1}(x_k - b)\big)\big)$$

Z definicji, odzworowanie Ψ^{-1} jest malejące, zatem $f_{a,b}(x) > x$ wtedy i tylko wtedy, gdy $x < b$, a nierówność $f^3(x) < x$ jest równoważna nierówności $x + f_{a,b}(x) + f^2_{a,b}(x) > 3b$.

Bez utraty ogólności możemy założyć, że $b < \frac{1}{2}$, czyli $3b - 1 < b$. Wybieramy zatem $x_0 \in (3b-1, b)$. Z założeń o postaci regularyzatora mamy

$$\lim_{x \to 1^+} \Psi(x) = \lim_{x \to 1^+} -r'(x) = -\infty,$$

więc

$$\lim_{a \to \infty} f_{a,b}(x_0) = \lim_{a \to \infty} \Psi^{-1}(\Psi(x_0) + a(x_0 - b)) = 1.$$

Dodatkowo $3b - x_0 < 1$, więc dla dużych a zachodzi $f_{a,b}(x_0) > 3b - x_0$, czyli $x_0 + f_{a,b}(x_0) > 3b$ a więc tym bardziej $x_0 + f_{a,b}(x_0) + f^2_{a,b}(x) > 3b$.

Zatem dla odpowiednio dużych a istnieje x_0 takie, że $f^3_{a,b}(x_0) < x_0 < f_{a,b}(x_0)$. Stąd też, na podstawie Lematu 5, istnieje trajektoria okresowa o okresie 3, co na podstawie twierdzeń Szarkowskiego i Li-Yorke'a pociąga za sobą istnienie trajektorii okresowych o wszystkich okresach naturalnych dodatnich oraz chaotyczność dynamiki (2) (o ile $b \neq \frac{1}{2}$). Q.E.D.

Wnioski i podsumowania

Teoria gier jest dość młodą dziedziną matematyki, ale jej zastosowania już wywarły wielki wpływ, zwłaszcza na biologię i nauki społeczne. Na przykładzie prostego problemu transportowego mieliśmy okazję przyjrzeć się nowemu podejściu do klasycznej teorii gier i zobaczyć jak algorytmiczna teoria gier bada procesy uczenia się graczy. Z samej gry mieszkańców Poglądowa trudno wyciągnąć jakiekolwiek wiążące wnioski dla sytuacji zachodzących w świecie rzeczywistym, gdyż zarówno funkcje kosztów jak i sama przestrzeń możliwych wyborów były bardzo uproszczone. Jednakże rezultaty, które zostały osiągnięte dla tak pozornie nieskomplikowanej sytuacji, mają dość doniosłe znaczenie.

Niestabilność, brak zbieżności do stanu równowagi i chaotyczność uczenia się być może nie stawiają pod znakiem zapytania obecnego w teorii gier i, bardziej ogólnie, w matematycznym modelowaniu zagadnień społecznych oraz w algorytmach *machine learning* paradygmatu poszukiwania równowag i używania ich jako rozwiązań oczekiwanych w danej sytuacji. Jednakże, skoro istnieją relatywnie proste kontrprzykłady kwestionujące samorzutną zbieżność układów dynamicznych powstałych na bazie gier, należy poważnie się zastanowić nad zakresem stosowalności klasycznego podejścia. Czy wyniki przedstawione powyżej są pewną osobliwością, właściwą tylko stosunkowo małej klasie gier i generowanych przez nie dynamik, czy też niestabilność równowag jest problemem, nad którym należy się pochylić przy analizowaniu dowolnej gry? Jak czujni muszą być twórcy algorytmów *machine learning*, by uniknąć protokołów, które nie gwarantują zbieżności do optymalnych rozwiązań? Czy przypadkiem to nie nagły, silny bodziec zachęcający do szybkiego dostosowania się do nowych informacji jest przynajmniej częściowym wyjaśnieniem gwałtownych i nieprzewidywalnych reakcji rynków na zewnętrzne impulsy, takich jak, np. kryzysy finansowe? Na te pytania (i na wiele innych) odpowiedzi mogą dostarczyć tylko dalsze badania algorytmicznej teorii gier.

Literatura

[Fei] M.J. Feigenbaum (1978) *Quantitative universality for a class of nonlinear transformations*. Journal of Statistical Physics. 19 (1): 25-52.

[FTRL1] T. Chotibut, F.Falniowski, M. Misiurewicz, G. Piliouras (2020) *The Route to Chaos in Routing Games : When Is Price of Anarchy Too Optimistic?* Advances in Neural Information Processing Systems 33 (NeurIPS 2020), 1-12.

[FTRL2] J. Bielawski, T. Chotibut, F.Falniowski, G.Kosiorowski, M. Misiurewicz, G. Piliouras (2021) *Follow-the-Regularized-Leader Routes to Chaos in Routing Games*. Proceedings of Machine Learning Research (139), 925-935.

[LY] T.Y. Li, J.A. Yorke (1975) *Period Three Implies Chaos*. American Mathematical Monthly. 82 (10): 985-992.

[LMGY] T. Y. Li, M. Misiurewicz, G. Pianigiani, and J. A. Yorke (1982) *Odd chaos*. Physics Letters A, 311 87(6): 271-273.

[McM] H.B. McMahan at al. (2013) *Ad Click Prediction: a View from the Trenches* KDD '13: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining August, 1222-1230

[Per] J.Perolat at al. (2022) *Mastering the Game of Stratego with Model-Free Multiagent Reinforcement Learning*. https://arxiv.org/abs/2206.15378

[Shar] A.N. Sharkovskii (1964) *Co-existence of cycles of a continuous mapping of the line into itself*, Ukrainian Math. J. 16:61-71.

[Yan] I.Yaniv (2004) *Receiving other people's advice: Influence and benefit*. Organizational Behavior and Human Decision Processes 93(1), 1-13.