

Wprowadzenie

Sztuczna inteligencja. Naukę o sztucznej inteligencji (AI) można zdefiniować jako budowę inteligentnych systemów i ich analizę. Naturalna definicja systemu to wszystko, co ma strumień wejściowy i wyjściowy. Inteligencja jest bardziej skomplikowana. Może mieć wiele twarzy, takich jak kreatywność, rozwiązywanie problemów, rozpoznawanie wzorców, klasyfikacja, uczenie się, indukcja, dedukcja, budowanie analogii, optymalizacja, przetrwanie w środowisku, przetwarzanie języka, wiedza i wiele innych. Jednak formalna definicja obejmująca każdy aspekt inteligencji wydaje się trudna. Ponadto inteligencja jest stopniowana: istnieje płynne przejście między systemami, co do których wszyscy zgodziliby się, że nie są inteligentne, a prawdziwie inteligentnymi systemami. Wystarczy spojrzeć na naturę, zaczynając na przykład od nieożywionych kryształów, następnie aminokwasów, następnie niektórych fragmentów RNA, następnie wirusów, bakterii, roślin, zwierząt, małych cząłkopszątnych, a następnie prawdziwie inteligentnych homo sapiens i ewentualnie kontynuowanych przez systemy AI lub ET. Tak więc najlepszym, czego możemy się spodziewać, jest częściowa lub całkowita relacja porządkująca na zbiorze systemów, która je porządkuje w odniesieniu do ich stopień inteligencji (podobnie jak testy inteligencji dla systemów ludzkich, ale dla ograniczonej klasy problemów). Mając tę kolejność, jesteśmy oczywiście zainteresowani dużymi elementami, tj. wysoce inteligentnymi systemami. Jeśli istnieje największy element, odpowiadałby on najbardziej inteligentnemu systemowi, jaki mógłby istnieć. Większość, jeśli nie wszystkie, znane aspekty inteligencji można sformułować jako ukierunkowane na cel lub, dokładniej, jako maksymalizujące pewną funkcję użyteczności. Dlatego wystarczy zbadać ukierunkowaną na cel sztuczną inteligencję. Na przykład (biologicznym) celem zwierząt i ludzi jest przetrwanie i rozprzestrzenianie się. Celem systemów sztucznej inteligencji powinno być bycie użytecznym dla ludzi. Problem polega na tym, że poza szczególnymi przypadkami nie znamy z góry ani funkcji użyteczności, ani środowiska, w którym agent będzie działał. Główna idea. Ta książka przedstawia teorię, która formalnie[^] rozwiązuje problem nieznanego celu i środowiska. Można ją postrzegać jako połączenie idei uniwersalnej indukcji, planowania probabilistycznego i uczenia się przez wzmocnienie lub jako połączenie teorii decyzji sekwencyjnych z algorytmiczną teorią informacji. Stosujemy ten model do niektórych aspektów inteligencji, w tym indukcji, gry, optymalizacji, wzmocnienia i uczenia nadzorowanego, i pokazujemy, jak rozwiązuje on te klasy problemów. To, wraz z ogólnymi twierdzeniami zbieżności, wspiera przekonanie, że skonstruowany uniwersalny system AI jest najlepszy w pewnym sensie, który zostanie wyjaśniony w następujący sposób, tj. że jest to najbardziej inteligentny, niezależny od środowiska system, jaki jest możliwy. Celem tej książki jest wprowadzenie uniwersalnego modelu AI i przedstawienie obszernej analizy.

Prostota i niepewność

Ta sekcja wprowadza zasadę brzytwy Ockhama, złożoność Kolmogorova i prawdopodobieństwa obiektywne/subiektywne. W końcu docieramy do problemu uniwersalnej predykcji i jego rozwiązania przez Solomonoffa.

Wprowadzenie

Ważnym i nietrywialnym aspektem inteligencji jest wnioskowanie indukcyjne. Mówiąc prościej, indukcja to proces przewidywania przyszłości na podstawie przeszłości, lub, mówiąc precyzyjniej, proces znajdowania reguł w (przeszłych) danych i wykorzystywania tych reguł do odgadywania przyszłych danych. Nietrywialnymi przykładami są prognozowanie pogody lub rynku akcji lub ciągłe serie liczbowe w teście IQ. Tworzenie dobrych prognoz odgrywa centralną rolę w inteligencji naturalnej i sztucznej w ogóle, a w szczególności w uczeniu maszynowym. Wszystkie problemy indukcyjne można sformułować jako zadania przewidywania sekwencji. Jest to na przykład oczywiste w przypadku

przewidywania szeregów czasowych, ale obejmuje również zadania klasyfikacyjne. Po zaobserwowaniu danych x_t w momentach $t < n$, zadaniem jest przewidzenie n -tego symbolu x_n z sekwencji $x_1 \dots x_{n-1}$. To podejście presekwencyjne pomija pośredni krok uczenia się modelu na podstawie zaobserwowanych danych $x_1 \dots x_{n-1}$, a następnie użycia tego modelu do przewidzenia x_n . Podejście presekwencyjne unika problemów spójności modelu, sposobu oddzielania szumu od użytecznych danych i wielu innych kwestii. Celem jest dokonywanie „dobrych” przewidywań, gdzie jakość przewidywań jest zwykle mierzona funkcją straty, która musi zostać zminimalizowana. Kluczowym założeniem dla dobrego definiowania i rozwiązywania problemów indukcyjnych jest zasada brzytwy Ockhama (sympcity), która mówi, że bytów nie należy mnożyć ponad konieczność. Można to interpretować jako zachowanie spójności najprostszej teorii z obserwacjami $x_1 \dots x_{1-t}$ i wykorzystanie tej teorii do przewidywania x_n . Zanim przedstawimy formalne rozwiązanie Solomonowa, musimy określić brzytwę Ockhama w kategoriach złożoności Kolmogorowa i wprowadzić pojęcia prawdopodobieństwa subiektywnego i obiektywnego.

Algorytmiczna teoria informacji

Intuicyjnie, ciąg jest prosty, jeśli można go opisać kilkoma słowami, jak „ciąg miliona jedynek”, i jest złożony, jeśli nie ma takiego krótkiego opisu, jak w przypadku losowego ciągu, którego najkrótszym opisem jest określenie go bit po bicie. Możemy ograniczyć dyskusję do ciągów binarnych, ponieważ dla innych (nieciągów) obiektów matematycznych możemy założyć pewne domyślne kodowanie jako ciągi binarne. Ponadto interesują nas tylko efektywne opisy, a zatem ograniczamy dekodery do maszyn Turinga. Wybierzmy jakąś uniwersalną (tzw. prefiksową) maszynę Turinga U z jednokierunkowymi taśmami wejściowymi i wyjściowymi binarnymi oraz dwukierunkową taśmą roboczą. Możemy następnie zdefiniować prefiksową złożoność Kolmogorowa ciągu binarnego x jako długość ℓ najkrótszego programu p , dla którego U wyprowadza ciąg binarny x

$$K(x) := \min_p \{ \ell(p) : U(p) = x \}$$

Proste ciągi znaków, takie jak 000...0, mogą być generowane przez krótkie programy i stąd mają niską złożoność Kolmogorowa, ale nieregularne (np. losowe) ciągi znaków są ich najkrótszym opisem i stąd mają wysoką złożoność Kolmogorowa. Ważną właściwością K jest to, że jest niemal niezależna od wyboru U . Ponadto dzieli wiele właściwości z entropią Shannona (miarą informacji) S , ale K jest pod wieloma względami lepsza od S . Krótko mówiąc, K jest doskonałą uniwersalną miarą złożoności, odpowiednią do kwantyfikacji brzytwy Ockhama. Istnieje (tylko) jedna poważna wada: K nie jest skończenie obliczalna. Dokładniej rzecz biorąc, funkcja K jest nazywana skończenie obliczalną (lub rekurencyjną), jeśli istnieje maszyna Turinga, która przy danym x oblicza $f(x)$, a następnie zatrzymuje się. Niektóre funkcje nie są skończenie obliczalne, ale nadal aproksymowalne w tym sensie, że istnieje nieprzerwana maszyna Turinga z nieskończonym ciągiem wyjściowym y_1, y_2, y_3 z $\lim_{t \rightarrow \infty} y_t = f(x)$. Jeśli dodatkowo ciąg wyjściowy jest monotonicznie rosnący/malejący, to f jest nazywane dolnym/górnym półobliczalnym (lub przeliczalnym/współprzeliczalnym). Na koniec nazywamy f szacowalnym, jeśli jakaś maszyna Turinga, podana x i precyzją ϵ , oblicza skończenie ϵ -aproksymację x . Główną algorytmiczną własnością K jest to, że jest współprzeliczalna, ale nie skończenie obliczalna.

Niepewność i prawdopodobieństwa

Dla obiektywisty, prawdopodobieństwa są rzeczywistymi aspektami świata. Wynik obserwacji lub eksperymentu nie jest deterministyczny, ale obejmuje fizyczne procesy losowe. Aksjomaty teorii prawdopodobieństwa Kolmogorowa formalizują właściwości, jakie powinny mieć prawdopodobieństwa. W przypadku niezależnych i identycznie rozłożonych (i.i.d.) eksperymentów prawdopodobieństwa przypisane zdarzeniom można interpretować jako częstotliwości graniczne (wid

częstotliwościowy) ale zastosowania nie ograniczają się do tego przypadku. Warunkowanie prawdopodobieństw i reguła Bayesa są głównymi narzędziami w obliczaniu prawdopodobieństw a posteriori z prawdopodobieństw poprzednich. Na przykład, biorąc pod uwagę początkową sekwencję binarną $x_1 \dots x_{n-1}$, jakie jest prawdopodobieństwo, że następny bit będzie 1? Prawdopodobieństwo zaobserwowania x_n w czasie n , biorąc pod uwagę wcześniejsze obserwacje $x_1 \dots x_{n-1}$, można obliczyć za pomocą mnożenia lub reguły łańcuchowej, jeśli znany jest prawdziwy rozkład generujący μ ciągów $x_1 x_2 x_3$: $\mu(x_n | x_{<n}) = \mu(x_{1:n}) / \mu(x_{<n})$ whcrc wc iutuduccd skróty $x_{1:n} \equiv x_1 x_2 \dots x_n$ i $x_{<n} \equiv x_1 x_2 \dots x_{n-1}$. Problem polega jednak na tym, że często nie znamy prawdziwego rozkładu μ (np. w przypadku prognozowania pogody i giełdy). Subiektywista używa prawdopodobieństw do scharakteryzowania stopnia wiary agenta w coś (lub prawdopodobieństwa czegoś), a nie do scharakteryzowania fizycznych procesów losowych. Jest to najbardziej odpowiednia interpretacja prawdopodobieństw w AI. Jest nieco zaskakujące, że można wykazać, że prawdopodobieństwa również respektują aksjomaty prawdopodobieństwa Kolmogorowa i regułę łańcuchową, zakładając tylko kilka prawdopodobnych reguł jakościowych, których powinny przestrzegać. Stąd, jeśli prawdopodobieństwo $x_{1:n}$ wynosi $\rho(x_{1:n})$, stopień wiary w x_n przy założeniu $x_{<n}$ jest, ponownie, dane przez regułę łańcuchową: $\rho(x_n | x_{<n}) = \rho(x_{1:n}) / \rho(x_{<n})$. Reguła łańcuchowa pozwala na obliczenie prawdopodobieństw a posteriori/prawdopodobieństw z prawdopodobieństw wcześniejszych, ale pozostawia otwartą kwestię, jak określić same prawdopodobieństwa wcześniejsze. W fizyce statystycznej zasada obojętności (zasada symetrii) i zasada maksymalnej entropii mogą być często wykorzystywane do określania prawdopodobieństw a priori, ale tylko brzytwa Ockhama jest wystarczająco ogólna, aby przypisać prawdopodobieństwa a priori w każdej sytuacji, zwłaszcza w celu poradzenia sobie ze złożonymi domenami typowymi dla sztucznej inteligencji.

Algorytmiczne prawdopodobieństwo i uniwersalna indukcja

Brzytwa Ockhama (odpowiednio zinterpretowana i będąca w kompromisie z zasadą obojętności Epikura) mówi nam, aby przypisać wysoką/niską prawdopodobieństwo a priori prostym/złożonym ciągom x . Używając K jako miary złożoności, każda monotoniczna malejąca funkcja K , np. $\rho(x) = 2^{-K(x)}$ spełniałaby to kryterium. Ale ρ musi również spełniać aksjomaty prawdopodobieństwa, więc musimy być nieco ostrożniejsi. Solomonoff zdefiniował uniwersalne a priori $M(x)$ jako prawdopodobieństwo, że wyjście uniwersalnej maszyny Turinga U zaczyna się od x , gdy na taśmie wejściowej podano uczciwe rzuty monetą. Formalnie M można zdefiniować jako

$$M(x) := \sum_{p : U(p)=x^*} 2^{-\ell(p)} \quad (1.1)$$

gdzie suma jest po wszystkich (tzw. minimalnych) programach p , dla których U wyprowadza ciąg zaczynający się od x . Ściśle rzecz biorąc, M jest tylko półmiarą, ponieważ nie jest znormalizowane do 1, ale jest to dopuszczalne/korygowalne. Wyprowadzamy następującą granicę:

$$\sum_{t=1}^{\infty} (1 - M(x_t | x_{<t}))^2 \leq -\frac{1}{2} \sum_{t=1}^{\infty} \ln M(x_t | x_{<t}) = -\frac{1}{2} \ln M(x_{1:\infty}) \leq \frac{1}{2} \ln 2 \cdot Km(x_{1:\infty})$$

gdzie $Km\{x_{1:\infty}\}$ jest długością najkrótszego (nieprzerwanego) programu obliczającego $x_{1:\infty}$. W pierwszej nierówności użyliśmy $(1-a)^2 \leq -1/2 \ln a$ dla $0 \leq a \leq 1$. W równości zamieniliśmy sumę na logarytm i wyeliminowaliśmy otrzymany iloczyn za pomocą reguły łańcuchowej. W ostatniej nierówności użyliśmy $M(x) \geq 2^{-Km(x)}$, co wynika z definicji (1.1) przez pominięcie wszystkich wyrazów w Σ_p z wyjątkiem najkrótszego p obliczającego x . Jeśli $x_{1:\infty}$ jest ciągiem obliczalnym, to $Km\{x_{1:\infty}\}$ jest skończony, co implikuje $M(x_t | x_{<t}) \rightarrow 1$ $\sum_{t=1}^{\infty} (1 - a_t)^2 < \infty \Rightarrow a_t \rightarrow 1$. Oznacza to, że jeśli środowisko jest obliczalną

sekwencją (cokolwiek, np. cyfry π lub e w reprezentacji binarnej), po zobaczeniu pierwszych kilku cyfr, M poprawnie przewiduje następną cyfrę z dużym prawdopodobieństwem, tj. rozpoznaje strukturę sekwencji. Załóżmy teraz, że prawdziwa sekwencja jest losowana z rozkładu μ , tj. prawdziwe (obiektywne) prawdopodobieństwo $x_{1:n}$ wynosi $\mu(x_{1:n})$, ale μ jest nieznanne. W jaki sposób późniejsze (subiektywne) przekonanie $M(x_n|x_{<n}) = M(x_n)/M(x_{<n})$ jest powiązane z prawdziwym (obiektywnym) prawdopodobieństwem późniejszym $\mu(x_n|x_{<n})$? Centralnym wynikiem Solomonoffa jest to, że późniejsze (subiektywne) przekonania zbiegają się do prawdziwych (obiektywnych) prawdopodobieństw późniejszych, jeśli te ostatnie są obliczalne. Dokładniej, pokazał, że

$$\sum_{t=1}^{\infty} \sum_{x_{<t} \in \{0,1\}^{t-1}} \mu(x_{<t}) \left(M(0|x_{<t}) - \mu(0|x_{<t}) \right)^2 \leq \frac{1}{2} \ln 2 \cdot K(\mu) + O(1). \quad (1.2)$$

Złożoność $K(\mu)$ jest skończona, jeśli μ jest funkcją obliczalną, ale suma nieskończona na lewej stronie może być skończona tylko wtedy, gdy różnica $M(0|x_{<t}) - \mu(0|x_{<t})$ dąży do zera dla $t \rightarrow \infty$ z μ -prawdopodobieństwem 1 (w.μ.p.l). Pokazuje to, że użycie M jako oszacowania dla μ może być rozsądnym rozwiązaniem.

Uogólnione uniwersalne (pół)miary

Można wyprowadzić uniwersalne wcześniejsze w inny sposób: Solomonoff definiuje nieco problematyczną mieszaną dla wszystkich obliczalnych rozkładów prawdopodobieństwa. Levin rozważa większą klasę $M_U := \{\nu_1, \nu_2, \dots\}$ wszystkich tak zwanych wyliczalnych półmiar. Niech $\mu \in M_U$ i przypiszemy (zgodnie z brzytwą Ockhama) wcześniejsze prawdopodobieństwo $2^{-K(\nu_a)}$ do ν_a . Wtedy wcześniejsza prawdopodobieństwo $x_{1:n}$ wynosi, zgodnie z elementarną teorią prawdopodobieństwa,

$$\xi_U(x_{1:n}) := \sum_{\nu \in M_U} 2^{-K(\nu)} \nu(x_{1:n}). \quad (1.3)$$

Można pokazać, że ξ_U pokrywa się z M w obrębie (nieistotnej) stałej mnożnikowej, tj. $M(x) \stackrel{\approx}{\leq} \xi_U(x)$

, gdzie $f(x) \stackrel{\approx}{\leq} g(x)$ skraca $f(x) = O(g(x))$, a $\stackrel{\approx}{\leq}$ oznacza $\stackrel{\approx}{\leq}$ i $\stackrel{\approx}{\geq}$. Można pokazać, że zarówno ξ_U , jak i M

są dolną półobliczalnością. Dominacja $M(x) \stackrel{\approx}{\leq} \xi_U(x) \geq 2^{-K(\mu)} \mu(x)$ jest głównym składnikiem dowodu (1.2). Zaletą ξ_U nad M jest to, że definicja natychmiast uogólnia się do dowolnych ważonych sum (pół)miar w M dla dowolnych policzalnych M . Większość dowodów przechodzi przez generyczne M i wagi. Więc co jest takiego szczególnego w klasie wszystkich przeliczalnych półmiar M_U ? Im większe M wybierzemy, tym mniej restrykcyjne będzie założenie, że M powinno zawierać prawdziwy rozkład μ , który będzie istotny w całym tekście. Dlaczego nie ograniczyć się do wciąż dość ogólnej klasy szacowanych lub skończenie obliczalnych (pół)miar? Dla każdej policzalnej klasy M spełniona jest mieszanina

$\xi(x) := \xi_M(x) := \sum_{\nu \in M} w_\nu \nu(x)$ z $w_\nu > 0$ ważna dominacja $\xi(x) \geq w_\nu \nu(x)$. Pytanie brzmi,

jakie właściwości posiada ξ . Wyróżniającą właściwością M_U jest to, że ξ_U samo w sobie jest elementem M_U . Z drugiej strony, w tej książce $\xi_M \in M$ samo w sobie nie jest ważną właściwością. Ważne jest, czy ξ jest obliczalne w jednym ze znaczeń zdefiniowanych powyżej. Istnieje przeliczalna półmiara (M), która dominuje nad wszystkimi przeliczalnymi półmiarami w M_U . Jak zobaczymy, nie ma szacowanej półmiary, która dominuje nad wszystkimi obliczalnymi miarami, i nie ma przybliżonej półmiary, która dominuje nad wszystkimi przybliżalnymi miarami. Z tego wynika, że dla uniwersalnej (pół)miary, która przynajmniej spełnia najłagodniejszą formę obliczalności, a mianowicie jest przybliżona, największą zdominowaną klasą wśród klas rozważanych w tej książce jest klasa przeliczalnych półmiar, ale istnieją jeszcze większe klasy. To jest powód, dla którego M_U i M odgrywają szczególną rolę w tej (i innych)

pracy. W praktyce jednak trzeba ograniczyć się do skończonego podzbioru skończone obliczalnych środowisk v , aby uzyskać skończenie obliczalne ξ .

Uniwersalna predykcja sekwencji

W dalszej części dokładniej zbadamy schematy predykcji sekwencji (SP) oparte na uniwersalnym priorytecie Solomonoffa $M \stackrel{\times}{=} \xi_U$ i na bardziej ogólnych mieszankach Bayesa ξ , głównie z perspektywy teorii decyzji. W szczególności pokazujemy, że są one optymalne w odniesieniu do różnych kryteriów optymalności.

Konfiguracja konwergencji

Niech $M := \{v_1, v_2, \dots\}$ będzie przeliczalnym zbiorem rozkładów prawdopodobieństwa kandydatów na ciągach znaków nad skończonym alfabetem X . Definiujemy średnią ważoną na M :

$$\xi(x_{1:n}) := \sum_{\nu \in M} w_\nu \cdot \nu(x_{1:n}), \quad \sum_{\nu \in M} w_\nu = 1, \quad w_\nu > 0. \quad (1.4)$$

Łatwo zauważyć, że ξ jest rozkładem prawdopodobieństwa, ponieważ wagi w_ν są dodatnie i znormalizowane do 1, a $\nu \in M$ są prawdopodobieństwami. ξ nazywamy uniwersalnym względem M , ponieważ mnoży on wszystkie rozkłady w M w tym sensie, że $\xi(x_{1:n}) \geq w_\nu \cdot \nu$ dla wszystkich $\nu \in M$. W poniższym przykładzie zakładamy, że M jest znane i zawiera prawdziwy, ale nieznan rozkład μ , tj. $\mu \in M$, a $x_{1:\infty}$ jest próbkowane z μ . Skracamy oczekiwania względem μ przez $E[\cdot]$; na przykład, $E[f(x_{1:n})] = \sum_{x_{1:n} \in X^n} \mu(x_{1:n}) f(x_{1:n})$. Używamy (całkowitej) względnej entropii D_n i kwadratu odległości euklidesowej S_n do pomiaru odległości między μ a ξ :

$$D_n := E \left[\ln \frac{\mu(x_{1:n})}{\xi(x_{1:n})} \right], \quad S_n := \sum_{t=1}^n E \left[\sum_{x'_t \in X} \left(\mu(x'_t | x_{<t}) - \xi(x'_t | x_{<t}) \right)^2 \right]. \quad (1.5)$$

Można pokazać następujący ciąg nierówności, który uogólnia wynik Solomonoffa (1.2): $S_n < D_n < \ln w_\mu^{-1} < \infty$. Skończoność S_∞ implikuje $\xi(x'_t | x_{<t}) - \mu(x'_t | x_{<t}) \rightarrow 0$ dla $t \rightarrow \infty$ dla dowolnego x'_t ($\sum_{t=1}^\infty s_t^2 < \infty \Rightarrow s_t \rightarrow 0$). Pokazujemy również, że $\sum_{t=1}^n E \left[\left(\frac{\sqrt{\xi(x_t | x_{<t})}}{\mu(x_t | x_{<t})} - 1 \right)^2 \right] \leq D_n \leq \ln w_\mu^{-1} < \infty$, co implikuje $\xi(x_t | x_{<t}) / \mu(x_t | x_{<t}) \rightarrow 1$ dla $t \rightarrow \infty$. Ta zbieżność uzasadnia przekonanie, że przewidywania oparte na (znanym) μ są asymptotycznie tak samo dobre, jak przewidywania oparte na (nieznanym) ξ , przy czym zbieżność jest szybka

Granice strat

Większość przewidywań jest ostatecznie wykorzystywana jako podstawa do podjęcia decyzji lub działania, które samo w sobie prowadzi do pewnej nagrody lub straty. Niech $\ell_{x_t y_t} \in [0, 1] \subset \mathbb{R}$ będzie otrzymaną stratą podczas wykonywania przewidywania/decyzji/działania $y_t \in Y$, a $x_t \in X$ będzie t -tym symbolem ciągu. Niech $y_t^\wedge \in Y$ będzie przewidywaniem (przyczynowego) schematu przewidywania Λ . Prawdziwe prawdopodobieństwo, że następnym symbolem będzie x_t , przy założeniu $x_{<t}$, wynosi $\mu(x_t | x_{<t})$. Oczekiwana strata podczas przewidywania y_t wynosi $E[\ell_{x_t y_t}]$. Całkowita μ -oczekiwana strata poniesiona przez schemat Λ w pierwszych n przewidywaniach wynosi

$$L_n^\Lambda := \sum_{t=1}^n \mathbf{E}[\ell_{x_t y_t^\Lambda}].$$

Celem jest zminimalizowanie oczekiwanej straty. Bardziej ogólnie, definiujemy schemat predykcji sekwencji A_p (później nazywany również SPp) $\operatorname{argmin}_{y_t \in \mathcal{Y}} \sum_{x_t} \rho(x_t | x_{<t}) \ell_{x_t y_t}$, który minimalizuje p -oczekiwaną stratę. Jeśli znane jest μ , Λ_μ jest oczywiście najlepszym schematem predykcji w sensie osiągnięcia minimalnej oczekiwanej straty ($L_n^{\Lambda_\mu} \leq L_n^\Lambda$ dla dowolnego Λ). Udowadniamy następujące ograniczenie straty dla uniwersalnego predyktora Λ_ξ

$$0 \leq L_n^{\Lambda_\xi} - L_n^{\Lambda_\mu} \leq D_n + \sqrt{4L_n^{\Lambda_\mu} D_n + D_n^2} \leq 2D_n + 2\sqrt{L_n^{\Lambda_\mu} D_n}. \quad (1.6)$$

Razem z $L_n \leq n$ i $D_\infty \leq \ln w_\mu^{-1} < \infty$, pokazuje to, że $\frac{1}{n} L_n^{\Lambda_\xi} - \frac{1}{n} L_n^{\Lambda_\mu} = O(n^{-1/2})$, tj. asymptotycznie Λ_ξ osiąga optymalną średnią stratę Λ_μ z szybką zbieżnością. Co więcej, $L_\infty^{\Lambda_\xi}$ jest skończone, jeśli $L_\infty^{\Lambda_\mu}$ jest skończone, a $L_n^{\Lambda_\xi} / L_n^{\Lambda_\mu} \rightarrow 1$, jeśli $L_\infty^{\Lambda_\mu}$ nie jest skończone. Ograniczenie (1.6) implikuje również $L_n^\Lambda \geq L_n^{\Lambda_\xi} - 2\sqrt{L_n^{\Lambda_\xi} D_n}$, co pokazuje, że żaden (przyczynowy) predyktor Λ nie osiąga znacząco mniejszej (oczekiwanej) straty niż Λ_ξ . Należy zauważyć, że dla $w_v = 2^{-K(v)}$, $D_n \leq \ln 2 \cdot K(\mu)$ ma „rozsądny” rozmiar. Można również udowodnić natychmiastowe granice strat.

Właściwości optymalności

Dla dowolnego predyktora Λ można wyprowadzić dolną granicę najgorszego przypadku, która asymptotycznie odpowiada górnej granicy (1.6). Dokładniej, niech Λ będzie dowolnym deterministycznym predyktorem nie wiedzącym, z którego rozkładu $\mu \in M$ pobrano próbkę obserwowanej sekwencji $x_1 x_2 \dots$. Predyktor Λ zna (zależy od) M , w_v i ℓ , i ma w czasie t dostęp do poprzednich wyników $x_{<t}$. Wówczas dla każdego n istnieje M i $\mu \in M$ i ℓ i wagi w_v , takie, że

$$L_n^\Lambda - L_n^{\Lambda_\mu} \geq \frac{1}{2} [S_n + \sqrt{4L_n^{\Lambda_\mu} S_n + S_n^2}], \quad \text{and } D_n / S_n \rightarrow 1 \text{ for } n \rightarrow \infty.$$

Dla uniwersalnego predyktora $\Lambda = \Lambda_\xi$ dolna granica obowiązuje nawet bez czynnika 1/2. Pokazuje to, że granica (1.6) jest dość ścisła w tym sensie, że żaden inny predyktor nie może prowadzić do znacznie mniejszych ograniczeń bez dokonywania dodatkowych założeń dotyczących M , w_v lub ℓ . Na przykład

dla logarytmicznych i kwadratowych funkcji straty zał $L_\infty^{\Lambda_\xi} - L_\infty^{\Lambda_\mu}$ jest skończony i ograniczony przez $\ln w_\mu^{-1}$. Innym rodzajem optymalności jest optymalność Pareto. Niech $F(\mu, \rho)$ będzie dowolną miarą wydajności ρ względem μ . Uniwersalny a priori ξ jest nazywany optymalnym Pareto względem FT jeśli nie ma ρ z $F(v, \rho) \leq F(v, \xi)$ dla wszystkich $v \in M$ i ścisłą nierównością dla co najmniej jednego v . Wykazujemy, że uniwersalny prior ξ jest optymalny w sensie Pareto względem kwadratu odległości S_n , względnej entropii D_n i strat L_n . To znaczy, dla wszystkich miar wydajności, które są istotne z punktu widzenia teorii decyzji (tj. dla wszystkich funkcji strat i) każda poprawa osiągnięta przez pewien predyktor Λ_p ponad Λ_ξ w niektórych środowiskach v jest równoważona przez pogorszenie w innych środowiskach. Istnieją miary wydajności niebędące miarami teorii decyzji względem których ξ nie jest optymalne w sensie Pareto. Optymalność Pareto jest raczej słabą koncepcją optymalności, ale podkreśla odrębność strategii mieszanych Bayesa. Optymalność Pareto ξ nadal pozostawia otwartą kwestię, jak wybrać klasę M i wagi w_v . Twierdziliśmy, że M_u jest największą M odpowiednią z obliczeniowego punktu widzenia. M_u jest również wystarczająco duża, jeśli przyjmiemy łagodne założenie, że ciągi są próbkowane z obliczalnego rozkładu prawdopodobieństwa. Pokazujemy, że w

klasie wyliczalnych funkcji wagowych o krótkim programie uniwersalne wagi $w_v = 2^{-k(v)}$ prowadzą do najmniejszych ograniczeń wydajności w obrębie stałej addytywnej (do $\ln w_v^{-1}$) we wszystkich wyliczalnych środowiskach. Ten argument uzasadnia wybór wcześniejszej (1.3) Solomonoffa-Levina spośród wszystkich możliwych mieszanin Bayesa?

Różne

Gry losowe. Ogólny limit strat (1.6) można na przykład wykorzystać do oszacowania czasu potrzebnego do osiągnięcia progu wygranej w grze losowej (zdefiniowanego jako sekwencja zakładów, obserwacji i nagród). W czasie t obstawiamy, w zależności od historii $x_{<t}$, pewną kwotę pieniędzy s_t , podejmujemy jakieś działanie y_t , obserwujemy wynik x_t i otrzymujemy nagrodę r_t . Nasz zysk netto, który chcemy zmaksymalizować, wynosi $p_t = r_t - s_t \in [p_{\max} - p_{\Delta}, p_{\max}]$. Strata, którą chcemy zminimalizować, może być utożsamiana z ujemnym (skalowanym) zyskiem, $\ell_{x_k y_t} = (p_{\max} - p_t)/p_{\Delta} \in [0, 1]$. System Λ_p działa tak, aby zmaksymalizować p -oczekiwany zysk. Niech $\bar{p}_n^{A_p}$ będzie średnim oczekiwanym zyskiem pierwszych n rund. Granica (1.6) pokazuje, że średni zysk systemu Λ_ξ zbiega się do najlepszego możliwego średniego zysku $\bar{p}_n^{A_\mu}$ osiągniętego przez schemat Λ_μ ($\bar{p}_n^{A_\xi} - \bar{p}_n^{A_\mu} = O(n^{-1/2}) \rightarrow 0$ dla $n \rightarrow \infty$). Jeśli w ogóle istnieje dochodowy schemat, to asymptotycznie uniwersalny schemat Λ_ξ stanie się dochodowy z tym

$$(2p_{\Delta}/\bar{p}_n^{A_\mu})^2 \cdot \ln 2 \cdot K(\mu)$$

samym średnim zyskiem. Ponadto pokazujemy za pomocą ξ_u że jest górną granicą liczby zakładów n potrzebnych do osiągnięcia strefy wygranej. Granica jest proporcjonalna do złożoności środowiska μ .

Ciągłe klasy prawdopodobieństwa M.

Do tej pory rozważaliśmy przeliczalne klasy prawdopodobieństwa M , co ma sens z punktu widzenia obliczeniowego. Z drugiej strony w estymacji parametrów statystycznych często występuje ciągła klasa hipotez (np. proces Bernoulliego(θ) z nieznaną $\theta \in [0, 1]$). Niech $\mathcal{M} := \{\mu_\theta : \theta \in \Theta \subseteq \mathbb{R}^a\}$ będzie rodziną rozkładów prawdopodobieństwa sparametryzowanych przez d -wymiarowy parametr ciągły θ . Niech $\mu \equiv \mu_{\theta_0} \in M$, będzie prawdziwym rozkładem generującym. Dla ciągłej gęstości wag $w(\theta) > 0$ sumy w (1.4) są naturalnie zastępowane przez całki: $\xi(x_{1:n}) := \int_{\Theta} w(\theta) \cdot \mu_\theta(x_{1:n}) d\theta$ gdzie $\int_{\Theta} w(\theta) d\theta = 1$. Najważniejszą własnością ξ w przypadku dyskretnym była dominacja $\xi(x_{1:n}) \geq w_v \cdot v(x_{1:n})$, którą uzyskano z (1.4) przez pominięcie sumy ponad v . Analogiczna konstrukcja tutaj polega na ograniczeniu całki ponad Θ do małego otoczenia N_δ z θ . Dla wystarczająco gładkiego μ_θ i $w(\theta)$ oczekujemy $\xi(x_{1:n}) \gtrsim |N_{\delta_n}| \cdot w(\theta) \cdot \mu_\theta(x_{1:n})$, gdzie $|N_{\delta_n}|$ jest objętością N_{δ_n} . To z kolei prowadzi do $D_n \lesssim \ln w_\mu^{-1} + \ln |N_{\delta_n}|^{-1}$, gdzie $w_\mu := w(\theta_0)$. N_{δ_n} powinno być największym możliwym obszarem, w którym $\ln \mu_\theta$ jest średnio w przybliżeniu płaskie. Dokładniej, uogólniając na przypadek nie-i.i.d., pokazujemy $D_n \leq \ln w_\mu^{-1} + \frac{d}{2} \ln \frac{n}{2\pi} + O(1)$, gdzie człon $O(1)$ zależy od gładkości μ_θ , mierzonej informacją Fishera. D_n nie jest już ograniczone przez stałą, ale nadal rośnie tylko logarytmicznie z n , intuicyjnym powodem jest konieczność opisanie θ z dokładnością $O(n^{-1/2})$. Tak więc ograniczenie (1,6) jest również stosowalne w przypadku klas prawdopodobieństwa o ciągłej parametryzacji.

Agenci racjonalni w znanych środowiskach probabilistycznych

Model agenta

Bardzo ogólnym frameworkiem dla systemów inteligentnych jest framework agentów racjonalnych. W cyklu k agent wykonuje działanie $y_k \in Y$ (wyjście), które skutkuje percepcją $x_k \in X$ (wejście), po którym

następuje cykl $k+1$ itd. Zakładamy, że przestrzenie działania i percepcji X i Y są skończone. Piszemy $p\{x_{<k}\} = y_{1:k}$, aby oznaczyć wyjście $y_{1:k}$ polityki agenta p na wejściu $x_{<k}$, i podobnie $q\{y_{1:k}\} = x_{1:k}$ dla środowiska q w przypadku środowisk deterministycznych. Politykę p i środowisko q zachowujące się w ten sposób nazywamy chronologicznymi. Należy zauważyć, że polityka i środowisko mogą zależeć od całej historii. Nie przyjmujemy tutaj żadnych założeń MDP ani POMDP i nie mówimy o stanach środowiska, tylko o obserwacjach. W bardziej ogólnym przypadku środowiska probabilistycznego, biorąc pod uwagę historię $y_{<k}y_k \equiv y_1 \dots y_{k-1}y_k \equiv y_1x_1 \dots y_{k-1}x_{k-1}y_k$, prawdopodobieństwo, że środowisko prowadzi do percepcji x_k w cyklu k wynosi (z definicji) $\mu(y_{<k}y_k)$ podkreślony argument y_k w μ jest zmienną losową, a pozostałe niepodkreślone argumenty $y_{<k}y_k$ reprezentują warunki? Rozkłady prawdopodobieństwa, takie jak μ , nazywamy chronologicznymi. Ponieważ polityki optymalizujące wartość zawsze mogą być wybrane jako deterministyczne, nie ma rzeczywistej potrzeby uogólniania ustawień na polityki probabilistyczne.

Funkcje wartości i optymalne zasady

Celem agenta jest maksymalizacja przyszłych nagród, które są dostarczane przez środowisko poprzez dane wejściowe x_k . Dane wejściowe $x_k \equiv r_k o_k$ są podzielone na regularną część o_k i pewną (być może pustą lub opóźnioną) nagrodę $r_k \in [0, r_{\max}]$. Używamy skrótu

$$\mu(y_{<k}y_{k:m}) = \mu(y_{<k}y_k) \cdot \mu(y_{1:k}y_{k+1}) \cdot \dots \cdot \mu(y_{<m}y_m),$$

która jest zasadniczo regułą łańcuchową, a $\epsilon - y_{x_{<1}}$ dla pustego ciągu. Definiujemy (całkowitą) wartość polityki p w środowisku μ lub krócej, wartość μ z p , jako sumę oczekiwanej nagrody μ

$$V_\mu^p := \sum_{x_{1:m}} (r_1 + \dots + r_m) \mu(y_{1:m}) | y_{1:m} = p(x_{<m}), \quad (1.7)$$

gdzie m jest czasem życia lub początkowym horyzontem agenta. Optymalna polityka p^μ maksymalizująca wartość V_μ^p to

$$p^\mu := \arg \max_p V_\mu^p, \quad V_\mu^* := V_\mu^{p^\mu} = \max_p V_\mu^p \geq V_\mu^p \quad \forall p.$$

Polityka p^μ , którą nazywamy modelami Al_μ , jest optymalna w tym sensie, że żadna inna polityka dla agenta nie prowadzi do wyższej μ -oczekiwanej nagrody. Wyraźne wyrażenia dla działania y_k w cyklu k μ -optymalnej polityki p^μ i ich wartość V_μ^* są

$$y_k = y_k^\mu := \arg \max_{y_k} \sum_{x_k} \max_{y_{k+1}} \sum_{x_{k+1}} \dots \max_{y_m} \sum_{x_m} (r_k + \dots + r_m) \cdot \mu(y_{<k}y_{k:m}), \quad (1.8)$$

$$V_\mu^* = \max_{y_1} \sum_{x_1} \max_{y_2} \sum_{x_2} \dots \max_{y_m} \sum_{x_m} (r_1 + \dots + r_m) \cdot \mu(y_{1:m}), \quad (1.9)$$

gdzie $y_{<k}$ jest faktyczną historią. Pokazujemy, że te definicje są spójne i poprawnie oddają nasz zamiar. Na przykład rozważ wyrażenie expectimax (1.9): Najlepsza oczekiwana nagroda jest uzyskiwana przez uśrednienie możliwych percepcji x_i i maksymalizację możliwych działań y_i . Należy to zrobić w kolejności chronologicznej $y_1x_1 \dots y_mx_m$, aby poprawnie uwzględnić zależności x_i i y_i od historii. To jest źródło

naprzemiennej sekwencji expectimax, która jest podobna do dobrze znanej sekwencji/drzewa/algorytmu minimax w teorii gier.

Sekwencyjna teoria decyzji i: Uczenie się przez wzmacnianie

Można powiązać (1.9) z równaniami Bellmana sekwencyjnej teorii decyzji, identyfikując kompletne historie $y_{x_{<k}}$ ze stanami $\mu(y_{x_{<k}})$ z macierzą przejść stanów, V_μ^* z funkcją wartości i y_k z działaniem w cyklu k . Ze względu na użycie kompletnej historii jako przestrzeni stanów, model Al_μ nie zakłada ani stacjonarności, ani własności Markowa, ani całkowitej dostępności środowiska. Każdy stan występuje co najwyżej raz w okresie istnienia systemu. Z tego i innych powodów jawne sformułowanie (1.8) jest tutaj bardziej naturalne i użyteczne niż wymuszanie pseudorekurencyjnej formy równania Bellmana. Ponieważ mamy na myśli uniwersalny system ze złożonymi interakcjami, przestrzenie działania i percepcji Y i X są ogromne (np. obrazy wideo), a każde działanie lub samo postrzeganie występuje zwykle tylko raz w okresie istnienia m agenta. Ponieważ nie ma (oczywistej) uniwersalnej relacji podobieństwa w przestrzeni stanów, skuteczna redukcja jej rozmiaru jest niemożliwa, ale nie ma zasadniczego problemu w określeniu y_k z (1.8), o ile μ jest znane i obliczalne, a X, Y i m są skończone. Rzeczy drastycznie się zmieniają, jeśli μ jest nieznane. Algorytmy uczenia się przez wzmacnianie są powszechnie używane w tym przypadku do nauki nieznanego μ lub bezpośrednio jego wartości. Udaje im się to, jeśli przestrzeń stanów jest mała lub została skutecznie zmniejszona przez techniki generalizacji lub aproksymacji funkcji. W każdym przypadku rozwiązania są albo ad hoc, albo działają tylko w ograniczonych domenach, mają poważne problemy z eksploracją przestrzeni stanów w porównaniu z eksploatacją, są podatne na rozbieżności lub mają nieoptymalne wskaźniki uczenia się. Jak dotąd nie ma uniwersalnego i optymalnego rozwiązania tego problemu. Głównym tematem tej książki jest przedstawienie nowego modelu i argumentowanie, że formalnie rozwiązuje on wszystkie te problemy w optymalny sposób. Prawdziwy rozkład prawdopodobieństwa μ nie zostanie poznany bezpośrednio, ale zostanie zastąpiony przez pewien uogólniony uniwersalny rozkład a priori ξ_U , który zbiega się do μ , podobnie jak w przypadku indukcji (SP).

Uniwersalny Agent Algorytmiczny AIXI

Opracowaliśmy wystarczająco dużo formalizmów, aby przedstawić uniwersalny model AIXI. Wszystko, co musimy zrobić, to odpowiednio uogólnić uniwersalny prior Solomonoffa M i zastąpić prawdziwe, ale nieznane prawdopodobieństwo μ w modelu Al_μ tym uogólnionym M . Podobnie jak w (1.1), definiujemy M jako ważoną sumę $2^{-\ell(q)}$ wszystkich programów chronologicznych (środowisk) q , które wyprowadzają $x_{1:k}$ ale z $y_{1:k}$ dostarczonym na taśmie wejściowej. To również uogólnia ξ_U (w obrębie nieistotnej stałej mnożnikowej):

$$\xi(y_{1:k}) = \xi_U(y_{1:k}) \stackrel{\propto}{=} M(y_{1:k}) := \sum_{q: q(y_{1:k})=x_{1:k}} 2^{-\ell(q)}. \quad (1.10)$$

Jeżeli nie wynika to jasno z kontekstu, dodajemy indeksy górne SP i AI do ξ , aby rozwiązać niejasności między (1.3) i (1.10). Zastąpienie μ przez ξ w (1.8) powoduje, że system AIXI generuje

$$y_k = y_k^\xi := \arg \max_{y_k} \sum_{x_k} \dots \max_{y_m} \sum_{x_m} (r_k + \dots + r_m) \cdot \xi(y_{x_{<k}}) \quad (1.11)$$

w cyklu k biorąc pod uwagę historię $y_{x_{<k}}$. Wartość ξV_ξ^p i wartość uniwersalna V_ξ^* są zdefiniowane jak w (1.7) i (1.9), przy czym μ zastąpiono przez ξ . Model AIXI i jego zachowanie są całkowicie zdefiniowane przez (1.10) i (1.11). Zależy to (w niewielkim stopniu) od wyboru uniwersalnej maszyny Turinga,

ponieważ $K()$ i $\ell()$ zależą od U i stąd są zdefiniowane tylko z dokładnością do członów rzędu pierwszego. Model AIXI zależy również od wyboru X i Y , ale nie spodziewamy się żadnego odchylenia, gdy przestrzenie są wybrane wystarczająco duże i proste, np. wszystkie ciągi o długości 2^{16} . Wybranie IN jako przestrzeni I/O byłoby idealne, ale czy w tym przypadku istnieją maksima (lub suprema) musi zostać wcześniej pokazane. Jedyną nietrywialną zależnością jest horyzont m . W idealnym przypadku chcielibyśmy wybrać $m = \infty$, ale jest kilka subtelności do rozwikłania później, które uniemożliwiają przynajmniej naiwną granicę $m \rightarrow \infty$. Tak więc pomijając m i nieistotne szczegóły, układ AIXI jest jednoznacznie zdefiniowany przez (1.10) i (1.11) bez regulowanych parametrów.

O optymalności AIXI

Uniwersalność i zbieżność ξ . Można pokazać, że również ξ zdefiniowane w (1.10) jest uniwersalne i szybko zbiega do μ analogicznie do przypadku indukcji (SP). Jeśli weźmiemy skończony iloczyn warunkowych ξ i zastosujemy regułę łańcuchową, zobaczymy, że również $\xi(\underline{y}^{k:k+h} < \underline{y}^{k:k+h})$ zbiega do $\mu(\underline{y}^{k:k+h} < \underline{y}^{k:k+h})$ dla $k \rightarrow \infty$. Daje to pewność, że wyjścia y_k^ξ modelu AIXI (1.11) mogą zbiegać się do wyjść y_k^μ modelu $Al\mu$ (1.8), przynajmniej dla ograniczonego horyzontu ruchomego h . Problemy ze stałym horyzontem m , a zwłaszcza $m \rightarrow \infty$, zostaną omówione na końcu tej sekcji.

Uniwersalnie optymalne systemy AI.

Nazywamy model AI uniwersalnym, jeśli jest niezależny od prawdziwego środowiska μ (bezstronny, niezależny od modelu) i jest w stanie rozwiązać każdy rozwiązywalny problem i nauczyć się każdego zadania, którego można się nauczyć. Ponadto, nazywamy model uniwersalny uniwersalnie optymalnym, jeśli nie ma programu, który mógłby rozwiązać lub nauczyć się znacznie szybciej (w kategoriach cykli interakcji). Ponieważ model AIXI jest bezparametrowy, ξ zbiega się do μ , model $Al\mu$ jest sam w sobie optymalny i nie spodziewamy się, że żaden inny model nie zbiegnie się szybciej do $Al\mu$ przez analogię do przypadku SP, spodziewamy się, że AIXI będzie uniwersalnie optymalny. To jest nasze główne twierdzenie. Dalsze wsparcie podano poniżej.

Relacja porządku inteligencji. Chcemy nazwać politykę p bardziej lub równie inteligentną niż polityka p' i zapisać $p \succeq p'$, jeśli p daje w każdym cyklu k i dla każdej ustalonej historii $y_{<k}$ wyższą (przyszłą) ξ -oczekiwaną sumę nagród niż p' . Jest to formalne ćwiczenie, aby pokazać, że $p^\xi \succeq p$ dla wszystkich p . Model AIXI jest zatem najinteligentniejszym agentem w.r.t. \succeq . Relacja \succeq jest uniwersalną relacją porządku w tym sensie, że jest wolna od jakichkolwiek parametrów (oprócz m) lub określonych założeń dotyczących środowiska. Dowód, że \succeq jest rozsądnym porządkiem inteligencji (co uważamy za prawdziwe) udowodniłby, że AIXI jest uniwersalnie optymalny.

Ograniczenia wartości

Wartości V_p^* związane z układami Alp odpowiadają mniej więcej ujemnej całkowitej stracie $-L_n^{Ap}(z_{n=m})$ układów SPp (=Ap). W przypadku SP interesowały nas małe ograniczenia dla żalu $L_n^{\Delta\xi} - L_n^{\Delta\mu}$. Niestety, proste ograniczenia wartości dla AIXI lub dowolnego innego systemu AI w kategoriach V_p^* analogiczne do ograniczenia strat (1.6) nie mogą być spełnione. Mamy nawet trudności ze sprecyzowaniem, czego możemy oczekiwać dla AIXI lub dowolnego systemu AI, który twierdzi, że jest uniwersalnie optymalny. W SP jedyną ważną własnością μ do udowodnienia ograniczeń strat jest jego złożoność $K(\mu)$. W przypadku AI nie ma żadnych użytecznych ograniczeń w kategoriach $K(ii)$. Musimy albo zbadać ograniczony problem lub klasy środowiskowe, albo rozważyć ograniczenia zależne od innych właściwości μ , a nie tylko od jego złożoności.

Wyniki optymalizacji związanej z wartością

Rozkład mieszanin ξ . W dalszej części rozważamy ogólne mieszaniny Bayesa ξ ponad klasami AI chronologicznych rozkładów prawdopodobieństwa ν :

$$\xi(\underline{y}_{1:m}) = \sum_{\nu \in \mathcal{M}} w_\nu \nu(\underline{y}_{1:m}) \quad \text{with} \quad \sum_{\nu \in \mathcal{M}} w_\nu = 1 \quad \text{and} \quad w_\nu > 0 \quad \forall \nu \in \mathcal{M}.$$

Definiujemy V_ξ^p , p^ξ i V_ξ^* jak w (1.7)-(1.9) z μ zastąpionym przez ξ . Strategię p^ξ nazywamy modelem AI ξ . Dla $\xi = \xi_u$ odzyskiwany jest model AI $XI \equiv AI\xi_u$. Jeśli μ jest nieznane, ale wiadomo, że należy do znanej klasy \mathcal{M} , naturalne jest postępowanie zgodnie z polityką p^ξ , która maksymalizuje V_ξ^p . (Prawdziwa μ -)oczekiwana nagroda przy postępowaniu zgodnie z polityką p^ξ wynosi V_ξ^p . Optymalna (ale niewykonalna) polityka p^μ przynosi nagrodę $V_\xi^{\mu} = V_\mu^*$. Teraz interesujące jest (a) czy istnieją polityki o wartości jednostajnie większej niż V_ξ^p i (b) jak blisko V_ξ^p jest do V_μ^* .

Liniowość i wypukłość V_p w p . Następujące właściwości V_p są kluczowe. V_p jest funkcją liniową w p , a V_p^* jest funkcją wypukłą w p w tym sensie, że

$$V_\xi^p = \sum_{\nu \in \mathcal{M}} w_\nu V_\nu^p \quad \text{and} \quad V_\xi^* \leq \sum_{\nu \in \mathcal{M}} w_\nu V_\nu^*.$$

Liniowość jest oczywista z definicji V_p , a wypukłość wynika łatwo z wypukłości \max_p i nieujemności wag w_ν . Jedną z luźnych interpretacji wypukłości jest to, że mieszanka nigdy nie może zwiększyć wydajności.

Optymalność Pareto AI ξ . Podobnie jak w przypadku SP, można pokazać, że p^ξ jest optymalne w sensie Pareto w tym sensie, że nie ma innej polityki p z $V_\nu^p \geq V_\nu^{p^\xi}$ dla wszystkich $\nu \in \mathcal{M}$ i ścisła nierówność dla co najmniej jednego ν . W szczególności AI XI jest optymalne w sensie Pareto.

Samoptymalizująca się polityka p^ξ względem wartości średniej. Ponieważ nie znamy z góry prawdziwego środowiska μ , jesteśmy zainteresowani, w jakich okolicznościach

$$\frac{1}{m} V_\nu^{p^\xi} \rightarrow \frac{1}{m} V_\nu^* \quad \text{for horizon } m \rightarrow \infty \quad \text{for all } \nu \in \mathcal{M}. \quad (1.12)$$

Należy zauważyć, że zarówno V_ν , jak i $p^\xi = p_m^\xi$ zależą od m . Najmniej, co musimy wymagać od \mathcal{M} , aby mieć szansę, że (1.12) jest prawdą, to aby w ogóle istniała polityka (ciąg) $\bar{p} = \bar{p}_m$ mająca tę własność, tj.

$$\exists \bar{p}: \frac{1}{m} V_\nu^{\bar{p}} \rightarrow \frac{1}{m} V_\nu^* \quad \text{for horizon } m \rightarrow \infty \quad \text{for all } \nu \in \mathcal{M}. \quad (1.13)$$

Wykazujemy, że ten konieczny warunek jest również wystarczający, tj. (1.13) implikuje (1.12). Jest to kolejna (asymptotyczna) własność optymalności polityki p^ξ . Jeśli uniwersalna konwergencja w sensie (1.13) jest w ogóle możliwa w klasie środowisk \mathcal{M} , wówczas polityka p^ξ zbiega się w tym samym sensie (1.12). Polityki \bar{p} z własnością taką jak (1.13) nazywamy samoptymalizującymi. Niestety, wynik nie jest stwierdzeniem asymptotycznej konwergencji pojedynczej polityki p^ξ ponieważ p^ξ zależy od m . Wynik mówi jedynie, że w podanych warunkach średnia wartość p_m^ξ jest dowolnie bliska optimum dla wystarczająco dużego (wstępnie wybranego) horyzontu m . Ta słabość zostanie rozwiązana w następujący sposób.

Zdyskontowana funkcja wartości przyszłej. Teraz przenosimy naszą uwagę z wartości całkowitej na wartości przyszłe (wartość do przejścia). Najpierw musimy pozbyć się parametru horyzontu m.

Eliminujemy horyzont, dyskontując nagrody $r_k \rightsquigarrow \gamma_k r_k$ przy $\gamma_k \geq 0$ i $\sum_{i=1}^{\infty} \gamma_i < \infty$ biorąc $m \rightarrow \infty$. Analogiem m jest teraz efektywny horyzont h_k^{eff} , który można zdefiniować jako $\sum_{i=k}^{k+h_k^{eff}} \gamma_i \approx \sum_{i=k+h_k^{eff}}^{\infty} \gamma_i$. Co więcej, renormalizujemy wartość V przez $\sum_{i=k}^{\infty} \gamma_i$ oznaczając ją jako $V_{k\gamma}$. Na koniec rozszerzamy definicję na polityki probabilistyczne π (co nie jest niezbędne). Definiujemy γ -dyskontowaną średnią ważoną wartość przyszłą polityki (probabilistycznej) π w środowisku ρ przy danej historii $y_{x_{<k}}$, lub krócej, wartość ρ π przy danej $y_{x_{<k}}$

jako

$$V_{k\gamma}^{\pi\rho}(y_{x_{<k}}) := \frac{1}{\Gamma_k} \lim_{m \rightarrow \infty} \sum_{y_{x_{k:m}}} (\gamma_k r_k + \dots + \gamma_m r_m) \rho(y_{x_{<k}} | y_{x_{k:m}}) \pi(y_{x_{<k}} | y_{x_{k:m}}),$$

$$\Gamma_k := \sum_{i=k}^{\infty} \gamma_i.$$

z

Polityka p^ρ jest zdefiniowana tak, aby maksymalizować przyszłą wartość $V_{k\gamma}^{\pi\rho}$:

$$p^\rho := \arg \max_{\pi} V_{k\gamma}^{\pi\rho}, \quad V_{k\gamma}^{*\rho} := V_{k\gamma}^{p^\rho} = \max_{\pi} V_{k\gamma}^{\pi\rho} \geq V_{k\gamma}^{\pi\rho} \forall \pi.$$

Ustawienie $\gamma_k = 1$ dla $k \leq m$ i $\gamma_k = 0$ dla $k > m$ zwraca stary niezdykontowany model z horyzontem m i

$V_{1\gamma}^{p\rho} = \frac{1}{m} V_{\rho}^p$. Zauważ, że $V_{k\gamma}$ zależy od zrealizowanej historii $y_{x_{<k}}$. Co ważniejsze, można wykazać, że p^ρ jest niezależne od k. Podobnie jak w przypadku niezdykontowanym, można udowodnić, że dla

każdego k i historii $y_{x_{<k}}$, $V_{k\gamma}^{\pi\rho}$ jest funkcją liniową w ρ , $V_{k\gamma}^{*\rho}$ jest funkcją wypukłą w ρ , a p^ρ jest

optymalne w sensie Pareto w tym sensie, że nie ma innej polityki π z $V_{k\gamma}^{\pi\nu} \geq V_{k\gamma}^{p^\rho\nu}$ dla wszystkich $\nu \in M$ i ścisłej nierówności dla co najmniej jednego ν . Na koniec, p^ρ jest samooptymalizujące (w odniesieniu do wartości zdyskontowanej), jeśli AA dopuszcza samooptymalizujące polityki:

$$\text{If } \exists \tilde{\pi} \forall \nu : V_{k\gamma}^{\tilde{\pi}\nu} \xrightarrow{k \rightarrow \infty} V_{k\gamma}^{*\nu} \text{ w.l.p.1} \implies V_{k\gamma}^{p^\rho\nu} \xrightarrow{k \rightarrow \infty} V_{k\gamma}^{*\nu} \text{ w.l.p.1.}$$

Kwalifikator prawdopodobieństwa odnosi się do historycznych percepcji $x_{<k}$. Historyczne działania $y_{<k}$ są dowolne. Należy zauważyć, że k jest rzeczywistą wartością bieżącą, mianowicie bieżącym numerem cyklu, podczas gdy m było wstępnie wybranym stałym horyzontem.

Procesy decyzyjne Markowa

Spośród wszystkich możliwych środowisk procesy Markowa (decyzyjne) są prawdopodobnie najbardziej intensywnie badane. μ nazywa się (całkowicie obserwowalnym stacjonarnym) procesem decyzyjnym Markowa (MDP), jeśli prawdopodobieństwo postrzegania $x_k \in X$, przy danej historii $y_{x_{<k}}$, zależy tylko od ostatniej akcji $y_k \in Y$ i ostatniej percepcji x_{k-1} , tj. jeśli $\mu(y_{x_{<k}} | y_k, x_k) = \mu(x_{k-1} | y_k, x_k)$. W tym przypadku x_k nazywa się stanem, X przestrzenią stanów, a $f^{\mu}(x_{k-1} | y_k, x_k)$ macierzą przejść.

Proces MDP μ nazywa się ergodycznym, jeśli istnieje polityka, w ramach której każdy stan jest odwiedzany nieskończenie często z prawdopodobieństwem 1. Jeśli proces $\text{MDP } \mu(x_{k-1}y_kx_k)$ jest niezależny od działania y_k , jest to proces Markowa; jeśli jest niezależny od ostatniego postrzeżenia x_{k-1} , jest to proces i.i.d. Stacjonarne procesy MDP μ z dyskontowaniem geometrycznym $\gamma_k = \gamma^k$ mają stacjonarne optymalne polityki p^μ odwzorowujące ten sam stan/percepcję x_k zawsze na to samo działanie y_k . Z drugiej strony, mieszanka ξ MDP sama w sobie nie jest MDP, tj. $\xi \notin \mathcal{M}_{\text{MDP}}$; co oznacza, że p^ξ ogólnie nie jest polityką stacjonarną. Można skonstruować samooptymalizujące polityki dla klasy ergodycznych procesów MDP względem wartości średniej $\frac{1}{m}V_m^p$ i jeśli $\frac{\gamma_{k+1}}{\gamma_k} \rightarrow 1$ również względem zdyskontowanej wartości przyszłej $V_{k\gamma}^{\mu p}$. Warunek konieczny $\frac{\gamma_{k+1}}{\gamma_k} \rightarrow 1$ zapewnia nieograniczenie rosnący efektywny horyzont h^{eff}_k . Istnienie samooptymalizujących się polityk dla ergodycznych MDP oznacza, że dla przeliczalnej klasy M ergodycznych MDP, polityki p^ξ_m maksymalizujące V^ξ i p^ξ maksymalizujące $V^{\mu^\xi}_{k\gamma}$ są samooptymalizujące w tym sensie, że

$$\forall \nu \in M: \frac{1}{m}V_{1m}^{p^\xi \nu} \xrightarrow{m \rightarrow \infty} \frac{1}{m}V_{1m}^{*\nu} \quad \text{and} \quad V_{k\gamma}^{p^\xi \nu} \xrightarrow{k \rightarrow \infty} V_{k\gamma}^{*\nu} \quad \text{if} \quad \frac{\gamma_{k+1}}{\gamma_k} \rightarrow 1. \quad (1.14)$$

Pokazujemy również, że jeśli M jest skończone, to prędkość pierwszej zbieżności wynosi co najmniej $O(m^{-1/3})$. Warunki $\Gamma_k < \infty$ i $\frac{\gamma_{k+1}}{\gamma_k} \rightarrow 1$ w ciągu dyskontowym są na przykład spełnione dla $\gamma_k = 1/k^2$, ale nie dla popularnego dyskonta geometrycznego $\gamma_k = \gamma^k$, który ma skończony horyzont efektywny. Granice (1.14) pokazują, że p^ξ jest samooptymalizujące dla procesów bandytów, procesów i.i.d. i zadań klasyfikacyjnych, ponieważ są one szczególnymi (zdegenerowanymi) przypadkami ergodycznych MDP. Istnienie samooptymalizujących polityk nie ogranicza się do (podklas ergodycznych) MDP. Pewne klasy POMDPS, k -tego rzędu ergodycznych MDP, środowiska faktoryzowalne, gry powtarzalne i problemy predykcyjne nie są MDP, ale mimo to dopuszczają samooptymalizujące polityki. Stąd odpowiednia polityka optymalnej mieszanki Bayesa p^ξ jest samooptymalizująca.

Wybór horyzontu

Jedyną znaczącą arbitralnością w modelu AIXI jest wybór okresu życia m lub w przypadku dyskontowanym w sekwencji dyskontowej γ_k . Nie będziemy omawiać wyborów ad hoc dla konkretnych problemów. Interesują nas wybory uniwersalne. W wielu przypadkach czas, w którym jesteśmy skłonni uruchomić system, zależy od jakości jego działań. Stąd czas życia, jeśli w ogóle skończony, nie jest znany z góry. Dyskontowanie geometryczne $r_k \sim r_k \cdot \gamma^k$ rozwiązuje problem matematyczny $m \rightarrow \infty$, ale nie jest rozwiązaniem rzeczywistym, ponieważ wprowadzono efektywny horyzont $h^{\text{eff}} \sim \ln \gamma^{-1} < \infty$. Niezmiennicza skala dyskontująca $r_k \sim r_k \cdot k^{-\alpha}$, z $\alpha > 1$ ma dynamiczny horyzont $h \sim k$. Ten wybór jest w pewnym stopniu atrakcyjny, ponieważ wydaje się, że ludzie w wieku k lat również zazwyczaj nie planują swojego życia na dłużej niż następne $\sim k$ lat. Spełnia również warunek $\frac{\gamma_{k+1}}{\gamma_k} \rightarrow 1$, konieczny dla AI ξ jako samooptymalizującego się w ergodycznych MDPS. Największy niższy półobliczalny horyzont z gwarantowaną skończoną sumą nagród $\Gamma_1 < \infty$ jest uzyskiwany przez dyskonto $r_k \sim r_k \cdot 2^{-K(k)}$, gdzie $K(k)$ jest złożonością Kolmogorowa k . Jest to prawdopodobnie najbardziej atrakcyjny uniwersalny dyskonto. Jest to podobne do dyskonta niemal harmonicznego $r_k \sim r_k \cdot k^{-(1+\epsilon)}$ ponieważ $2^{-K(k)} \leq 1/k$ dla większości k i $2^{-K(k)} \geq c/(\log^2 k)$ dla pewnej stałej c . Nie jesteśmy pewni, czy wybór horyzontu ma marginalne znaczenie, o ile jest on wystarczająco duży, czy też okaże się centralnym tematem dla modelu AIXI lub dla aspektu planowania dowolnego uniwersalnego systemu AI w ogólności. Większość, jeśli nie wszystkie, problemy w

projektowaniu agentów równoważących eksplorację i eksploatację znikają dzięki wystarczająco dużemu wyborowi (efektywnego) horyzontu i wystarczająco ogólnemu priori.

Ważne klasy środowiskowe

W tej i następnej sekcji definiujemy $\xi = \xi_U \cong M$ jako klasę a priori Solomonoffa, tj. $AI\xi = AIXI$. Każda podsekcja stanowi streszczenie tego, co zostanie zrobione w odpowiedniej sekcji.

Wprowadzenie

Aby zapewnić dalsze wsparcie dla uniwersalności i optymalności teorii $AI\xi$, stosujemy $AI\xi$ do szeregu klas problemów. Obejmują one przewidywanie sekwencji, gry strategiczne, minimalizację funkcji, a zwłaszcza sposób, w jaki $AI\xi$ uczy się uczyć pod nadzorem. Dla niektórych klas podajemy konkretne przykłady, aby rzucić światło na zakres klasy problemów. Najpierw formułujemy każdą klasę problemów w jej naturalny sposób (gdy μ^{problem} jest znany), a następnie konstruujemy formułę w modelu $AI\mu$ i udowadniamy jej równoważność. Następnie rozważamy konsekwencje zastąpienia μ przez ξ . Głównym celem jest zrozumienie, dlaczego i w jaki sposób problemy są rozwiązywane przez $AI\xi$. Podkreślamy tylko szczególne aspekty każdej klasy problemów. Celem jest lepsze zobrazowanie elastyczności modelu $AI\xi$.

Przewidywanie sekwencji (SP)

Używanie modelu $AI\mu$ do przewidywania sekwencji (SP) jest identyczne z bayesowskim przewidywaniem sekwencji $SP\mu$. Można by się spodziewać, że przy użyciu modelu $AI\xi$ do przewidywania sekwencji, odzyska się dokładnie uniwersalny schemat przewidywania sekwencji $SP\xi$, ponieważ $AI\xi$ było unifikacją modelu $AI\mu$, i idei uniwersalnego prawdopodobieństwa ξ . Niestety tak nie jest. Jednym z powodów jest to, że ξ jest tylko rozkładem prawdopodobieństwa w danych wejściowych x , a nie w danych wyjściowych y . Jest to również jeden z powodów trudności w dowodzeniu granic strat/wartości dla $AI\xi$. Niemniej jednak twierdzimy, że $AI\xi$ jest równie dobrze przystosowane do przewidywania sekwencji jak $SP\xi$. W bardzo ograniczonym zakresie dowodzimy (słabej) granicy błędu dla $AI\xi$, co daje nadzieję, że ogólny dowód jest osiągalny.

Gry strategiczne (SG)

Bardzo ważną klasą problemów są gry strategiczne (SG). Ograniczamy się do deterministycznych, ściśle konkurencyjnych gier strategicznych, takich jak szachy. Jeśli środowisko jest graczem minimax, sam model $AI\mu$ redukuje się do strategii minimax. Powtarzające się gry o stałej długości są szczególnym przypadkiem faktoryzowalnego μ . Szkicowane są konsekwencje zmiennej długości gry. Model $AI\xi$ musi nauczyć się reguł rozważanej gry, ponieważ nie ma wcześniejszych informacji o tych regułach. Opisujemy, w jaki sposób $AI\xi$ faktycznie uczy się tych reguł.

Minimalizacja funkcji (FM)

Wiele problemów mieści się w kategorii „minimalizacji funkcji ograniczonej zasobami” (FM). Należą do nich problem komiwojażera, minimalizowanie kosztów produkcji, wynajdywanie nowych materiałów, a nawet produkcja, np. ładnych obrazów, które są (subiektywnie) oceniane przez człowieka. Zadanie polega na (w przybliżeniu) zminimalizowaniu pewnej funkcji $f : Y \rightarrow Z$ w minimalnej liczbie wywołań funkcji. Zobaczymy, że zachłanny model próbujący zminimalizować / w każdym cyklu zawodzi. Chociaż model chciwy nie ma nic wspólnego z technikami downhill lub gradientowymi (nie ma nic takiego jak gradient lub kierunek dla funkcji powyżej Y), które są znane z tego, że zawodzą, odkrywamy te same trudności. FM ma już prawie pełną złożoność ogólnej AI. Powodem jest to, że FM może aktywnie wpływać na proces gromadzenia informacji poprzez swoje próby y_k (podczas gdy SP i CF= klasyfikacja

nie mogą). Omawiamy szczegółowo optymalny model $FM\mu$ i jego pomysłowość w wyborze $y \in Y$. Następnie omawiamy subtelnosci przy używaniu $AI\xi$ do minimalizacji funkcji.

Nadzorowane uczenie się na przykładach (EX)

Uczenie się przez wzmacnianie, jak w przypadku modelu $AI\xi$, jest ważną techniką uczenia się, ale nie jedyną. Aby zwiększyć szybkość uczenia się, konieczne jest nadzorowane uczenie się, tj. uczenie się poprzez zdobywanie wiedzy lub uczenie się od konstruktywnego nauczyciela. Pokazujemy, jak $AI\xi$ uczy się uczyć pod nadzorem. W rzeczywistości bardzo szybko ustanawia nadzorowane uczenie się w ciągu $O(1)$ cykli.

Inne aspekty inteligencji

Na koniec przedstawiamy krótki przegląd innych ogólnych aspektów, idei i metod w AI oraz ich związku z modelem $AI\xi$. Niektóre aspekty są bezpośrednio zawarte w modelu $AI\xi$, podczas gdy inne są lub powinny się pojawiać.

Aspekty obliczeniowe

Do tej pory pokazaliśmy uniwersalny charakter modelu $AIXI$, ale całkowicie zignorowaliśmy aspekty obliczeniowe. Zaczynamy od opracowania algorytmu ML, który jest w stanie rozwiązać każdy dobrze zdefiniowany problem p tak szybko, jak najszybszy algorytm obliczający rozwiązanie dla p , z wyjątkiem czynnika $1+\epsilon$ i członów addytywnych niższego rzędu. Na podstawie podobnego pomysłu konstruujemy następnie obliczalną wersję modelu $AIXI$.

Najszybszy i najkrótszy algorytm dla wszystkich dobrze zdefiniowanych problemów

Wprowadzenie. Szeroką klasę problemów można sformułować w następujący sposób. Biorąc pod uwagę formalną specyfikację $f: X \rightarrow Y$ problemu zależnego od pewnego parametru $x \in X$, jesteśmy zainteresowani szybkim algorytmem obliczającym rozwiązanie $y \in Y$. Przeszukiwanie Levina jest (w ramach dużego stałego czynnika) najszybszym algorytmem odwracającym funkcję $g: Y \rightarrow X$, jeśli g można szybko ocenić. Przeszukiwanie Levina może również obsługiwać ograniczone czasowo problemy optymalizacji. Rozkład na czynniki pierwsze, kolorowanie grafów i przypisywanie prawdziwości to przykłady problemów odpowiednich dla przeszukiwania Levina, jeśli chcemy znaleźć rozwiązanie, ponieważ weryfikacja jest szybka. Przeszukiwanie Levina nie może rozstrzygnąć odpowiadających im problemów decyzyjnych. Nie ma również zastosowania np. do mnożenia macierzy i uczenia się przez wzmacnianie, ponieważ zadanie weryfikacji g jest tak trudne, jak zadanie obliczeniowe. Twierdzenie Bluma o przyspieszeniu pokazuje, że istnieją typy problemów f dla których istnieje (nieobliczalny) ciąg algorytmów zwiększających szybkość (o rosnącym rozmiarze), ale nie ma najszybszego algorytmu.

W przedstawionym tutaj podejściu rozważamy tylko te algorytmy, które dowodzą, że rozwiązują dany problem i mają szybkie (tj. szybko obliczalne) ograniczenie czasowe. Ani same programy, ani dowody nie muszą być znane z góry. Przy tych ograniczeniach konstruujemy asymptotycznie najszybszy algorytm, oszczędzając czynnik $1 + \epsilon$, który rozwiązuje każdy dobrze zdefiniowany problem f

Szybki algorytm $M_{p^*}^\epsilon$. Niech p^* będzie danym algorytmem obliczającym $p^*(x)$ z x lub, bardziej ogólnie, specyfikacją funkcji f . Jednym ze składników naszego najszybszego algorytmu $M_{p^*}^\epsilon$ do obliczania $p^*(x)$ jest wyliczenie dowodów o rosnącej długości w pewnym formalnym systemie aksjomatycznym. Jeśli dowód faktycznie udowadnia, że pewne p jest funkcjonalnie równoważne p^* , a p ma ograniczenie czasowe t_p , krotka (p, t_p) jest dodawana do listy L . Program p w L z aktualnie najmniejszym ograniczeniem czasowym $t_p(x)$ jest wykonywany. Z konstrukcji wynika, że $p(x)$ jest identyczny z $p^*(x)$.

Sztuczka, aby osiągnąć krótki czas wykonania, polega na odpowiednim zaplanowaniu wszystkiego, aby nie stracić zbyt wiele wydajności przez obliczanie wolnych p i t_p przed znalezieniem p . Bardziej formalnie mówimy, że program „ p oblicza funkcję f ”, gdy uniwersalna maszyna Turinga odniesienia U na wejściu (p,x) oblicza $f(x)$ dla wszystkich x . Oznacza się to jako $U(p,x) = f(x)$. Aby móc mówić o dowodach, potrzebujemy formalnego systemu logicznego $(\forall, \lambda, y_i, c_i, f_i, R_i, \rightarrow, \wedge, =, \dots)$ oraz aksjomatów i reguł wnioskowania. Dowód to ciąg formuł, gdzie każda formuła jest albo aksjomatem, albo wywnioskowana z poprzednich formuł w sekwencji poprzez zastosowanie reguł wnioskowania. Musimy tylko wiedzieć, że dowodzenie, maszyny Turinga i czas obliczeniowy mogą być sformalizowane, a zbiór (poprawnych) dowodów jest przeliczalny. Mówimy, że p jest dowodliwie równoważne p^* , jeśli formuła $[\forall y: U(p,y) = U(p^*,y)]$ może być udowodniona. Ustalmy $\varepsilon \in (0,1/2)$. $M^{\varepsilon_{p^*}}$ uruchamia trzy algorytmy A, B i C równolegle:

$M^{\varepsilon_{p^*}}(x)$

Inicjuj współdzielone zmienne Przeprowadź wszystkie dowody.

$L := \{\}, t_{fast} := \infty, p_{fast} = p^*$

Uruchom algorytmy A, B i C równolegle, odpowiednio przy użyciu względnych zasobów obliczeniowych ε, ε i $1 - 2\varepsilon$

[A]

Przejrzyj wszystkie dowody.

Jeśli dowód dla pewnego (p,t) że $p(\cdot)$ jest równoważne (oblicza) $p^*(\cdot)$ i ma ograniczenie czasowe $t(\cdot)$, to dodaj (p,t) do L

[B]

Oblicz wszystkie $t(x)$ równolegle dla wszystkich $(p,t) \in L$ z względnym czasem obliczeń $2^{-\ell(p) - \ell(t)}$..
jeśli dla pewnego t , $t(x) < t_{fast}$, wtedy $t_{fast} := t(x)$ i $p_{fast} := p$ continue

[C]

uruchom U na (p_{fast},x) .

Dla każdego kroku czasowego zmniejsz t_{fast} o 1.

jeśli U się zatrzyma, wydrukuj wynik $U(p_{fast},x)$

i przerwij obliczenia A, B i C.

Zauważ, że A i B kończą się tylko wtedy, gdy zostaną przerwane przez C. Oczywiście jest, że $M^{\varepsilon_{p^*}}$ jest równoważne (oblicza) p^* . Pokażemy, że czas obliczeniowy $M^{\varepsilon_{p^*}}$ jest ograniczony przez

$$time_{M^{\varepsilon_{p^*}}}(x) \leq (1 + \varepsilon) \cdot t_p(x) + \frac{d_p}{\varepsilon} \cdot time_{t_p}(x) + \frac{c_p}{\varepsilon},$$

$$d_p = 3 \cdot 2^{\ell(p) + \ell(t_p)}, \quad c_p = 3 \cdot 2^{\ell(\text{proof } p) + 1} \cdot O(\ell(\text{proof } p)^2)$$

gdzie p jest dowolnym algorytmem, który w sposób udowodniony oblicza tę samą funkcję co p^* , a czas obliczeń jest w sposób udowodniony ograniczony funkcją $t_p(x)$ dla wszystkich x , a $time_{t_p}(x)$ jest czasem potrzebnym do obliczenia ograniczenia czasowego $t_p(x)$. Znane ograniczenia czasowe dla praktycznych problemów często można obliczyć szybko, tj. $time_{t_p}(x)/time_p(x)$ często zbiega bardzo szybko do zera.

Ponadto, z praktycznego punktu widzenia, ograniczenia dowodzenia są często dość słabe. Dlatego skonstruowaliśmy dla wszystkich tych problemów rozwiązanie, które jest asymptotycznie tylko o czynnik $1+\epsilon$ wolniejsze niż (w sposób udowodniony) najszybszy algorytm. Z drugiej strony, w przypadku realistycznych problemów, zwykle dominują wyrazy niższego rzędu, co ogranicza praktyczne zastosowanie M_p^ϵ .

Złożoność arytmetyczna i najkrótszy algorytm. Naturalną definicją złożoności (Kolmogorowa) funkcji f jest długość najkrótszego programu obliczającego $f : K\{f\} := \min_p \{\ell(p) : U(p,x) = f(x) \forall x\}$. Niestety, K cierpi na to, że nie jest nawet przybliżona, ponieważ równość funkcjonalna programów jest w ogólności nierozstrzygalna. Niech p^* będzie formalną specyfikacją lub programem dla f . Użycie $K(p^*)$ również nie jest odpowiednią alternatywą, ponieważ zależy ona zasadniczo od wyboru p^* , ponieważ np. „martwy kod” w p^* przyczynia się do $K(p^*)$. Zadowolającym rozwiązaniem jest wzięcie długości najkrótszego programu, który jest udowodnialnie równoważny p^* :

$$K''(p^*) := \min_p \{\ell(p) : \text{a proof of } [\forall y: U(p, y) = U(p^*, y)] \text{ exists}\}.$$

K'' (podobnie jak K) jest pólobliczalny w górnym zakresie. Niech p' będzie krótkim opisem p^* . Teraz zajmijmy się czasem obliczeń p' . Czy moglibyśmy uzyskać coraz wolniejsze algorytmy, kompresując p^* coraz bardziej? Co ciekawe, nie jest to prawdą. Wymyślanie złożonych (długich) programów nie jest konieczne do konstruowania asymptotycznie szybkich algorytmów, przy podanych założeniach dowodzenia, w przeciwieństwie do twierdzenia Bluma. Pokazujemy, że istnieje program p , równoważny p^* z

$$(i) \quad \ell(\tilde{p}) \leq K''(p^*) + O(1),$$

$$(ii) \quad \text{time}_{\tilde{p}}(x) \leq (1 + \epsilon) \cdot t_p(x) + \frac{d_p}{\epsilon} \cdot \text{time}_{t_p}(x) + \frac{c_p}{\epsilon},$$

gdzie p jest dowolnym programem, który jest udowodniony jako równoważny p^* z czasem obliczeń udowodnionym jako krótszym niż $t_p(x)$. Oznacza to, że \tilde{p} jest jednocześnie jednym z najkrótszych i najszybszych programów.

Uogólnienia. Algorytm $M_{p^*}^\epsilon$ można zmodyfikować, aby obsługiwał strumienie wejścia/wyjścia, definiowalne przez maszynę Turinga z jednokierunkowymi taśmami wejściowymi i wyjściowymi (oraz dwukierunkowymi taśmami roboczymi) odbierającymi strumień wejściowy i generującymi strumień wyjściowy, jak ma to miejsce w konfiguracji agenta.

Model AIXI ograniczony czasowo

Główną wadą modelu AIXI jest to, że jest on nieobliczalny. Aby przezwyciężyć ten problem, konstruujemy zmodyfikowany algorytm AIXItl, który jest nadal lepszy od każdego innego agenta ograniczonego czasem t i długością. Czas obliczeniowy AIXItl jest rzędu $t \cdot 2^t$. Zmniejszenie dużego czynnika 2^t zgodnie z linią z poprzedniej podsekcji jest możliwe, ale nie zostanie tutaj przedstawione.

Nieefektywność AIXI. $\xi^{AI} = \xi^{AI_U}$ nie jest obliczalny, ale jedynie wyliczalną półmiarą. Stąd wynik y_k modelu AIXI jest tylko asymptotycznie obliczalny (przybliżony). AIXI daje algorytm, który generuje sekwencję próbnych wyników, ostatecznie zbieżnych do prawidłowego wyniku y_k , ale nigdy nie można być pewnym, czy już go osiągnęło. Poza tym, konwergencja jest niezwykle powolna, więc ten typ asymptotycznej obliczalności nie ma bezpośredniego praktycznego zastosowania. Ponadto,

zastąpienie ξ^{AI} przez wersje ograniczone czasowo, które nadają się do przewidywania sekwencji, nie udaje się w przypadku modelu AIXI. Prowadzi to do problemów poruszanych dalej.

Ograniczenia czasowe i skuteczność. Niech \tilde{p} będzie polityką, która oblicza akceptowalny wynik w rozsądnym czasie \tilde{t} na cykl interakcji. Tego rodzaju założenie obliczalności, a mianowicie, że komputer ogólnego przeznaczenia o wystarczającej mocy i odpowiednim programie jest w stanie zachowywać się w sposób inteligentny, stanowi podstawę badań nad sztuczną inteligencją. Tutaj nie ma potrzeby omawiania, co dokładnie rozumie się przez „rozsądny czas/inteligencję” i „wystarczającą moc”. Interesuje nas, czy istnieje obliczalna wersja systemu AIXI, która jest lepsza lub równa dowolnej polityce p z czasem obliczeń na cykl wynoszącym co najwyżej \tilde{t} . Można realistycznie mieć nadzieję na skonstruowanie układu $AIXI_{\tilde{t}}$ o czasie obliczeniowym $c \cdot \tilde{t}$ na cykl dla pewnej stałej c . Pomysł polega na uruchomieniu wszystkich programów p o długości $\leq \tilde{l} := \ell(\tilde{p})$ i czasie $\leq \tilde{t}$ na cykl i wybraniu najlepszego wyniku w sensie maksymalizacji wartości uniwersalnej V_{ξ}^* . Całkowity czas obliczeniowy wynosi $c \cdot \tilde{t}$ z $c \approx 2^{\tilde{t}}$. Niestety, V_{ξ}^* nie może być użyte bezpośrednio, ponieważ ta miara jest sama w sobie tylko półobliczalna, a jakość przybliżenia przy użyciu obliczalnych wersji ξ^{AI} przy danym czasie rzędu $c \cdot \tilde{t}$ jest niedokładna. Z drugiej strony musimy użyć miary, która zbiega się do V_{ξ}^* dla $\tilde{t}, \tilde{l} \rightarrow \infty$, ponieważ chcemy, aby model AIXI_{tilde} zbiegał się w tym przypadku z modelem AIXI.

Prawidłowe przybliżenia. Sugerujemy następujące rozwiązanie spełniające powyższe warunki: Głównym pomysłem jest rozważenie rozszerzonych chronologicznych polityk przyrostowych p , które oprócz regularnego wyniku y_k^p oceniają swój wynik z w_k^p . Model $AIXI_{\tilde{t}}$ wybiera wyjście $\hat{y}_k = y_k^p$ politykę p z najwyższą oceną w_k^p . Polityka p może sugerować dowolne wyjście y_k^p ale nie wolno mu oceniać siebie dowolnie wysoko w_k^p , jeśli chce się, aby w_k^p było wiarygodnym kryterium wyboru najlepszego p . Należy wymagać, aby żadna polityka p nie mogła twierdzić, że jest lepsza niż jest w rzeczywistości. Definiujemy logiczny predykat $VA(p)$, nazywany poprawnym przybliżeniem, który jest prawdziwy wtedy i tylko wtedy, gdy p zawsze spełnia $w_k^p \leq V_{\xi}^p(y_{<k})$, tj. nigdy nie przecenia siebie. $V_{\xi}^p(y_{<k})$ to ξ^{AI} to oczekiwana przyszła nagroda przy polityce p . Poprawne polityki p mogą być następnie (częściowo) uporządkowane względem ich oceny w_k^p .

Uniwersalny ograniczony czasowo system AIXI_{tilde}. Poniżej opisujemy algorytm p^* leżący u podstaw systemu $AIXI_{\tilde{t}}$. Opiera się ona w zasadzie na wyborze najlepszych algorytmów p_k^* spośród ograniczonych czasem \tilde{t} i długością \tilde{l} polityk p , dla których istnieje dowód P z $VA(p)$ o długości $\leq l_p$.

1. Utwórz wszystkie ciągi binarne o długości l_p i zinterpretuj każdy z nich jako kodowanie dowodu matematycznego w tym samym formalnym systemie logicznym, w którym sformułowano $VA(\cdot)$. Weź te ciągi, które są dowodami $VA(p)$ dla pewnego p i zachowaj odpowiadające im programy p .
2. Wyliminuj wszystkie p o długości $> l$.
3. Zmodyfikuj zachowanie wszystkich pozostałych p w każdym cyklu k w następujący sposób: Nic się nie zmienia, jeśli p wyprowadza pewne $w_k^p y_k^p$ w \tilde{t} i krokach czasowych. W przeciwnym wypadku zatrzymaj p i zapisz $w_k = 0$ i pewne dowolne y_k na taśmie wyjściowej p . Niech P będzie zbiorem wszystkich tych zmodyfikowanych programów.
4. Rozpocznij pierwszy cykl: $k:=1$.

5. Uruchom każdy $p \in P$ na rozszerzonym wejściu $\hat{y}_{\leq k}$, gdzie wszystkie wyjścia są przekierowywane na taśmę pomocniczą: $p(\hat{y}_{\leq k}) = w_1^p y_1^p \dots w_k^p y_k^p$. Ten krok jest wykonywany przyrostowo przez dodanie \hat{y}_{k-1} dla $k > 1$ do taśmy wejściowej i kontynuowanie obliczeń z poprzedniego cyklu.

6. Wybierz program p o najwyższej ocenie w_k^p : $p_k^* := \operatorname{argmax}_p w_k^p$.

7. Zapisz $\hat{y}_k := y_k^{p_k^*}$ na taśmie wyjściowej.

8. Odbierz wejście \hat{x}_k ze środowiska.

9. Rozpocznij następny cykl: $k := k + 1$, przejdź do kroku 5.

Właściwości algorytmu p^* . Niech p będzie dowolną rozszerzoną polityką chronologiczną (inkrementalną) o długości $\ell(p) \leq \bar{l}$ i czasie obliczeń na cykl $t(p) \leq \bar{t}$, dla której istnieje dowód VA(p) o długości $\leq l_p$. Algorytm p^* , zależny od \bar{l}, \bar{t} i l_p , ale nie od p , ma zawsze wyższą ocenę niż jakikolwiek taki p . Czas konfiguracji p^* wynosi $t_{\text{setup}}(p^*) = O(l_p^2 \cdot 2^p)$, a czas obliczeń na cykl wynosi $t_{\text{cycle}}(p^*) = O(2^{\bar{l}} \cdot \bar{t})$. Ponadto dla $\bar{t}, \bar{l}, l_p \rightarrow \infty$, polityka p^* zbiega się do zachowania modelu AIXI. Mówiąc w skrócie, oznacza to, że jeśli w ogóle istnieje obliczalne rozwiązanie jakiegoś problemu AI, to jawnie skonstruowany algorytm p^* jest takim rozwiązaniem. Twierdzenie to jest dość ogólne, jednak istnieją pewne ograniczenia i otwarte pytania dotyczące czasu konfiguracji, konieczności, aby zasady oceniały własne wyniki, prawdziwej, ale niemożliwej do (efektywnego) udowodnienia VA(p) oraz „niespójnych” zasad.

Dyskusja

Co osiągnięto. Zaproponowaliśmy eleganckie matematyczne podstawy sztucznej inteligencji. Dokładniej rzecz biorąc, opracowaliśmy teorię dla racjonalnych agentów działających optymalnie w każdym środowisku. W ten sposób poruszyliśmy różne obszary naukowe, w tym uczenie się przez wzmacnianie, algorytmiczną teorię informacji, teorię złożoności obliczeniowej, teorię prawdopodobieństwa, teorię decyzji sekwencyjnych i wiele innych. Przedstawiliśmy teorię decyzji sekwencyjnych w bardzo ogólnej formie i zjednoczyliśmy ją z teorią indukcji uniwersalnej Solomonoffa, obie okazały się optymalne w swojej własnej domenie. Powstały w rezultacie bezparametrowy model AIXI stanowi agenta, dla którego przedstawiliśmy mocne argumenty, że zachowuje się optymalnie w każdym środowisku. Dla ograniczonych klas środowiskowych i mieszanin Bayesa ξ pokazaliśmy, że AI ξ jest samoopimalizująca i optymalna w sensie Pareto. Omówiliśmy wybór horyzontu i zmotywowaliśmy do użycia niegeometrycznego dyskontowania nagród. Omówiliśmy również szereg ważnych klas problemów, w tym przewidywanie sekwencji, gry strategiczne, minimalizację funkcji i uczenie nadzorowane. Podsumowując, pokazuje to, że sztuczną inteligencję można ująć w ramy eleganckiej teorii matematycznej. Pewien postęp został również poczyniony w kierunku eleganckiej obliczeniowej teorii inteligencji. AIXItl ma optymalny rząd czasu obliczeniowego, poza dużą stałą mnożnikową, której moglibyśmy się pozbyć kosztem (niestety jeszcze większej) stałej addytywnej..

Porównanie z innymi podejściami. Istnieje wiele innych podejść do AI, zbyt wiele, aby wymienić je wszystkie. Wśród modeli, które mogą uczyć się z doświadczenia, są „klasyczne” algorytmy uczenia się przez wzmacnianie, takie jak uczenie się różnic czasowych [SB98], adaptacyjne warianty wyszukiwania Levina, przewidywanie z poradą eksperta, uczenie się przez wzmacnianie oparte na rynku/gospodarce itp. Wszystkie te modele zostały zaimplementowane i są stosowane w ograniczonych domenach,

często z rozsądną wydajnością. W przeciwieństwie do tego $AIXI(tl)$ zachowuje się optymalnie w każdym (całkowicie ogólnym) środowisku, jest wydajny pod względem danych, ma zdolność generalizacji, rozwiązuje problem eksploracji w porównaniu z eksploatacją itd., ale nie jest wykonalny obliczeniowo bez dalszych przybliżeń.

Outlook &: pytania otwarte. Głównym wyzwaniem teoretycznym jest wyprowadzenie dobrych nieasympotycznych ograniczeń wartości lub powiązanych wielkości dla $AI\xi$ i $AIXI$, w idealnym przypadku tak silnych, jak w przypadku przewidywania sekwencji. Głównym wyzwaniem praktycznym jest skalowanie modelu $AI\xi$ w dół, np. poprzez użycie bardziej ograniczonych form ξ , tak jak zasada minimalnej długości opisu robi to dla indukcji uniwersalnej. Model $AIXI(tl)$ to inne, bardzo ogólne podejście do modelu obliczeniowego. Niestety, cierpi na ten sam duży czynnik 2^i w czasie obliczeń, co wyszukiwanie Levina dla problemów inwersji [Lev73b, Lev84]. Z drugiej strony wyszukiwanie Levina zostało zaimplementowane i pomyślnie dostosowane i zastosowane do różnych problemów, a stała mnożnikowa może zostać wyeliminowana, jak w $M_{p^*}^\epsilon$ lub zredukowana przez maszynę Gödla. Przegląd istniejących podejść sugeruje, że podczas gdy $AIXI$ jest elegancką teorią matematyczną, która wydaje się spełniać wszystkie formalne potrzeby, obliczeniowa AI może być chaotyczna. Na przykład algorytmy specjalnego przeznaczenia do wstępnego przetwarzania danych wejściowych i końcowego przetwarzania danych wyjściowych prawdopodobnie będą konieczne w każdym wydajnym systemie AI. Innym problemem jest włączenie dodatkowej wiedzy. Zasadniczo nie ma potrzeby modyfikowania $AIXI$, ponieważ każda wcześniejsza wiedza to otwarte pytania. Głównym wyzwaniem teoretycznym jest wyprowadzenie dobrych nieasympotycznych ograniczeń wartości lub powiązanych wielkości dla $AI\xi$ i $AIXI$, w idealnym przypadku tak silnych, jak w przypadku przewidywania sekwencji. Głównym wyzwaniem praktycznym jest skalowanie modelu AI^\wedge w dół, np. poprzez użycie bardziej ograniczonych form ξ , tak jak zasada minimalnej długości opisu dla indukcji uniwersalnej. Model $AIXI(tl)$ jest innym, bardzo ogólnym podejściem do modelu obliczeniowego. Niestety cierpi na ten sam duży czynnik 2^i w czasie obliczeń, co wyszukiwanie Levina dla problemów inwersji. Z drugiej strony, wyszukiwanie Levina zostało zaimplementowane i pomyślnie zaadaptowane i zastosowane do różnych problemów, a stała mnożenia może zostać wyeliminowana jak w $M_{p^*}^\epsilon$ lub zredukowana przez maszynę Gödla. Badanie istniejących podejść sugeruje, że podczas gdy $AIXI$ jest elegancką teorią matematyczną, która wydaje się spełniać wszystkie formalne potrzeby, obliczeniowa AI może być chaotyczna. Na przykład algorytmy specjalnego przeznaczenia do wstępnego przetwarzania danych wejściowych i końcowego przetwarzania danych wyjściowych prawdopodobnie będą konieczne w każdym wydajnym systemie AI. Innym problemem jest włączenie dodatkowej wiedzy. Zasadniczo nie ma potrzeby modyfikowania $AIXI$, ponieważ każda wcześniejsza wiedza może być po prostu przedstawiona jako pierwsze dane wejściowe x_1 w dowolnym formacie. Dopóki algorytm interpretujący dane ma rozmiar $O(1)$, $AIXI$ „zrozumie” dane po kilku cyklach. Innym ważnym problemem jest sam proces szkolenia. Przez proces szkoleniowy rozumiemy sekwencję zadań od prostych do złożonych do rozwiązania, przy czym prostsze pomagają w nauce bardziej złożonych. Te i wiele innych zagadnień koncepcyjnych, praktycznych i filozoficznych, w tym współbieżne działania i percepcje, wybór przestrzeni wejścia/wyjścia, traktowanie zaszyfrowanych informacji, osobliwości śmiertelnych ucieleśnionych agentów, paradoks wolnej woli, istnienie obiektywnych prawdopodobieństw, test Turinga, istnienie wydajnych i eleganckich uniwersalnych teorii inteligencji związanych z nieobliczalnymi środowiskami Penrose'a i „liczbą mądrości” Q Chaitina zostaną omówione później.