Prawdopodobieństwo a informacja

Piotr MIGDAŁ[*]

[*] ICFO–The Institute of Photonic Sciences, Castelldefels (Barcelona), obecnie freelancer z analizy danych
http://migdal.wikidot.com

Filozof może się zadowolić tym, że *„wie, że nic nie wie”*. Fizyk zaś potrzebuje wiedzieć, *ile* nie wie.

Pojęcie *entropii* ma swoje źródło w termodynamice i jest związane z tym, jak bardzo nie znamy dokładnego mikrostanu układu. Tylko gdy wiemy o nim wszystko, możemy w pełni wykorzystać jego energię, przekształcając ją na inne formy. Gdy nie – część energii pozostaje niejako uwięziona.

Entropia jest równie cenna w teorii informacji – pozwala ściśle zmierzyć, jak dobrze możemy skompresować daną wiadomość oraz jak bardzo możemy niwelować szum przy jej przesyłaniu. Przydaje się ona również jako miara losowości i korelacji w różnych narzędziach statystycznych.

O ile entropia jest używana w języku potocznym jako chaos i nieuporządkowanie, to sama wielkość (a konkretniej, *entropia Shannona*) jest ściśle określona następującym wzorem:

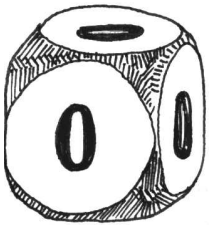
$$(*) \qquad H(p_1,\ldots,p_n)=\sum_{i=1}^{n} p_i \log\left(\frac{1}{p_i}\right),$$

gdzie $p_1,\ldots,p_n$ są prawdopodobieństwami możliwych wyników przeprowadzanego eksperymentu. Będziemy używać logarytmu o podstawie 2, co odpowiada mierzeniu entropii w *bitach*. Powyższy wzór ma wiele zastosowań i interpretacji.

Częściej zapisujemy H jako $-\sum_{i=1}^{n} p_i \log(p_i)$, równoważnie, ale, moim zdaniem, mniej dydaktycznie.

Korzystanie z logarytmu o innej podstawie, np. logarytmu naturalnego powoduje przemnożenie wyniku przez stałą, a zatem odpowiada tylko zmianie jednostek: $\ln(x)=\ln(2)\log_2(x)$. Np. dla podstawy e jednostką jest *nit*.

Ale zanim do nich przejdziemy, spójrzmy na kilka prostych przykładów. Gdy mamy jedną możliwość, entropia wynosi $1\log(1)=0$ – wszak nie ma tu miejsca na losowość. Gdy rzucamy uczciwą monetą, entropia to $\frac{1}{2}\log(2)+\frac{1}{2}\log(2)=1$. Entropia jest największa, gdy dla ustalonej liczby zdarzeń wszystkie są równoprawdopodobne – wtedy wynosi ona $\log(n)$. Dla kostki do gry z 6 ścianami to $\log(6)\approx 2{,}6$.



Warto pamiętać, że entropia jest zawsze miarą niewiedzy. Stąd np. kostka, na której widzimy wyrzucone dwa oczka, ma entropię zero. Dowiadujemy się dokładnie tyle, o ile entropia zmalała, tu:

$$H\left(\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6},\frac{1}{6}\right)-H(0,1,0,0,0,0)=\log(6)-\log(1)\approx 2{,}6.$$

Gdyby ktoś nam powiedział „wypadło jedno, dwa lub trzy oczka”, nasza wiedza końcowa byłaby niepełna i dowiedzielibyśmy się $\log(6)-\log(3)$, czyli 1 bit informacji.

Dlaczego potrzebujemy w tym wzorze logarytmu? Gdy mamy dwa niezależne zdarzenia, chcemy, by ich entropia była sumą entropii składników. Powiedzmy, że chcemy dowiedzieć się, jaki jest znak zodiaku naszego obiektu westchnień oraz czy nas kocha. Nie powinno grać roli, czy dowiemy się jednej informacji naraz czy po kawałku. Oznaczymy prawdopodobieństwa znaków zodiaku jako $p_1,\ldots,p_{12}$ oraz uczucie do nas jako q_1,q_2. Tym samym prawdopodobieństwa poszczególnych, niezależnych zdarzeń to iloczyny $p_i q_j$. Zatem jak jest z entropią?

Czytelnikowi, który zastanawia się, czy owe dane są niezależne, polecam zapoznać się z http://bit.ly/dating-zodiac.

$$\begin{aligned}\sum_{i=1}^{12}\sum_{j=1}^{2} p_i q_j \log(1/(p_i q_j)) &= \sum_{i=1}^{12}\sum_{j=1}^{2} p_i q_j (\log(1/p_i)+\log(1/q_j)) \\ &= \sum_{i=1}^{12}\sum_{j=1}^{2} p_i q_j \log(1/p_i) + \sum_{i=1}^{12}\sum_{j=1}^{2} p_i q_j \log(1/q_j) \\ &= \sum_{i=1}^{12} p_i \log(1/p_i) + \sum_{j=1}^{2} q_j \log(1/q_j),\end{aligned}$$

zatem entropia niezależnych zdarzeń dodaje się. W szczególności: n rzutów uczciwą monetą to n bitów entropii.

Czytelnik Wnikliwy może sprawdzić, że $-\log\sum_{i=1}^{n} p_i^2$ również ma opisaną obok własność addytywności. Ma ją także cała rodzina *entropii Rényiego*, będących uogólnieniem entropii Shannona, przy czym nie mają one interpretacji jako informacja.

Na wzór na entropię Shannona można patrzeć również jak na średnią uzyskaną informację. Zobaczmy to na przykładzie gry w *dwadzieścia pytań*, w której jedna

Rozwiązanie zadania M 1475.
Ponieważ ϕ jest różnowartościowa, to dla każdego $k \geqslant 1$ spełniona jest nierówność

$$\sum_{i=1}^{k} \phi(i) \geqslant 1 + 2 + \ldots + k.$$

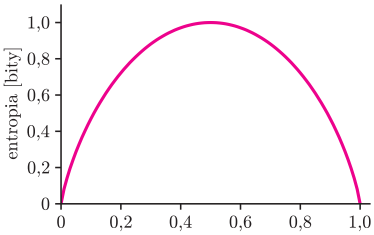
Oznacza to, że liczby

$$s_k = \sum_{i=1}^{k} (\phi(i) - i)$$

są nieujemne. Niech $a_i = \phi(i) - i$ oraz $b_i = 1/i^2$. Stosując przekształcenie Abela, otrzymujemy równość

$$\begin{aligned} a_1b_1 + a_2b_2 + \ldots + a_nb_n &= \\ &= a_1(b_1 - b_2) + (a_1 + a_2)(b_2 - b_3) + \\ &\quad + (a_1 + a_2 + a_3)(b_3 - b_4) + \ldots + \\ &\quad + (a_1 + \ldots + a_{n-1})(b_{n-1} - b_n) + \\ &\quad + (a_1 + \ldots + a_n)b_n = \\ &= \sum_{i=1}^{n-1} s_i(b_i - b_{i+1}) + s_nb_n. \end{aligned}$$

Prawa strona tej równości jest nieujemna, zatem lewa też. Stąd bezpośrednio wynika teza.



Entropia przy dwóch możliwościach, $H(p, 1-p)$, np. rzutu nieuczciwą monetą (p to prawdopodobieństwo orła) lub odpowiedzi na pytanie (p to prawdopodobieństwo „tak"). Maksimum, równe 1 bitowi, osiąga dla $p = 1/2$.

Pozwolę sobie zachęcić Czytelnika do doświadczalnego zmierzenia (na znajomych) entropii pytań oraz entropii pomyślanych obiektów.

Odpowiedź na pytanie ze strony 1

Nie spełnia rzekomej reguły Pappusa np. koło obracane względem swojej średnicy – ponieważ środek ciężkości nie rusza się, kula musiałaby mieć zarówno objętość, jak i pole powierzchni równe zeru.

Wystarczającym założeniem, by reguła Pappusa była twierdzeniem, jest wymaganie, by podczas ruchu żaden z punktów nie był odwiedzony dwukrotnie.

osoba wymyśla jakąś rzecz, a druga ma za zadanie zgadnąć, o co chodzi, zadając po kolei pytania na „tak" lub „nie". Można się zastanowić, czy warto zadawać pytania, które są z grubsza *pół na pół* (np. „Czy to jest żywe?"), czy też takie, w których jest niewielka szansa, że dużo się rozjaśni (np. „Czy to element biżuterii?").

Powiedzmy, że na starcie jest m możliwych obiektów i każdy z nich jest równoprawdopodobny. Jeśli zadamy pytanie, dla t obiektów odpowiedzią jest „tak", dla $m - t$ pozostałych – „nie". Wiąże się to bezpośrednio z prawdopodobieństwami odpowiedzi:

$$p_{\text{tak}} = \frac{t}{m}, \qquad p_{\text{nie}} = \frac{m-t}{m}.$$

Zatem po uzyskaniu odpowiedzi zbiór możliwych obiektów zmniejsza się do $p_{\text{tak}}m$ albo $p_{\text{nie}}m$. Po zadaniu dwóch pytań (dla uproszczenia przyjmując, że po zadaniu pierwszego drugie ma to samo prawdopodobieństwo udzielenia twierdzącej odpowiedzi) dostajemy 4 możliwości: $p_{\text{tak}}p_{\text{tak}}m$, $p_{\text{nie}}p_{\text{tak}}m$, $p_{\text{tak}}p_{\text{nie}}m$ albo $p_{\text{nie}}p_{\text{nie}}m$. Wygramy w sytuacji, gdy po iluś pytaniach zredukujemy liczbę możliwości do tylko jednej. O ile może być w tym trochę szczęścia, to przy większej liczbie pytań (a mamy do dyspozycji ich aż 20), możemy śledzić, co się *średnio* stanie. Skoro przy dodawaniu kolejnych pytań prawdopodobieństwa się mnożą, to wielkością, którą chcemy uśredniać, jest wspomniany logarytm prawdopodobieństwa. Po jednym pytaniu dostajemy

$$H(p_{\text{tak}}, p_{\text{nie}}) = p_{\text{tak}} \log\left(\frac{1}{p_{\text{tak}}}\right) + p_{\text{nie}} \log\left(\frac{1}{p_{\text{nie}}}\right),$$

– to samo co $(*)$. Po zadaniu k pytań dowiadujemy się średnio kH bitów informacji. Wielkość ta to $\sum_p p \log(1/p)$, gdzie p jest prawdopodobieństwem uzyskania poszczególnego ciągu odpowiedzi. Jeśli jej wartość dojdzie do $\log(m)$, to średnio $p = 1/m$, a zatem liczba możliwych obiektów zostanie zredukowana do jednej, czyli: zgadliśmy! Gdy zadamy pytanie *pół na pół*, to niezależnie od odpowiedzi dowiadujemy się $\log 2$ czyli 1 bit informacji. A co gdy, powiedzmy, zadamy pytanie, na które prawdopodobieństwo odpowiedzi „tak" jest równe tylko $1/1000$? Jest niewielka szansa, że dowiemy się bardzo dużo ($\log(1000) \approx 10$), ale najprawdopodobniej dowiemy się niezbyt wiele. A sumarycznie, będzie lepiej czy gorzej? Zobaczmy!

$$0{,}001 \log\left(\frac{1}{0{,}001}\right) + 0{,}999 \log\left(\frac{1}{0{,}999}\right) \approx 0{,}0100 + 0{,}0014 = 0{,}0114,$$

czyli bardzo niewiele! Zresztą, *pół na pół* średnio dostarcza najwięcej informacji – patrz wykres na marginesie.

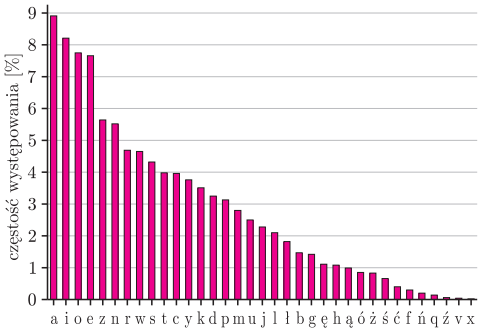
Entropia jest też mocno związana z tym, jak dobrze potrafimy skompresować dane. *Podstawowe twierdzenie Shannona* mówi, że jeśli mamy ciąg n liter, każda niezależna od innych i z entropią H, to nie możemy ich skompresować do długości krótszej niż nH. Choć stoją za tym pewne szczegóły techniczne, główną ideę można oddać, zliczając ciągi. Liczba słów o długości n z alfabetu z d literami to $d^n = 2^{n \log(d)}$. Jeśli chcielibyśmy zakodować owe słowa za pomocą zer i jedynek tak, by każde słowo miało swój kod binarny, potrzebujemy $m = n \log(d)$ bitów. A co, jeśli nie wszystkie litery są równie często używane?

Gdy przesyłamy wiele liter, niektóre ciągi staną się skrajnie mało prawdopodobne. Dla ustalenia uwagi, zróbmy to na ciągu z zer i jedynek, z prawdopodobieństwami p i q, odpowiednio. Typowy ciąg o długości n będzie miał np zer i nq jedynek. Prawdopodobieństwo tego ciągu to $p^{np}q^{nq}$. O ile, rzecz jasna, często można otrzymać ciągi z mniejszą lub większą liczbą zer i jedynek, to czym dłuższy ciąg, tym udział zer i jedynek będzie się bardziej zbliżał do p i q, odpowiednio; innych ciągów jest skrajnie niewiele. Skoro typowe ciągi są mniej więcej równoprawdopodobne, a całkowite prawdopodobieństwo musi się sumować do jedności, to liczba tych typowych ciągów to około $p^{-np}q^{-nq}$. Czyli potrzebujemy

$$m = n\left(p \log\left(\frac{1}{p}\right) + q \log\left(\frac{1}{q}\right)\right)$$

bitów do ich opisu. Przy obliczaniu prawdopodobieństwa jesteśmy dość liberalni – wszak np. dla $p = 0{,}8$ i $q = 0{,}2$ zamiana nawet jednego znaku zmienia

Na zachętę (gzip, kody i poezja): `http://jvns.ca/blog/2015/02/22/how-gzip-uses-huffman-coding/`. Dla zainteresowanych: J. Thomas, T. Cover, *Elements of Information Theory*.



Częstość występowania liter w języku polskim, na podstawie Korpusu IPI PAN.

prawdopodobieństwo o czynnik 4. Niemniej, jest to kompensowane przez niewielkie zmiany n (czasem potrzeba kilku mniej lub więcej bitów do zakodowania owego ciągu). Innymi słowy, liczba typowych ciągów zer i jedynek rośnie jak 2^{nH}.

Zastosowania? Zobaczmy, jak dobrze można skompresować tekst! W języku polskim używa się 35 liter (wliczając `q`, `x` i `v` z zapożyczonych słów). Jednak niektóre litery występują znacznie częściej niż inne, np. `a` stanowi 9% liter, podczas gdy `ź` – tylko 0,06%. Możemy obliczyć, że entropia liter to 4,56 bitów. Dla porównania, taką samą entropię jak litery w języku polskim miałby hipotetyczny język z tylko $2^{4,56} \approx 24$ równo występującymi literami. Ale jak to się ma do kompresji? Porównując entropię z log(35), możemy zobaczyć, że, gdy znamy częstość znaków, potrafimy skompresować tekst do 89% objętości. W praktyce kompresja tekstu jest znacznie lepsza – zarówno dlatego, że nieskompresowany znak zwykle zajmuje 8 bitów (tu możemy kompresować do 57%), jak i to, że znaki nie są losowe, tj. łączą się w sylaby, słowa, a te – w teksty, które często używają podobnych słów.

Jako autor wiem, że ten tekst był *skompresowany*, ale, mam nadzieję, przekazał jakąś *informację* o entropii.

7. Międzynarodowy Turniej Fizyków

W dniach 6–11 kwietnia 2015 roku Wydział Fizyki Uniwersytetu Warszawskiego był gospodarzem prestiżowych zawodów 7. Międzynarodowego Turnieju Fizyków, w którym brały udział reprezentacje 11 krajów z całego świata. Turniej wygrała reprezentacja Ukrainy z Charkowskiego Uniwersytetu Narodowego im. Wasyla Karazina, drugie miejsce zajęli studenci Duńskiego Uniwersytetu Technicznego (DTU), a trzecie reprezentacja Francji z École Polytechnique.

Międzynarodowy Turniej Fizyków to zawody skierowane do studentów, mające na celu rozwój naukowy uczestników przez pracę w grupie, tworzenie prezentacji naukowych, ich wygłaszanie, merytoryczne krytykowanie oraz odpieranie krytyki.

W tym roku studenci z Chin, Danii, Francji, Iranu, Rosji, Singapuru, Szwecji, Szwajcarii, Ukrainy, Wielkiej Brytanii i Polski przez cały rok poprzedzający zawody rozwiązywali 17 niecodziennych problemów fizycznych, np. czy można woltomierzem określić, czy ziemniak jest już ugotowany, dlaczego wstążka, po której przeciągnie się ostrzem, zaczyna się skręcać albo jak zmienia się kolor kamieni, gdy spadnie na nie deszcz. Zawody mają formę specyficznej konferencji naukowej, a poszczególne drużyny odgrywają role referentów, oponentów i recenzentów. Wybrany spośród członków drużyny referent prezentuje zagadnienia wybrane przez drużynę oponentów. Następnie po wygłoszonej prezentacji rozpoczyna się dyskusja, w której oponenci wskazują mocne i słabe strony prezentacji, punktują nieścisłości, dążąc do lepszego zrozumienia omawianego rozwiązania. Następnie recenzenci oceniają dyskusję, wskazują, które aspekty prezentacji zostały należycie przedyskutowane, które zostały niesłusznie pominięte, a które fragmenty wymagały większej uwagi. Po około godzinnej zaciętej dyskusji, w której nie brak kąśliwych uwag i ciętych ripost, jury zaczyna zadawać dociekliwe pytania wszystkim występującym na polu walki, a na koniec wystawia oceny. Po krótkiej przerwie drużyny zamieniają się rolami i zaczyna się kolejna rozgrywka. Zawody są niezwykle emocjonujące, o czym świadczy to, że wielu widzów, którzy przyszli tylko na moment, żeby zobaczyć, na czym polega Turniej, zostawało do końca zawodów.

Nieskromnie mówiąc, zagraniczni goście byli pod wrażeniem nowej siedziby Wydziału Fizyki Uniwersytetu Warszawskiego oraz poziomu organizacji Turnieju. Poza zawodami uczestnicy, dzięki współpracy z Urzędem M. St. Warszawy, Tramwajami Warszawskimi, Fundacją Universitatis Varsoviensis i Kancelarią Prezydenta RP, mieli możliwość zwiedzania muzeów, Warszawy, poznawania miasta podczas pieszych spacerów oraz przejażdżek tramwajem, a także oglądania nocnej panoramy Warszawy z XXX piętra Pałacu Kultury. Międzynarodowy Komitet Organizacyjny uznał, iż wyznaczone zostały nowe standardy organizacyjne Turnieju: był to najlepiej zorganizowany Turniej w historii, co nie byłoby możliwe bez wsparcia ze strony społeczności Wydziału Fizyki oraz prorektora Uniwersytetu Warszawskiego, profesora Alojzego Nowaka.

Jan Stefan BIHAŁOWICZ