ZAGADNIENIA
NAUKOZNAWSTWA
1–4 (215–218), 2018
PL ISSN 0044 – 1619

*Ks. Jakub S. Płonka**
Wydział Filozofii
Katolicki Uniwersytet Lubelski Jana Pawła II

Zmniejszenie entropii i prostota jako zasady maszynowego rozpoznawania wzorców w ujęciu Satosi Watanabego

Abstrakt. Rozpowszechnienie maszynowego rozpoznawania wzorców zachęca filozofów do systematycznego namysłu nad tą dziedziną wiedzy. Jednym z pionierów badań w tym obszarze był Satosi Watanabe, często komentowany w literaturze tematu. Artykuł omawia zasady zmniejszenia entropii i prostoty w kontekście maszynowego rozpoznawania wzorców. Choć pojęcie entropii używane było pierwotnie w dziedzinie termodynamiki, jednak po odpowiednim przekształceniu, omówionym w artykule, może być stosowane w opisie zasad rozpoznawania wzorców. Artykuł przedstawia przykłady takiego zastosowania oraz powiązanie pojęcia entropii z prostotą. W ujęciu Watanabego, prostota jest przekładana głównie na prostotę krzywej wielomianowej, artykuł omawia jednak zagadnienie prostoty, umieszczając je w nieco szerszym kontekście.

Słowa kluczowe: rozpoznawanie wzorców, aspekty filozoficzne, zmniejszenie entropii, prostota, Watanabe

Decrease in entropy and simplicity as the rules of mechanical pattern recognition in Satosi Watanabe's research

Abstract. Philosophers are motivated to do research concerning pattern recognition because of wide range of its applications. One of the pathfinders of research in that area was Satosi Watanabe, who has been frequently commented in the literature concerning this subject. The rule of decrease in entropy and the rule of simplicity are described in the context of pattern recognition. Although the concept of entropy had been initially used in the area of thermodynamics, it could be adopted also in the field of pattern recognition. The concept of entropy should be then suitable transformed. A few of examples of the entropy concept application and the relationship between entropy and simplicity are discussed in the article. Simplicity considered by Watanabe should be treated mainly as polynomial curve simplicity, however the issue is described in the wider context.

Key words: pattern recognition, philosophical aspects, decrease in entropy, simplicity, Watanabe

Wprowadzenie

Maszynowe rozpoznawanie wzorców jest obecnie dobrze rozwiniętą dziedziną informatyki i automatyki, powiązaną z wieloma innymi dyscyplinami. Do dziedziny rozpoznawania wzorców zaliczamy nie tylko rozpoznawanie obrazów, ale także rozpoznawanie głosu i innych sygnałów. Powszechność stosowania maszynowego rozpoznawania wzorców wpłynęła na niektórych naukowców, którzy dokonali filo-

* Adres do korespondencji: ul. Radziszewskiego 7, 20-039 Lublin, e-mail: jakub.s.plonka@gmail.com

zoficznego namysłu nad tą dyscypliną nauki. Obecnie w opracowaniu filozoficznych aspektów maszynowego rozpoznawania wzorców wyróżniają się M. Pelillo (Pelillo 2014) i T. Scantamburlo (Scantamburlo 2013), tematem tym zajmują się też m.in. R. Duin, E. Pękalska (Duin, Pękalska 2007), S. Bartlett (Bartlett 2015).

Wyżej wspomniani autorzy nie są jednak pierwszymi, którzy dokonali filozoficznego namysłu, dotyczącego tej dziedziny wiedzy. Wśród pionierów opracowania niniejszego tematu można wyróżnić Satosi Watanabego (1910–1993), japońskiego fizyka, inżyniera i filozofa. W literaturze tematu można odnaleźć wiele odniesień do prac tego autora. Wskazanym zatem jest przedstawienie choćby niektórych jego prac dla zdobycia wiedzy, dotyczącej filozoficznych aspektów maszynowego rozpoznawania wzorców.

Jednym z ważniejszych tematów, poruszonych przez Watanabego, jest zagadnienie zmniejszenia entropii i prostoty, jako zasad maszynowego rozpoznawania wzorców. Sam Watanabe używa słów „minimum, minimize” (Watanabe 1981, 381), w tym artykule jest jednak stosowany termin „zmniejszenie”. Entropia jest pojęciem stosowanym pierwotnie w dziedzinie termodynamiki, choć zostało ono również z powodzeniem zastosowane w teorii informacji. Pojęcie prostoty kieruje myśl autora bardziej w kierunku zagadnień filozoficznych. Watanabe wiąże ze sobą oba te pojęcia, wskazując ich istotną rolę w omawianej dziedzinie. Zasada prostoty jest dość szeroko omawiana w kontekście nauk przyrodniczych, takich jak fizyka i astronomia. Warto prześledzić stosowanie tej zasady na gruncie maszynowego rozpoznawania wzorców, które jest aktualnie już dobrze rozwiniętą i ciągle dynamicznie rozwijającą się dyscypliną wiedzy.

Omawiane zagadnienie można umieścić w szerokim kontekście filozoficznym. Watanabe odnotowuje we wstępie do opisu zagadnienia zmniejszenia entropii: „Celem rozpoznawania wzorców jest odkrycie formy lub konkretniej struktury w systemie, składającym się z podsystemów. Antonimem formy lub struktury jest amorfizm lub chaos.” (Watanabe 1985, 137). Watanabe świadomie odnosi pojęcie formy m.in. do hylemorfizmu Arystotelesa (Watanabe 1985, 7), a także wiąże formę ze strukturą i z zasadą zmniejszenia entropii (Watanabe 1981, 381), omawia rozpoznawanie wzorców w kontekście problemów starożytnej i średniowiecznej metafizyki, sporu o uniwersalia (Watanabe 1985, 45–65).

Ponieważ pojęcie entropii jest stosowane w prawach termodynamiki, nasuwa się epistemologiczne pytanie o relacje naszego sposobu poznawania, powiązanego przez Watanabego z pojęciem entropii, do porządku rzeczy, obserwowanego we wszechświecie. Watanabe nie stroni od uwag przyrodniczych w swoich pracach. Przy omawianiu zagadnienia grupowania danych, wspomina o fenomenie malejącej entropii w przypadku skupiania się rozproszonej materii w gwiazdach. Zaznacza przy tym, że zjawiska tego nie można traktować tylko czysto geometrycznie, bez uwzględnienia sił i energii, odpowiedzialnych za ten proces, dlatego skupienie rozproszonej materii w gwiazdach nie łamie praw termodynamiki (Watanabe

1985, 160). Wspomina też o zjawisku „generowania struktur" przez żywe organizmy (Watanabe 1985, 137). Czyni to teorię Watanabego bardziej wszechstronną i stanowi dobry punkt wyjścia do dalszej refleksji nad znaczeniem entropii dla całokształtu naszego istnienia (Rifkin, Howard 2008).

Artykuł jest skoncentrowany jednak na omówieniu zasad prostoty i zmniejszenia entropii; zasady te z kolei odnoszą czytelnika do następnych, klasycznych zagadnień filozoficznych.

Różne ujęcia entropii

Ponieważ Watanabe w swoich książkach odwołuje się do powszechnie stosowanych wzorów na entropię, konieczne jest niniejsze, skrótowe przedstawienie rozwoju myśli naukowej na ten temat. Po raz pierwszy pojęcie entropii wprowadził Clausius w roku 1865. W dziedzinie termodynamiki entropia oznacza przede wszystkim miarę nieuporządkowania cząstek. Zależność przyrostu entropii od ciepła, dostarczonego do układu oraz temperatury bezwzględnej, definiujemy następująco[1]:

$$\Delta S = \frac{\Delta Q}{T}, \tag{1}$$

gdzie ΔS jest przyrostem entropii, ΔQ – ciepłem, dostarczonym do układu, T – temperaturą bezwzględną (Maes, Netočný 2014, 188; Harrison 1975, 42). Prawo to opiera się na obserwacji wzrostu prędkości cząstek przy dostarczeniu ciepła do układu. Im większa prędkość cząstek, tym większe ich nieuporządkowanie. Ten typ entropii jest odmienny od omawianej w dalszej części artykułu, entropii struktury.

Powyższe prawo Clausiusa jest ujęciem typowo makroskopowym (Harrison 1975, 45). Pod koniec dziewiętnastego wieku Boltzmann, Gibbs, Maxwell czynili wysiłki, aby pojęcie entropii oprzeć o perspektywę atomistyczną, mikroskopową i statystyczną (Watanabe 1985, 139). W równaniach zaczęto stosować funkcję logarytmiczną, która zapewnia odpowiednie dodawanie prawdopodobieństw. Jeśli prawdopodobieństwo jakiegoś stanu układu jest bliskie jedności, czyli jest ono bardzo prawdopodobne, wówczas funkcja logarytmiczna przyjmuje wartość bliską zeru, a co za tym idzie, entropia układu jest mała. Jeśli jednak w danym układzie prawdopodobieństwo danego stanu jest małe, a więc istnieje szereg, mało praw-

---

[1] Podany wzór, jak również dwa następne, są stosowane w literaturze, dotyczącej termodynamiki. Podane prawa pochodzą z pracy Harrisona (Harrison 1975, 41–59). W literaturze, podanej przez niego znajdują się dwadzieścia trzy pozycje, takie, jak np. Termodynamika A. W. Portera z 1946 r., Termodynamika H. Zeisego z 1944 r., prace L. Boltzmanna, dotyczące teorii ciepła i rachunku prawdopodobieństwa z 1877 r., jego wykłady o teorii gazu, wydane w 1923 r. Ponieważ Harrison w dalszej części odnosi się do nowszych ujęć entropii w fizyce, znajdują się także wśród bibliografii prace, takie, jak np. artykuł z czasopisma *Physics Today* z roku 1972, autorstwa I. Prigogine, G. Nicolisa, A. Babloyantza, dotyczący termodynamiki ewolucji.

dopodobnych stanów, wówczas funkcja logarytmiczna przyjmuje większą wartość, co do wartości bezwzględnej, a także entropia przyjmuje większe wartości.

Jednym z pierwszych, nowych ujęć entropii był wzór Boltzmanna, w którym tę wielkość wyrażono przy pomocy funkcji logarytmicznej i za pomocą całek (Bobylev, Cercignani 1999, 603; Harrison 1975, 46):

$$S = -k \iiint f(\vec{v}) \log f(\vec{v}) d\vec{v}, \tag{2}$$

gdzie S oznacza entropię, k – stałą Boltzmanna, $f(\vec{v})d\vec{v}$ liczbę molekuł, których prędkość zawiera się w przedziale $d\vec{v}$. W ujęciu statystycznym, analogicznie do powyższego wzoru Boltzmanna, entropię można wyrazić w postaci bardziej powszechnie stosowanego wzoru, który jest sumą następujących iloczynów (Peters 1975, 72–74; Harrison 1975, 51; Ramaswamy 2001, 3):

$$S = -k \sum_i P(\Omega_i) \ln P(\Omega_i), \tag{3}$$

gdzie $P(\Omega_i)$ jest rozkładem prawdopodobieństwa zajmowania poszczególnego mikrostanu. Równanie to można też zapisać w postaci następującej (Ramaswamy 2001, 4; Harrison 1975, 46):

$$S = k \log W, \tag{4}$$

gdzie W jest liczbą fizycznie odrębnych mikrostanów.

Wzory te, z różnicą co do stałej k Boltzmanna i podstawy funkcji logarytmicznej, przypominają wzory, określające entropię w dziedzinie teorii informacji, lub ilość informacji, zawartą w danym układzie. Matematyczna definicja ilości informacji jest następująca: ilością informacji, zawartej w zbiorze $X = \{x_1, x_2, \ldots x_n\}$, gdzie prawdopodobieństwem elementu x_i jest liczba p_i, nazywamy liczbę (Strawiński 1991, 110–111):

$$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i \,. \tag{5}$$

Wielkość $H(X)$ nazywana jest czasem całkowitą entropią informacyjną układu, reprezentowanego przez zbiór stanów x_i. Jest to wzór, przedstawiający teorię Shannona.

Wzór ten, z innymi oznaczeniami, z różnicą w podstawie logarytmu, stosuje też Watanabe, zaznaczając jednocześnie, że przy równych prawdopodobieństwach entropia, określona jako entropia typu Hartleya, będzie wyrażona jeszcze prostszym wzorem (Watanabe 1985, 122–123):

$$S = \log n \tag{6}$$

Można zauważyć podobieństwo między wzorami na entropię informacyjną (5), (6) a wcześniejszymi wzorami na entropię układu cząstek (3), (4), w zależności od mikrostanów. Zazwyczaj entropię informacyjną i ilość możliwej informacji, zawartej w jakimś układzie, rozpatruje się na przykładzie ciągu n znaków. W takim przykładzie liczba znaków n alfabetu, odpowiadałaby liczbie mikrostanów, w których mogą znaleźć się cząstki, a prawdopodobieństwo mikrostanu, prawdopodobieństwu pojawienia się danego znaku alfabetu. Watanabe skorzysta z powyższych zależności w innym, rozpatrywanym przez niego problemie, ilości możliwych krzywych, przechodzących przez podane punkty, stosując podobne wzory, jak w przypadku entropii informacyjnej n-wyrazowego ciągu znaków (Watanabe 1985, 123). Prezentuje on również wiele użytecznych wzorów do obliczenia entropii, zaczynając od wzoru analogicznego do wzorów (3) i (5), ale również omawia wzory z zastosowaniem całek, zastosowanych w równaniu (2) oraz wzory oparte na rachunku macierzowym, analogiczne do wzoru (7), przedstawionego poniżej (Watanabe 1985, 145–152). Obszernie omawia również matematyczne właściwości wzorów na entropię (Watanabe 1969, 15–27). Niniejszy artykuł pomija te szczegółowe rozważania, koncentrując uwagę bardziej na wątkach filozoficznych.

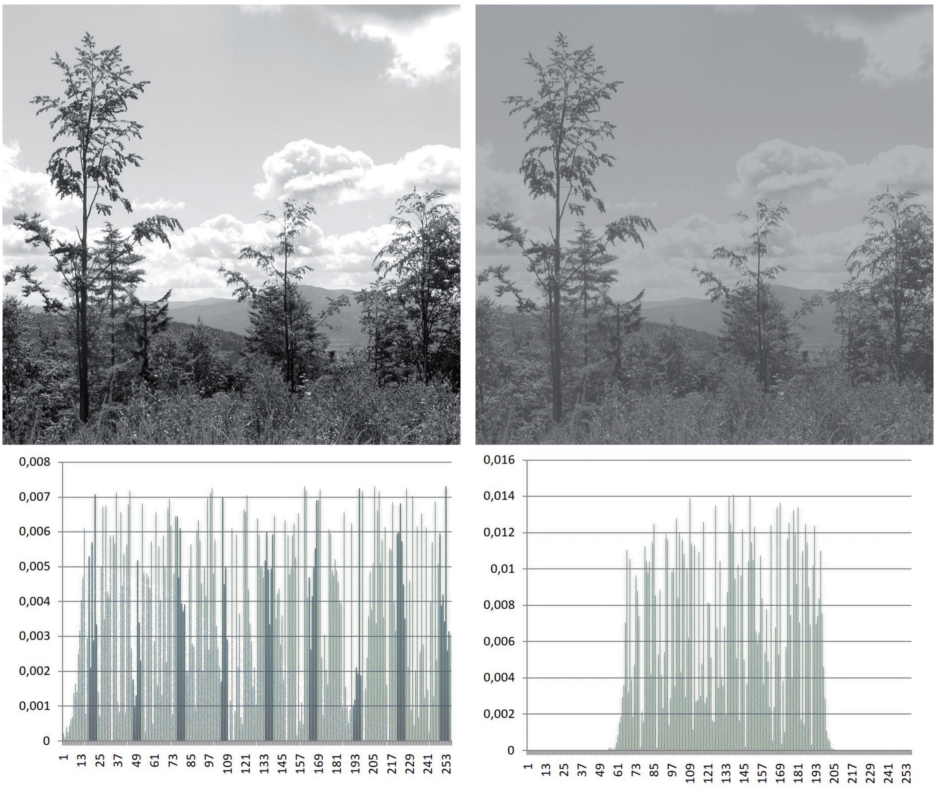
Entropia w przetwarzaniu obrazów

Dla uniknięcia niejasności należy podkreślić, że w literaturze, dotyczącej przetwarzania obrazów, występuje często inny sposób liczenia (choć oparty na tych samych wzorach) i pojmowania entropii w stosunku do entropii struktury, opisanej w dalszej części artykułu. Entropię obrazu można odnosić do częstotliwości występowania kolorów lub odcieni szarości poszczególnych jego pikseli (Santhi, Wahida Banu 2015, 78–81). Podobnie można czynić z wariancją, o której jest mowa w dalszej części artykułu.

Zauważmy, że jest to ujęcie, co do istoty, analogiczne z termodynamicznym. Jeśli np. skalę szarości poszczególnych pikseli porównamy do prędkości poszczególnych cząstek w gazie, skalując o odpowiednie wielkości, otrzymamy analogiczny sposób liczenia entropii, jak dla entropii Boltzmanna. Możemy wówczas otrzymać takie same wartości entropii, niezależnie od tego, czy na obrazie są widoczne rozpoznawane przez człowieka obiekty, czy nie.

Dla tak policzonej entropii możemy zaobserwować, jak na rysunku 1, że większy kontrast obrazu wiąże się z większym zróżnicowaniem wartości skali szarości poszczególnych pikseli, a więc większą entropią i wariancją tych wartości. W tym więc wypadku, zwiększenie entropii powoduje łatwiejsze uchwycenie wzorców przez ludzkie oko (Santhi, Wahida Banu 2015, 78–86). Tak dzieje się również w wielu wypadkach przetwarzania obrazów np. dla tomografii komputerowej lub zdjęć lotniczych. W literaturze możemy odnaleźć wiele innych zagadnień związanych ze zwiększaniem entropii (Gull, Skilling 1985, 288–289). Nie możemy zatem mówić o jej zmniejszaniu, jeśli nie jest doprecyzowany sposób jej liczenia, jako zasadzie rozpoznawania wzorców. Trzeba także powiedzieć, że nie jest to entropia, która odnosi się do ilościowego ujęcia zwartości, „siły" (Watanabe 1985, 143) samej struktury, rozpoznanej przez człowieka. Powstaje zatem pytanie, czy istnieje taki sposób ujęcia entropii, który pozwala adekwatnie powiązać ją ze strukturą?

Rysunek 1. Przykłady obrazów o większym i mniejszym kontraście oraz ich histogramy. Średnia skali szarości (wartości 0–255) dla lewej fotografii wynosi 167; wariancja 5518,52; entropia 0,53, średnia skali szarości dla prawej fotografii wynosi 146; wariancja 1390,79; entropia 0,26.



Entropia i struktura

Opisy entropii, stosowane w dziedzinie termodynamiki, nie wystarczają jednak do tego, aby opisać jej stosowalność w dziedzinie rozpoznawania wzorców. Watanabe uznaje wzory Boltzmanna, Gibbsa i Maxwella za zbyt uzależnione od pojęć termodynamicznych, choć jak przyznaje, są one skonstruowane z perspektywy bardziej mikroskopowej, atomistycznej. Do rozpatrywania entropii rozpoznawanych wzorców, np. obrazów, potrzebna jest bardziej entropia struktury.

Zdaniem Watanabego pierwszy krok do uniezależnienia entropii od pojęć termodynamicznych uczynił von Neumann w roku 1932, podając wzór[2] na „entropię mikroskopijną" (Cyranski 1984, 175; Watanabe 1985, 139):

$$S(\sigma) = -\mathrm{Trace}\,\sigma\ \ln\sigma \tag{7}$$

Jest to zależność, wyrażona w rachunku macierzowym, gdzie σ jest nieujemną macierzą hermitowską, której ślad[3] wynosi jeden. Biorąc pod uwagę diagonalne elementy macierzy σ, którymi są poszczególne rozkłady prawdopodobieństwa $p_1, p_2, \ldots, p_n$, powyższy wzór (7) można sprowadzić do bardziej powszechnie używanego wzoru, jak (3) i (5). Wzór (7) jest jednak formą bardziej ogólną i został tu podany ze względu na znaczenie, podkreślone przez Watanabego. Von Neumann stosował jednak powyższe prawo do podkreślenia nieodwracalności pewnych fizycznych procesów i nie odnosił się do zagadnienia strukturalności samej materii.

Watanabe zaznacza swój wkład w rozpoczęcie badań nad entropią struktury pod kierunkiem Heisenberga, którego był uczniem. Heisenberg stwierdził, że wysoka współzależność cząstek elementarnych, którą to wcześniej dostrzegł Bohr, musi być kluczem do wyjaśnienia stanu materii również w stanie równowagi, przy zerowej temperaturze. Watanabe zaproponował stosowanie pojęcia „mikroskopijnej entropii" poszczególnych cząstek, według wzoru (7) von Neumanna, do wyrażenia stopnia współzależności materii w każdej temperaturze. Watanabe zaproponował również stosowanie nazwy „entropia budulcowego elementu"[4] (Watanabe 1985, 140), dla podkreślenia, że nie chodzi tu o entropię opartą na wielkościach termodynamicznych.

Termodynamika nie wyjaśnia nam struktury kryształów czy budowy ciał przy niskiej temperaturze. Należało wydobyć te cechy entropii, które są miarą współzależności układów statycznych cząstek. Watanabe przedstawia streszczenie swoich artykułów, poruszających ten temat, na przykładzie prawdopodobieństwa występowania poszczególnych cząstek elementarnych w zależności od występowania innych. Po uwzględnieniu sił jądrowych, wykres gęstości prawdopodobieństwa występowania danej cząstki się zmienia. Można wykazać, że uwzględnienie w obliczeniach realnie występującej siły jądrowej, obok innych praw, jak zasada Pauliego, wzmacnia strukturalność materii i współzależność cząstek (Watanabe 1985, 142).

Rozważania, dotyczące struktury materii i współzależności cząstek, są jednak tylko wstępem do ukazania tego, czym jest entropia struktury, jeśli chodzi o rozpoznawanie wzorców. Pewną strukturę można zauważyć już na poziomie

[2] Watanabe podaje ten wzór w formie: $S = -\,\mathrm{Trace}\ \sigma \log \sigma$, wynika to z różnych konwencji oznaczeń logarytmów.

[3] Ang. trace, czyli suma elementów diagonalnych.

[4] Watanabe podaje dwie nazwy: angielską „building block entropy" oraz niemiecką: „Bausteinentropie".

prostych rozważań, dotyczących uporządkowania cząstek i wysnuć prosty wniosek, że odczytanie wzorca będzie się wiązać ze zmniejszeniem entropii (Cohen, Stewart 1994, 234)[5]. Watanabe przypomina, że pierwszym, który zauważył proces nabywania informacji przy spadku entropii był Szilard w 1929 r., natomiast jego opracowania zapoczątkowały badania nad powiązaniem entropii z wewnętrzną współzależnością i organizacją danego wzorca (Watanabe 1969, 50).

Jak twierdzi japoński fizyk, zadaniem rozpoznawanie wzorców jest odkrywanie struktury lub inaczej formy, którym przeciwstawić można pojęcia chaosu lub amorfizmu (Watanabe 1985, 137). W swym artykule zaznacza również: „W jednym ze sposobów zwięzłego ujęcia istoty rozpoznawania wzorców można stwierdzić, że jest ono pojęciową adaptacją danych empirycznych w celu uchwycenia w nich formy. Forma oznacza strukturę, która zawsze wymaga małych wartości entropii. W terminologii matematycznej, może to być ujęte następująco: rozpoznawanie wzorców jest poszukiwaniem minimum entropii, przy czym system kategorii, stosowany w opisie danych, musi być odpowiednio dopasowany i entropia odpowiednio zdefiniowana” (Watanabe 1981, 381).

Watanabe dwukrotnie na początku artykułu podkreśla potrzebę odpowiedniego zdefiniowania entropii (Watanabe 1981, 381). Kluczowe jest zatem właściwe zrozumienie, jak należy obliczać entropię w kontekście rozpoznawania wzorców oraz jak można obliczać „siłę struktury”[6] i wewnętrznego zorganizowania jakiegoś wzorca, np. obrazu.

Watanabe wychodzi z następującego założenia: „Istnienie struktury oznacza to, że znajomość, dotycząca części, pozwala nam łatwo odgadnąć resztę, dotyczącą całości. (…). Jeśli nie ma korelacji lub współzależności między częściami, entropia całości jest sumą entropii części. Wobec tego, musi zostać przeprowadzone porównanie między entropią całości, a sumą entropii części. Im silniejsza struktura, tym bardziej znacząco suma entropii części będzie przeważać nad entropią całości” (Watanabe 1985, 142).

Można zatem podać następujący wzór, ujmujący zależność między entropią a „siłą struktury” (Watanabe 1985, 143):

$$J = \text{siła struktury} = (\text{suma entropii części}) - (\text{entropia całości}) \qquad (8)$$

Autor sam zaznacza, że istnienie struktury wiąże się zasadniczo z istnieniem małych wartości entropii, jednakże tylko powyższe rozumowanie i zależność, pokazana we wzorze (8), może przybliżać nas do właściwego rozumienia tego związku.

[5] Autorzy przedstawiają następujący przykład: „Załóżmy, że w pewnej chwili wszystkie molekuły tlenu w pokoju są skupione w jednym końcu, a molekuły azotu w drugim. Jest to uporządkowany stan termodynamiczny. Jednak po bardzo krótkim czasie przypadkowe zderzenia wymieszają wszystkie molekuły mniej lub bardziej jednorodnie w całym pokoju. Jest to ortodoksyjny obraz nieubłaganego wzrostu entropii”.

[6] Autor używa określenia: „strength of structure”.

Autor przedstawia również przykład liczenia siły struktury dla systemu, składającego się z dwóch części: $x^{(1)}$ oraz $x^{(2)}$. Siła struktury jest liczona wówczas w następujący sposób (Watanabe 1985, 143):

$$J = -\sum p\big(x^{(1)}\big)\log p\big(x^{(1)}\big) - \sum p\big(x^{(2)}\big)\log p\big(x^{(2)}\big) \\ + \sum p(x^{(1)}, x^{(2)}) \log p(x^{(1)}, x^{(2)})) \tag{9}$$

Dalsze przykłady liczenia entropii struktury można przeczytać w (Watanabe 1969, 52–89). Autor analizuje współzależność dwóch części, których wzajemne powiązanie wyraża się za pomocą prawdopodobieństwa warunkowego – znajomość jednej części determinuje drugą. Siła struktury jest liczona, jak we wzorze (8) (Watanabe 1969, 54). Japoński fizyk podaje również przykład liczenia siły struktury dla bardzo prostego obrazu, składającego się tylko z czterech kwadratów, z których każdy może być biały lub czarny. Struktura zaczyna się pojawiać, gdy jakiś któryś z szesnastu możliwych przypadków jest bardziej prawdopodobny niż inne, a znając wartość jednego kwadratu składowego, możemy łatwiej odgadnąć pozostałe (Watanabe 1969, 61–63).

Watanabe, pisząc potrzebne komentarze, dotyczące wzoru (8) nadmienia, że w fizycznych systemach w stanie równowagi, struktura zdaje się być rosnącą funkcją entropii cząstkowych. Następnie odnosi się do drugiego rodzaju systemów: „W niefizycznych aplikacjach często zdarza się, że entropie części nie zmieniają się znacznie między poszczególnymi przypadkami. W ten sposób strukturalność okazuje się być malejącą funkcją entropii całkowitej. Jest to powód, dla którego nie jest błędnym twierdzenie, że w tych aplikacjach, im mniejsza entropia, tym silniejsza struktura (…). Wracając do rozpoznawania wzorców, możemy przyjąć za zasadę heurystyczną adaptację naszych ram pojęciowych tak, aby zmaksymalizować funkcję struktury, albo – co jest równoważne w wielu przypadkach – minimalizować entropię” (Watanabe 1985, 143–144).

Szukając wzoru, który wyraża siłę struktury, entropię musimy więc liczyć w ściśle określony sposób. Sposób liczenia jest także przystosowany do naszych własnych ram pojęciowych, według których usiłujemy odczytać jakąś strukturę z danych. Watanabe wyraźnie pisze o zmniejszaniu entropii, jako zasadzie rozpoznawania wzorców, mając na uwadze głównie wydobywanie struktury z chaosu, a więc dokonywanie uporządkowania w otaczającym nas świecie. Przełożenie tego poglądu na wielkości matematyczne, nie jest jednak oczywiste oraz wymaga adaptacji do sposobu spostrzegania człowieka.

Wzór (8) stanowi pewną ogólną koncepcję, rachunkowego ujęcia „siły”, lub inaczej mówiąc zwartości struktury. Jest to propozycja dość konsekwentnie rozwijana i uzasadniana przez autora w wielu pracach przez kilkanaście lat. Odpowiedź na pytanie, czy przedstawiona zależność daje adekwatne wyniki w każdym przypadku, również skomplikowanych, złożonych wzorców, wymaga szerszego opracowania.

Wyrażenie „siła struktury" jest wymiennie używane przez japońskiego fizyka z terminem „organizacja", a entropia może być także określona przez indeterminację. „Miarę organizacji" należy również liczyć jako różnicę między indeterminacją poszczególnych części oraz indeterminacją całości. Wzór, określający tę zależność, jest identyczny z wzorem (8) (Watanabe 1969, 52). Pojęcie indeterminacji nawiązuje także do małych wartości prawdopodobieństwa jakiegoś stanu, liczonego jak we wzorach (3)–(6).

Zależność między zmniejszeniem entropii a prostotą

Rozpoznawanie wzorców może być ujmowane jako uproszczenie złożoności i różnorodności w takim sensie, że wiąże ze sobą wielość lub złożoność w jakąś jednostkę, w coś co stanowi zwartą strukturę. Watanabe odnosi się w swoich pracach do filozofii Platona (Watanabe 1985, 7, 47–51), dlatego należy odczytywać to stwierdzenie w kontekście ważnego zagadnienia metafizyki, relacji jedności do wielości. Japoński fizyk mówi o dwóch znaczeniach „widzenia jednego w wielości". Pierwsze to dostrzeganie jednego obiektu w zbiorze części, jak np. dostrzeganie twarzy w zbiorze punktów o różnych odcieniach szarości, natomiast drugie to ujmowanie różnych obiektów we wspólnej klasie, np. kwiaty (Watanabe 1985, 8). Jednak nie każdy aspekt maszynowego rozpoznawania wzorców może być ujęty w ten sposób.

Zasada prostoty może być ponadto, z epistemologicznego punktu widzenia, ujęta jako nawyk, rozwinięty przez mentalny proces, dokonujący uproszczenia w powyższym sensie (Watanabe 1985, 121). Można pokazać, że jest ona również używana jako pewnego rodzaju heurystyczna wytyczna, nie tylko w przypadku rozpoznawania wzorców, ale również w podobnym rozumowaniu, jakim jest indukcja (Watanabe 1985, 97–115).

W celu ukazania związku prostoty ze zmniejszeniem entropii, Watanabe dokonuje następującego rozumowania. Rozważamy prosty przypadek, w którym dane eksperymentalne są wartościami numerycznymi y w funkcji zmiennej x. Uogólnienie skończonej ilości danych do postaci ciągłej funkcji $y = f(x)$ jest rodzajem indukcji[7]. Za taki sam proces uogólnienia może być brane również rozpoznawanie wzorców, w którym dane punkty są danymi mierzalnymi i formuła $y = f(x)$ jest definicją jakiejś klasy wzorca. Domyślnie przyjmuje się, że ilość danych wzrasta wraz z czasem.

Celem jest zatem znalezienie takiej funkcji $y = f(x)$, dla której spełnione byłoby równanie $y_i = f(x_i)$, $i = 1, 2 \dots N$, gdzie para (x_i, y_i) jest i-tym spośród N punktem

[7] Za indukcję może być zatem uznana interpolacja lub aproksymacja, znana w metodach obliczeń numerycznych. Interpolacja dokładnie odwzorowuje tzw. węzły interpolacji, czyli zmierzone doświadczalnie dane. Aproksymacja dopuszcza jakiś błąd w odwzorowaniu przez funkcję aproksymującą zbioru danych. Otrzymujemy w wyniku funkcję ciągłą na podstawie skończonej liczby dyskretnych danych. Na temat aproksymacji i interpolacji zob. np. (Phillips 2003).

danych. Watanabe zaznacza, że możliwe jest dopuszczenie pewnej niedokładności, w której funkcja przebiega przez wyznaczone punkty z pewnym odchyleniem, wówczas równość będzie rozumiana w sensie aproksymacji (Watanabe 1985, 121).

Przyjmujemy, że używamy wielomianowych funkcji $f(x)$, lub mówiąc dokładniej, takich funkcji, dla których obowiązuje zależność:

$$\frac{d^s y}{dx^s} = 0 \tag{10}$$

dla określonego s[8]. Można też pokazać, że s jest liczbą o 1 większą od stopnia wielomianu.

Jeśli dziedzinę funkcji (wartości x) podzielimy w P krokach, podobnie uczynimy z przeciwdziedziną, to liczba możliwych krzywych, przechodzących przez te poszczególne poddziedziny wynosi P^s. Otrzymujemy w ten sposób pewną entropię informacyjną, podobną do entropii ciągu znaków. Tak jak dla ciągu n znaków entropia, według wzoru (6) wynosi log n, analogicznie entropia informacyjna dla ilości możliwych krzywych P^s, wynosi log P^s, zatem po przekształceniu (Watanabe 1981, 382):

$$S \text{ (entropia)} = s \log P \tag{11}$$

Domyślnie przyjęto, że każda z krzywych jest tak samo prawdopodobna (Watanabe 1985, 122), wobec czego można zastosować łatwiejszy wzór na entropię typu Hartleya (6), niż bardziej skomplikowany wzór (5), w którym liczymy pojedyncze prawdopodobieństwa z osobna.

Jeśli mamy t eksperymentalnych danych (punktów), wówczas mamy t ograniczeń, węzłów, i t punktów, przez które musi przechodzić funkcja. Wówczas entropia wyniesie

$$S \text{ (entropia)} = (s - t) \log P \tag{12}$$

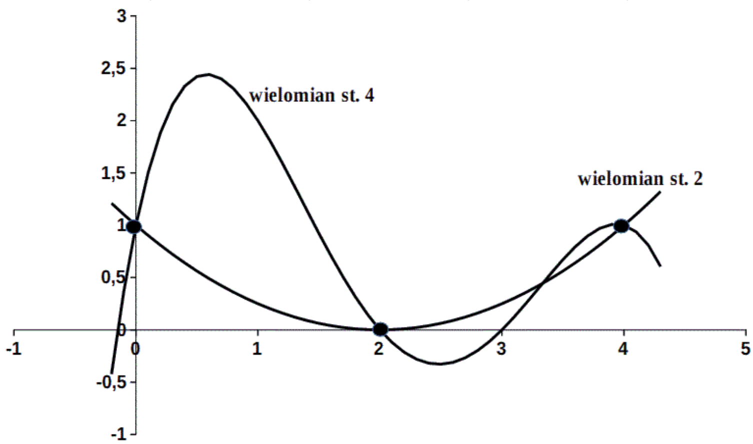
Ponieważ t oraz P w tym równaniu są ustalone, więc zasada zmniejszenia entropii sugeruje wybranie wielomianu z najmniejszym s. Stopień wielomianu wynosi co najwyżej $s - 1$. Watanabe wyciąga tu ważny wniosek: „Tak więc zasada minimum entropii może być przełożona na zasadę prostoty, ponieważ im mniejszy stopień ($s - 1$) wielomianu, tym prostsza jest krzywa” (Watanabe 1981, 382).

Trzeba zauważyć, że Watanabe założył tutaj, że prostota jest utożsamiana z prostotą krzywej wielomianowej. Prostotę i złożoność według powyższego rozumowania, można zatem przedstawić, jak na rysunku 2.

---

[8] Czyli pochodna rzędu s jest równa zero. Taka zależność nie obowiązuje np. dla funkcji $y = e^x$.

Rysunek 2. Przykład prostszej i bardziej złożonej (wielomian wyższego stopnia) krzywej, która charakteryzuje zależność zmierzonych, dyskretnych danych. Na wykresie widzimy dwa wielomiany interpolujące. Wybieramy prostszą krzywą, przebiegającą przez podane punkty.



Warto zauważyć dość istotne kryterium wyboru, jeśli chodzi o ustalenie funkcji ciągłej, która ma charakteryzować skończoną, dyskretną liczbę danych eksperymentalnych. Przez skończoną liczbę punktów można poprowadzić nieskończoną ilość różnych krzywych. Każda z tych krzywych będzie indukcyjnym wnioskiem eksperymentalnym. Jak zauważył H. Reichenbach, przy wyborze jednej z tych krzywych stosujemy regułę „prostoty indukcyjnej". Wybór prostoty indukcyjnej można nazwać według niego swego rodzaju wiarą w to, że najprostsza krzywa dostarczy nam najlepszych przewidywań (Reichenbach 1949, 375). Warto zaznaczyć, że również dla uniknięcia tzw. efektu Rungego, który zwiększa ryzyko błędu interpolacji, wybierany jest prostszy wielomian (Burckel 1979, 256–292).

Warto w tym kontekście zastanowić się nad negatywnymi konsekwencjami zbyt pochopnego uproszczenia, zmierzonych lub dostrzeganych przez człowieka danych. Możemy np. jakąś krzywą potraktować jako prostą, jeśli akurat w danym przedziale nie ma wyraźnej różnicy między nimi. Analogicznie możemy fałszywie uprościć dostrzeganą przez nas rzeczywistość, gdy np. dwa, blisko stojące drzewa potraktujemy jako jedno, lub kształt ptaka uznamy za jeszcze jedną, spośród wielu gałązek w drzewie.

Różne ujęcia prostoty

Ujmowanie prostoty jako prostoty krzywej, jako swego rodzaju niezłożoności graficznej, z jednej strony mieści się w tym, co również w filozofii przyrody zwykło się określać jako prostotę, z drugiej strony może rodzić poważne wątpliwości. Warto zatem choć skrótowo prześledzić różne sposoby ujmowania prostoty i spróbować umiejscowić poglądy japońskiego fizyka w szerszym kontekście. W tym celu warto skorzystać z artykułu D. Sztejnbarg (1932). Watanabe sam odwołuje się do artykułu J. Pearla, jednakże wydaje się dystansować od niektórych tez tam zawartych (Watanabe 1981, 382), a temat prostoty jest omówiony w tym artykule w nieco węższych perspektywach (Pearl 1978, 255–264).

Według Sztejnbarg, jako proste można po pierwsze traktować prawo o nieskomplikowanej strukturze. Można wyliczyć następujące podtypy tego typu prostoty: przejrzystość budowy; prawo, któremu odpowiada krzywa prostszego kształtu; prawo, którego zastosowanie wymaga mniej skomplikowanych operacji matematycznych (np. równanie pierwszego i drugiego stopnia); prostsze jest to prawo, w którym występuje mniejsza liczba kategorii zmiennych (Sztejnbarg 1932, 37–38).

Widzimy zatem, że Watanabe wpisuje się przede wszystkim w pierwsze z wymienionych przez Sztejnbarg, rozumienie prostoty, szczególnie jeśli chodzi o wzmiankę o krzywej prostego kształtu. Możemy jednak odnaleźć zgodność opisu prostoty również w innych punktach.

Warto zwrócić uwagę na czwarte z kolei, z zaproponowanych przez Sztejnbarg, sposobów rozumienia prostoty. Proste są prawa „intuicyjne" i „oczywiste", którym towarzyszy „poczucie powszechnej rozpoznawalności". Watanabe twierdzi natomiast, że „rozpoznawanie wzorców wymaga umysłowej adaptacji, aby dana struktura mogła zaistnieć w naszej percepcji, bądź w mechanicznej symulacji" (Watanabe 1985, 137). Można zatem sądzić, że ten typ prostoty jest również uwzględniony w opracowaniach japońskiego fizyka, gdyż umysł ludzki łatwiej zaadoptuje wzorce ze swej natury łatwiejsze w odbiorze.

W wymienionym jako siódme rozumieniu prostoty, prostymi nazywa się te prawa, które „stwierdzają prosty stan rzeczy"[9], względnie dotyczą „prostych, nierozkładalnych elementów" (Sztejnbarg 1932, 40) Taki typ rozumienia prostoty również jest obecny w myśli Watanabego, na co może wskazywać ujęcie rozpoznawania wzorców jako wysiłku, zmierzającego do wydobycia jednolitej struktury wśród zmierzonych danych. Również cały rozdział książki „rozpoznawanie wzorców jako dostrzeganie jedności w wielości" (Watanabe 1985, 1–20) sugeruje powyższe ujęcie.

Artykuł Sztejnbarg opisuje ponadto inne rozumienie prostoty. Dane prawo możemy nazywać prostym, jeśli daje się ono utrzymać przy mniejszej liczbie założeń dodatkowych, również metafizycznych; bądź dane prawo daje się utrzymać przy prostszych założeniach (Sztejnbarg 1932, 39). Do takiego typu prostoty Watanabe nie odwołuje się bezpośrednio, przynajmniej w rozważaniach dotyczących

---

[9] W niektórych punktach autorka, dla uniknięcia błędu błędnego koła, odwołuje się do punktu pierwszego (prostota jako przejrzystość struktury). Uniknięcie tego błędu jest jednak w tym wypadku trudne, co widać w kolejnych przytaczanych przykładach. Pomocne może być odwołanie się do wariancji, wielkości z dziedziny statystyki, omówione w dalszej części artykułu.

zasad zmniejszenia entropii i prostoty. Ewentualne dalsze wnioski, po przeanalizowaniu jego prac wymagałyby osobnego opracowania[10].

Ponadto Sztejnbarg wymienia jeszcze następujące typy rozumienia prostoty: prostszym jest to prawo, które prowadzi do prostszych (w sensie prostoty struktury) konsekwencji; prostymi są takie prawa, które wynikają z innych „intuicyjnych" i „oczywistych" praw; proste prawa to te, które są nierozkładalne, nieredukowalne do innych elementów; proste jest to prawo, które jest uogólnieniem samych tez spostrzegawczych w przeciwieństwie do tego, które jest uogólnieniem wniosków z tez spostrzegawczych; proste jest to prawo, które nie zawiera terminów mętnych, niejasnych (Sztejnbarg 1932, 39–41). W teorii Watanabego raczej nie odnajdujemy tych ujęć prostoty, przynajmniej jeśli chodzi o rozpatrywane zagadnienie.

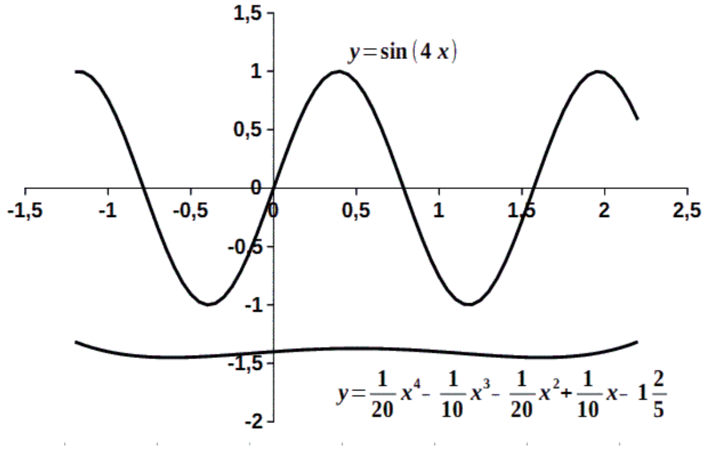
Trzeba oczywiście powiedzieć, że istnieje dużo wątpliwości, dotyczących właściwego ujęcia prostoty, w tym co do jej rozumienia jako geometrycznej prostoty krzywej wielomianowej. Z jednej strony może się wydawać, jak zauważa Harré (Harré 1959, 233), że nie byłoby zbyt trudnym podanie pewnego kryterium prostoty dla krzywych, opartego np. na stopniu zmiany krzywizny, ilości punktów zwrotnych i stacjonarnych. Jednakże nawet i te parametry, jak zauważa Strawiński, określimy dokładnie jedynie na podstawie samego równania. „Jeżeli twierdzimy, że linia prosta jest prostsza niż parabola, czy elipsa, a te z kolei są prostsze od krzywych trzeciego stopnia, to odwołujemy się do własności odpowiednich równań" (Strawiński 1991, 103).

Gdyby uwzględnić samo kryterium geometryczne, to również wg Strawińskiego, byłoby one niezwykle trudne w zastosowaniu np. przy porównaniu prostoty elipsy i paraboli bez odwołania się do ich równań. Gdyby próbować wyliczyć różne typy krzywych, trzeba dokonać dość skomplikowanego podziału. Np. Newton podzielił krzywe trzeciego rzędu na cztery typy, dzielące się następnie na siedemdziesiąt dwie postacie. Do szczegółowej analizy użył wielu pojęć: średnicy, osi wierzchołka, środka, parametru, gałęzi parabolicznej, hiperbolicznej; wprowadził ponadto punkty osobliwe, odosobnione (Strawiński 1991, 104). Również znalezienie kryterium prostoty dla trójwymiarowych tworów geometrycznych stwarza wiele problemów. Trudno np. zdecydować, czy o prostocie bryły geometrycznej ma decydować ilość elementów składowych, czy ilość operacji symetrii, czy może oba te kryteria jednocześnie. Całkowicie inaczej ocenilibyśmy prostotę sześcianu i prostotę kuli (Strawiński 1991, 105–108).

W wypadku samych wielomianów przyjęcie zasady, że im mniejszy stopień wielomianu, tym większa prostota krzywej, nie budzi takich kontrowersji, jak w przypadku porównania krzywych innego rodzaju. Trudność w jednoznacznym określeniu prostoty jakiejś krzywej może ilustrować również rysunek 3.

[10] Wątki, dotyczące zagadnień metafizycznych są obecne szczególnie w rozdziale pierwszym i trzecim cytowanej książki (Watanabe 1985, 1–20, 45–74).

Rysunek 3. W jakimś przedziale krzywa wielomianowa może bardziej przypominać prostą i być jednocześnie oparta na równaniu, złożonym z wielu członów. Inna krzywa, mająca „wizualnie proste" równanie, może odznaczać się wysokim stopniem krzywizny. Jednakże trzeba też powiedzieć, że funkcja może mieć jednak przejrzysty wizualnie zapis, ale być wewnętrznie bardziej złożoną od innych, np. gdyby funkcję sinus przybliżyć szeregiem wielomianowym, miałby on dość skomplikowaną postać.



Przy definiowaniu prostoty możemy odwołać się także do pojęcia z dziedziny statystyki, wariancji. Odwołanie to może także pomóc w uniknięciu błędnego koła w definicji. Wariancja jest wartością oczekiwaną różnic, podniesionych do kwadratu, poszczególnych wartości zmiennej od wartości oczekiwanej tej zmiennej. Wartość oczekiwana dla zmiennej, będącej statystyczną populacją, jest średnią arytmetyczną. Jeśli zatem przez X oznaczymy zmienną, przez $E(X) = \mu$, oznaczymy wartość oczekiwaną zmiennej, to wariancję obliczymy ze wzoru (Dodge 2008, 558):

$$Var(X) = E[(X-\mu)^2] = E(X)^2 - \mu^2 \tag{13}$$

Wariancja jest zatem miarą rozrzutu danej zmiennej (Dodge 2008, 555). Obliczenie wariancji dla zmiennej, będącą skalą szarości poszczególnych pikseli, zaprezentowano w opisie do rysunku 1.

W powyższych rozważaniach możemy zauważyć, że Watanabe powiązał entropię z prostotą, odwołując się ostatecznie do jakiegoś rodzaju graficznej niezłożoności (prostoty) krzywej. Powstaje zatem pytanie, czy możemy entropię, powiązać z wariancją i przy jej pomocy doprecyzować pojęcie prostoty. Jest to zagadnienie dość złożone i przykład obliczeń do rysunku 1 nie powinien stanowić podstawy do zbyt pochopnych uogólnień.

W ogólnym wypadku trzeba powiedzieć, że zależność entropii od wariancji zależy od rozkładu prawdopodobieństwa zmiennej. Zależność ta się różni, choć czasem nieznacząco, np. dla normalnego i trójkątnego rozkładu prawdopodobieństwa (Novitskii, Korol, Ivanova 1973, 42–43). Są także przypadki, w których wariancja i entropia dają różne wyniki. Rozpatrując informatywność zmierzonych danych, analiza za pomocą wariancji i entropii może dać różne wyniki, dzieje się tak np. dla danych eksponencjalnych (Ebrahimi, Maasoumi, Soofi 1999, 77).

Powstaje również pytanie o ostateczne odwołanie, o ostateczne kryterium, które ma stanowić o prostocie: czy ma to być odwołanie do prostoty krzywej, jakiejś formy jej graficznej, wizualnej niezłożoności, odwołanie do wyniku jakiegoś konkretnego działania lub jeszcze do innej wielkości. Powstaje zatem zagadnienie tzw. filozoficznego trylematu: możemy arbitralnie przyjąć jakieś kryterium za ostateczne, zgodzić się, że kryteria będziemy uzasadniać w nieskończoność, bądź przyjąć, że w uzasadnieniu wrócimy do pierwotnego argumentu, zatoczymy koło (Weisberg 2015).

Zmniejszenie entropii w innych aspektach maszynowego rozpoznawania wzorców

Maszynowe rozpoznawanie wzorców ma zasadniczo za zadanie przyporządkowanie jakiegoś wzorca, na podstawie zmierzonych danych, do jakiejś konkretnej klasy wzorców. Decyzję podejmuje przygotowany wcześniej klasyfikator, np. sieć neuronowa. Przygotowanie sieci neuronowej wymaga wcześniejszego uczenia, inaczej trenowania, na podstawie wybranych danych wzorcowych, służących do ukształtowania odpowiedniego działania sieci neuronowej lub innego klasyfikatora.

Jeśli zdefiniujemy entropię jako różnorodne, nieuporządkowane odpowiedzi klasyfikatora, przyporządkowującego dane wzorce do odpowiednich klas, to zasadą procesu uczenia klasyfikatora okazuje się również być zmniejszenie entropii (Watanabe 1981, 382). Uczenie, o którym mowa, jest dostosowaniem klasyfikatora do takiego działania, aby rozpoznawał wzorce, wcześniej istniejące w naszych pojęciach. Niezależnie od tego, czy nasza pojęciowa klasyfikacja wzorców jest poprawna, czy nie, trenujemy klasyfikator do jakiegoś określonego zachowania. Uczenie „złej" odpowiedzi jest też uczeniem.

Trzeba jednak zaznaczyć, że proces uczenia nie może być wiązany z malejącą entropią dla każdego zbioru arbitralnie wybranych zachowań lub schematów. Watanabe podaje przykład zachowania szczura w pomieszczeniu: gdybyśmy rozpatrywali schemat zachowania szczura przez ujęcie parzystości i nieparzystości jego kroków oraz zwrotów w prawo i lewo, nie dostrzeglibyśmy żadnej regularności. Dlatego uczenie powinno być scharakteryzowane przez istnienie jakiegoś zbioru kategorii, behawioralnych schematów, po zastosowaniu których możemy obserwować zmniejszenie entropii (Watanabe 1981, 383). Zmniejszenie entropii, o którym tu mowa, odwołuje się szczególnie do prostoty, jako poczucia powszechnej rozpoznawalności według klasyfikacji D. Sztejnbarg, ale również do przejrzystości budowy i nierozkładalności elementów.

Zasadę zmniejszenia entropii odszukać możemy również w wielu elementach wstępnej obróbki danych[11]. Jedną z metod wstępnej obróbki danych jest metoda osi głównej[12], która może prowadzić do zredukowania liczby wymiarów zebranych danych.

Watanabe ilustruje to w sposób następujący. Dany jest zbiór N wektorów (lub punktów) danych $x_i^{(a)}$, $a = 1, 2, \dots N$, $i = 1, 2, \dots n$, w n-wymiarowej przestrzeni danych. Jeśli każdy z tych wektorów znormalizujemy, tzn., że będzie obowiązywała zależność:

$$\sum_{i=1}^{n}(x_i^{(\alpha)})^2 = 1 \tag{14}$$

Możemy wówczas wprowadzić „wagę":

$$p_i = \frac{1}{N}\sum_{\alpha=1}^{N}(x_i^{(\alpha)})^2 \tag{15}$$

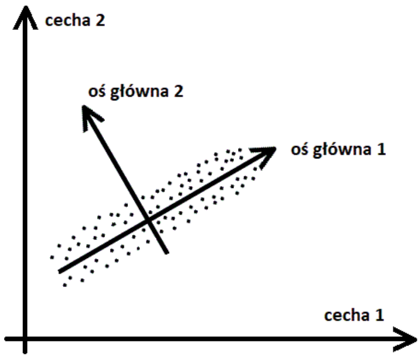
Możemy interpretować p_i jako znaczenie i-tej osi w reprezentowaniu zbioru N wektorów (Watanabe 1985, 154). Przy znormalizowanych wektorach, wielkość ta jest analogiczna do prawdopodobieństwa. Możemy zatem policzyć entropię jako miarę różnorodności znaczenia poszczególnych osi w reprezentowaniu wektorów danych, według wzoru, analogicznego do wzoru (5) (Watanabe 1981, 384–385):

$$S = -\sum_{i=1}^{n} p_i \log p_i \tag{16}$$

Zmniejszenie entropii będzie w tym wypadku polegać na stosowaniu macierzy obrotów ortogonalnych w taki sposób, aby zmniejszyć entropię i wydobyć najbardziej znaczące osie danych. W przypadku dwóch wymiarów jest to obrót na płaszczyźnie. Również w tym wypadku zmniejszenie entropii jest powiązane z prostotą, rozumianą jako przejrzystość budowy, mało skomplikowana struktura. Geometryczną interpretację metody osi głównej można przedstawić, jak na rysunku 4.

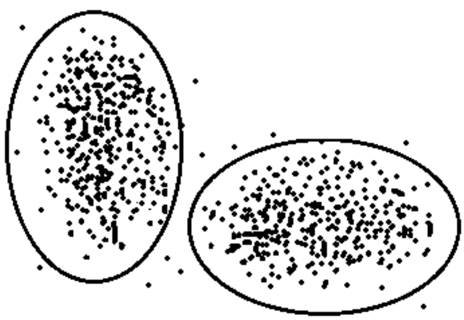
[11] Ang. preprocessing.
[12] Ang. principal axis method.

Rysunek 4. Zarówno oś cechy 1, jak i oś cechy 2 nie oddają właściwego rozróżnienia punktów danych, jednak oś główna 1 rozróżnia je bardzo efektywnie, natomiast oś główna 2 ma w tym względzie niewielkie znaczenie, które można pominąć



Zasada zmniejszenia entropii jest obecna również na etapie tzw. klasteryzacji[13], inaczej grupowania danych. Watanabe proponuje dynamiczny model łączenia danych w poszczególne grupy. Dla zobrazowania zagadnienia porównuje grupowanie danych m.in. do łączenia się ze sobą kropelek oleju na powierzchni stygnącej wody (Watanabe 1985, 162). Grupowanie danych wprowadza uporządkowanie i strukturę wśród rozproszonych punktów. Grupowanie danych można przedstawić graficznie, jak na rysunku 5.

Rysunek 5. Graficzne przedstawienie grupowania danych do względnie jednorodnych klas



Przy dzieleniu danych zasada zmniejszenia entropii jest obecna w inny sposób. Watanabe podaje przykład podziału grupy ludzi na mniejsze grupy ze względu na jakieś kryterium, np. na wzajemną znajomość danych osób. Takie kryterium

[13] Ang. clustering.

jest dość naturalne, natomiast nie jest ono jednorodne, ponieważ może istnieć znajomość pomiędzy osobami z poszczególnych podgrup po podziale. W takim wypadku zasadą podziału będzie maksymalizacja spójności wewnątrz podgrupy i zmniejszenie spójności międzygrupowej. W takim wypadku zastosowana będzie funkcja entropii, wyrażająca współzależność. Punktem wyjścia jest w tym wypadku wzór (8), wyrażający entropię struktury (Watanabe 1985, 143, 148–149, 170).

Podobnie przy rozpatrywaniu funkcji lub płaszczyzny dyskryminacyjnej[14], dzielącej dane na poszczególne klasy, rozpatrywana jest funkcja, która wyraża różnicę entropii, podobnie jak we wzorze (8) (Watanabe 1985, 143, 146–147, 179).

Watanabe omawia ponadto zagadnienie zmniejszenia entropii w dziedzinach pokrewnych w stosunku do rozpoznawania wzorców, takich jak kompresja informacji i teoria automatyki (Watanabe 1985, 183–198). Tematy te wymagają osobnego opracowania.

Podsumowanie

Przedstawione przykłady pokazują, że zmniejszenie entropii oraz prostota są jednymi z ważniejszych zasad maszynowego rozpoznawania wzorców. Zasada prostoty może być wielorako rozumiana, odwołanie Watanabego do prostoty krzywej wielomianowej może budzić wiele kontrowersji. Można jednak wskazać na pewną ograniczoną zgodność, co do rozumienia prostoty, pomimo trudności w określeniu, jak można ją precyzyjnie ująć. Watanabe wyraźnie jednak zaznacza, że rozpoznanie danego wzorca zależy również od spostrzeżeń człowieka, od tego, co według własnych interpretacji doda on do spostrzeganych obiektów. Przykładem może być przypisanie jakiejś chmurze kształtu np. twarzy lisa, lub dostrzeganie ruchu na jakimś rysunku (Watanabe 1985, 160). Rodzi to oczywiście dalsze pytania o pochodzenie wzorców w naszym umyśle i tym samym tematyka artykułu ociera się o ważne i historycznie doniosłe zagadnienie filozofii.

Można powiedzieć, że rozpoznawanie wzorców polega m.in. na formułowaniu i modyfikowaniu punktu odniesienia w taki sposób, aby zmniejszyć entropię. Entropia musi być z kolei właściwie zdefiniowana, w zależności od samego punktu odniesienia (Watanabe 1981, 387), na przykład dla rozpoznania struktury w jakimś obrazie potrzebne jest odpowiednie zdefiniowanie entropii struktury. Nie każdy zatem sposób ujęcia entropii, np. w przetwarzaniu obrazu i zwiększaniu jego kontrastu, nadaje się do sformułowania tezy artykułu. Z tego względu epistemologiczne pytanie o relację naszego sposobu poznawania z porządkiem praw fizycznych, w szczególności zasad termodynamiki, jest bardzo złożone, bo zależy od sposobu ujęcia entropii.

[14] Ang. discriminant surface.

Zasadę prostoty rozważano w kontekście różnych nauk przyrodniczych. Jak wykazuje niniejszy artykuł, jest ona obecna również w dziedzinie maszynowego rozpoznawania wzorców i wniosek ten może stanowić dalszy impuls do namysłu nad znaczeniem prostoty w całokształcie wiedzy ludzkiej. Gdyby rozpoznawanie wzorców rozpatrywać jedynie w kontekście zmniejszenia entropii, bez odwołania np. do prostoty, moglibyśmy mieć wrażenie pewnego błędnego koła: entropia z definicji jest miarą nieuporządkowania, więc wprowadzenie uporządkowania również, z definicji, zmniejszałoby entropię. Stwierdzenie to byłoby zbytnim uproszczeniem, ponieważ pojęcie entropii może być różnorodnie rozumiane, co pokazano na przykładach.

W powyższych rozważaniach bardziej chodzi jednak o ponowny impuls do namysłu nad ważnym stwierdzeniem Quine'a: „Nie jest zaskoczeniem, że twórcy teorii szukają prostoty. Jeśli dwie teorie są równie odporne na zarzuty pod różnymi względami, z pewnością prostsza z nich będzie preferowana na podstawie zarówno estetyki, jak i dogodności. Jest jednak niezwykłe to, że prostsza teoria jest nie tylko bardziej pożądana, ale i bardziej prawdopodobna" (Quine 1964, 47).

Teza ta nie dotyczy tylko maszynowego rozpoznawania wzorców, lub jakiegoś wąskiego zagadnienia w ramach tej dziedziny, ale całokształtu wiedzy ludzkiej, dlatego obszar dalszych badań jest bardzo szeroki. Ograniczając się do zagadnień związanych z tematyką artykułu, można wskazać np. na zagadnienie zmniejszenia entropii w zastosowaniu sieci neuronowych (Santos, Alexandre, de Sá 2005).

Osobnym, ważnym zagadnieniem jest możliwość zbytniego uproszczenia, pewnego redukcjonizmu w kontekście maszynowego rozpoznawania wzorców. Źle stosowana zasada prostoty mogłaby doprowadzić do ignorowania niektórych istotnych informacji, zawartych w mierzonych danych. Temat ten tylko pośrednio jest obecny w rozważaniach Watanabego i wymaga osobnych badań.

Literatura

Bartlett, S.J., 2015, *The species problem and its logic. Inescapable Ambiguity and Framework-relativity*, Dokument internetowy: https://arxiv.org/ftp/arxiv/papers/ 1510/1510.01589.pdf [dostęp: 13.06.2018].

Bobylev, A., Cercignani, C., 1999, *On the Rate of Entropy Production for the Boltzmann Equation*, "Journal of Statistical Physics", 94.3-4: 603–618.

Burckel, R., 1979, *An introduction to classical complex analysis*, Birkhäuser Basels.

Cohen, J., Stewart, I., 1994, *Załamanie chaosu*, tłum. M. Tempczyk, Warszawa: Prószyński i S-ka.

Cyranski, J., 1984, *Von Neumann Entropy as Information Rate*, "International Journal of Theoretical Physics", 24.2: 175–178.

Dodge, Y., 2008, *The Concise Encyclopedia of Statistics*, Springer, New York.

Duin, R., Pękalska, E., 2007, *The Science of Pattern Recognition. Achievements and Perspectives*, w: Duch W., Mańdziuk J. (red.) *Challenges for Computational Intelligence*, Springer, Berlin, Heidelberg, 221–259.

Ebrahimi, N, Maasoumi, E., Soofi, E.S., 1999, *Measuring Informativeness of Data by Entropy and Variance,* w: Slottje D. (red.), *Advances in Econometrics, Income Distribution and Scientific Methodology*, Physica-Verlag HD, 61–77.

Gull, S.F., Skilling, J., 1985, *The Entropy of an Image*, w: Smith C.R., Grandy W.T. (red.), *Maximum-Entropy and Bayesian Methods in Inverse Problems. Fundamental Theories of Physics (A New International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application)*, vol. 14, Springer, Dordrecht.

Harré, R., 1959, *Simplicity as a criterion of induction*, "Philosophy", 34.130: 229–234.

Harrison, M., 1975, *Entropy concepts in physics*, w: Kubát L., Zeman J. (red.), *Entropy and Information in Science and Philosophy*, Praga: Publishing House of the Czechoslovak Academy of Sciences, 41–59.

Howard, T., Rifkin, J., 2008, *Entropia, nowy światopogląd*, tłum. B. Baczyńska, Katowice: KOS.

Maes, C., Netočný, K., 2014, *A Nonequilibrium Extension of the Clausius Heat Theorem*, "Journal of Statistical Physics", 154: 188–203.

Novitskii, P.V., Korol, E.I., Ivanova, V.Ya., 1973, *Dependence of variance of estimates of RMS value and entropy value of error on number of observations*, "Measurment Techniques", 16: 41–45.

Pearl, J., 1978, *On the connection between the complexity and credibility of inferred models*, "International Journal of General Systems", 4: 255–264.

Pelillo, M., 2014, *Philosophical Aspects of Pattern Recognition*, ICPR 2014, Dokument internetowy: www.dsi.unive.it/~pelillo/ICPR2014_Tutorial.pdf [dostęp: 6.04.2018].

Peters, J., 1975, *Entropy and Information: Conformities and Controversies*, w: Kubát L., Zeman J. (red.), *Entropy and Information in Science and Philosophy*, Praga: Publishing House of the Czechoslovak Academy of Sciences.

Phillips, G.M., 2003, *Interpolation and approximation by polynomials*, Springer, New York.

Quine, W., 1964, *On simple theories of a complex world*, w: Gregg J., Harris F. (red.), *Form and strategy in science*, Dordrecht: Springer, 47–50.

Reichenbach, H., 1949, *Experience and prediction*, Chicago.

Ramaswamy, S., 2001, *Ludwig Boltzmann and entropy*, "Resonance", 6.9: 3–5.

Santhi, K., Wahida, Banu R.S.D., 2015, *Contrast enhancement by modified octagon histogram equalization*, "Signal, Image and Video Processing", 9: 73–87.

Santos, J.M., Alexandre, L.A., de Sá, J.M., 2005, *Neural Network Classification Using Error Entropy Minimization*, w: Apolloni B., Marinaro M., Tagliaferri R. (red.), *Biological and Artificial Intelligence Environments*, Springer, Dordrecht.

Scantamburlo, T., 2013, *Philosopfical Aspects in Pattern Recognition Research*, Dokument internetowy: http://dspace.unive.it/bitstream/handle/10579/4639/phdthesis_Teresa Scantamburlo.pdf?sequence=1 [dostęp: 6.04.2018].

Strawiński, W., 1991, *Prostota, Redukcja, Jedność nauki*, Warszawa: FEA.

Sztejnbarg, D., 1932, *Zagadnienie indeterminizmu na terenie fizyki współczesnej*, „Przegląd Filozoficzny", 35.2-3: 34–69.

Watanabe, S., 1969, *Knowing and Guessing*, Nowy York, Londyn, Sydney, Toronto: John Wiley & Sons.

Watanabe, S., 1981, *Pattern recognition as a quest for minimum entropy*, "Pattern Recognition", 13.5: 381–387.

Watanabe, S., 1985, *Pattern Recognition: Human and Mechanical*, New York, Cichester, Brisbane, Toronto, Singapore: John Wiley & Sons.

Weisberg, J., 2015, *Formal epistemology*, Dokument internetowy: https://plato.stanford.edu/entries/formal-epistemology/ [dostęp: 12.06.2018].