

Ondrej Majer

Teoria zdarzeń sekwencyjnych

Artykuł dotyczy problematyki poszukiwania prawidłowych zdarzeń sekwencyjnych w procesie obserwacji oraz formułowania hipotez o dalszym pojawianiu się prawidłowych zdarzeń sekwencyjnych w tym procesie, a ponadto podejmuje kwestię oceny tych hipotez. Przez proces rozumiem ciąg zjawisk przebiegających w czasie, a przez zdarzenie sekwencyjne — każdą (czasowo) powiązaną i ograniczoną część procesu. Innymi słowy — zdarzenia sekwencyjne są zdarzeniami wypełniającymi pewien interwał czasowy; są zdarzeniami, w których można mówić o momencie ich początku, o odcinku czasowym, kiedy (nieprzerwanie) przebiegają, i o momencie, w którym następuje ich koniec.

W pierwszej części artykułu przedstawiony jest model obserwatora, który łączy (rozstrzyga) zadanie poszukiwania i aktualizowania prawidłowości. Druga część artykułu jest poświęcona problematyce oceny hipotez za pomocą logiki predykatów 1. rzędu. Tu zdefiniowany jest język zdarzeń sekwencyjnych i aksjomaty teorii zdarzeń sekwencyjnych. W artykule udowodniono również, że przedstawiona teoria jest niesprzeczna (ma model). Następnie dla pewnej klasy wyrażeń języka zdarzeń sekwencyjnych zdefiniowana została funkcja oceniania. Ta funkcja określa (podobnie jak Carnapa funkcja potwierdzania) stopień, w którym aktualnie obserwowane zdarzenie (wyrażone za pomocą pewnej formuły) potwierdza hipotezę (również wyrażoną za pomocą pewnej formuły). Na zakończenie rozważa się, jaki typ prawdopodobieństwa przedstawia skonstruowana funkcja oceniania.

1. MODEL OBSERWATORA

W tej części opisany jest model, który posłuży jako podstawa dla rozważania całej problematyki zdarzeń sekwencyjnych.

Punktem wyjścia jest wyobrażenie obserwatora, którego celem jest znalezienie prawidłowości nieznanego procesu mającego jakościowy charakter i przebiegającego w czasie. Chociaż będę posługiwać się spersonifikowanym terminem 'obserwator', to nie mam tu koniecznie na myśli jakiegoś człowieka wykonującego czynność obserwacji. Tej możliwości zdecydowanie nie wykluczam, nie interesują mnie jednak 'ludzkie' (tzn. psychologiczne lub fizjologiczne) aspekty obserwacji. Obserwatorem zgodnie z tym wyobrażeniem może więc być także pewien przyrząd. Przez obserwatora rozumiem bowiem pewien algorytmiczny sposób ewidencji i opracowania informacji przychodzącej w czasie, przy czym nieistotne jest, kto (lub co) tę czynność wykonuje.

Obserwator śledzi przebieg nieznanego procesu w *dyskretnych momentach czasowych* (nie mam więc na uwadze ciągłych procesów reprezentowanych zmianą jakiejś wielkości w czasie), a rezultatem obserwacji w danym momencie jest *elementarny stan procesu*. Przyjmuję, że ten elementarny stan nie jest dalej strukturalizowany i że ma charakter jakościowy. Typowym przykładem procesu może być następstwo stanów jakiegoś systemu biologicznego. Zakładam również, że obserwator dysponuje jakimś systemem znaków, który umożliwia reprezentację zjawisk obserwowanych w przebiegu procesu. Wyrażenia tego systemu są łańcuchami znaków oznaczających elementarne zjawiska (stany). (Formalnie można powiedzieć, że zdarzenia sekwencyjne mogą być reprezentowane jako półgrupa z asocjatywną niekomutatywną operacją uszeregowania, gdzie generatorem tej półgrupy jest zbiór odpowiadający stanom elementarnym.)

Proces można teraz wyspecyfikować jako czasowe następstwo (ze stałym początkiem, ale potencjalnie nieskończone) stanów elementarnych. Z definicji zdarzeń sekwencyjnych wynika, że nieistotne są aspekty takie, jak długość interwałów między stanami elementarnymi, regularność ewentualnie nieregularność tych interwałów, ich absolutne zlokalizowanie w czasie itp. Jedyne, co jest istotne, to ich następstwo. (Formalnie można proces przedstawić jako ciąg znaków reprezentujących stany elementarne, gdzie zbiorem indeksów (zbiorem reprezentującym momenty czasowe) jest zbiór liczb naturalnych.)

Wracając do zadania obserwatora — zadanie to polega na wynajdywaniu w obserwowanym procesie zdarzeń sekwencyjnych, które są prawidłowe i przewidywaniu ich występowania w dalszym przebiegu procesu. Zadanie to polega na dokonaniu pewnego rodzaju indukcji. Nim ono zostanie wyspecyfikowane bardziej szczegółowo, należy wcześniej wyjaśnić założenia, które dla rozwiązania tego zadania są niezbędne.

Większość filozofów zgodna jest w tym, że ważność sądów indukcyjnych i stosowanie metod indukcyjnych w rozwiązywaniu praktycznych zadań jest uwarunkowane założeniem o jedności naszego świata. Założenie to, zwane Zasadą Jedności,¹

¹ W angielskich tekstach oprócz terminu *Principle of Uniformity* stosuje się także termin *Principle of Uniformity of Nature* (czeski ekwiwalent: *Princip uniformity světa*).

zostało po raz pierwszy sformułowane przez Davida Hume'a w 1770 roku² w związku z wnioskowaniem indukcyjnym i od tego czasu Zasada stała się przedmiotem wielu dyskusji i sporów. Jej pierwotne sformułowanie brzmiało: „*the future will remember the past*” (przyszłość podobna jest do przeszłości).

Problematyczność tej zasady wynika z jej zbyt ogólności i nieokreśloności, co nie raz było przedmiotem krytyki. Na przykład C. Howson i P. Urbach o Zasadzie Jedności piszą:

[...] w postaci, w jakiej [Zasada Jedności] jest sformułowana, jest ona nie do użycia, bowiem nie specyfikuje w jakich aspektach przyszłość miałaby być podobna do przeszłości. Aby Zasada Jedności pełniła zamierzoną rolę musiałaby być, w celu jej zastosowania, w każdym konkretnym przypadku sformułowana specyficznie [...] lecz skoro tylko Zasada Jedności świata byłaby sformułowana dostatecznie pewnie i mogła w ten sposób być łącznikiem między danymi obserwacyjnymi a pewnymi ogólnymi prawami, to jej nieadekwatność jako podstawa dla wnioskowania naukowego staje się widoczna, albowiem jej ważność jest w tym momencie tak samo wątpliwa jak ważność teorii naukowej, którą miała zwieńczyć.³

R. Carnap krytykował obrońców Zasady za starania o zasadniczo bardziej specyficzne sformułowanie uważając, że:⁴

[Wnioskowania indukcyjne] byłyby prawdziwe, o ile świat, jako całość, wykazywałby pewien rodzaj jedności w tym znaczeniu, że jakiś typ zdarzeń, pojawiający się bardzo często w określonych warunkach w przeszłości, będzie w takich samych warunkach bardzo często pojawiać także i w przyszłości.

W miejsce potocznej formy Zasady (świat jest jednolity) Carnap proponuje stosować kwantytatywną jej formę w jednym z następujących wariantów (dla przejrzystości oznaczyłem je C1 i C2):⁵

(C1) „Stopień jedności świata jest wysoki.”

lub

(C2) „O ile relatywna liczebność własności w długim początkowym odcinku następstw jest wysoka (na przykład r), to tak samo będzie wysoka (w przybliżeniu równa r) w dostatecznie długiej kontynuacji tych następstw.”

Dalej Carnap zajmuje się problemem weryfikacji Zasady Jedności: Zasada jest twierdzeniem syntetycznym,⁶ tzn. twierdzeniem zależnym od stanu świata, ale zasad-

² D. Hume, *An Inquiry of Human Understanding*. Cytowane według polskiego wydania *Traktatu o naturze ludzkiej*, Warszawa 1963, w przekładzie Czesława Znamierowskiego, I, III, 12, s. 178.

³ C. Howson, P. Urbach, *Scientific Reasoning: The Bayesian Approach* (Chicago: Open Court), Chicago 1993, s. 4.

⁴ R. Carnap, *Logical Foundations of Probability*, Chicago 1962, s. 178.

⁵ Tamże, s. 179.

⁶ W znaczeniu Carnapowskiego rozróżnienia zdań na analityczne i syntetyczne.

niczo różnym od innych hipotez faktualnych np. praw fizykalnych — nie można jej bowiem testować empirycznie. Testowanie jest metodą indukcyjną i jako takie oparte jest na Zasadzie Jedności. Dochodzimy zatem do błędnego koła. Dlatego też Carnap proponuje prawdopodobieństwowe sformułowanie Zasady — nie jest pewne, że świat jest jednolity, ale jest to prawdopodobne w tym sensie, że jedność świata przejawia się w długim ciągu obserwacji. Dokładniej:⁷

- (C3) „Na podstawie dostępnej ewidencji jest bardzo *prawdopodobne*, że stopień jedności świata jest wysoki.”
- (C4) „Na podstawie ewidencji o tym, że relatywna liczebność własności w długim początkowym odcinku następstw jest wysoka (na przykład r), jest bardzo *prawdopodobne*, że ta liczebność tak samo będzie wysoka (w przybliżeniu równa r) w dostatecznie długiej kontynuacji tych następstw.”

Jako rozwiązanie problemów związanych z podstawami indukcji i weryfikacji Zasady Jedności proponuje Carnap w *Logical Foundations* swój system logiki indukcyjnej — teorię funkcji potwierdzania (*confirmation function*). W tym systemie każdy wniosek indukcyjny ma czysto logiczny charakter i jest ważny (lub nieważny) *a priori* — nie wymaga żadnego empirycznego potwierdzenia.

Carnapowi ze zrozumiałych powodów⁸ nie powiodło się znalezienie jedynej funkcji potwierdzania, która byłaby możliwa do zastosowania we wszystkich typach sytuacji indukcyjnych. Swoją teorię następnie rozszerza w książce *Continuum of Inductive Methods*, gdzie definiuje funkcję potwierdzania w zależności od parametru λ , który ma wartości rzeczywiste. W rezultacie nie istnieje jedna indukcyjna miara, ale cała klasa miar (dla każdej wartości λ jedna). Parametr ten określa jak szybko ewidencja uzyskana w trakcie obserwacji przeniesiona zostanie do potwierdzania hipotez. Ten parametr może być rozumiany także jako kwantytatywne określenie stopnia jedności świata albo — dokładniej — stopnia jedności świata dla danej sytuacji indukcyjnej. Główne cechy tego rozwiązania — odmiennosc stopnia jedności dla różnych sytuacji indukcyjnych i możliwość ich parametryzacji — będę starał się zachować w sformułowaniu Zasady Jedności dla prezentowanego modelu obserwatora.

Jak wyżej powiedziano, typ zadania, jakie obserwator ma rozwiązywać (z przeszłych obserwacji procesu ma wyprowadzać prawidłowości ważne w dalszym przebiegu procesu), jest indukcyjny. Należy więc przyjąć jakąś formę Zasady Jedności. Nie trzeba jej jednak zakładać w całkowicie ogólnej formie, bowiem typ opisywanych procesów jest dość specyficzny. Wystarczy założyć, że proces obserwowany przez obserwatora jest w **pewnej mierze** jednolity. Analogicznie do wprowadzonego przez Carnapa parametru λ , określającego miarę jedności świata w danej sytuacji in-

⁷ Tamże, s. 180, kursywa R.C.

⁸ Szczegółowa dyskusja o problemach logiki indukcyjnej nie mieści się w ramach tej pracy. Autor (razem z T. Childersem) zajmował się tą problematyką w wykładzie *Is Logical Theory of Probability Dead?*

dukcyjnej, proponuję wprowadzić parametry obserwacji, które będą odpowiadać mierze jedności danego konkretnego procesu. Ściślej — parametry te można interpretować jako subiektywne oszacowanie przez obserwatora miary jedności danego konkretnego procesu.

Pewna różnica, w której proponowane pojęcie jedności odmienne jest od Carnapowskiego — to czasowa ograniczoność ze względu na aktualny moment czasowy. Carnap mówi w (C2) i (C4) o „*długim początkowym odcinku następstw*” oraz o „*dotatecznie długiej kontynuacji tych następstw*”, zatem nie rozróżnia między «starą» a «nową» informacją; z punktu widzenia jedności wszelkie świadectwa z przeszłości są tak samo ważne, niezależnie od tego, jak długo przed aktualnym momentem zostały uzyskane. W wypadku tych rozważań zakładam jedność tylko w czasowo ograniczonej formie, dla przyszłości informacja uzyskana w odległej przeszłości jest mniej ważna, niż informacja aktualna. Jednocześnie nie zakładam, że uzyskane prawidłowości będą ważne w nieograniczenie długiej kontynuacji procesu, lecz jedynie w pewnej bliskiej przyszłości. Przyjęte warunki umożliwiają sformułowanie specyficznego wariantu Zasady Jedności, na którym oparty będzie proponowany model obserwatora. Będę go nazywał Zasadą Lokalnej Jedności:

„Zdarzenia, które często powtarzały się w niedawnej przeszłości będą często powtarzały się także w bliskiej przyszłości”.

„Obserwator zakłada, że obserwowany proces jest lokalnie jednolity: występowanie prawidłowości stwierdzonych lub potwierdzonych w niedawnej przeszłości jest zakładane także w bliskiej przyszłości”.

Te dwa twierdzenia są pochodne wobec sformułowań Carnapa (C2), (C4) i na pierwszy rzut oka wydają się zbyt nieokreślone i ogólne. Umyślnie nie używam tu terminów (relatywnych) „liczebność” lub „prawdopodobieństwo”. Określenia „prawidłowości stwierdzone lub potwierdzone” i „zakładane występowanie prawidłowości” w trakcie budowania modelu nabiorą jednak ścisłego znaczenia i będą zastąpione kwantytatywnymi wielkościami. Jednakże w miejsce liczebnego potwierdzenia lub doniosłości ewidencji posłużę się ogólniejszą funkcją. W odniesieniu do funkcji określającej stopień, w jakim zakładane jest przyszłe występowanie prawidłowości, nie przyjmuję *a priori*, że jest prawdopodobieństwowa; jej powiązanie z prawdopodobieństwem zostawiam do rozważenia później.

Zasada Lokalnej Jedności jest podstawowym założeniem filozoficznym, służącym do konstruowania modelu.

Wracając do obserwatora i jego zadania zakładam, że nie dysponuje on żadną aprioryczną informacją o elementarnych stanach procesu, a dokładniej — przed rozpoczęciem obserwacji nie zna lub nie bierze pod uwagę ważności jakichkolwiek relacji między stanami elementarnymi. Nie może więc badać danego procesu na podstawie jakichś zewnętrznych (z punktu widzenia obserwowanego procesu) kryteriów.

(Np. nie może testować rozkładu występowania jakiejś własności elementarnych zjawisk w procesie.) To założenie ma wyznaczyć typ zadań, którym chcę się niżej zająć.

Z poprzednich ustaleń wynika, że jedynym typem relacji, który obserwator będzie śledził w procesie, jest następstwo. Które następstwa będą dla obserwatora prawidłowe? Zakładam, że te, które się powtarzają, a założenie to jest uzasadnione przez uprzednie przyjęcie Zasady Lokalnej Jedności.

Kolejnym założeniem będzie możliwość subiektywnego (z punktu widzenia obserwatora) oszacowania jedności procesu. Obserwator nie będzie rejestrować i przetwarzać wszystkich zdarzeń sekwencyjnych zachodzących w procesie. Jeżeli będzie przestrzegać założenia o lokalnej jedności, to w każdym aktualnym momencie nie wszystkie zdarzenia będą tak samo istotne jak w chwili, kiedy były obserwowane i rejestrowane; niektóre z nich są ważne (aktualne), niektóre mniej, a jeszcze inne już w ogóle nie mogą być brane pod uwagę. Obserwator musi zatem dokonywać selekcji informacji, którą będzie oceniać i przechowywać. Tę selekcję obserwator przeprowadza korzystając z parametrów obserwacji — subiektywne oszacowanie jedności procesu rozumiane jest jako wybór wartości tych parametrów.

Założenie to można by potraktować czysto technicznie i pojmować wybór parametrów jako rezultat ograniczonych możliwości obserwatora (np. ograniczona pojemność pamięci) — bądź jako ograniczenie spersonalizowanego obserwatora, bądź też jako techniczne ograniczenie jakiegoś przyrządu. Takie podejście jednakże byłoby bezzasadne i zbyt mocno zawężające — jeśli dopuszcza się możliwość, aby obserwator na podstawie własnego oszacowania miary jedności procesu mógł sam wprowadzać limit uzależniony od wielkości przetworzonej oraz przechowanej informacji, mimo że jego fizyczna pojemność pozwalałaby na więcej.

Powyższe stwierdzenie może się jawić jako zaprzeczenie założenia o braku apriorycznej informacji: skoro obserwator nie ma przed rozpoczęciem obserwacji żadnej informacji o procesie, to jak może szacować miarę jedności procesu? W założeniu tym nie mówi się jednak o procesie jako całości, lecz o relacjach między elementarnymi zjawiskami. Można wyobrazić sobie sytuację, w której nie istnieje żadna informacja o elementarnych zjawiskach, lecz jest do dyspozycji jakaś informacja umożliwiająca obserwatorowi oszacowanie jedności. Dalszą możliwością byłoby zastanowienie się nad zmianą parametrów w przebiegu obserwacji. Możliwość ta jednak zalicza się już do zakresu realizacji modelu i z teoretycznego punktu widzenia nie wnosi żadnej zasadniczej zmiany, dlatego też nie jest tu przedmiotem zainteresowania.

Przyjmuję także, że selekcja informacji jest czasowo uwarunkowana w tym znaczeniu, że bardziej aktualna (w danym momencie) informacja ma większą wagę niż informacja mniej aktualna. Czasowe uwarunkowanie zapewnia większą zdolność do przystosowania się obserwatora w procesach dynamicznych, tzn. w procesach, których prawidłowość zmienia się w czasie. Czasowo ograniczona selekcja informacji przejawia się w obserwacji procesu i w przechowywaniu wydobytych prawidłowości.

Wprowadzonym wyżej założeniom chciałbym teraz nadać konkretniejszą postać.

Proces jest dla obserwatora ciągiem elementarnych stanów. W danym momencie obserwacji procesu obserwator nie jest w stanie obejmować całego szeregu elementarnych stanów, które zaszły w procesie od rozpoczęcia obserwacji. Obserwatora interesować będzie jedynie aktualna część procesu — elementarny stan w danym momencie i elementarne stany bezpośrednio poprzedzające. Taki zwarty odcinek aktualnych stanów elementarnych będzie nazywać *aktualnym odcinkiem procesu* (w danym momencie czasowym).

Informację o aktualnym odcinku procesu obserwator musi w jakiś sposób przechowywać (ponieważ bezpośrednio obserwowany może być tylko jeden stan elementarny w danym momencie obserwacji). Miejsce, gdzie ta informacja jest przechowywana, będzie nazywać *pamięcią krótkookresową* lub *lokalną*.⁹ Jest widoczne, że w każdym momencie aktualnym odcinkiem jest inna część procesu, w związku z tym zmienia się i zawartość pamięci krótkookresowej — obejmuje ona aktualny stan i pewną liczbę poprzedzających stanów elementarnych. Długość aktualnego odcinka (liczba elementarnych stanów), który pamięć krótkookresowa może ogarnąć, nazywać będą *długością pamięci krótkookresowej*. Długość pamięci krótkookresowej będzie pierwszą charakterystyką obserwatora.

Jak powiedziano, celem obserwatora jest wyszukiwanie zdarzeń sekwencyjnych, które są prawidłowe. Będzie on je wybierać z aktualnego odcinka, który jest jego jedyną aktualną informacją o procesie. Kryterium prawidłowości będzie liczebność w aktualnym odcinku. Zatem w danym momencie będą wybrane, jako prawidłowe, te zdarzenia sekwencyjne, których liczebność w aktualnym odcinku jest dostatecznie wysoka. Minimalna liczebność zdarzeń w aktualnym odcinku procesu (liczebność dostatecznie wielka, aby zdarzenie sekwencyjne zostało uznane za prawidłowe) jest kolejną charakterystyką obserwatora.

Prawidłowe zdarzenia sekwencyjne obserwator przechowuje w *pamięci długookresowej*. Ze względu na przyjęte założenie o selekcji informacji (danej w subiektywnym oszacowaniu jedności procesu) jest oczywiste, że obserwator nie będzie trwale przechowywać wszystkich wybranych prawidłowości. Musi rozstrzygać o tym, które ze zdarzeń, wybranych jako prawidłowe, będzie przechowywać. Jest widoczne, że tych rozstrzygnięć może dokonywać jedynie na podstawie obserwacji dalszych pojawiających się wybranych prawidłowości w procesie. Z założenia o lokalnej jedności wynika, że obserwacyjnymi kryteriami będą liczebność oraz bliskość w czasie. Można powiedzieć, że dla obserwatora ważniejsze są liczniejsze zdarzenia niż te mniej liczne i zdarzenia niedawne (ze względu na aktualny moment) w stosunku do zdarzeń odległych w czasie.

Oba te kryteria zawiera charakterystyka, którą będą nazywać *aktualnością zdarzeń*. Aktualność jest wielkością zależną od czasu, którą obserwator śledzi dla każdego zdarzenia wybranego w procesie obserwacji. Wartość aktualności zwiększa się dla

⁹ Terminem tym będę posługiwać się nie wglębiając się w to, w jaki konkretny sposób obserwator tę informację przechowuje.

danego zdarzenia, o ile to zdarzenie wystąpi w danym momencie w procesie (w danym czasie jest zdarzeniem aktualnym). W przeciwnym wypadku w danym momencie wartość aktualności się zmniejsza. Stałe dla zwiększania ewentualnie zmniejszania aktualności są ostatnią charakterystyką obserwatora.¹⁰ Będę je nazywał *stałą aktualizacji resp. stałą zapominania*. Wartość aktualności jest kryterium, według którego obserwator rozstrzyga, czy będzie dane zdarzenie przechowywać, czy też nie. Jeśli wartość aktualności opadnie poniżej pewnej zadanej granicy, to dane zdarzenie zostaje wyeliminowane z długookresowej pamięci («zapominane»). Oczywiście w dalszym przebiegu obserwacji zdarzenie to może być ponownie wybrane (w standardowy sposób) jako prawdziwe.

Ze względu na przyjęte rozumienie aktualności dla obu parametrów sens mają tylko nieujemne wartości. Jako skrajny przypadek (zerowa wartość zapominania) może wystąpić sytuacja, w której obserwator przechowa wszystkie wybrane zdarzenia, lub że go z dwóch «składowych» aktualności (liczebność i bliskość w czasie) zainteresuje jedynie pierwsza z nich.

Niektóre konfiguracje parametrów w sposób widoczny nie rokują dobrze — np. jeśli stała zapominania jest większa niż stała aktualizacji (zdarzenie jest zapominane natychmiast po jego wyborze). Bez uszczerbku dla ogólności można dalej założyć, że stałe te są liczbami całkowitymi (decydujący jest ich wzajemny stosunek, nie zaś ich absolutna wielkość) i że graniczna wartość dla wyeliminowania zdarzenia jest równa 1.

Powyższe rozważania można by nazwać nieformalną częścią budowy modelu obserwatora. Teraz przejdę do jego formalnego opisu.¹¹

Jak już stwierdzono, zjawiska elementarne są reprezentowane przez skończony zbiór znaków. Nad tym zbiorem nie jest definiowana żadna struktura. Obserwację procesu przedstawia ciąg, którego kolejne wyrazy są elementami tego skończonego zbioru. Indeksy tych wyrazów odpowiadają momentom.

Oznaczenie

Zbiór wszystkich łańcuchów nad zbiorem znaków

Niech A jest zbiorem skończonym; zbiór wszystkich łańcuchów znaków z A można oznaczyć jako zbiór

$$A^* =_{\text{def}} \{a_1 \dots a_k; a_i \in A, k \in \mathbb{N}\}, \text{ gdzie } \mathbb{N} \text{ jest zbiorem liczb naturalnych.}$$

¹⁰ Z matematycznego punktu widzenia tak definiowana jest aktualność dla danego zdarzenia po częściach linearyjnej funkcji. Można by oczywiście zdefiniować ją nieliniarnie, jednakże w tej pracy interesuje mnie przede wszystkim zasada aktualizacji, nie zaś pełna dyskusja o możliwościach jej realizacji.

¹¹ Ten model powstał na Wydziale Matematyczno-Fizycznym Uniwersytetu Karola w Pradze w latach 1985—86 na seminarium doc. Hedrlína, którego autor artykułu był uczestnikiem. Sformalizowany opis po raz pierwszy został opublikowany w pracy dyplomowej autora, w 1986 roku.

Długość łańcucha

Przez długość łańcucha rozumiana jest liczba znaków, które ten łańcuch tworzą: dla $a \in A^*$, $a = a_1 \dots a_k$, $a_i \in A$, jest

$$|a| =_{\text{def}} k.$$

Relacja podłańcuch

Dla dowolnych dwóch łańcuchów $a, b \in A^*$, $a = a_1 \dots a_k$, $b = b_1 \dots b_m$ definiujemy:

a jest *podłańcuchem* b ($a \ll b$) wtedy i tylko wtedy, gdy $a_1 = b_1, \dots, a_k = b_{1+k-1}$, dla jakiegoś i , $1 \leq i \leq m - k + 1$

a jest *lewym podłańcuchem* b ($a \ll_L b$) wtedy i tylko wtedy, kiedy $a_1 = b_1, \dots, a_k = b_k$, $k \leq m$

a jest *prawym podłańcuchem* b ($a \ll_R b$) wtedy i tylko wtedy, kiedy $a_1 = b_{m-k+1}, \dots, a_k = b_m$, $k \leq m$.

Liczebność łańcuchów w innym łańcuchu

Dla dowolnych dwóch łańcuchów $a, b \in A^*$, $a = a_1 \dots a_k$, $b = b_1 \dots b_m$, oznaczamy $f(a, b)$ liczbę wystąpień łańcucha a w łańcuchu b :

$$f(a, b) = |\{i: 1 \leq i \leq m - k + 1 \text{ a } a_1 = b_i, a_2 = b_{i+1}, \dots, a_k = b_{i+k-1}\}|$$

Definicja 1.1 (Proces nad zbiorem zjawisk elementarnych)

Procesem nad zbiorem skończonym A nazwiemy ciąg

$$S_A = \{s_i \in A; i \in \mathbb{N}\}, \text{ gdzie } \mathbb{N} \text{ jest zbiorem liczb naturalnych;}$$

A interpretujemy jako zbiór zjawisk elementarnych, a \mathbb{N} jako zbiór momentów.

Jak już powiedziano, wielkość aktualnej części procesu, która może być przetworzona w danym czasie, jest ograniczona. W związku z tym wprowadzone zostały pojęcia: odcinek aktualny i pamięć lokalna. Oba pojęcia różnią się jedynie w tym, że odcinek aktualny jest jakąś częścią procesu bezpośrednio poprzedzającą czasowy moment, natomiast pamięć lokalna jest odcinkiem aktualnym, który obserwator ma (w danym czasowym momencie) do dyspozycji. Formalnie oba pojęcia mogą być przedstawione tak samo.

Definicja 1.2 (Pamięć lokalna i odcinek aktualny)

Aktualnym odcinkiem długości K resp. *lokalną pamięcią* długości K w procesie S_A w czasie t nazwiemy łańcuch

$$u_{K,t} = s_{t-K+1} \dots s_t.$$

Pojęcia „odcinek aktualny” można używać i bez podania długości jako powołanie się na część procesu, bezpośrednio poprzedzającą czasowy moment.

Selekcja prawidłowości w odcinku aktualnym będzie reprezentowana **funkcją wybierania**. Jej podstawowym parametrem jest minimalna liczebność $f_{\min} \geq 1$. Podłańcuch odcinka aktualnego, który pojawia się w pamięci lokalnej (odpowiedniej dłu-

gości) przynajmniej tyle razy, ile wynosi wartość parametru f_{\min} , jest przez funkcję wybierania włączany do pamięci długookresowej. Widoczne jest, że wystarczy zajmować się w danym czasie tylko tymi łańcuchami, które są «najnowsze», tj. są aktualnym odcinkiem (w czasie t są one postaci $s_{t-1} \dots s_t$, $0 \leq t \leq K-1$).

Definicja 1.3 (Funkcja wybierania)

Funkcją wybierania z minimalną liczebnością f_{\min} nazwiemy funkcję

$$F : A^* \rightarrow \sigma(A^*),$$

gdzie $a \in F(u)$ wtedy i tylko wtedy, kiedy a jest prawym podłańcuchem u i $f(a, u) \geq f_{\min}$, i $\sigma(A^*)$ jest podzbiorem zbioru A^* .

Warto zauważyć, że dla $f_{\min} = 1$ wybierze $F(u)$ wszystkie prawe podłańcuchy u .

Wyselekcjonowane prawidłowości przechowywane są w pamięci długookresowej. Formalnie będzie ona definiowana przez funkcję wybierania w przebiegu obserwacji w danym czasowym momencie jako zbiór łańcuchów wybranych. (Na początku obserwacji zbiór ten jest oczywiście pusty.)

Definicja 1.4 (Pamięć długookresowa)

Pamięcią długookresową w czasie t przy obserwacji procesu S_A z funkcją wybierania F i lokalną pamięcią długości K nazywamy zbiór $M_t \subseteq A^$ definiowany indukcyjnie:*

$$M_t = \emptyset \text{ dla } t < K$$

$$M_t = M_{t-1} \cup F(s_{K-t+1, t}).$$

Aktualność jako czasowo uwarunkowana ważność łańcuchów w pamięci będzie reprezentowana indukcyjnie definiowaną funkcją przyporządkowującą każdemu łańcuchowi w pamięci wartości w postaci liczb całkowitych. Wartość tej funkcji dla danego łańcucha będzie wzrastać w każdym momencie, kiedy ten łańcuch pojawi się w procesie (tj. pojawi się jako prawy podłańcuch lokalnej pamięci).

Definicja 1.5 (Aktualność łańcucha w pamięci długookresowej)

Aktualnością dla obserwacji procesu S_A z pamięcią długookresową M_t i z nieujemnymi parametrami w postaci liczb całkowitych d^+ resp. d^- dla aktualizacji, resp. zapominania nazywam funkcję

$$a_t : M_t \rightarrow \mathcal{V}, \text{ gdzie } \mathcal{V} \text{ jest zbiorem liczb rzeczywistych,}$$

definiowaną

$$a_0(r) = 0 \text{ dla wszystkich } r \in A^* ;$$

dla $t > 0$

$$a_t(r) = 0 \text{ dla } r \notin M_t$$

$$a_t(r) = a_{t-1}(r) + d^+ \text{ o ile } r \in M_t \text{ i } r \text{ jest aktualnym odcinkiem w czasie } t$$

$a_t(\mathbf{r}) = a_{t-1}(\mathbf{r}) - d^-$ w przeciwnym razie.

Zapominanie dla pamięci M_t będzie definiowane jako wyeliminowanie łańcuchów, których aktualność opada poniżej zadanej granicy (bez szkody dla ogólności możemy za tę granicę przyjąć aktualność zerową).

Definicja 1.6 (Pamięć długookresowa z zapominaniem)

Pamięcią długookresową z zapominaniem w czasie t obserwacji procesu S_A z funkcją wybierania F , długością lokalnej pamięci K i aktualnością a nazywamy zbiór $M_t \subseteq A^*$ definiowany:

$$M_t = \emptyset \text{ dla } t < K$$

$$M_t = (M_{t-1} \setminus M_{t-1}') \cup F(s_{K-t+1,t}), \text{ gdzie } M_{t-1}' = \{r \in M_{t-1}; a_{t-1}(r) \leq 0\}.$$

W dalszej części artykułu przez pamięć długookresową będą rozumieć pamięć długookresową z zapominaniem.

W ten sposób kończy się formalna część opisu modelu obserwatora. Teraz można uściślić określenie pojęcia obserwatora z wstępu artykułu: przez obserwatora rozumieć ewidencję i przetwarzanie czasowego następstwa zjawisk elementarnych według właśnie opisanych algorytmów.

Obserwatora resp. obserwację charakteryzują cztery parametry: długość pamięci lokalnej K , minimalna liczebność dla wyboru f_{\min} , stała dla aktualizacji d^+ oraz stała dla zapominania d^- . Tę czwórkę parametrów będą nazywać *konfiguracją obserwatora* (resp. *obserwacją*).

Niewątpliwie przy niektórych wartościach d^+ i d^- wybrane łańcuchy są eliminowane z pamięci natychmiast w następnym kroku. Aby łańcuch r w pamięci «prze-trwał» jest oczywiście konieczne, by jego długość była większa niż iloraz d^+/d^- . W przeciwnym przypadku, o ile $|r| \geq d^+/d^-$, to łańcuch r zostanie «zapomniany», mimo że będzie pojawiać się w procesie tak często, jak mu na to jego długość będzie dozwalać, tj. także i w procesie złożonym jedynie z łańcuchów r . Zatem sensowne jest zajmowanie się konfiguracjami z parametrem $K \geq 1$, a dla konfiguracji z $d^- = 0$ można bez uszczerbku dla ogólności przyjąć $d^+ = 1$ resp. dla konfiguracji z $d^- > 0$ można puszczać $d^- = 1$.

Podstawowa konfiguracja obserwatora

Na pewno nie wszystkie konfiguracje parametrów obserwacji są sensowne. W dalszym ciągu będą zajmować się więc tylko niektórymi klasami konfiguracji, które z punktu widzenia celów tego opracowania są znaczące.

$(K, 1, d^+, 0)$ — obserwator wybiera jako prawidłowości wszystkie łańcuchy długości mniejszej lub równej K , które pojawiły się w procesie, ich aktualności w takim razie odpowiadają d^+ — wielokrotności ich liczebności w dotychczasowym przebiegu procesu. Jeśli wykluczyć jako nieinteresujący przypadek $d^+ = 0$ (aktualność nie niesie

żadnej informacji), to można bez szkody dla ogólności przyjąć $d^+=1$, a więc przyjąć aktualność równą liczebności. Ten typ obserwacji odpowiada pojęciu Zasady Jedności określonej w (C4) — zakładam jedność w długim przebiegu (*long run*); interesujące zatem są jedynie liczebności, nie zaś czasowa bliskość. Przy tej konfiguracji nie zachodzi żaden wybór informacji, (z wyjątkiem ograniczenia długości pamięci lokalnej). Aktualność łańcuchów (dla łańcuchów odpowiedniej długości) bezpośrednio odpowiada tu liczebności łańcuchów w ciągu. Ponieważ moim zamiarem jest, aby model był dostatecznie ogólny i uwzględniał również klasyczne pojęcie uczenia się oparte na liczebnościach, dlatego tę konfigurację można traktować jako przypadek skrajny. Obserwator działający z taką konfiguracją przeprowadza w pewnym sensie «obiektywną» obserwację — nie jest tu dołączony jakikolwiek (przy danej wielkości pamięci lokalnej) subiektywny czynnik wyboru.

$(K, f_{\min}, 1, 0)$ — obserwator wybiera łańcuchy długości najwyżej K , które w procesie pojawiały się dostatecznie «gęsto», ich aktualność odpowiada liczebności od momentu włączenia do pamięci długookresowej. Także ta konfiguracja nie uwzględnia czasowej bliskości — pracuje jedynie z liczebnościami — jednak w odróżnieniu od poprzedniej tu wybór informacji dokonywany jest w pamięci lokalnej; aktualność zatem nie odpowiada liczebności w całym ciągu, lecz w odcinku ciągu od momentu wybrania łańcucha.

(K, f_{\min}, d^+, d^-) — ogólna konfiguracja. Powyższe przypadki zawierają się w niej oczywiście przy odpowiednich wartościach parametrów.

Przejściu od obiektywnej obserwacji bez wyboru informacji do obserwacji z subiektywnym wyborem odpowiada przejście od obiektywnej ewidencji o zdarzeniach — liczebności — do aktualności jako subiektywnej ewidencji. Można pokazać,¹² że przy tym przejściu zostaje zachowana jedna istotna własność liczebności, umożliwiającą skonstruowanie na podstawie liczebności miary prawdopodobieństwa: subaddytywność ze względu na relację lewy podłańcuch.

Kolejną kwestią jest zadanie predykcji, dokładniejsza ocena hipotez na podstawie dostępnej ewidencji. To zadanie można pojmować algebraicznie jako zdefiniowanie algebry zdarzeń i miary prawdopodobieństwa nad tą algebrą, jednakże w tym artykule nie będę się tym zajmować, a zainteresowanych odsyłam do innej mojej pracy.¹³

2. JĘZYK A TEORIA ZDARZEŃ SEKWENCYJNYCH

Wpierw chciałbym sformalizować intuicję dotyczącą zdarzeń sekwencyjnych środkami logiki predykatów. Zostaną tu zdefiniowane: język zdarzeń sekwencyjnych i aksjomaty teorii zdarzeń sekwencyjnych. Zamierzam udowodnić, że teoria określona przez te aksjomaty jest niesprzeczna. Następnie zdefiniowana zostanie dla pewnej części zdań języka prawdopodobieństwowa (dwuargumentowa) funkcja oceniania,

¹² Patrz O. Majer, *Problematika sekvencyjnych udalostí*, Praha 1998, Rozdział 2.

¹³ Patrz *ibidem*.

która dla pary zdań języka daje warunkowe prawdopodobieństwo, z jakim pojawienie się zdarzenia określonego pierwszym zdaniem języka warunkuje następne pojawienie się zdarzenia opisanego drugim zdaniem.

Na koniec pozostawiam do dyskusji pytanie, jakiego typu jest wykryte prawdopodobieństwo. Fakt, że definiuję funkcję prawdopodobieństwową nad zdaniami języka nie jest oczywiście wystarczający, aby wykryte prawdopodobieństwo było prawdopodobieństwem logicznym. Pokażę też, że pojęcie tak skonstruowanej funkcji oceny podpada pod subiektywne pojęcie prawdopodobieństwa.

Język zdarzeń sekwencyjnych L_{seq} będzie dwuargumentowym językiem logiki predykatów pierwszego rzędu. Zawierać będzie dwa rodzaje indywiduów — zdarzenia i momenty. Język będzie zawierać także trójargumentowy predykat O , reprezentujący pojawianie się zdarzenia w czasowym interwale oraz dwuargumentowy predykat \leq dla uporządkowania momentów czasowych. L_{seq} zawiera operator $*$ dla operacji (czasowego) łańcuchowania zdarzeń, zmienne dla zdarzeń i momenty, a także zwyczajne spójniki logiczne i kwantyfikatory.

Definicja 4.1 (Symbole języka zdarzeń sekwencyjnych)

Język L_{seq} zawiera następujące symbole:

x, x', x_1, \dots zmienne dla zdarzeń,

t, t', t_1, \dots zmienne dla momentów,

λ stała dla neutralnego zdarzenia,

O losowy predykatowy symbol pojawiania się zdarzeń między dwoma momentami,

\leq predykatowy symbol dla uporządkowania momentów,

$*$ operator dla łańcuchowania zdarzeń,

spójniki logiczne $\neg, \&, \vee$, kwantyfikatory \forall, \exists , i nawiasy $(,)$.

Definicja 4.2 (Formacyjne reguły języka zdarzeń sekwencyjnych)

a) e-termny

- (i) każda zmienna dla zdarzenia jest e-termem, stała λ jest e-termem;
- (ii) jeśli S, S' są dowolnymi e-termami, to także $S*S'$ jest e-termem;
- (iii) nic innego oprócz wyrażenia utworzonego według punktów (i) oraz (ii) nie jest e-termem;

b) formuły atomowe

- (i) wyrażenie $O(_, _)$ ze zmiennymi dla momentów na pierwszym i drugim miejscu i z e-termem na trzecim miejscu jest formułą atomową;
- (ii) wyrażenie $\leq(_, _)$ ze zmiennymi dla momentów na pierwszym i drugim miejscu jest formułą atomową;
- (iii) nic innego oprócz wyrażenia utworzonego według punktów (i) oraz (ii) nie jest formułą atomową;

c) poprawnie utworzone formuły

- (i) każda formuła atomowa jest poprawnie utworzoną formułą;

- (ii) jeśli A , B są poprawnie utworzonymi formułami, to także $\neg A$, $(A \vee B)$, $(A \& B)$ są poprawnie utworzonymi formułami;
- (iii) jeśli $A(x)$, $B(t)$ są poprawnie utworzonymi formułami z wolnymi zmiennymi x , *resp.* t , to także „ $x A(x)$ ”, $\exists x A(x)$, $\forall t B(t)$, $\exists t B(t)$ są poprawnie utworzonymi formułami;
- (iv) nic innego oprócz wyrażenia utworzonego według punktów (i), (ii) oraz (iii) nie jest poprawnie utworzoną formułą;

Uwaga: w ramach języka L_{seq} będę używał spójników \Rightarrow i \Leftrightarrow , zdefiniowanych w ogólnie przyjęty sposób.

Definicja 4.3 (Predykaty dedukowane)

Będę stosował następujące dedukowane predykaty (S , S' oznaczają dowolne e-term):

$$A(t, S) \text{ wtedy i tylko wtedy, gdy } \exists t' t \leq t' \& O(t, t', S)$$

(pojawienie się zdarzenia S zaczynającego się bezpośrednio po czasowym momencie t);

$$B(t, S) \text{ wtedy i tylko wtedy, gdy } \exists t'' t'' \leq t \& O(t'', t, S)$$

(pojawienie się zdarzenia S kończącego się dokładnie w czasie t);

$$\ll_L(S, S') \text{ wtedy i tylko wtedy, gdy } \forall t A(t, S') \Rightarrow A(t, S)$$

(predykat lewe podzdarzenie — pojawieniu się zdarzenia S' za każdym razem towarzyszy pojawienie się zdarzenia S);

$$x \equiv x' \text{ wtedy i tylko wtedy, gdy } \forall t, t' O(t, t', x) \Leftrightarrow O(t, t', x')$$

(tożsamość zdarzenia definiowana jako tożsamość pojawienia się w tych samych interwałach czasowych);

$$S \sim S' \text{ wtedy i tylko wtedy, gdy } \neg(S \ll_L S') \& \neg(S' \ll_L S)$$

(predykat dla sprzeczności zdarzeń);

$$t \equiv t' \text{ wtedy i tylko wtedy, gdy } (t \leq t' \& t' \leq t)$$

(tożsamość czasowych momentów).

Znaczenie

Czas w predykatkach A , B będzie zapisywany jako indeks dolny tj. $A_t(s)$, $B_t(s)$ zamiast $A(t, s)$, $B(t, s)$. W uporządkowaniu posłużyłem się notacją niestałą tj. $s \ll_L s'$ zamiast $\ll_L(s, s')$. Spójnik \Rightarrow definiowany jest w zwyczajny sposób.

Nad językiem zdarzeń sekwencyjnych można teraz zdefiniować aksjomaty teorii zdarzeń sekwencyjnych.

Aksjomaty teorii zdarzeń sekwencyjnych

$$(E) \quad \forall t_1, t_2 \exists x O(t_1, t_2, x) \quad \text{(istnienie)}$$

W każdym interwale czasowym zachodzi jakieś zdarzenie. Semantycznie — dla każdego interwału czasowego istnieje obiekt, który jest zdarzeniem w tym interwale. Ten postulat formalnie może wydawać się zbyt silny, w przyjętej interpretacji odpowiada wyobrażeniu, że obserwator w każdym momencie coś obserwuje. To, co obserwuje nie musi jednak być zdarzeniem, które sobie zapamiętuje i które przy jego dalszym pojawianiu się będzie zdolny zidentyfikować. Te zdarzenia «bez identyfikacji» będą reprezentowane neutralnym symbolem λ , który pozwala rozumieć pewne odcinki (zdarzeń) jako szумы.

(λ) $\forall x (x \neq \lambda \Rightarrow (\lambda \sim x))$ (zdarzenie neutralne)

Stając dla zdarzenia neutralnego wprowadzam jako reprezentację zdarzenia «niekonkretnego» (szumowego). Zdarzenia odmienne od neutralnego będą oznaczać jako standardowe. Zdarzenie neutralne pozwala reprezentować zdarzenia, które są z jakiegoś punktu widzenia bez znaczenia, nieinteresujące, nieinformatywne i nie trzeba ich zatem wyróżniać jako pewnej klasy. Aksjomat (λ) mówi, że zdarzenie neutralne jest sprzeczne ze standardowym (a więc znanym, konkretnym) zdarzeniem. λ -zdarzenie będzie mieć znaczenie przy ocenie zdarzenia, które będzie konstruowane w dalszej części: umożliwi przypisanie szumowym częściom wartości 0, bowiem z punktu widzenia prawdopodobieństwa opartego na klasycznych liczebnościach ich wartość powinna być niezerowa.

(O) $\forall t_1, t_2, x (O(t_1, t_2, x) \& (x \neq \lambda) \Rightarrow \neg(\exists t_1', t_2' (t_1 \leq t_1') \& (t_2' \leq t_2) \& \neg(t_1 = t_1') \vee \neg(t_2' = t_2)) \& O(t_1', t_2', x))$ (pojawianie się)

Pojawianie się zdarzenia w jakimś interwale oznacza pojawienie się dokładnie w tym interwale, nie zaś «gdzieś» w tym interwale. Dokładniej — o ile zdarzenie pojawia się w danym interwale, to nie pojawia się w żadnym własnym podinterwale tego interwału. Z reguły tej ze zrozumiałych powodów wyłączone są zdarzenia neutralne. Wywnioskowane predykaty **A**, **B** określają więc pojawienie się bezpośrednio po, ewentualnie przed danym momentem czasowym. Dane pojęcie różni się od rozumienia analogicznych predykatów np. w logikach czasu (*tense logics*), gdzie „przed” i „po” oznacza niekiedy kiedykolwiek przed ewentualnie po. Należy zauważyć, że właśnie to uściślenie pojęcia pojawiania się jest powodem, dla którego został wprowadzony jako podstawowy predykat pojawiania się w interwale, co wcale nie jest jednakowe z predykatem pojawiania się przed momentem (*resp.* po momencie).

(D) $\forall t_1, t_2, x \neg(x \equiv \lambda) \& O(t_1, t_2, x) \Rightarrow \neg(\exists t_1', t_2' ((t_1 \leq t_1') \& (t_2' \leq t_2) \& O(t_1', t_2', \lambda)))$ (dziedziczność)

Dla każdego czasowego interwału spełnione jest: jeżeli pojawi się w nim zdarzenie standardowe, to w żadnym jego podinterwale nie może pojawić się zdarzenie neutralne. O ile by w odpowiedni sposób rozszerzyć pojęcie podzdarzenia, to aksjomat ten stwierdza, że wszystkie (nie zaś tylko lewe) podzdarzenia zdarzenia standardowego są standardowe. Dla lewego podzdarzenia (D) wynika z aksjomatu (λ).

$$(N) \quad \forall x \exists x_1, \dots, x_k ((x \sim x_i \text{ dla } i=1..k) \& (x_i \neq x_j \Rightarrow x_i \sim x_j \text{ dla } i, j=1..k) \& (\forall x' x' \sim x \Rightarrow \Rightarrow \exists i ((x_i \ll_L x') \vee (x' \ll_L x_i))) \quad (\text{negacja})$$

Dla każdego zdarzenia x istnieje skończony zbiór zdarzeń nie dających się z nim połączyć taki, że zdarzenia z tego zbioru są wzajemnie sprzeczne i zbiór ten jest maksymalny w tym sensie, że każde inne zdarzenie nie dające się połączyć z x jest podzdarzeniem lub nadzdarzeniem jakiegoś zdarzenia ze zbioru. Ten aksjomat pozwala równoważnie określić negację zdarzeń jako dysjunkcję nie dających się połączyć zdarzeń. Aksjomat ponadto stwierdza, że każdy zbiór nawzajem nie dających się połączyć zdarzeń jest skończony. (Warunku powyższego nie można oczywiście wyrazić bezpośrednio w języku pierwszego rzędu.) Mówi także o typie uporządkowania \ll_L — każde zdarzenie ma tylko skończoną liczbę nawzajem nie dających się połączyć nadzdarzeń, albo jeszcze inaczej — każde zdarzenie może mieć kontynuację na wiele nawzajem nie dających się pogodzić sposobów. Ten aksjomat można odczytać tak, że w każdym momencie mamy do dyspozycji pełne wyliczenie wszystkich zdarzeń, które mogą wystąpić. Warunek ten jest znowu osłabiony istnieniem λ -zdarzenia.

$$(CH) \quad \forall x, y (x * y \sim \lambda) \Rightarrow (x \ll_L x * y) \quad (\text{ułańcuchowanie})$$

Aksjomat ten ujmuje współzależność operatora ułańcuchowania i relację podzdarzenia. Jeżeli ułańcuchowimy zdarzenie x ze zdarzeniem y i wynikiem tego ułańcuchowania nie jest zdarzenie dające się połączyć ze zdarzeniem nieznanym, to x jest lewym podzdarzeniem tego ułańcuchowania.

$$(T) \quad (a) \quad \forall t \leq t \quad (\text{uporządkowanie momentów})$$

$$(b) \quad \forall t, t', t'' (t \leq t' \& t' \leq t'') \Rightarrow t \leq t''$$

$$(c) \quad \forall t, t' (t \leq t' \vee t' \leq t)$$

Relacja \leq jest dobrym uporządkowaniem momentów.

Definicja 4.4 (Teoria zdarzeń sekwencyjnych)

Teorię nad językiem zdarzeń sekwencyjnych charakteryzowaną aksjomatami $E, O, D, N, CH, \lambda, S, T$ będę nazywał teorią zdarzeń sekwencyjnych.

Następne twierdzenie głosi, że teoria ta ma model i jest niesprzeczna.

Twierdzenie 4.5 (Niesprzeczność teorii zdarzeń sekwencyjnych)

Teoria zdarzeń sekwencyjnych ma model.

Dowód:

Weźmy zbiór skończony A , jego elementy nazwijmy znakami. Oznaczmy A^* zbiór wszystkich skończonych łańcuchów nad A . Nad zbiorem A^* definiujemy w ten sam sposób jak w pierwszej części artykułu relację lewy podłańcuch. Niech M jest podzbiorem skończonym A^* , domkniętym ze względu na relację brania podłańcucha, tj. w każdym łańcuchu zawiera wszystkie jego lewe podłańcuchy. Zbiór M będzie reprezentować zdarzenie standardowe, łańcuchom poza M będą odpowiadać neutralne

(nieznane) zdarzenia. Zdarzenia będą w modelu teorii reprezentowane klasami równoważności nad zbiorem A^* według należenia do zbioru M :

$$|a| = \{a\} \text{ dla } a \in M$$

$$|a'| = A^* - M \text{ dla } a' \in A^* - M.$$

Zgodnie z obiegową konwencją nie będę dalej rozróżniać między $|a|$ jako oznaczeniem klasy równoważności a oznaczeniem reprezentanta klasy — a więc (dowolnego) łańcucha zawartego w odpowiedniej klasie. N niech oznacza zbiór liczb naturalnych. Weźmy dalej ciąg $\{s_i\}$, $s_i \in A$, $i=0, 1, \dots$. Modelem dla teorii zdarzeń sekwencyjnych będzie $M=(A, M, \{s_i\})$.

Waluacja zmiennych

Waluacja zmiennych v będzie wyrażać w liczbach naturalnych zmienne dla zdarzeń w klasach równoważności i zmienne dla momentów.

$$[x]_v = \{a\} = |a|, \text{ dla jakiegoś } a \in A^*$$

$$[t]_v \in N$$

Przy końcu dowodu zakładam, że była obrana jakaś waluacja v , przy tym będę opuszczać stosowny indeks.

Interpretacja stałej λ

$$[\lambda] = A^* - M = |a'| \text{ dla jakiegoś } a' \in A^* - M$$

λ jest naturalnie interpretowana jako klasa równoważności wszystkich łańcuchów poza zbiorem M reprezentującym zdarzenia standardowe.

Interpretacja operatora ułańcuchowania

$$\text{Dla } [x] = |a| \text{ i } [y] = |b| \text{ jest } [x*y] = |a*b|.$$

Wynikiem ułańcuchowania zdarzeń jest klasa reprezentowana ułańcuchowaniem reprezentantów pierwotnych klas.

Interpretacja formuł atomowych

$M \cdot t_1 \leq t_2$ wtedy i tylko wtedy, gdy $[t_1] \leq [t_2]$, gdzie \leq jest zwykłym nieostrym uporządkowaniem liczb naturalnych.

$M \cdot O(t_1, t_2, x)$ wtedy i tylko wtedy, gdy $[t_1] \leq [t_2]$ a $|s_{[t_1]} * s_{[t_1]+1} * \dots * s_{[t_2]}| = [x]$, jeśli $[x] \in M$ (x jest przy danej waluacji interpretowane jako zdarzenie znane), to formuła $O(t_1, t_2, x)$ jest spełniona, jeżeli tworzy łańcuch $[x]$ odcinek $s_{[t_1]} \dots s_{[t_2]}$ ciągu s , jeżeli nie, to $[x] = A^* - M$ i formuła $O(t_1, t_2, x)$ jest spełniona, jeżeli w odcinku $s_{[t_1]} \dots s_{[t_2]}$ ciągu s nie pojawi się żaden z łańcuchów z M , zatem $s_{[t_1]} * s_{[t_1]+1} * \dots * s_{[t_2]} \notin M$, zatem $|s_{[t_1]} * s_{[t_1]+1} * \dots * s_{[t_2]}| = A^* - M = [x]$.

Interpretacja formuł złożonych

$M \cdot \neg C$ wtedy i tylko wtedy, gdy nie jest spełnione $M \cdot C$

$M \cdot C \& D$ wtedy i tylko wtedy, gdy $M \cdot C$ i równocześnie $M \cdot D$

$M \cdot \forall x C(x)$ wtedy i tylko wtedy, gdy $M \cdot C(x)$

$M \cdot \forall C(x)$ wtedy i tylko wtedy, gdy $M \cdot \vee C(x)$ dla każdej waluacji zmiennych v

Prawomocność aksjomatów

(E) $\forall t_1, t_2 \exists x O(t_1, t_2, x)$ — oczywiste;

(λ) $\forall x (\neg(x \equiv \lambda) \Rightarrow (\lambda \sim x))$ — wynika z interpretacji (wyprowadzonej w modelu) relacji lewy podłańcuch: o ile $\neg(x \equiv \lambda)$, to $[x] \neq [\lambda]$, więc $[x] \in M$ a $[\lambda] = A^* \cdot M$, zatem w jakimś odcinku ciągu s może pojawić się łańcuch $[x]$ nawet gdyby w pewnym większym odcinku s pojawił się łańcuch z $A^* \cdot M = [\lambda]$ i przeciwnie. λ i x nie dają się więc połączyć.

(O) $\forall t_1, t_2, x (O(t_1, t_2, x) \& (x \neq \lambda) \Rightarrow \neg(\exists t_1', t_2' (t_1 \leq t_1') \& (t_2' \leq t_2) \& (\neg(t_1 = t_1') \vee \neg(t_2' = t_2)) \& O(t_1', t_2', x)))$.

Gdyby ten aksjomat nie był spełniony, to musiałby istnieć łańcuch (reprezentujący zdarzenie standardowe), który jest swoim własnym podłańcuchem, co jest niemożliwe.

(D) $\forall t_1, t_2, x \neg(x \equiv \lambda) \& O(t_1, t_2, x) \Rightarrow \neg(\exists t_1', t_2' ((t_1 \leq t_1') \& (t_2' \leq t_2) \& O(t_1', t_2', \lambda)))$

— oczywiste z powodu dziedziczności relacji podłańcuch dla klasy równoważności w modelu M .

(N) $\forall x \exists x_1, \dots, x_k ((x \sim x_i \text{ dla } i=1..k) \& (x_i \neq x_j \Rightarrow x_i \sim x_j \text{ dla } i, j=1..k) \& (\forall x' x' \sim x \Rightarrow \exists i ((x_i \ll_L x') \vee (x' \ll_L x_i)))$.

Oznaczmy zbiór komplementarnych zdarzeń dla $[x]$ jako $C[x]$ (więc $C[x]$ jest zbiorem klas równoważności);

— jeżeli jest $x \equiv \lambda$ to przyjmijmy $C[\lambda] = \{ |a|; a \in A \text{ i } \{a\} \in M \}$ (zbiór wszystkich znanych zdarzeń reprezentowanych przez łańcuch długości 1),

— jeżeli jest $[x] = |a|$; $a \in A \cap M$ (x odpowiada zdarzeniu reprezentowanemu łańcuchem długości 1), to definiujemy $C[x] = \{ |\lambda| \} \cup \{ |a'|; a' \in A, a' \neq a \}$ (dopełnieniem jest zbiór łańcuchów dla zdarzenia neutralnego i wszystkie zdarzenia reprezentowane łańcuchem długości 1 różnymi od a);

— inaczej można definiować dla $[x] = |a_1 \dots a_n|$, $a_j \in A$ (x odpowiada zdarzeniu reprezentowanemu łańcuchem długości n) $C[x] = \{ |\lambda| \} \cup \{ |a_1 \dots a_{n-1} a'|; a' \in A, a' \neq a_n \}$ (łańcuchy długości n które różnią się od $[x]$ znakiem na n -tym miejscu).

(CH) $\forall x, y x \ll_L x^* y$

— oczywiste ze względu na interpretację operatora ułańcuchowienia i definicję relacji podłańcucha w modelu —

(T) (a) $\forall t t \leq t$

(b) $\forall t, t', t'' (t \leq t' \& t' \leq t'') \Rightarrow t \leq t''$

(c) $\forall t, t' (t \leq t' \vee t' \leq t)$

— oczywiste.

Funkcja oceniania

Głównym celem jest zdefiniowanie dla pewnego sensownego fragmentu języka zdarzeń sekwencyjnych funkcji oceniania. Ta funkcja miałaby być subiektywną (w sensie subiektywnego pojęcia prawdopodobieństwa) analogią funkcji potwierdzania (*confirmation function*, *c-function*), którą konstruuje Carnap w *Logical Foundations of Probability*. Jest to więc funkcja określająca stopień, w jakim ewidencja (dostępna informacja) wyrażona formułą jakiegoś języka potwierdza hipotezę (wyrażoną formułą tegoż języka).

Logiczne pojęcie prawdopodobieństwa opiera się na założeniu, że prawdopodobieństwo zdarzenia jest dane przez strukturę języka, którego używamy w opisie prawdopodobieństwowej przestrzeni właściwej temu zdarzeniu. Carnap konstruuje funkcję potwierdzania jako standardowe prawdopodobieństwo warunkowe, przy czym za wyjściowe niewarunkowe prawdopodobieństwo przyjmuje logiczną miarę (*measure function*, *m-function*), której wartości są dane przez zastosowanie zasady symetrii na atomach języka.

W budowanej przeze mnie teorii trzeba rozstrzygnąć, czy włączyć funkcję oceniania do języka, czy pozostawić ją na metapoziomie. W literaturze występują oba podejścia: Carnapa funkcja potwierdzania pozostaje na metapoziomie, inni autorzy (patrz np. Fagin, Halpern, Megiddo: *A Logic for Reasoning about Probabilities*) włączają prawdopodobieństwo do języka przedmiotowego. Argumentem przeciw drugiej możliwości jest ogromny wzrost złożoności takiego języka. Można także argumentować, że tu zamiarem nie jest stworzenie języka o ocenianiu zdarzeń sekwencyjnych wyrażających się jakąś hipotezą, ale stworzenie języka o zdarzeniach sekwencyjnych i konstrukcja reguł oceny zdań tego języka rozumianych jako hipotezy. Z tego powodu w tych rozważaniach funkcja oceniania nie będzie elementem języka, ale będzie (tak samo jak Carnapa funkcja potwierdzenia) definiowana na metapoziomie.

Redukcja języka zdarzeń sekwencyjnych

Język L_{seq} jest niewątpliwie dogodny jako narzędzie do formułowania ogólnych twierdzeń o zdarzeniach sekwencyjnych, jednakże dla celów predykcji jest zbyt szeroki. Zadania predykcji wyspecyfikowano jako oszacowanie kontynuacji obserwowanego procesu na podstawie aktualnie obserwowanego stanu i zapisów z poprzednich obserwacji. Potrzeba zatem tylko formuły opisującej stan procesu ze względu na pewien moment. Dla tych zadań użyta zostanie jedynie część języka L_{seq} — formuły homogeniczne w czasie t .

Definicja 4.6 (Formuła homogeniczna)

Formułą języka L_{seq} jest **A-formuła homogeniczna w czasie t** , wtedy i tylko wtedy, gdy nie zawiera kwantyfikatorów, a jedynymi jej atomowymi podformułami są atomowe formuły w postaci $A_i(_)$.

Formułą języka L_{seq} jest **B-formuła homogeniczna w czasie t**, wtedy i tylko wtedy, gdy nie zawiera kwantyfikatorów, a jedynymi jej atomowymi podformułami są atomowe formuły w postaci $B_t(_)$.

Funkcję oceniania będę konstruował jako dwuargumentową funkcję nad formułami, gdzie pierwsza formuła reprezentuje aktualną informację (zdarzenie obserwowane), a druga hipotezę.

Widoczne jest, że aktualnym stanem jest zdarzenie obserwowane bezpośrednio przed czasem t, co w języku L_{seq} odpowiada atomowo homogenicznej B-formule $B_t(x)$, natomiast hipotezą — zdarzenia, które mogą pojawić się bezpośrednio po momencie t, co odpowiada jakiejś A-formule homogenicznej w czasie t. Wyjściowa miara w naszym pojęciu nie jest (w odróżnieniu od konstrukcji Carnapa) dana przez strukturę użytego języka. Odwołując się do Zasady Lokalnej Jedności i do modelu obserwatora można powiedzieć, że wyjściowa miara jest subiektywną kwantytatywną oceną zdarzeń w pewnym czasie. Konkretnym przykładem miary może być funkcja aktualności definiowana w pierwszej części artykułu, tj. funkcja, której wartość dla danego zdarzenia (reprezentowanego łańcuchem) zwiększa się przy pojawieniu się tego zdarzenia i zmniejsza się w przeciwnym wypadku. Jednakże teraz nie ograniczając się do modelu obserwatora, będę wręcz starał się o największe uogólnienie. Pozostaną jedynie takie postulaty co do tej miary, które są konieczne dla zapewnienia wymaganych własności oceniania hipotez.

W języku L_{seq} nie można odsyłać bezpośrednio do zdarzenia, jako ocenę zdarzenia w danym czasie rozumieć więc będę ocenę jego pojawienia się po danym momencie. Formalnie wyjściową miarą może być dowolna funkcja m definiowana nad atomowymi A-formułami języka zdarzeń sekwencyjnych, spełniająca następujące warunki.

Definicja 4.7 (Miara nad językiem zdarzeń sekwencyjnych)

Miarą w liczbach rzeczywistych nad językiem zdarzeń sekwencyjnych nazwiemy funkcję m ze zbioru atomowych A-formuł języka zdarzeń sekwencyjnych spełniającą następujące warunki (S, S' są dowolnymi e-termami):

(i) nieujemność

$$m(A_t(S)) \geq 0, \text{ gdzie } t \text{ jest momentem czasowym i } S \text{ dowolnym e-termem;}$$

(ii) zerowa miara zdarzenia neutralnego

$$m(A_t(\lambda))=0;$$

(iii) miara zdarzeń równoważnościowych

o ile spełnione jest $\forall t (A_t(S) \Leftrightarrow A_t(S'))$, to $m(A_t(S))=m(A_t(S'))$;

(iv) subaddytywność ze względu na ułańcuchowanie ze zbiorem zdarzeń nawzajem nie dających się połączyć

$$m(A_t(S)) \geq \sum_{S' \in H'} m(A_t(S*S')),$$

dla każdego momentu czasowego t i dla każdego zbioru zdarzeń nawzajem nie dających się połączyć H' .

Uwaga: O funkcji aktualności zdefiniowanej w modelu obserwatora można dowieść, że spełnia warunki definicji 4.7.¹⁴

Miara wyjściowa jest definiowana dla atomowej A-formuły. Jak na podstawie tej informacji ustalać ocenę dla formuły bardziej złożonej? Następujące twierdzenie pokazuje, że można każde złożone zdarzenie reprezentowane A-homogeniczną formułą rozłożyć na dysjunkcję elementarnych nie dających się połączyć zdarzeń, reprezentowanych przez atomowe A-formuły.

Twierdzenie 4.8 (Kanoniczna postać formuły A-homogenicznej)

Każda A-homogeniczna formuła jest równoważna dysjunkcji $A_t(S_1) \vee \dots \vee A_t(S_n)$, $n \geq 1$, gdzie $S_i = x_i$ lub $S_i = \lambda$ i dla $i \neq j$ spełnione jest $x_i \sim x_j$.

Dowód:

Przekład na postać kanoniczną będzie przebiegać w kilku krokach. Użyte zostaną zdania w formach normalnych dla rachunku kwantyfikatorów. Wpierw należy pozbyć się negacji. Posłużmy się twierdzeniem w dysjunktywnej formie normalnej, każdą A-homogeniczną formułą można określić jako koniunkcję dysjunkcji:

$$(\text{TX}^{1,1} \vee \dots \vee \text{TX}^{1, n_1}) \& \dots \& (\text{TX}^{k,1} \vee \dots \vee \text{TX}^{k, n_k}), \text{ gdzie } \text{TX}^{i,j} \text{ jest bądź } A_t(S^{i,j}) \text{ bądź } \neg A_t(S^{i,j}) \text{ i } S^{i,j} \text{ jest e-termem.}$$

Następnie konieczne jest przeprowadzenie przemianowania zmiennych; jeżeli jest e-term $S^{i,j}$ tworzony łańcuchem zmiennych i nie zawiera stałej λ , to można dla niego wprowadzić nową zmienną $x^{i,j}$. Jeżeli zawiera stałą λ , to można go oczywiście zgodnie z aksjomatami (A) i (D) zastąpić stałą λ .

Dalej można na podstawie aksjomatu teorii zdarzeń sekwencyjnych o negacji zastąpić negację dysjunkcjami. Po odpowiednim przekształceniu przybierze postać

$$A_t(S^{1,1}) \vee \dots \vee A_t(S^{1, n_1}) \& \dots \& A_t(S^{m,1}) \vee \dots \vee A_t(S^{m, m_k}), \text{ gdzie } S^{i,j} = y^{i,j} \text{ lub } S^{i,j} = \lambda.$$

W kolejnym kroku trzeba zredukować koniunkcję. Posłużę się w tym celu przekładem na koniunktywną formę normalną i następne oczywiste spostrzeżenie:

$$A_t(S) \& A_t(S') \& S \sim S' \Leftrightarrow A_t(\lambda) \& \neg A_t(\lambda)$$

(koniunkcja nie dających się połączyć zdarzeń
jest równoważna formule kontradiktorycznej),

$$A_t(S) \& A_t(S') \& S \ll_L S' \Leftrightarrow A_t(S')$$

(koniunkcja zdarzenia S i jego nadzdarzenia S'
jest równoważna nadzdarzeniu S').

¹⁴ Dowód można znaleźć w : *ibidem*.

Rezultatem jest dysjunkcja koniunkcji, gdzie każda koniunkcja jest bądź jednoelementowa bądź ma postać $A_i(\lambda) \& \neg A_i(\lambda)$. Jest niewątpliwe, że o ile pozostanie w ostatecznej koniunktywnej normalnej formie przynajmniej jeden niekontrykcyjny element, to wszystkie kontrykcyjne koniunkcje $A_i(\lambda) \& \neg A_i(\lambda)$ można pominąć. Jeżeli pozostaną tylko kontrykcyjne koniunkcje, to można pominąć wszystkie oprócz jednej i tak przekształcenie do formy normalnej jest gotowe. Jeżeli nie, to otrzymujemy wyrażenie

$$A_1^1(S^1) \vee \dots \vee A_r^m(S^m).$$

Pozostaje zredukować $S^1 \dots S^m$ do zbioru nawzajem nie dających się połączyć zdarzeń. To ponownie można przekształcić na podstawie dalszego oczywistego spostrzeżenia:

$$A_i(S) \vee A_i(S') \& S \ll_L S' \Leftrightarrow A_i(S)$$

(dysjunkcja zdarzenia S i jego nadzdarzenia S'
jest równoważna zdarzeniu S).

Jak już powiedziano, przy ogólnej formule języka L_{seq} nie ma sensu definiowanie funkcji oceniania dla predykcji. Formalnie funkcja ta będzie definiowana dla pary formuł, z których jedna opisuje aktualny stan procesu w danym momencie (wyrażony atomową homogeniczną B-formułą $B_i(x)$), a druga wyraża hipotezę o jego dalszym przebiegu (wyrażoną homogeniczną A-formułą).

Definicja 4.9 (Funkcja oceniania)

Założmy nieujemną miarę nad językiem zdarzeń sekwencyjnych m , spełniającą warunek z definicji 4.8, atomową homogeniczną B-formułę $B_i(S)$ i homogeniczną A-formułę F_i . Jeżeli F_i nie jest kontrykcyjna, to przez $N[F_i] = A_i(S_1) \vee \dots \vee A_i(S_n)$ oznaczmy jej kanoniczną formę. Funkcję oceniania można dalej zdefiniować w sposób następujący:

(i) dla F_i kontrykcyjnej

$$o(F_i, B_i(E)) =_{\text{def}} 0;$$

(ii) jeżeli żadne S_i w normalnej formie F_i nie jest λ

$$o(F_i, B_i(S)) =_{\text{def}} \sum_{i=1}^n \frac{m(A_i(S * S_i))}{m(A_i(S))}.$$

Własności funkcji oceniania

Na koniec chciałbym dowieść kilku własności funkcji oceniania a także przedyskutować, co te własności oznaczają z punktu widzenia teorii prawdopodobieństwa.

Uzasadnienie przedstawionego modelu predykcji jest takie samo jak w standardowych rozważaniach o prawdopodobieństwie, natomiast interesujące jest rozstrzygnięcie w warunkach niepewności. Należy sobie zatem postawić dwa pytania: czy zaproponowana funkcja oceniania jest odpowiednia dla tego zadania i jak wiąże się ze

standardowym prawdopodobieństwem? (Tzn. czy sama jest prawdopodobieństwowa, a jeśli nie jest, to jaki jest jej stosunek do prawdopodobieństwa?). Zajmę się teraz pierwszym pytaniem pozostawiając drugie do późniejszego rozważenia.

Stosowność funkcji nie jest niewątpliwie możliwa ściśle do udowodnienia, o stosowności różnych funkcji jako instrukcji dla rozstrzygnięcia można dyskutować (i w literaturze wiele jest takich dyskusji). Zamiarem moim jest sprawdzenie warunków, które dla funkcji rozstrzygnięcia w sytuacji niepewności postulował Carnap.¹⁵ Warunki w następującym twierdzeniu oczywiście nie mogą być identyczne z tymi, jakie bierze pod uwagę Carnap, ponieważ używany jest tu inny język. Są to Carnapowskie warunki przeformułowane na język L_{seq} . Spełnienie tych warunków dla zaproponowanej funkcji oceniania oznacza także, że jest ona prawdopodobieństwowa w sensie formalnym.

Twierdzenie 4.10 (Własności funkcji oceniania)

Niech H, H', E są dowolnymi e-termami, F_t dowolną A-homogeniczną formułą i B_t dowolną B-homogeniczną formułą atomową. Spełnione jest:

- (i) $0 \leq o(F_t, B_t(E)) \leq 1$ (zakres wartości);
- (ii) jeżeli $B_t(E) \Leftrightarrow B_t(E')$, to $o(F_t, B_t(E)) = o(F_t, B_t(E'))$ (ewidencja równoważna);
- (iii) jeżeli $F_t \Leftrightarrow F'_t$, to $o(F_t, B_t(E)) = o(F'_t, B_t(E))$ (hipotezy równoważne);
- (iv) jeżeli spełnione jest $F_t \sim F'_t$ (F_t, F'_t są nie dającymi się połączyć zdarzeniami), to $o(F_t \vee F'_t, B_t(E)) = o(F_t, B_t(E)) + o(F'_t, B_t(E))$ (addytywność);
- (v) jeżeli F_t jest formułą kontradykcyjną, to $o(F_t, B_t(E)) = 0$ (hipoteza kontradykcyjna);
- (vi) $o(A_t(H * H'), B_t(E)) \leq o(A_t(H), B_t(E))$ (monotoniczność oceniania ze względu na łańcuchowanie hipotez);
- (vii) jeżeli jest F_t formułą tautologiczną, to $o(F_t, B_t(E)) = 1$ (hipoteza tautologiczna).

Dowód:

(i) wynika natychmiast z definicji funkcji oceniania i z warunku (iv) dla miary m w definicji 4.8 (subaddytywność ze względu na łańcuchowanie ze zbiorem nawzajem nie dających się połączyć zdarzeń): $m(A_t(S)) \geq \sum_{S' \in H'} m(A_t(S * S'))$ dla każdego momentu t i dla każdego zbioru nawzajem nie dających się połączyć zdarzeń H' ;

(ii) — jeżeli $B_t(E) \Leftrightarrow B_t(E')$, to według warunku (iii) z definicji miary (miara równoważnych zdarzeń) spełnione jest $m(A_t(E)) = m(A_t(E'))$; (ii) następnie wynika z definicji funkcji oceniania;

(iii) — jeżeli spełnione jest $F_t \Leftrightarrow F'_t$, to obie formuły muszą mieć identyczną postać kanoniczną aż do permutacji homogenicznych A-formuł i przemianowania zmien-

¹⁵ Carnap nazywa je warunkami adekwatności (*conditions of adequacy*). Patrz: Carnap, *Logical Foundations...*, s. 315.

nych; każda homogeniczna A-formuła $A_t(S)$ z kanonicznej postaci F_t odpowiada następnie jakiejś homogenicznej A-formule $A_t(S')$ z kanonicznej postaci F'_t , tak, że $A_t(S) \leftrightarrow A_t(S')$ dla każdego czasowego momentu t . Dalej, ponownie według warunku (iii), z definicji miary spełnione jest $m(A_t(S)) = m(A_t(S'))$ i (iii) wynika z definicji funkcji oceniania;

(iv) — jeżeli spełnione jest $F_t \sim F'_t$, to musi być spełnione: każda homogeniczna A-formuła $A_t(S)$ z wyrażonym F_t w kanonicznej postaci jest nie dającą się połączyć z dowolną homogeniczną A-formułą $A_t(S')$ z wyrażenia F'_t w kanonicznej postaci, a więc kanoniczna postać dysjunkcji $F_t \vee F'_t$ musi być dysjunkcją kanonicznych postaci F_t i F'_t . Z definicji funkcji oceniania wynika zatem wymagana własność;

(v) — z definicji;

(vi) — według warunku (iv) z definicji 4.8 (subaddytywność ze względu na łańcuchowanie zbioru nawzajem nie dających się połączyć zdarzeń) spełnione jest $m(A_t(S)) \geq \sum_{S' \in H'} m(A_t(S * S'))$ dla każdego momentu czasowego t i dla każdego zbioru nawzajem nie dających się połączyć zdarzeń H' , więc specjalnie dla $H' = \{S'\}$ spełnione jest $m(A_t(S)) \geq m(A_t(S * S'))$ (monotonia dla miary m ze względu na łańcuchowanie). Z definicji oceniania:

$$o(A_t(H * H'), B_t(E)) = \sum_{i=1}^n \frac{m(A_t(E * H * H'))}{m(A_t(E))} \leq \sum_{i=1}^n \frac{m(A_t(E * H))}{m(A_t(E))} = o(A_t(H), B_t(E)).$$

(vii) — jeżeli F_t jest formułą tautologiczną, to jest równoważna formule $A_t(\lambda) \vee \neg A_t(\lambda)$.

Według aksjomatu (N) teorii zdarzeń sekwencyjnych spełnione jest $\exists x_1, \dots, x_k ((\lambda \sim x_i \text{ dla } i=1..k) \& (x_i \neq x_j \Rightarrow x_i \sim x_j \text{ dla } i, j=1..k) \& (\forall x' x' \sim \lambda \Rightarrow \exists i ((x_i <_{\mathcal{L}} x') \vee (x' <_{\mathcal{L}} x_i))))$ (istnieje zbiór zdarzeń nie dających się połączyć z λ , te zdarzenia nawzajem nie dają się połączyć i dany zbiór jest maksymalny ze względu na inkluzję), widoczne jest również, że $A_t(x_1) \vee \dots \vee A_t(x_k)$ jest kanoniczną formą $\neg A_t(\lambda)$; z twierdzenia (iv) tego lematu otrzymujemy:

$$o(F_t, B_t(S)) = o(A_t(\lambda) \vee \neg A_t(\lambda), B_t(S)) = o(A_t(\lambda), B_t(S)) + o(\neg A_t(\lambda), B_t(S))$$

(bowiem $A_t(\lambda)$ i $\neg A_t(\lambda)$ nie dają się połączyć), dalej z definicji funkcji oceniania

$$o(F_t, B_t(S)) = \frac{m(A_t(S * \lambda))}{m(A_t(S))} + \sum_{i=1}^k \frac{m(A_t(S * x_i))}{m(A_t(S))}.$$

Według punktu (iv) definicji miary nad językiem zdarzeń sekwencyjnych spełnione jest $m(A_t(S * \lambda)) = m(A_t(S)) - \sum_{S' \in H_{\max}} m(A_t(S * S'))$; dla każdego maksymalnego (ze względu na relację brania podłańcucha) zbioru nawzajem nie dających się połączyć zdarzeń H_{\max} , dla $H_{\max} = \{x_1, \dots, x_k\}$ otrzymamy:

$$o(F_i, B_i(S)) = \frac{m(A_i(S))}{m(A_i(S))} - \sum_{i=1}^k \frac{m(A_i(S * x_i))}{m(A_i(S))} + \sum_{i=1}^k \frac{m(A_i(S * x_i))}{m(A_i(S))} = 1.$$

Na koniec należy rozważyć pytanie, jakiego typu jest skonstruowana funkcja oceniania. Jest widoczne, że w tym wypadku głównymi kandydatami są logiczne i subiektywne pojęcie prawdopodobieństwa.

Istota logicznego pojęcia polega na zastosowaniu zasady symetrii — równomierne rozdzielanie prawdopodobieństwa między atomy języka opisujące obiekty przestrzeni probabilistycznej (zdarzenia, zjawiska). Rozdzielenie prawdopodobieństwa jest następnie zależne od złożoności i struktury stosowanego języka. Tak np. Wittgenstein w jednym z pierwszych pomysłów logicznego prawdopodobieństwa wprowadzonego do języka logiki zdań proponuje równomierne rozdzielanie prawdopodobieństwa w wierszach tabeli prawdziwościowej (patrz *Tractatus*). Pojęcie logicznego prawdopodobieństwa może jednak różnicować się ze względu na odmienne rozumienie atomów języka. Np. Carnap w 1962 r. doszedł do dwóch różnych miar (m' i m^*) logicznego prawdopodobieństwa na podstawie dwóch różnych sposobów rozumienia atomów języka (opisy stanów — *state descriptions* — w pierwszym wypadku i opisy struktury — *structure descriptions* — w drugim wypadku).

Subiektywne pojęcie zakłada, że wartości prawdopodobieństwowe określają subiektywne stopnie przeświadczenia (*degrees of belief*) i zależą tylko od podmiotu — nie jest konieczne wyprowadzanie ich z relatywnej liczebności ani ze struktury języka. Jedynym warunkiem, jaki te stopnie muszą spełniać, jest niesprzeczność, tj. wymóg, aby funkcja, której wartości odpowiadają stopniom przeświadczenia pewnego podmiotu, spełniała aksjomaty prawdopodobieństwa. Połączenie subiektywnego pojęcia prawdopodobieństwa i procesu obserwacji jest sednem subiektywnego Bayesjanizmu. Doktryna ta zakłada, że podmiot regularnie aktualizuje swoje stopnie przeświadczenia w zgodzie z faktami stwierdzonymi w procesie obserwacji. Aktualizacja przebiega jako obliczanie warunkowego prawdopodobieństwa na podstawie twierdzenia Bayesa.

Jak już powiedziano, przedstawiona koncepcja nie odpowiada logicznemu pojęciu prawdopodobieństwa, ponieważ nie odwołuje się w żadnej formie do zasady symetrii. Funkcja oceniania odpowiada subiektywnemu podejściu — prawdopodobieństwo jest ustalane na podstawie subiektywnego oceniania zdarzeń wyjściowych (miary nad formułami języka zdarzeń sekwencyjnych). Jednak w odróżnieniu od pojęcia Bayesowskiego aktualizacja oceniania nie dokonuje się z zastosowaniem czysto prawdopodobieństwowego mechanizmu w procesie uczenia się. W tym artykule również nie zakładam, że ta aktualizacja w ogóle musi przebiegać na podstawie jakiegoś procesu uczenia się. Prawdopodobieństwowy charakter funkcji oceniania zależy tylko od warunków nakładanych na wyjściową funkcję miary. To pojęcie jest w pewnym sensie bardziej «subiektywne» niż Bayesowskie, pozostawia bowiem podmiotowi większą swobodę w wyborze miary wyjściowej *resp.* jej wartości. Gdyby przyjęto założenie, że jakiś proces uczenia zachodzi, to warunki nakładane na ten proces są

całkowicie dowolne — ostateczna miara nie musi być prawdopodobieństwowa, musi jedynie spełniać warunki z definicji 4.8. Jeśli rozważy się przedstawiony model z punktu widzenia zadania predykcji, to można powiedzieć, że proces uczenia się nie przebiega w takich samych ramach jak proces oceniania hipotez. Przebiega równoległe poza tymi ramami i stosuje się w nim inne niż prawdopodobieństwowe środki, jedyny zaś związek zachodzi za pośrednictwem wspomnianych już warunków nakładanych na końcową miarę.

Z języka czeskiego tłumaczył Bogusław Szubert

Za przejrzanie tłumaczenia i pomoc w ustaleniu terminologii dziękuję mgr. Jerzemu Karczewskiemu z Katedry Biofizyki i Biologii Komórki Uniwersytetu Śląskiego.

Bogusław Szubert

LITERATURA

- R. Carnap, *Logical Foundations of Probability*, Chicago 1962
C. Howson, P. Urbach, *Scientific Reasoning: The Bayesian Approach* (Chicago: Open Court), Chicago 1993
O. Majer, *Rozbor dlouhých slov v pologrupách. Diplomová práce*, Matematicko-fyzikální fakulta University Karlovy v Praze, Praha 1986.
O. Majer, *Problematika sekvencijnych udalostí. Kandidátská disertační práce*, Filozofický ústav Akademie věd České republiky, Praha 1998.