



MathUp

Konferencja Zastosowań
Matematyki

Edycja VIII

Matematyka – lekarstwo na niepewność jutra
– MathUp 2026

Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki
Centrum Nauczania Matematyki i Fizyki
Politechnika Łódzka

SPIS TREŚCI

Spis treści / 1

1. Gertruda Gwóźdź-Łukawska, Elżbieta Galewska, Elżbieta Kotlicka-Dwurzniak, Monika Potyrała
Matematyczny przepis na algorytm odwagi. / 2
2. Julia Glicner, Amelia Głowala
Od rzutu monetą do Wall Street, czyli jak zrozumieć i wycenić ryzyko za pomocą modelu Coxa-Rossa-Rubinsteina. / 6
3. Kacper Bazyl, Bartosz Górski
Procesy decyzyjne Markowa i ich zastosowanie w grze blackjack. / 14
4. Jędrzej Niedzielski, Nina Zugaj
Kod pod lupą: algorytmy w roli cyfrowych detektywów. / 21
5. Marcin Król
Weryfikacja dezinformacji z użyciem teorii Dempstera-Shafera. / 28
6. Hanna Klus
Topologiczna detekcja bifurkacji w modelach terapii adaptacyjnej nowotworów. / 35
7. Maria Jędrzejewska
From Equation to Dose: The Mathematics Behind Personalized Medicine / 40
8. Jan Kocierz
Jak matematyka uratowała zdjęcia Jowisza? / 47
9. Alan Marciniak
Wpływ geometrii na luminescencję, czyli co Schrödinger robi w owocach? / 53
10. Szymon Woźniak, Łukasz Zychowicz
Kodowanie kształtu: Matematyka w projektowaniu parametrycznym. / 61

MATEMATYCZNY PRZEPIS NA ALGORYTM ODWAGI

Gertruda GWÓŹDŹ-ŁUKAWSKA, Elżbieta GALEWSKA,
Elżbieta KOTLICKA-DWURZNIK, Monika POTYRAŁA

Politechnika Łódzka, Łódź

Wstęp

Odwaga zaczyna się od pytania

Odwaga najczęściej kojarzona jest z działaniem w warunkach niepewności, podejmowaniem ryzyka oraz przekraczaniem własnych ograniczeń. Rzadziej dostrzega się jej związek z matematyką. Tymczasem artykuły zgromadzone w niniejszej monografii pokazują, że matematyka odgrywa istotną rolę wszędzie tam, gdzie konieczne jest podejmowanie odpowiedzialnych decyzji, przewidywanie skutków działań oraz zrozumienie procesów zachodzących w złożonych systemach. Wspólnym motywem wielu prezentowanych prac jest próba oswojenia niepewności poprzez wykorzystanie narzędzi matematycznych do opisu, analizy i prognozowania zjawisk występujących w nauce, technice, medycynie oraz gospodarce.

Rozważania te stanowią naturalną kontynuację tematów podejmowanych w poprzednich monografiach konferencji MathUp, poświęconych roli matematyki w lepszym rozumieniu rzeczywistości oraz odkrywaniu zarówno oczywistych, jak i zaskakujących zastosowań królowej nauk.

Jak matematyka mierzy ryzyko

Szczególnym obszarem, w którym niepewność stanowi nieodłączny element procesu decyzyjnego, są finanse. Artykuł Julii Glicner i Amelii Głowali poświęcony został wycenieniu europejskich opcji kupna w wielookresowym modelu Coxa-Rossa-Rubinsteina. Autorki pokazują, że cena instrumentu finansowego nie jest rezultatem intuicyjnych przewidywań inwestora, lecz konsekwencją matematycznej konstrukcji rynku pozbawionego arbitrażu. Wykorzystując pojęcie portfela replikującego oraz prawdopodobieństwa neutralnego względem ryzyka, przeprowadzają czytelnika przez proces wyceny opcji, ilustrując go szczegółowym przykładem obliczeniowym. Praca stanowi przystępne wprowadzenie do matematyki finansowej i pokazuje, że nawet w świecie pozornie nieprzewidywalnych zdarzeń możliwe jest budowanie modeli umożliwiających racjonalną ocenę ryzyka.

Gdy decyzje stają się grą matematyki

Podejmowanie decyzji w warunkach niepewności stanowi również temat artykułu Kacpra Bazyla i Bartosza Górskiego poświęconego analizie strategii w blackjacku. Autorzy pokazują, że popularna gra karciana może być traktowana jako model problemu decyzyjnego, w którym konieczne jest wybieranie działań na podstawie niepełnych informacji i przewidywanych konsekwencji.

Wykorzystując uproszczony model rozgrywki, analizują zależności między aktualną sytuacją gracza a oczekiwanym wynikiem poszczególnych decyzji. Dzięki temu pokazują, jak matematyka pozwala przejść od intuicyjnych wyborów do racjonalnie uzasadnionej strategii.

Choć punktem wyjścia jest gra losowa, przedstawione podejście ma znacznie szersze znaczenie. Podobne mechanizmy występują bowiem w ekonomii, zarządzaniu, informatyce czy sztucznej inteligencji, wszędzie tam, gdzie decyzje podejmowane są w warunkach ryzyka i niepewności. Artykuł stanowi interesujący przykład wykorzystania matematyki do analizy wyborów oraz pokazuje, jak modele i algorytmy wspierają proces podejmowania decyzji.

W poszukiwaniu prawdy ukrytej w danych

We współczesnym świecie ogromne znaczenie ma także umiejętność analizy informacji. Temu zagadnieniu poświęcony został artykuł Jędrzeja Niedzielskiego i Niny Zugaj dotyczący wykrywania plagiatu i automatycznego rozpoznawania podobieństw pomiędzy tekstami oraz kodem źródłowym. Autorzy przedstawiają proces przygotowania danych do analizy, a następnie opisują metody wykorzystywane do oceny podobieństwa dokumentów. Czytelnik może poznać zarówno techniki oparte na analizie sąsiedztwa wyrazów i odległości między tekstami, jak i bardziej zaawansowane reprezentacje numeryczne dokumentów. Praca pokazuje, że skuteczna analiza podobieństwa nie opiera się na „przecuciach” komputera, lecz na precyzyjnie definiowanych własnościach matematycznych, które można mierzyć, porównywać i interpretować.

Jak ocenić wiarygodność danych

Działanie w warunkach niepewności wymaga nie tylko gromadzenia informacji, ale także umiejętności oceny ich wiarygodności. Temu zagadnieniu poświęcony jest artykuł Marcina Króla, w którym teoria Dempstera-Shafera służy jako narzędzie wspomagające analizę danych pochodzących z różnych źródeł.

Autor pokazuje, że w wielu sytuacjach dostępne informacje są niepełne, niepewne lub nawet częściowo sprzeczne. W takich przypadkach konieczne staje się wykorzystanie metod matematycznych pozwalających określić stopień przekonania co do prawdziwości poszczególnych hipotez oraz połączyć niezależne przesłanki w spójną ocenę sytuacji.

Przedstawione podejście umożliwia analizę niepewności oraz fuzję danych pochodzących z wielu źródeł, co pozwala wskazać najbardziej prawdopodobne wnioski wraz z oceną ich wiarygodności. Artykuł stanowi interesujący przykład zastosowania matematyki w procesie podejmowania decyzji i pokazuje, że odpowiednie modele mogą wspierać ocenę informacji wszędzie tam, gdzie działamy w warunkach niepewności.

Walka z chorobą na poziomie liczb

Problem podejmowania decyzji w sytuacjach, gdy stawką jest zdrowie i życie człowieka, pojawia się w artykule Hanny Klus dotyczącym terapii adaptacyjnej nowotworów. Autorka analizuje możliwość wczesnego wykrywania momentu, w którym stosowane leczenie zaczyna tracić skuteczność. W tym celu wykorzystuje narzędzia analizy topologicznej pozwalające badać strukturę zmian zachodzących w układzie opisującym rozwój choroby.

Zaprezentowane wyniki wskazują, że odpowiednio skonstruowane wskaźniki topologiczne mogą sygnalizować pogorszenie odpowiedzi na leczenie wcześniej niż standardowa obserwacja poziomu biomarkera PSA. Artykuł stanowi interesujący przykład wykorzystania zaawansowanych metod matematycznych w medycynie oraz pokazuje, że matematyka może wspierać lekarzy nie tylko w analizie danych, ale także w podejmowaniu decyzji dotyczących dalszego przebiegu terapii.

Od równań do dawkowania leków

Matematyka odgrywa coraz większą rolę w medycynie, pomagając w projektowaniu skutecznych i bezpiecznych terapii. Artykuł Marii Jędrzejewskiej poświęcony jest farmakokinetyce, czyli dziedzinie badającej zmiany stężenia leku w organizmie po jego podaniu.

Autorka przedstawia podstawowe modele matematyczne wykorzystywane do opisu tego procesu. Zastosowane w nich równania różniczkowe pozwalają opisywać wchłanianie, dystrybucję i eliminację leków oraz przewidywać zmiany ich stężenia w czasie. W szczególności dowiadujemy się, w jaki sposób wyznaczane jest okno terapeutyczne, czyli zakres stężeń zapewniających skuteczność leczenia przy jednoczesnym ograniczeniu ryzyka działań niepożądanych.

Artykuł stanowi interesujący przykład zastosowania matematyki w naukach medycznych i pokazuje, że odpowiednio skonstruowane modele mogą wspierać rozwój medycyny personalizowanej, uwzględniającej indywidualne cechy pacjenta.

Jak matematyka pomaga przesyłać informacje z odległych zakątków kosmosu

Algebraiczna teoria kodowania stanowi podstawę niezawodnej komunikacji cyfrowej, szczególnie w warunkach silnych zakłóceń, takich jak przestrzeń kosmiczna. Kody korekcyjne, wykorzystujące odpowiednią strukturę algebraiczną i nadmiarowość, pozwalają odzyskać dane uszkodzone podczas transmisji. Jan Kocierz w swoim artykule przybliżył to zagadnienie na przykładzie sondy Voyager 1, która przesyłała na Ziemię pierwsze precyzyjne zdjęcia Jowisza.

Autor pokazuje, jak narzędzia matematyczne – w szczególności struktury algebraiczne nad ciałem $\mathbb{Z}_2$ oraz liniowe kody binarne – umożliwiły skuteczną ochronę przesyłanych danych. Historia Voyagera 1 pokazuje, że pozornie abstrakcyjne pojęcia matematyczne mogą stać się kluczowym elementem zaawansowanych technologii, umożliwiając przesyłanie informacji na ogromne odległości.

Gdy światło rodzi się z geometrii

W świecie nanotechnologii nawet najmniejsze zmiany rozmiaru materiału mogą prowadzić do istotnych zmian jego właściwości. Temat ten podejmuje Alan Marciniak w artykule poświęconym grafenowym kropkom kwantowym i ich zdolności do emitowania światła. Autor pokazuje, że kolor emitowanego promieniowania zależy od geometrii nanostruktury oraz praw mechaniki kwantowej opisujących zachowanie cząstek w bardzo małych skalach.

Przedstawione modele matematyczne pozwalają wyjaśnić zależność między rozmiarem kropki kwantowej a energią emitowanych fotonów. Szczególnie interesujące jest zestawienie przewidywań teoretycznych z wynikami eksperymentów dla grafenowych kropek kwantowych otrzymywanych metodami zielonej syntezy.

Artykuł pokazuje, że matematyka stanowi kluczowe narzędzie projektowania nowoczesnych materiałów, a zrozumienie zjawisk zachodzących na poziomie nanometrów otwiera drogę do zastosowań w biosensorach, diagnostyce medycznej i technologiach pozyskiwania energii. Jest to interesujący przykład połączenia teorii, eksperymentu i nowoczesnych zastosowań technologicznych.

Gdy budynki rodzą się z algorytmów

W monografii nie zabrakło również zagadnień związanych z inżynierią i projektowaniem. Rozdział autorstwa Szymona Woźniaka i Łukasza Zychowicza przybliżył rolę matematyki

w projektowaniu parametrycznym, które stało się jednym z najważniejszych kierunków rozwoju współczesnej architektury i budownictwa. Autorzy pokazują, że spektakularne realizacje architektoniczne nie są wyłącznie efektem twórczej wizji projektanta, lecz rezultatem zastosowania geometrii, algorytmów oraz metod cyfrowego modelowania.

Przywołane przykłady znanych budowli pokazują, że geometria wpływa nie tylko na wygląd obiektu, ale również na jego wytrzymałość, aerodynamikę, efektywność materiałową czy możliwości wykonawcze. Szczególnie interesująca jest przedstawiona w artykule zmiana perspektywy projektowej: współczesny projektant coraz częściej nie tworzy pojedynczego kształtu, lecz definiuje zbiór zależności i parametrów, które pozwalają generować, analizować oraz optymalizować różne warianty rozwiązania.

Gdzie kończy się intuicja, zaczynają się obliczenia

Choć tematyka poszczególnych rozdziałów jest bardzo różnorodna, wszystkie prezentowane opracowania łączy przekonanie, że matematyka nie jest jedynie abstrakcyjnym zbiorem pojęć i wzorów. Stanowi uniwersalne narzędzie pozwalające opisywać rzeczywistość, wspomagać procesy decyzyjne oraz lepiej rozumieć złożone zjawiska występujące w otaczającym świecie. Matematyczne modele pomagają analizować rynki finansowe, rozpoznawać wzorce ukryte w danych, monitorować przebieg terapii i projektować coraz bardziej zaawansowane konstrukcje.

Przepis na algorytm odwagi

Niniejsza monografia pokazuje matematykę jako naukę aktywnie uczestniczącą w rozwiązywaniu problemów współczesności. Zebrane w niej artykuły dowodzą, że odwaga nie polega na ignorowaniu ryzyka, lecz na jego rozumieniu. W tym sensie matematyka staje się swoistym algorytmem odwagi: pozwala przekraczać granice poznania, formułować trafniejsze wnioski oraz poszukiwać rozwiązań tam, gdzie sama intuicja okazuje się niewystarczająca. To właśnie dzięki królowej nauk możliwe jest podejmowanie świadomych decyzji w świecie pełnym niepewności i wyzwań.

Podobnie jak monografie MathUp z lat 2024 [1] i 2025 [2], również niniejszy tom pokazuje, że matematyka nie tylko opisuje rzeczywistość, ale także pozwala odkrywać jej nowe wymiary, łącząc teoretyczne podstawy z wyzwaniami współczesnego świata.

Literatura

- [1] Gwóźdź-Łukawska G. (red.), Potyrała M. (red.), MathUp Jak matematyka pomaga w lepszym zrozumieniu rzeczywistości – MathUp2024
[dostęp https://mathup.p.lodz.pl/shared/documents/Monografia_MathUp_2024.pdf]
- [2] Gwóźdź-Łukawska G. (red.), Galewska E., Kotlicka-Dwurznic E., Potyrała M., Matematyka – od zastosowań oczywistych do zdumiewających, MathUp 2025, Politechnika Łódzka, Łódź,
[dostęp: https://mathup.p.lodz.pl/shared/documents/Monografia_MathUp_2025.pdf]

OD RZUTU MONETĄ DO WALL STREET – ZASTOSOWANIE MODELU COXA-ROSSA-RUBINSTEINA W WYCENIE INSTRUMENTÓW POCHODNYCH.

Julia GLICNER, Amelia GŁOWALA

Politechnika Łódzka, Łódź

Wstęp

Współczesne rynki finansowe charakteryzują się wysokim stopniem złożoności oraz permanentną ekspozycją na ryzyko rynkowe. Choć uwaga większości inwestorów indywidualnych koncentruje się wokół klasycznych instrumentów udziałowych (akcji) bądź powszechnie oferowanych przez brokerów instrumentów pochodnych o charakterze spekulacyjnym, takich jak kontrakty na różnice kursowe (CFD), relatywnie niewielka część uczestników rynku swobodnie posługuje się opcjami. Wynika to z faktu, iż instrumenty te są zaawansowane matematycznie i wymagają specyficznego podejścia analitycznego.

Celem niniejszej pracy jest formalne udowodnienie, że inwestowanie i zarządzanie portfelem opcyjnym nie opiera się na intuicji czy spekulacji, lecz stanowi dyscyplinę opartą na ścisłym aparacie matematycznym. W pracy przedstawiamy mechanizm obiektywnej wyceny i neutralizacji ryzyka przy użyciu wielookresowego modelu dwumianowego Coxa-Rossa-Rubinsteina (CRR), stanowiącego dyskretny odpowiednik ciągłych modeli inżynierii finansowej.

Oznaczenia

W celu zapewnienia przejrzystości wywodu matematycznego, poniżej podajemy kluczowe oznaczenia stosowane w strukturze wielookresowego modelu dwumianowego oparte o źródło [1]:

S_0 – bieżąca (początkowa) cena instrumentu bazowego w chwili $t = 0$, przy założeniu, że $S_0 > 0$,
 S_t – cena instrumentu bazowego w dyskretnym momencie czasu t (gdzie $t = 0, 1, \dots, N$), będąca elementem procesu stochastycznego,
 K – cena wykonania opcji, określająca stały poziom transakcji gwarantowanej przez kontrakt, przy założeniu, że $K > 0$,
 T – całkowity horyzont czasowy, oznaczający moment wygaśnięcia kontraktu ($t = N$), gdzie $T > 0$,
 N – całkowita liczba okresów w modelu dyskretnym, gdzie $N \in \mathbb{N}_+$,
 C_0 – bieżąca wartość (cena sprawiedliwa) europejskiej opcji kupna w chwili początkowej $t = 0$, przy czym $C_0 \geq 0$,
 P_0 – bieżąca wartość (cena sprawiedliwa) europejskiej opcji sprzedaży w chwili początkowej $t = 0$, przy czym $P_0 \geq 0$,
 C_t^u – wartość opcji w węźle sieci odpowiadającym multiplikatywnemu wzrostowi ceny aktywa w kroku t ,
 C_t^d – wartość opcji w węźle sieci odpowiadającym multiplikatywnemu spadkowi ceny aktywa w kroku t ,

u – współczynnik stochastycznego wzrostu ceny instrumentu bazowego dla stanu up, spełniający kryterium $u > 1$,
 d – współczynnik stochastycznego spadku ceny instrumentu bazowego dla stanu down, spełniający kryterium $0 < d < 1$,
 r – stopa wolna od ryzyka dla pojedynczego okresu rozliczeniowego, reprezentująca stały koszt pieniądza w czasie, gdzie $r > 0$,
 p – prawdopodobieństwo neutralne wobec ryzyka (prawdopodobieństwo arbitrażowe), definiowane syntetycznie wzorem $p = \frac{1+r-d}{u-d}$, przy założeniu, że $0 < p < 1$,
 $1 - p$ – dopełniające prawdopodobieństwo przejścia procesu cenowego do stanu spadkowego (down) [1], gdzie $0 < 1 - p < 1$,
 Δ – współczynnik zabezpieczenia, określający dynamiczną liczbę jednostek aktywa bazowego niezbędną do replikacji profilu opcji,
 B – pozycja gotówkowa (wartość kredytu lub lokaty bankowej) wchodząca w skład portfela replikującego,
 $\binom{N}{k}$ – symbol Newtona, reprezentujący współczynnik kombinatoryczny liczby alternatywnych ścieżek rynkowych w strukturze drzewa.

Fundamenty „Rynku Doskonałego”

Matematyczna wycena instrumentów pochodnych wymaga przyjęcia wyidealizowanych założeń.

W ten sposób został skonstruowany abstrakcyjny model tzw. rynku doskonałego. Koncepcja ta, ugruntowana w [2] bazuje na następujących postulatach mikroekonomicznych i operacyjnych:

- **Brak tarć rynkowych:** Koszty transakcyjne, prowizje maklerskie, spready cenowe oraz wszelkie obciążenia fiskalne (w tym podatki od zysków kapitałowych) są równe zero.
- **Doskonała płynność oraz bezwarunkowa swoboda alokacji kapitału:** Zakłada się nieskończoną podzielność instrumentów finansowych, umożliwiającą operowanie na ułamkowych wolumenach aktywów pierwotnych, a także pełną dostępność do nielimitowanej i wolnej od restrykcji depozytowych krótkiej sprzedaży.
- **Symetria stóp procentowych:** Nominalna stopa wolna od ryzyka (r) dla lokowania nadwyżek kapitałowych (rynkowe lokaty) oraz pozyskiwania finansowania dłużnego (kredyty) jest stała, przyjęta ogólnie i identyczna dla wszystkich uczestników handlu.

Abstrakcyjny charakter tych założeń jest matematycznie niezbędny do zapewnienia cechy zupełności rynku. W rynku zupełnym każdy instrument pochodny może być idealnie zreplikowany za pomocą kombinacji instrumentów pierwotnych. Istnienie tak zdefiniowanego portfela replikującego implikuje bezpośrednio spełnienie warunku braku arbitrażu, co definitywnie wyklucza możliwość generowania dodatnich przepływów pieniężnych przy zerowym nakładzie kapitału początkowego i całkowitym braku ekspozycji na ryzyko strat rynkowych.

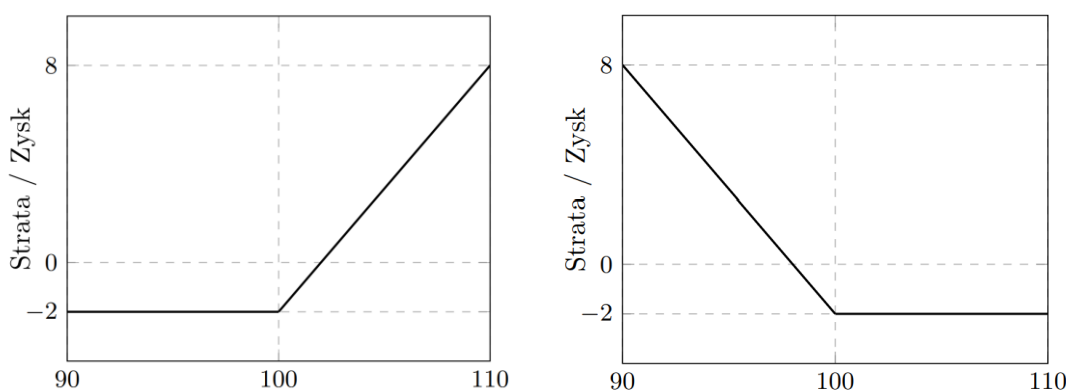
Klasyfikacja opcji i parytet kupna-sprzedaży

Opcja jest pochodnym instrumentem finansowym (derywatem), którego profil wypłaty jest asymetryczny – reprezentuje prawo, a nie obowiązek kupującego do realizacji transakcji w ściśle określonym terminie wygaśnięcia T po z góry ustalonej cenie wykonania K . Za to jednostronne zobowiązanie kupujący płaci wystawcy (sprzedawcy) bezzwrotną kwotę, zwaną premią opcyjną. Rozróżniamy dwa podstawowe rodzaje opcji rynkowych:

- **Opcja kupna (Call):** Daje nabywcy prawo zakupu instrumentu bazowego. Stanowi pozycję o charakterze pro-wzrostowym, analogiczną rynkowo do wpłaty zadatku zabezpieczającego stałą cenę dobra.
- **Opcja sprzedaży (Put):** Daje nabywcy prawo sprzedaży instrumentu bazowego. Działa na wzór polisy ubezpieczeniowej, gwarantując stałą cenę zbytu w przypadku załamania rynku.

W warunkach wolnych od arbitrażu ceny opcji Call (C_0) oraz Put (P_0) o identycznych parametrach (K, T) są ze sobą sztywno powiązane matematyczną relacją znaną jako **parytet kupna-sprzedaży**

$$P_0 = C_0 - S_0 + \frac{K}{(1+r)^N}. \quad (1)$$



Rys. 1. Profile wypłat opcji typu europejskiego kolejno Call i Put (wykresy asymetrycznych profili zysków/strat w zależności od ceny S w momencie wygaśnięcia). Źródło: [1]

Portfel replikujący i prawo jednej ceny

Podstawą nowoczesnej teorii wyceny derywatów jest konstrukcja tzw. portfela replikującego. Zamiast prognozować przyszłe trendy cenowe, budujemy syntetyczny ekwiwalent opcji złożony z określonej liczby akcji instrumentu bazowego (Δ) oraz pozycji gotówkowej (B), reprezentującej zaciągnięty kredyt bądź ulokowaną lokatę.

Konstrukcja ta opiera się na prawie jednej ceny: jeśli w momencie wygaśnięcia T portfel replikujący generuje identyczne przepływy pieniężne (wypłaty) jak wyceniana opcja, niezależnie od tego, czy rynek odnotował wzrost, czy spadek, to w chwili początkowej wartość premii opcyjnej musi być bezwzględnie równa wartości początkowej tego portfela. Dla pojedynczego okresu relację tę opisuje układ równań bilansowych:

$$\begin{cases} \Delta \cdot uS_0 + B(1+r) = C_1^u \\ \Delta \cdot dS_0 + B(1+r) = C_1^d \end{cases} \quad (2)$$

Model dwumianowy Coxa-Rossa-Rubinsteina (CRR)

Model CRR zakłada dyskretną ewolucję ceny instrumentu bazowego w czasie. W każdym pojedynczym kroku cena początkowa S_0 może zmienić się wyłącznie na dwa sposoby: wzrosnąć do poziomu uS_0 lub spaść do poziomu dS_0 . Aby wykluczyć strukturę arbitrażową, parametry te muszą spełniać fundamentalną zasadę braku arbitrażu na rynku.

Założenie CRR:

$$d < 1 + r < u.$$

Rozwiązując układ równań (2) z niewiadomymi Δ oraz B , otrzymujemy precyzyjne miary struktury portfela zabezpieczającego:

$$\Delta = \frac{C_1^u - C_1^d}{(u - d)S_0} \quad (3)$$

$$B = \frac{1}{1 + r} \cdot \frac{uC_1^d - dC_1^u}{u - d}. \quad (4)$$

Współczynnik Δ określa dynamiczną liczbę akcji, jaką wystawca opcji musi posiadać w danym węźle sieci, aby w pełni zneutralizować ryzyko portfela. Podstawiając (3) i (4) do wzoru na bieżącą wartość portfela $C_0 = \Delta S_0 + B$, dochodzimy do kluczowego pojęcia prawdopodobieństwa arbitrażowego p , neutralnego wobec ryzyka:

$$C_0 = \frac{1}{1 + r} [pC_1^u + (1 - p)C_1^d], \quad (5)$$

gdzie parametr p zdefiniowany jest wzorem:

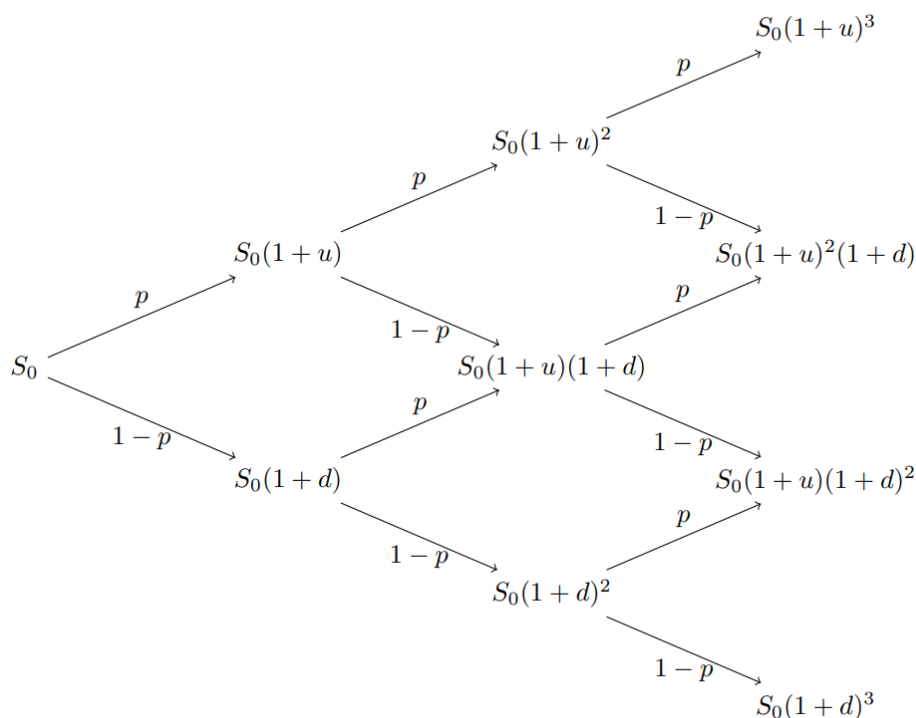
$$p = \frac{1 + r - d}{u - d}. \quad (6)$$

Należy podkreślić, że p nie odzwierciedla subiektywnego, probabilistycznego przekonania inwestorów o kierunku ruchu rynku, lecz jest parametrem ściśle syntetycznym. Uogólniając model na przypadek wielookresowy (N okresów), oczekiwana wartość końcowych wypłat opcji podlega rozkładowi dwumianowemu Bernoulliego, a ich obecna wartość po zdyskontowaniu przyjmuje postać jawną zarówno dla opcji kupna (7) jak i sprzedaży (8):

$$C_0 = \frac{1}{(1 + r)^N} \sum_{k=0}^N \binom{N}{k} p^k (1 - p)^{N-k} \max\{0, u^k d^{N-k} S_0 - K\}, \quad (7)$$

$$P_0 = \frac{1}{(1 + r)^N} \sum_{k=0}^N \binom{N}{k} p^k (1 - p)^{N-k} \max\{0, K - u^k d^{N-k} S_0\}. \quad (8)$$

Rysunek 2 wizualizuje dyskretną strukturę stochastyczną opisywaną przez powyższe równania. Schemat ilustruje ewolucję ceny instrumentu bazowego w uogólnionym modelu wielookresowym, stanowiąc graficzną podstawę do agregacji końcowych wypłat i wyznaczenia sprawiedliwej premii opcyjnej.



Rys. 2. Schemat uogólnionego wielookresowego drzewa dwumianowego. Źródło: opracowanie własne

Analiza praktyczna: model 3-okresowy

W celu weryfikacji operacyjnej modelu, rozważmy wycenę europejskiej opcji kupna przy następujących danych wejściowych:

- początkowa cena akcji: $S_0 = 100$ PLN,
- cena wykonania kontraktu: $K = 105$ PLN,
- czynniki ewolucji rynkowej: $u = 1.1$; $d = 0.9$,
- stopa wolna od ryzyka: $r = 2\%$ na okres rozliczeniowy,
- horyzont czasowy: $N = 3$ okresy.

Zgodnie ze wzorem (6), prawdopodobieństwo neutralne wobec ryzyka wynosi:

$$p = \frac{1.02 - 0.9}{1.1 - 0.9} = \frac{0.12}{0.20} = 0.6.$$

Proces wyceny realizowany jest algorytmem indukcji wstecznej (od momentu wygaśnięcia $t = 3$ do momentu początkowego $t = 0$). Stosując dyskontowanie wsteczne dla poszczególnych węzłów pośrednich, obliczamy wartości oczekiwane premii:

- dla $t = 2$ w stanie uu : $C_{uu} = \frac{1}{1.02} [0.6 \cdot 28.10 + 0.4 \cdot 3.90] = 18.06$ PLN,
- dla $t = 1$ w stanie u : $C_u = \frac{1}{1.02} [0.6 \cdot 18.06 + 0.4 \cdot 2.29] = 11.52$ PLN.

Ostatecznie, w punkcie startowym ($t = 0$), sprawiedliwa matematycznie cena opcji wynosi:

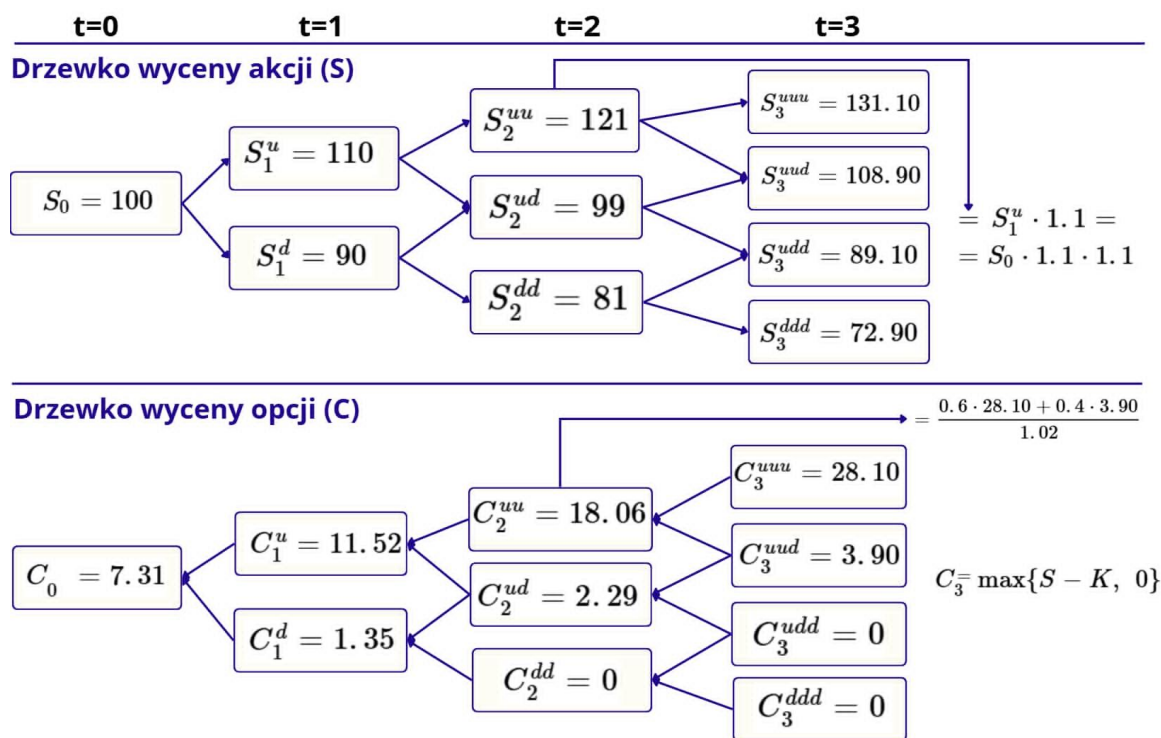
$$C_0 = \frac{1}{1.02} [0.6 \cdot 11.52 + 0.4 \cdot 1.35] = 7.31 \text{ PLN.}$$

Scenariusz rynkowy	Węzeł	$t = 0$	$t = 1$	$t = 2$	$t = 3$ (S_3)	Wypłata końcowa (C_3)
uuu (wzrost – wzrost – wzrost)	S_{uuu}	$S = 100$ $C = 7.31$	$S = 110$ $C = 11.52$	$S = 121$ $C = 18.06$	133.10	28.10
uud (wzrost – wzrost – spadek)	S_{uud}	$S = 100$ $C = 7.31$	$S = 110$ $C = 11.52$	$S = 121$ $C = 18.06$	108.90	3.90
udd (wzrost – spadek – spadek)	S_{udd}	$S = 100$ $C = 7.31$	$S = 110$ $C = 11.52$	$S = 99$ $C = 2.29$	89.10	0.00
ddd (spadek – spadek – spadek)	S_{ddd}	$S = 100$ $C = 7.31$	$S = 90$ $C = 1.35$	$S = 81$ $C = 0.00$	72.90	0.00

Tab. 1. Ewolucja cen instrumentu bazowego (S) i wartości opcji (C) w czasie. Źródło: opracowanie własne

W momencie początkowym ($t = 0$) inwestor ponosi stały koszt w postaci premii opcyjnej, wynoszącej 7.31 PLN. W scenariuszu wzrostowym (uuu), gdy cena instrumentu bazowego osiągnie poziom 133.10 PLN, kupujący realizuje zysk netto w wysokości 20.79 PLN. Wynika to bezpośrednio z różnicy między wypłatą końcową (28.10 PLN) a początkowym kosztem opcji (7.31 PLN). Z kolei w przypadku znacznych spadków cen rynkowych wyraźnie uwidacznia się asymetria ryzyka tego instrumentu. Maksymalna strata inwestora jest z góry ograniczona wyłącznie do zapłaconej na starcie premii (7.31 PLN), co oznacza, że dalsze załamanie rynku nie pogarsza już jego sytuacji kapitałowej.

Poniższy schemat ilustruje symulację możliwych ścieżek cenowych instrumentu bazowego oraz przypisane im końcowe wartości wypłat dla kolejnych dyskretnych momentów w czasie ($t = 0$, $t = 1$, $t = 2$, $t = 3$).



Rys. 3. Trzyokresowe drzewo dwumianowe wyceny akcji oraz opcji Call. Źródło: opracowanie własne

Podsumowanie i wnioski końcowe

Analiza trójetapowego drzewa dwumianowego oraz struktury rynku doskonałego pozwala na sformułowanie kilku kluczowych wniosków dotyczących natury instrumentów pochodnych. Przede wszystkim wykazaliśmy, że wartość opcji nie jest determinowana przez subiektywne prognozy inwestorów co do przyszłych trendów giełdowych, lecz opiera się na obiektywnym prawdopodobieństwie arbitrażowym, zależnym wyłącznie od zmienności rynkowej i stopy wolnej od ryzyka. Model udowadnia również, że za pomocą odpowiednio skonstruowanego portfela replikującego i dynamicznego współczynnika zabezpieczenia, wystawca opcji jest w stanie w pełni zneutralizować ryzyko rynkowe, niezależnie od tego, czy rynek ostatecznie rośnie, czy spada.

Przeprowadzone w artykule obliczenia bezsprzecznie potwierdzają skuteczność modelu Coxa-Rossa-Rubinsteina w dokładnym mierzeniu i ograniczaniu ryzyka finansowego. Zastąpienie spekulacji ścisłymi regułami matematycznymi sprawia, że inwestorzy zyskują precyzyjne narzędzia do świadomego zarządzania kapitałem. Współczesne instytucje i fundusze inwestycyjne powszechnie wykorzystują ten mechanizm, co pozwala im zachować stabilność nawet w warunkach drastycznej niepewności na giełdzie.

Tytułowa droga „od rzutu monetą do Wall Street” doskonale obrazuje ewolucję podejścia do rynków kapitałowych. Opcje, choć często przez niewiedzę utożsamiane z niebezpieczną spekulacją, w rzeczywistości dają uczestnikom rynku potężną dźwignię finansową przy jednoczesnej żelaznej ochronie wpłaconego kapitału. Zrozumienie ich matematycznej natury udowadnia, że współczesne inwestowanie to nie kwestia pecha czy szczęścia, lecz precyzyjna nauka. W dobie powszechnego dostępu do technologii algorytmicznych, mechanizmy wyceny zarezerwowane niegdyś wyłącznie dla elitarnych analityków z Wall Street stają się dziś intuicyjne i dostępne dla każdego świadomego uczestnika rynku.

Literatura

- [1] Weron A., Weron R., Inżynieria finansowa. Wycena instrumentów pochodnych. Symulacje komputerowe. Statystyka rynku, Wydawnictwa Naukowo-Techniczne, Warszawa 1998
- [2] Black, F., & Scholes, M. The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3), 637–654, 1973

PROCESY DECYZYJNE MARKOWA I ICH ZASTOSOWANIE W GRZE BLACKJACK

Kacper BAZYL, Bartosz GÓROWSKI

Politechnika Krakowska im. Tadeusza Kościuszki, Kraków

Wstęp

Blackjack jest jedną z najpopularniejszych karcianych gier losowych na świecie, w której decyzje graczy mają istotny wpływ na wynik rozgrywki. Losowy charakter blackjaka znacznie utrudnia wyznaczenie strategii optymalnej. Możemy jednak zauważyć, że w każdej chwili nasza decyzja zależy wyłącznie od aktualnego stanu gry, co sprawia, że naturalnym modelem opisującym problem jest proces decyzyjny Markowa.

Celem pracy jest opracowanie modelu blackjaka oraz wyznaczenie strategii maksymalizującej oczekiwaną wartość wygranej gracza. By to osiągnąć, zdefiniujemy najpierw przestrzenie stanów i akcji, a także funkcje przejścia i nagrody. Następnie zastosujemy algorytm iteracji wartości do wyznaczenia optymalnej strategii. W kolejnych rozdziałach pracy przedstawimy przyjęte w modelu zasady gry w blackjaka, niezbędne matematyczne definicje, opis konstrukcji modelu i analizę wyników, z uwzględnieniem możliwości przyszłego rozwoju.

Przedstawienie problemu

Blackjack jest karcianą grą losową, w której każdy z graczy gra przeciwko krupierowi. Celem gry jest uzyskanie sumy punktów wyższej niż suma punktów krupiera, nie przekraczając jednocześnie 21 punktów. Gra rozpoczyna się od rozdania dwóch kart każdemu z graczy oraz krupierowi, przy czym tylko jedna z kart krupiera jest odsłonięta. Uczestnicy powiększają swój wynik poprzez dobieranie kart z wierzchu talii. Karty 2-10 mają wartość równą swojemu numerowi; król, dama oraz walet warte są 10. As interpretowany może być zarówno jako 1, jak i 11. Na początku każdej rundy gracz może podjąć decyzję dobrania kolejnej karty, co w konsekwencji podniesie jego wynik lub doprowadzi do natychmiastowej przegranej wskutek przekroczenia wartości 21. Gdy każdy z graczy zakończy proces dobierania, krupier odkrywa swoje karty oraz zaczyna dobierać nowe do momentu uzyskania wartości co najmniej 17. Suma wartości kart krupiera również nie może przekroczyć 21; skutkowałoby to wygraną wszystkich graczy. Gdy zakończy, punkty każdego gracza z osobna zostają porównane z wynikiem krupiera i gracze o większej liczbie punktów niż krupier wygrywają.

Można zauważyć losowość rozważanej gry oraz podział na dyskretne tury, w których podejmowane są decyzje. Przy pewnych założeniach struktura gry posiada własność Markowa, co pozwala nam na opisanie jej za pomocą procesu decyzyjnego Markowa. Najpierw jednak przypomnijmy najważniejsze matematyczne definicje potrzebne do skonstruowania modelu.

Wprowadzenie teoretyczne

Łańcuch Markowa

Ciąg zmiennych losowych $(X_n)_{n=0}^{\infty}$ o wartościach w przeliczanym zbiorze S (przestrzeni stanów), określony na przestrzeni probabilistycznej (Ω, Σ, P) , nazywamy **łańcuchem Markowa** wtedy i tylko wtedy, gdy dla każdego $n \in \mathbb{N}$ i każdego ciągu $s_0, s_1, \dots, s_n \in S$ mamy

$$P(X_n = s_n \mid X_{n-1} = s_{n-1}, \dots, X_1 = s_1, X_0 = s_0) = P(X_n = s_n \mid X_{n-1} = s_{n-1}),$$

przy założeniu, że $P(X_{n-1} = s_{n-1}, \dots, X_1 = s_1, X_0 = s_0) > 0$. Powyższą własność nazywamy **własnością Markowa** [1].

Intuicyjnie oznacza ona, że stan procesu w momencie n jest zależny tylko od stanu w momencie $n - 1$. Innymi słowy, ciąg spełniający własność Markowa nie zależy od całej swojej przeszłości, a tylko od poprzedniego momentu.

Proces decyzyjny Markowa

Niech (Ω, Σ, P) będzie przestrzenią probabilistyczną. **Proces decyzyjny Markowa** definiujemy jako piątkę

$$(X, A, A(x), t, r),$$

gdzie X to przestrzeń możliwych stanów określonych na przestrzeni (Ω, Σ, P) , A jest przestrzenią możliwych decyzji, $A(x)$ to zbiór decyzji podejmowalnych w stanie $x \in X$, ponadto

$$\begin{aligned} t: X \times X \times A \ni (x, x', a) &\rightarrow P(x' \mid x, a) \in [0, 1], \\ r: X \times X \times A \ni (x, x', a) &\rightarrow r(x, x', a) \in \mathbb{R} \end{aligned}$$

są odpowiednio funkcją przejścia ze stanu x do stanu x' za pomocą akcji a oraz funkcją nagrody za powyższe przejście.

W stanie $x \in X$, dostępnymi akcjami są elementy zbioru $A(x)$. Po wybraniu akcji $a \in A(x)$, przejście do kolejnego stanu zadane jest rozkładem prawdopodobieństwa $P(\cdot \mid x, a)$ [2]. Nagrody są sposobem oceny użyteczności akcji w danym stanie. Przejścia dla kolejnych stanów gry po przyjęciu strategii możemy interpretować jako łańcuchy Markowa, w których przechodzimy pomiędzy stanami z określonym prawdopodobieństwem, przy założeniu, że stan w chwili $\tau \in \mathbb{N}$ jest zależny tylko od stanu w chwili $\tau - 1$. Oznacza to, że nasz model spełnia własność Markowa. Celem rozważania procesu decyzyjnego Markowa może być wyznaczenie optymalnej strategii π , gdzie $\pi(x)$ jest optymalną decyzją, którą należy podjąć w stanie x , aby zmaksymalizować wartość oczekiwaną sumy zdyskontowanych nagród. Zagadnienie to nazywać będziemy **problemem decyzyjnym Markowa**.

Konstrukcja modelu

Grę w blackjacka dla skończonej talii kart modelować możemy za pomocą rozkładu hipergeometrycznego. Jeżeli w talii znajduje się N kart, w tym M wyróżnionych, to prawdopodobieństwo wylosowania dokładnie m takich kart w n próbach wynosi

$$P(m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}.$$

W rzeczywistości zwykle używa się od sześciu do ośmiu talii kart. Wiemy jednak, że jeśli $\lim_{N \rightarrow \infty} \frac{M}{N} = p \in (0, 1)$, to rozkład hipergeometryczny jest zbieżny do rozkładu dwumianowego:

$$\lim_{N \rightarrow \infty} P(m) = C_n^m p^m (1-p)^{n-m}.$$

Oznacza to, że dla dużych populacji losowanie bez zwracania można z dobrym przybliżeniem zastąpić losowaniem ze zwracaniem.

Niech N oznacza łączną liczbę kart używanych w danej rozgrywce. Wówczas $M = \frac{N}{13}$ określa liczbą kart o numerze wartości 1-9 w talii. Zwiększając liczbę talii kart, którymi gramy, otrzymujemy

$$\lim_{N \rightarrow \infty} \frac{M}{N} = \lim_{N \rightarrow \infty} \frac{N}{13N} = \lim_{N \rightarrow \infty} \frac{1}{13} = \frac{1}{13} \in (0,1).$$

Dla kart o wartości dziesięć otrzymujemy wynik $\frac{4}{13}$. Założenia powyższego twierdzenia są zatem asymptotycznie spełnione. Stąd w naszym modelu przyjmujemy założenie nieskończonej liczby kart w talii. Liczba kart w typowej grze w kasynie jest na tyle duża, że błąd, który uzyskujemy z przybliżenia, jest niewielki. Przykładowo, błąd względny dla jednego przejścia między przybliżeniem a prawdziwym prawdopodobieństwem wyciągnięcia karty o wartości 10 w drugiej turze, zakładając, że została taka już uprzednio wyciągnięta, wynosi

$$\left| \frac{16}{52} - \frac{8 \cdot 16 - 1}{52 \cdot 8 - 1} \right| \cdot \frac{52 \cdot 8 - 1}{8 \cdot 16 - 1} \approx 0,5\%.$$

Oczywiście, błąd ten rośnie proporcjonalnie do długości gry.

Przejdźmy teraz do opisu modelu. Niech stanami nieterminalnymi będą trójki (x_1, x_2, x_3) takie, że:

- $x_1 \in \{12, 13, \dots, 21\}$ reprezentuje wartość kart w dłoni gracza,
- $x_2 \in \{1, 2, \dots, 10\}$ reprezentuje wartość karty odsłoniętej przez krupiera,
- $x_3 \in \{0, 1\}$ określa, czy w ręce gracza znajduje się as o wartości 11.

Ograniczenie wartości x_1 do przedziału od 12 do 21 wynika z faktu, że dobieranie przy sumie punktów mniejszej niż 12 zawsze jest korzystne, ponieważ przekroczenie wartości 21 jest wówczas niemożliwe. Kluczowym w obraniu strategii jest określenie obecności asa w ręce gracza, gdyż ten może zredukować wynik o 10 punktów z uwagi na dwojaką naturę interpretacji wartości asa. Łącznie mamy zatem $10 \cdot 10 \cdot 2 = 200$ stanów nieterminalnych. Dodatkowo definiujemy zbiór stanów absorbujących $B = \{fura, wygrana, przegrana, remis\}$. Zbiór akcji dla każdego stanu nieterminalnego x ma postać $A(x) = \{dobranie, pas\}$. W prawdziwej grze dostępne są również takie akcje jak split oraz double down, jednak w celu uproszczenia modelu zostają one pominięte. Pomijany jest również aspekt obstawiania żetonów.

W przeciwieństwie do gracza, krupier nie może podejmować decyzji dotyczących dalszej gry. Zgodnie z zasadami blackjacka dobiera on kolejne karty do momentu osiągnięcia co najmniej 17 punktów. Dzięki temu możliwe jest wyliczenie za pomocą łańcuchów Markowa rozkładu prawdopodobieństwa jego wyniku końcowego, zależnego tylko od jego początkowo odkrytej karty. Rozważmy wektor, w którym każda współrzędna odpowiada prawdopodobieństwu osiągnięcia przez krupiera kolejnego z możliwych wyników. Niech ponadto A będzie macierzą przejścia pomiędzy kolejnymi stanami. Pamiętając, że maksymalną wartością osiągalną przez krupiera jest 26, rozkład prawdopodobieństwa dla wyniku po n -tym dobraniu obliczamy jako

$$[p_n(1), p_n(2), \dots, p_n(25), p_n(26)]A = [p_{n+1}(1), p_{n+1}(2), \dots, p_{n+1}(25), p_{n+1}(26)],$$

gdzie $p_n(m)$ oznacza prawdopodobieństwo otrzymania wyniku m w n -tej turze. Zgodnie z zasadami gry, po osiągnięciu wyniku co najmniej 17 punktów krupier kończy dobieranie kart. Dzięki temu, po dostatecznie wielu krokach, rozkład procesu odpowiada rozkładowi końcowego wyniku krupiera.

Oznaczmy teraz wynik gracza i krupiera jako odpowiednio g oraz k . Niech ponadto K_i oznacza zdarzenie odpowiadające krupierowi otrzymującemu i punktów. Korzystając z rozłączności zdarzeń, dla ustalonego $g < 22$ otrzymujemy

$$\begin{aligned} P(\text{gracz wygra}) &= P(g > k \vee k > 21) = P\left(\bigcup_{i=17}^{g-1} K_i \cup \bigcup_{i=22}^{26} K_i\right) = \\ &= \sum_{i=17}^{g-1} P(K_i) + \sum_{i=22}^{26} P(K_i). \end{aligned}$$

Podobnie obliczamy prawdopodobieństwa remisu oraz przegranej. W przypadku wybrania przez gracza akcji pas, model przechodzi do jednego ze stanów terminalnych: wygrana, przegrana lub remis, z prawdopodobieństwami wyznaczonymi na podstawie rozkładu końcowego wyniku krupiera. Funkcja nagrody przyjmuje odpowiednio wartości 1, -1 oraz 0 dla tych stanów terminalnych. Oczekiwana wartość nagrody jest następnie obliczana zgodnie z równaniem Bellmana [3]. Dla doboru liczone są prawdopodobieństwa przejść do kolejnych stanów. Z uwagi na założenie nieskończoności talii, prawdopodobieństwa doboru poszczególnych kart z talii są stałe i wynoszą:

$$\begin{aligned} P(\text{wyciągnięcie karty o wartości } 10) &= \frac{4}{13}, \\ P(\text{wyciągnięcie karty o wartości } i) &= \frac{1}{13}, \quad i \in \{1, 2, \dots, 9\}. \end{aligned}$$

Algorytm iteracji wartości

Algorytm iteracji wartości rozwiązuje problem decyzyjny Markowa, używając programowania dynamicznego. Iteracyjnie rozwiązuje on tzw. równania Bellmana dla stanu x , definiowane jako

$$U(x) = \max_{a \in A(x)} \left\{ r(x, x', a) + \gamma \sum_{x'} P(x' | x, a) U(x') \right\}.$$

Wartość $U(x)$ rozumieć możemy jako „użyteczność” stanu x w realizacji naszego celu. Wyznacza ona ważoną prawdopodobieństwem przydatność przyszłych decyzji. Parametr γ , nazywany współczynnikiem dyskontowania, określa preferencję naszego modelu względem długotrwałych nagród. Dla małych γ , drugi składnik sumy maleje w wartości bezwzględnej, co skutkuje faworyzacją nagród krótkoterminowych. Przyjmujemy $\gamma = 0,999$, aby ograniczyć wpływ dyskontowania i przybliżyć problem bez dyskontowania. Pseudokod algorytmu przedstawiony został poniżej [3]:

wejście Proces decyzyjny Markowa ze stanami X , akcjami $A(x)$, modelem przejść $P(x' | x, a)$, nagrodami $r(x, x', a)$; współczynnik dyskontowania γ , tolerancja błędu ϵ , wielkość błędu w danej iteracji δ

wyjście Funkcja użyteczności U

1: Połóż $U(x) \leftarrow 0$ oraz $U'(x) \leftarrow 0$ dla każdego $x \in X$

2: **powtarzaj**

3: $U \leftarrow U'$

4: $\delta \leftarrow 0$

5: **dla każdego** $x \in X$ **wykonaj**

- 6: $U'(x) \leftarrow \max_{a \in A(x)} \{r(x, x', a) + \gamma \sum_{x'} P(x' | x, a) U(x')\}$
 7: $\delta \leftarrow \max\{\delta, |U'(x) - U(x)|\}$
 8: **zakończ dla**
 9: **dopóki** $\delta > \epsilon(1 - \gamma)\gamma$
 10: **zwróć** U

Algorytm rozpoczyna się od dowolnej funkcji wartości $U(x)$, a następnie iteracyjnie aktualizuje jej wartości zgodnie z równaniami Bellmana, aż do zejścia poniżej założonego poziomu błędu. Po wykonaniu algorytmu iteracji wartości, optymalną strategię dla stanu x otrzymujemy wybierając tę akcję, która maksymalizuje oczekiwaną sumę zdyskontowanych nagród, czyli

$$\pi(x) = \arg \max_{a \in A(x)} \left\{ \sum_{x'} P(x' | x, a) U(x') \right\}.$$

Wyniki

Po zaimplementowaniu algorytmu otrzymaliśmy optymalną strategię przedstawioną poniżej.

Bez asa		Karta krupiera									
		2	3	4	5	6	7	8	9	10	A
Punkty gracza	12	H	H	S	S	S	H	H	H	H	H
	13	S	S	S	S	S	H	H	H	H	H
	14	S	S	S	S	S	H	H	H	H	H
	15	S	S	S	S	S	H	H	H	H	S
	16	S	S	S	S	S	H	H	H	H	S
	17	S	S	S	S	S	S	S	S	S	S
	18	S	S	S	S	S	S	S	S	S	S
	19	S	S	S	S	S	S	S	S	S	S
	20	S	S	S	S	S	S	S	S	S	S
	21	S	S	S	S	S	S	S	S	S	S

Tab. 1. Optymalna strategia w stanach bez asa.

Z asem		Karta krupiera									
		2	3	4	5	6	7	8	9	10	A
Punkty gracza	12	H	H	H	H	H	H	H	H	H	H
	13	H	H	H	H	H	H	H	H	H	H
	14	H	H	H	H	H	H	H	H	H	H
	15	H	H	H	H	H	H	H	H	H	H
	16	H	H	H	H	H	H	H	H	H	H
	17	H	H	H	H	H	H	H	H	H	H
	18	S	S	S	S	S	S	S	H	H	S
	19	S	S	S	S	S	S	S	S	S	S
	20	S	S	S	S	S	S	S	S	S	S
	21	S	S	S	S	S	S	S	S	S	S

Tab. 2. Optymalna strategia w stanach z asem.

Tabele 1 i 2 opisują maksymalizującą oczekiwaną wygraną decyzję w każdym z 200 nieterminalnych stanów modelu, gdzie H oznacza dobieranie, a S spasowanie. Są one zgodne z intuicją – dla niskiej sumy punktów preferowaną akcją jest dobranie, natomiast dla wyższych wartości dominującą strategią jest spasowanie. Uzyskana strategia zgadza się na 197 z 200 stanów z ogólnodostępnymi tabelami „strategii podstawowej” w grze blackjack. Jedyne rozbieżności występują dla stanów (15,1,0), (16,1,0) oraz (18,1,1) i wynikają z uwzględnienia tylko akcji dobierania i pasowania.

Podsumowanie

W pracy opracowaliśmy uproszczony model gry w blackjacka w oparciu o proces decyzyjny Markowa, uwzględniający 2 akcje i 204 stany, zależne od sumy punktów gracza, karty krupiera i obecności asa. W dalszej kolejności zaimplementowaliśmy algorytm iteracji wartości w celu uzyskania strategii maksymalizującej oczekiwaną wartość nagrody. Porównaliśmy wyniki modelu z tabelami strategii podstawowej, uzyskując zgodność na poziomie 98,5%. Jednocześnie należy podkreślić, że przyjęte przez nas uproszczenia ograniczają użyteczność otrzymanych rezultatów w prawdziwej rozgrywce i traktujemy je jako punkt wyjścia do dalszego rozwoju. Przedstawiony model jest ograniczony jedynie do dwóch akcji, używa pewnych uproszczeń obliczeniowych i nie bierze pod uwagę aspektu obstawiania żetonów. W celu poprawy modelu oraz lepszego odwzorowania rzeczywistej gry należałoby rozważyć dokładny model przejść oparty na hipergeometrycznym rozkładzie prawdopodobieństwa bez przybliżenia nieskończonej talii, a także zaimplementować akcje *double down* oraz *split*. Ponadto można rozważyć uwzględnienie wysokości zakładu jako części procesu decyzyjnego.

Literatura

- [1] Jakubowski J., Sztencel R.: Wstęp do teorii prawdopodobieństwa, SCRIPT, Warszawa, 2010
- [2] Feinberg E., Schwartz A.: Handbook of Markov Decision Processes, Kluwer Academic Publishers, 2002
- [3] Russell S., Norvig P.: Artificial Intelligence A Modern Approach, Prentice, 2010

KOD POD LUPĄ: ALGORYTMY W ROLI CYFROWYCH DETEKTYWÓW

Jędrzej NIEDZIELSKI, Nina ZUGAJ

Politechnika Łódzka, Łódź

Wstęp

Współczesna cyfrowa rzeczywistość opiera się na masowym przetwarzaniu danych tekstowych oraz kodu źródłowego, co niesie za sobą ryzyko nadużyć związanych z kradzieżą własności intelektualnej. Wykrywanie plagiatu stało się kluczowym elementem ochrony praw autorskich w dobie powszechnego dostępu do informacji. Plagiat w najprostszym ujęciu definiowany jest jako przypisywanie sobie cudzej twórczości jako własnej lub przywłaszczenie czyjegoś dzieła.

O ile dla człowieka intuicyjne rozpoznanie zapożyczeń, synonimów czy przeformułowań nie stanowi większego problemu, o tyle dla maszyn operujących wyłącznie na poziomie binarnym jest to zadanie niezwykle skomplikowane. Niniejszy rozdział przedstawia matematyczne i algorytmiczne metody identyfikacji plagiatu, stanowiące istotny element cyfrowej kryminalistyki – od analizy czysto sekwencyjnej po zaawansowane przestrzenie wektorowe. Są one omawiane najpierw na tekstach, a następnie przeniesione na kod.

Reprezentacja danych i przygotowanie tekstu

Z punktu widzenia procesora komputera, każdy analizowany tekst to jedynie sekwencja zer i jedynek. Wystarczy, że nieuczciwy autor podmieni jedno słowo na synonim, a struktura binarna pliku ulega drastycznej zmianie. Z tego powodu maszyna nie może porównywać tekstów bezpośrednio. Aby algorytmy mogły spełnić swoją rolę, dane tekstowe muszą zostać poddane wstępnej obróbce, na którą składają się:

- tokenizacja – proces polegający na pocięciu ciągłego tekstu na pojedyncze słowa (tokeny) przy jednoczesnym oczyszczeniu go ze znaków interpunkcyjnych,
- lematyzacja – zabieg polegający na sprowadzeniu słów do ich form podstawowych, co sprawia, że leksemy o wspólnym rdzeniu są traktowane jako to samo słowo. Dzięki temu algorytmy stają się niewrażliwe na odmianę gramatyczną.

Po zakończeniu tych operacji reprezentacja literowa przestaje mieć znaczenie, komputer przypisuje każdemu unikalnemu słowu identyfikator liczbowy i operuje wyłącznie na strukturach numerycznych. Ze względu na to, że przeszukiwanie tradycyjnych list indeksowanych cechuje się nieefektywną złożonością obliczeniową $O(n)$, standardem staje się stosowanie tablic mieszających¹, w których kluczami są unikalne, stałe wartości wyliczone przez funkcję haszującą. Pozwala to na bezpośrednie wyznaczanie adresów pamięci i optymalizację czasu dostępu.

Algorytmy analizy sekwencyjnej

¹ Tablica mieszająca (ang. hash table) – struktura danych, która przyporządkowuje klucze do wartości za pomocą funkcji haszującej. Zapewnia ona teoretyczny czas dostępu do danych w klasie złożoności obliczeniowej $O(1)$, co drastycznie przyspiesza przeszukiwanie dużych zbiorów informacji.

Pierwszy etap detekcji kradzieży własności intelektualnej w warstwie tekstowej stanowią metody analizy sekwencyjnej. Koncentrują się one na matematycznej ocenie bliskości elementów słownikowych oraz ciągłości konfiguracji syntaktycznych.

N-gramy

Metoda N-gramów opiera się na analizie krótkich sekwencji sąsiadujących ze sobą słów poprzez przesuwanie nad tekstem stałego okna badawczego. Algorytm ten wykorzystuje matematyczny aparat łańcuchów Markowa. Modele te charakteryzują się „krótką pamięcią” – stan przyszły zależy wyłącznie od stanu bezpośrednio go poprzedzającego, a cała wcześniejsza historia systemu staje się bezużyteczna. Własność tę udowadnia formalne równanie łańcucha:

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_1 = x_1) = P(X_n = x_n | X_{n-1} = x_{n-1}),$$

gdzie:

X_n – zmienna losowa reprezentująca stan systemu w kroku n ,

x_n – konkretna wartość (stan), jaką przyjmuje zmienna losowa w kroku n ,

P – prawdopodobieństwo wystąpienia określonego zdarzenia warunkowego.

W kontekście lingwistycznym algorytm wyznacza prawdopodobieństwo warunkowe pojawienia się konkretnego słowa (w_i) przy określonej szerokości okna pamięci określonej indeksem wzoru:

$$P(w_i | w_{i-1}, \dots, w_{i-n-1}),$$

gdzie:

w_i – aktualnie analizowane słowo w tekście,

$w_{i-1}, \dots, w_{i-n-1}$ – sekwencja słów bezpośrednio poprzedzających słowo w_i , stanowiąca kontekst o długości n .

Szerokość stosowanego okna determinuje głębokość analizowanej zależności wstecznej.

Najdłuższy Wspólny Podciąg – LCS

Jeśli plagiat modyfikuje tekst, dodając losowe przymiotniki lub usuwając co drugie słowo, metoda N-gramów zawodzi z powodu przerywania ciągłości badanych par czy trójek słów. W takich sytuacjach zastosowanie znajduje algorytm LCS (*Longest Common Subsequence*). Szuka on chronologicznego „szkieletu” tekstu – badane elementy nie muszą leżeć bezpośrednio obok siebie, kluczowe jest jedynie zachowanie ich pierwotnej kolejności.

Wykorzystując technikę programowania dynamicznego², algorytm rozwiązuje globalny problem poprzez podział na mniejsze podproblemy (porównywanie prefiksów) przy użyciu zależności rekurencyjnej:

$$LCS(i, j) = \begin{cases} LCS(i-1, j-1) + 1, & \text{dla } x_i = y_j \\ \max\{LCS(i-1, j), LCS(i, j-1)\} & \text{dla } x_i \neq y_j \end{cases}$$

gdzie:

² Programowanie dynamiczne to technika optymalizacji algorytmów, która polega na rozwiązywaniu złożonego problemu poprzez rozbicie go na mniejsze podproblemy.

$LCS(i, j)$ – długość najdłuższego wspólnego podciągu dla prefiksów ciągów o długościach odpowiednio i oraz j ,
 x_i – i -ty element (np. token lub znak) pierwszego porównywanego ciągu X ,
 y_j – j -ty element drugiego porównywanego ciągu Y .

Metryki odległości edycyjnej i teoria mnogości

W celu skutecznego wykrycia zaawansowanych manipulacji w szyku zdań czy strukturze tekstu należy odejść od prostej analizy liniowej. Do tego stosuje się podejście optymalizacyjne, które wykorzystuje matematyczne miary odmienności oraz teorię mnogości do oceny relacji między badanymi dokumentami.

Odległość Levenshteina

Gdy struktura zdań pozostaje zbliżona, lecz słowa są celowo podmieniane, skutecznym narzędziem staje się odległość Levenshteina. Metryka ta określa minimalną liczbę operacji edycyjnych takich jak wstawienie, usunięcie lub zamiana słowa – niezbędnych do przekształcenia jednego ciągu w drugi. Na poziomie zaawansowanej analizy porównawczej jednostkami mogą być całe słowa zamiast pojedynczych liter.

Dla zilustrowania działania metody rozważmy dwa przykładowe zdania:

- zdanie pierwsze (X): "To zwykle oszustwo"
- zdanie drugie (Y): "To jawne oszustwo"

Wyznaczenie odległości polega na obliczeniu minimalnego kosztu przekształcenia ciągu słów zdania X w zdanie Y .

Niech $lev(i, j)$ oznacza odległość Levenshteina pomiędzy i -elementowym prefiksem pierwszego zdania a j -elementowym prefiksem drugiego zdania. Wartość tę wyznacza się rekurencyjnie za pomocą metody programowania dynamicznego:

$$lev(i, j) = \begin{cases} i & \text{dla } j = 0 \\ j & \text{dla } i = 0 \\ \min \begin{cases} lev(i-1, j) + 1 \\ lev(i, j-1) + 1 \\ lev(i-1, j-1) + cost \end{cases} & \text{dla } i, j > 0 \end{cases}$$

Poszczególne wiersze wewnątrz funkcji $\min$ odpowiadają kosztom kolejnych operacji edycyjnych: usunięcia słowa ($lev(i-1, j) + 1$), jego wstawienia ($lev(i, j-1) + 1$) oraz zamiany ($lev(i-1, j-1) + cost$). Wartość zmiennej $cost$ wynosi 0 w przypadku identycznych słów ($x_i = y_j$) lub 1, gdy słowa się różnią ($x_i \neq y_j$).

Na podstawie powyższego wzoru konstrukcja macierzy kosztów przebiega następująco. Ruch z góry (wstawienie) oraz z lewej strony (usunięcie) nakładają na wynik twardą karę o wartości +1. Ruch po przekątnej (zamiana) generuje koszt +1 tylko wtedy, gdy słowa są różne; dla słów identycznych koszt wynosi 0. Symbol ε w macierzy oznacza ciąg pusty (brak słowa) i służy do reprezentacji warunków brzegowych (dla $i = 0$ lub $j = 0$).

	ε	To	jawne	oszustwo
ε	0	1	2	3
To	1	0	1	2
zwykłe	2	1	1	2
oszustwo	3	2	2	1

$lev(3,3)$
 $= \min(2 + 1, 2 + 1, 1 + 0)$
 $= 1$

Rys. 1. Wizualizacja procesu wyznaczania odległości Levenshteina dla zdań „To zwykłe oszustwo” oraz „To jawne oszustwo”.

Dla przykładu, wartość w prawej dolnej komórce $lev(3,3)$ reprezentuje ostateczny koszt przekształcenia pełnych zdań i wyliczana jest jako:

$$lev(3,3) = \min(lev(2,3) + 1, lev(3,2) + 1, lev(2,2) + cost)$$

Ponieważ trzecie słowa w obu zdaniach są identyczne („oszustwo” = „oszustwo”), koszt zamiany wynosi $cost = 0$. Podstawiając wartości z poprzednich komórek macierzy, otrzymujemy:

$$lev(3,3) = \min(2 + 1, 2 + 1, 1 + 0) = 1$$

Ostateczna wartość w prawym dolnym rogu macierzy stanowi jawny rachunek za dokonane modyfikacje – im mniejsza wartość, tym porównywane zdania są do siebie bardziej podobne.

Współczynnik Jaccarda

Zupełnie inną perspektywę oferuje analiza oparta na teorii zbiorów z wykorzystaniem współczynnika Jaccarda. W przeciwieństwie do odległości Levenshteina, podejście to traktuje dokumenty wyłącznie jako zbiór słów, całkowicie ignorując ich kolejność oraz krotność występowania. Współczynnik podobieństwa Jaccarda, oznaczany jako $J(A, B)$, definiuje się jako stosunek liczebności części wspólnej zbiorów do liczebności ich sumy:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

W celach analitycznych bezpośrednią miarę podobieństwa często przekształca się w formalną metrykę przestrzeni matematycznej. Służy do tego odległość Jaccarda $J(A, B)$, stanowiąca dopełnienie współczynnika do jedności [1]:

$$d_j(A, B) = 1 - J(A, B) = 1 - \frac{|A \cap B|}{|A \cup B|}$$

gdzie:

A, B – badane dokumenty reprezentowane jako zbiory unikalnych słów,

$J(A, B)$ – klasyczny współczynnik podobieństwa Jaccarda,

$d_j(A, B)$ – odległość Jaccarda pomiędzy dokumentami A i B .

Algorytm ten jest bezbłędny w wykrywaniu oszustw polegających na losowym przestawianiu całych akapitów lub zdań.

Analiza statystyczno-wektorowa – TF-IDF

Najbardziej zaawansowane systemy antyplagiatowe odchodzą od traktowania wszystkich słów w tekście na równi. Powszechne występujące spójniki (np. „i”, „oraz”) nie niosą ze sobą takiego samego ładunku informacyjnego, jak rzadko występujące słowa kluczowe. Filtrem statystycznym rozwiązującym ten problem jest TF-IDF. Waga danego słowa ($w_{i,j}$) zależy od częstości jego występowania w tekście ($tf_{i,j}$) oraz od odwrotnej częstości występowania tego słowa we wszystkich dokumentach bazy (idf_i), wyliczanej jako logarytm ze stosunku całkowitej liczby dokumentów (N) do liczby dokumentów zawierających to słowo (df_i) [2]:

$$w_{i,j} = tf_{i,j} \cdot \log \frac{N}{df_i},$$

gdzie:

N – całkowita liczba wszystkich dokumentów znajdujących się w bazie systemowej,

$tf_{i,j}$ – częstość występowania i -tego słowa w j -tym dokumencie,

df_i – liczba dokumentów w bazie, które zawierają w sobie przynajmniej jedno wystąpienie i -tego słowa,

$w_{i,j}$ – ostateczna waga statystyczna i -tego słowa w j -tym dokumencie.

Dzięki tej operacji lingwistyka ustępuje miejsca algebrze liniowej – każdy dokument staje się wektorem w wielowymiarowej przestrzeni Hilberta³. Aby wyeliminować wpływ długości dokumentu, systemy nie mierzą odległości euklidesowej, lecz obliczają cosinus kąta między wektorami, zwany w [4] podobieństwem cosinusowym:

$$\cos(\theta) = \frac{\vec{A} \circ \vec{B}}{\|\vec{A}\| \|\vec{B}\|},$$

gdzie:

$\vec{A}, \vec{B}$ – wektory cech reprezentujące porównywane dokumenty w wielowymiarowej przestrzeni Hilberta,

$\vec{A} \circ \vec{B}$ – iloczyn skalarny wektorów $\vec{A}$ oraz $\vec{B}$,

$\|\vec{A}\|, \|\vec{B}\|$ – normy (długości) poszczególnych wektorów w przestrzeni,

$\cos(\theta)$ – współczynnik podobieństwa cosinusowego pomiędzy dwoma wektorami.

Jeżeli wyliczony cosinus kąta θ zbliża się do jedności, to w przybliżeniu możemy uznać wektory za współliniowe i wówczas uzyskuje się jednoznaczny dowód na istnienie bliskiego podobieństwa semantycznego.

Zastosowanie algorytmów w analizie kodu źródłowego

³ W kontekście systemów n -wymiarowych, takich jak model wektorowy tekstu, przestrzeń Hilberta uogólnia pojęcie przestrzeni euklidesowej do dowolnej (również nieskończonej) liczby wymiarów, zachowując możliwość rygorystycznego obliczania odległości oraz iloczynu skalarnego między wektorami cech.

Praktyczne zastosowanie przedstawionych metod numerycznych ujawnia swoją specyfikę podczas bezpośredniej analizy struktury kodów źródłowych. Weryfikacja programistyczna wymaga zmagania się z unikalnymi cechami składniowymi. Drobne zmiany nazw zmiennych mogą całkowicie zmienić reprezentację tekstową, nie wpływając na działanie programu.

Wydajność algorytmów można zaobserwować na przykładzie analizy dwóch funkcji sprawdzających czy liczba jest pierwsza o średniej długości około 165 znaków, zaprezentowanych odpowiednio na listingu 1 oraz listingu 2.

```
def is_prime(n):
    if n <= 1:
        return False
    for i in range(2, int(n**0.5) + 1):
        if n % i == 0:
            return False
    return True
```

Lst. 1. Implementacja funkcji `is_prime` w języku Python.

```
def check_prime(number):
    if number <= 1:
        return False
    for j in range(2, int(number**0.5) + 1):
        if number % j == 0:
            return False
    return True
```

Lst. 2. Implementacja zmodyfikowanej funkcji `check_prime` w języku Python.

W sytuacji, gdy algorytm porównuje funkcję pierwotną, widoczną na listingu 1, z wersją zmodyfikowaną, widoczną na listingu 2, jedynie pod kątem identyfikatorów zmiennych i pętli (podmiana nazw takich jak `is_prime` na `check_prime` czy `n` na `number`), odległość Levenshteina wyznacza stałą wartość kosztu edycyjnego równą 27. Najdłuższy wspólny podciąg LCS identyfikuje w takim układzie 152 znaki wspólnego rdzenia instrukcji, a metoda N-gramów wskazuje stabilne podobieństwo sekwencyjne na poziomie 53%. Filtr statystyczny TF-IDF skutecznie izoluje wagi unikalnych nazw własnych, pomijając powtarzalne instrukcje sterujące.

W przypadku zestawienia struktur całkowicie odmiennych semantycznie – na przykład funkcji odwracania tekstów, przedstawionej na listingu 3, z algorytmem wyszukiwania wartości maksymalnej, zobrazowanym na listingu 4 – wskaźniki korelacji drastycznie maleją.

```
def reverse_string(text):
    reversed_text = ""
    for char in text:
        reversed_text = char + reversed_text
    return reversed_text
```

Lst. 3. Implementacja funkcji odwracania tekstu `reverse_string` w języku Python.

```
def find_maximum(elements):
    if not elements:
        return None
    max_val = elements[0]
    for item in elements:
        if item > max_val:
            max_val = item
    return max_val
```

Lst. 4. Implementacja funkcji wyszukiwania wartości maksymalnej `find_maximum` w języku Python.

Przy długościach operacyjnych rzędu 141 i 194 znaków, odległość Levenshteina pomiędzy kodem z listingu 3 a kodem z listingu 4, wzrasta do poziomu 128 jednostek edycyjnych, podciąg LCS drastycznie kurczy się do 76 znaków, a pokrycie metodą N-gramów osiąga marginalną wartość 2%.

Przetwarzanie tak zdefiniowanych danych w masowej skali systemów akademickich wiąże się z wysoką złożonością obliczeniową tradycyjnych algorytmów opartych na programowaniu dynamicznym. Generowanie dynamicznych macierzy porównawczych dla tysięcy plików źródłowych tworzy wąskie gardła wydajnościowe. Z tego powodu nowoczesne platformy antyplagiatowe odchodzą od interpretowanych środowisk skryptowych na rzecz natywnych rozwiązań implementowanych w językach kompilowanych, takich jak Go [5]. Wykorzystywanie zaawansowanych mechanizmów współbieżności, w tym goroutines – lekkich wątków w języku Go, umożliwia równoległe transformacje matematyczne, jednoczesne wyliczanie podobieństwa cosinusowego oraz szybką redukcję czasu przetwarzania strumieni danych w wielkich bazach kodów źródłowych [6].

Podsumowanie

W obszarach, w których maszynom brakuje ludzkiej intuicji niezbędnej do naturalnego zrozumienia tekstu, z pomocą przychodzi rygorystyczny aparat matematyczny. Przejście od struktur dyskretnych i kombinatoryki w algorytmach N-gramów, LCS czy odległości Levenshteina, poprzez teorię miary w zbiorach Jaccarda, aż do struktur algebraicznych w metodzie TF-IDF, pozwala na skuteczną walkę z nieuczciwością. Plagiat w świetle metod numerycznych nie jest subtelnym niuansiem językowym, lecz mierzalną anomalią statystyczną, z którą algorytmy radzą sobie doskonale.

Literatura

- [1] Jaccard P.: The Distribution of the Flora in the Alpine Zone, New Phytologist, vol. 11, no. 2, pp. 37-50, 1912
- [2] Salton G., Buckley C.: Term-weighting Approaches in Automatic Text Retrieval, Information Processing & Management, vol. 24, no. 5, pp. 513-523, 1988
- [3] Wagner R. A., Fischer M. J.: The String-to-String Correction Problem, Journal of the ACM, vol. 21, no. 1, pp. 168-173, 1974
- [4] Manning C. D., Raghavan P., Schütze H.: Introduction to Information Retrieval, Cambridge University Press, 2008
- [5] Lion D., Chiu A., Stumm M., Yuan D.: Investigating Managed Language Runtime Performance: Why JavaScript and Python are 8x and 29x Slower than C++, yet Java and Go can be Faster?, 2022 USENIX Annual Technical Conference (USENIX ATC 22), pp. 835-852, 2022
- [6] Malhotra A.: Concurrency Patterns in Golang: Real-World Use Cases and Performance Analysis, Journal of Computer Science and Technology Studies, vol. 7, no. 6, pp. 782-788, 2025

WERYFIKACJA DEZINFORMACJI Z UŻYCIEM TEORII DEMPSTERA-SHAFERA

Marcin KRÓL

Politechnika Krakowska im. Tadeusza Kościuszki, Kraków

Wstęp

Współczesny świat, zdominowany przez Internet i media cyfrowe, charakteryzuje się ogromnym przepływem informacji. Niestety, istotna część z tych treści ma charakter dezinformacyjny lub jest ukierunkowana na wywołanie sensacji. W takiej rzeczywistości kluczowe stają się nie tylko pozyskiwanie danych, ale przede wszystkim ocena ich wiarygodności. Jednym z podejść do tego problemu jest analiza dostępnych przesłanek (dowodów) oraz ich łączenie w celu oszacowania stopnia przekonania, co do prawdziwości danej informacji.

Podstawy teoretyczne

Rama rozróżnienia i funkcja podstawowego przyporządkowania prawdopodobieństwa

Modelem matematycznym stosowanym przy tej koncepcji jest teoria Dempstera-Shafera, stworzona początkowo przez A.P. Dempstera, a następnie rozwinięta przez G. Shafera w XX wieku. Za jej pomocą możemy wnioskować o pewnych hipotezach nawet w warunkach niepewności, wykorzystując tym samym w pełni wiedzę, którą posiadamy, do precyzyjnych wyników matematycznych a następnie do konkluzji dotyczących naszych badań lub obserwacji [1].

Podstawowym zagadnieniem w teorii D-S jest skończony zbiór Θ będący zbiorem hipotez, nazywany również ramą rozróżnienia. Elementy tego zbioru będą reprezentować pojedynczy wynik systemu lub jedną możliwą decyzję. Przy tej definicji należy zwrócić szczególną uwagę na to, że nie będziemy ograniczać się jedynie do pojedynczych elementów zbioru Θ , ale także do wszystkich jego podzbiorów, co będzie kluczowe przy uwzględnianiu w obliczeniach niepewności. Stanowi to fundamentalne rozróżnienie względem klasycznej teorii bayesowskiej. W tradycyjnym rachunku prawdopodobieństwa cała masa decyzyjna musi zostać rozdzielona między pojedyncze, rozłączne zdarzenia elementarne. Teoria Dempstera-Shafera przełamuje to ograniczenie, umożliwiając przypisywanie wag dowodowych bezpośrednio do całych podzbiorów ramy rozróżnienia [2].

Formalne przejście do analizy niepewności wymaga wprowadzenia narzędzia, które w sposób ilościowy pozwoli opisać siłę posiadanych poszlak. W tym celu zdefiniujemy funkcję $m: 2^\Theta \rightarrow [0,1]$ taką, że

- 1) $m(\emptyset) = 0$,
- 2) $\sum_{A \subseteq \Theta} m(A) = 1$.

Funkcję tę nazywamy funkcją podstawowego przyporządkowania prawdopodobieństwa (PPP) dla Θ . Przypisuje ona masę prawdopodobieństwa każdemu podzbiorowi zbioru Θ . Opisuje w ten sposób ile całkowitego dowodu wspiera zbiór A nie rozstrzygając wewnętrznie, który element w A jest prawdziwy. Wartości te są przypisywane na podstawie m.in. opinii eksperckich, badań statystycznych lub danych pomiarowych. Są one niezbędne dla dalszych obliczeń i analiz [3].

W celu zobrazowania wcześniejszych zależności, rozważmy przykład z kulą umieszczoną w pudełku. Osoba, która ją tam wkładała, posiada 20% przekonania, że była ona biała oraz 50% pewności, iż była czarna. Rozważmy zatem $\Theta = \{Czarna, Biała\}$. Wówczas rozkład podstawowego przyporządkowania prawdopodobieństwa hipotez przedstawia się następująco:

Hipoteza	Wartość $m(\cdot)$
Ani Biała, ani Czarna	0
Biała	0,2
Czarna	0,5
Biała lub Czarna	0,3

Tab. 1. Przykładowy rozkład wartości funkcji PPP. Źródło: opracowanie własne

Widać stąd, że suma wszystkich wartości funkcji jest równa 1. Dane te nie są jednak wystarczające, by wysnuwać wnioski. Aby to zrobić, konieczne staje się wprowadzenie kolejnych narzędzi opartych na funkcji podstawowego przyporządkowania prawdopodobieństwa oraz strukturze ramy rozróżnienia.

Miary niepewności: funkcje przekonania i wiarygodności

Pierwszym z omawianych narzędzi matematycznych jest funkcja $Bel: 2^\Theta \rightarrow [0,1]$ zdefiniowana za pomocą wzoru

$$Bel(A) = \sum_{B \subseteq A} m(B).$$

Nazywamy ją funkcją przekonania związaną z funkcją m . Miara ta pozwala określić łączny stopień pewności, jaki na podstawie posiadanych danych można przypisać do konkretnej hipotezy lub ich zbioru. Funkcja Bel dla danego podzbioru sumuje wartości PPP przypisane bezpośrednio do niego, jak również do wszystkich jego podzbiorów. Odzwierciedla ona zatem minimalny, całkowicie pewny dowód popierający dane stwierdzenie.

Drugim, dopełniającym narzędziem jest funkcja $Pl: 2^\Theta \rightarrow [0,1]$, zdefiniowana wzorem

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B).$$

Nazywamy ją funkcją wiarygodności odpowiadającą m . Opisuje ona maksymalny stopień przekonania, jaki mógłby zostać przypisany do danej hipotezy, gdyby pojawiły się nowe, sprzyjające jej dowody. Matematycznie funkcja ta sumuje wartości PPP wszystkich tych zbiorów, które mają chociażby jeden element wspólny z analizowaną hipotezą.

Zaprezentowana konstrukcja matematyczna obu miar jednoznacznie wskazuje, że funkcja przekonania skupia się na dowodach potwierdzających konkretny stan, podczas gdy funkcja

wiarygodności określa brak dowodów jemu przeciwnych. Każda z tych funkcji operuje na tym samym zbiorze wartości funkcji PPP.

Mając przypisane wartości podstawowego przyporządkowania prawdopodobieństwa dla poszczególnych hipotez, możemy obliczyć odpowiadające im wartości funkcji przekonania oraz wiarygodności. Na podstawie tego dostaniemy przedziały przekonania tych hipotez, będące przedziałami postaci $[Bel(A), Pl(A)]$. Ta wiedza umożliwi nam wysnuwanie wstępnych wniosków o prawdziwości lub fałszywości hipotez [4].

Dla pełnego zobrazowania omówionych zależności, rozważmy teraz zbiór hipotez $\theta = \{A, B, C\}$. W celu wyznaczenia granic niepewności, określono na początek wartości funkcji podstawowego przyporządkowania prawdopodobieństwa. Następnie, zgodnie ze wzorami, wyznaczone zostały wartości funkcji przekonania i wiarygodności. Wyniki tych obliczeń przedstawia tabela 2.

Hipoteza	$m(\cdot)$	$Bel(\cdot)$	$Pl(\cdot)$
$\emptyset$	0	0	0
$\{A\}$	0,3	0,3	0,62
$\{B\}$	0,2	0,2	0,53
$\{C\}$	0,1	0,1	0,35
$\{A, B\}$	0,15	0,65	0,9
$\{B, C\}$	0,08	0,38	0,7
$\{A, C\}$	0,07	0,47	0,8
$\{A, B, C\}$	0,1	1	1

Tab. 2. Rozkład wartości funkcji PPP, Bel i Pl dla hipotez. Źródło: opracowanie własne

Analiza danych zestawionych w tabeli 2 pozwala na sformułowanie istotnych wniosków dotyczących struktury niepewności rozważanego problemu. Szczególną uwagę należy zwrócić na przedział przekonania hipotezy C . Dolna granica (pełny dowód) wynosi zaledwie 0,1. Z kolei górna granica (wiarygodność) osiągnęła wartość 0,35. Wartości te są bezpośrednim wynikiem agregacji odpowiednich wielkości funkcji podstawowego przyporządkowania prawdopodobieństwa. Uzyskany w ten sposób przedział $[0,1; 0,35]$ sugeruje, że hipoteza ta charakteryzuje się niskim poparciem dowodowym, co w kontekście decyzyjnym przemawia za jej nieprawdziwością. Jest on bowiem odległy od 1, choć jego długość, równy 0,25, oznacza operowanie na umiarkowanym poziomie niepewności. Im większą wiedzą posługujemy się, licząc wartości funkcji Bel oraz Pl , tym węższe przedziały przekonania otrzymujemy. Jednak pozyskując dane tylko z jednego źródła, długości owych przedziałów niekiedy będą niesatysfakcjonujące przy interpretacji. Przykładem takiego stanu są hipotezy A i B , w których przedziały przekonania okazują się jeszcze szersze niż dla hipotezy C . Dla A jest on dodatkowo skoncentrowany wokół wartości 0,5. Takie rozlokowanie niepewności uniemożliwia poprawną diagnozę tej hipotezy [5]. Z kolei w ujęciu alternatywnym, rozważając łączną hipotezę $\{A, B\}$, zauważamy zupełnie inną strukturę poszlak. Jest ona bowiem silnie wspierana przez

zgromadzone dowody, co oznacza, że pomimo dużej niepewności do poszczególnych wariantów z osobna, zgromadzone dane potwierdzają prawdziwość którejś z tych opcji. Ogólne zasady interpretacji wskazują, że im krótszy jest przedział przekonania, tym wyższa jest precyzja i pewność wnioskowania. Szerokie przedziały wskazują na dominację niedookreśloności i wymagają dostarczenia dodatkowych informacji źródłowych. Jeśli dla pewnej hipotezy dolna granica wiarygodności Bel przekracza 0,5, oznacza to, że ponad połowa pełnej puli mas dowodowych wspiera tę samą hipotezę lub jej podzbiory. Daje to silną przewagę nad resztą hipotez, gdyż funkcja przekonania Bel mierzy sumę bezpośrednich dowodów wspierających dane zdarzenie. W przypadku, gdy górna granica wiarygodności Pl jest niższa od 0,5, to nawet przy uwzględnieniu całej niedookreślonej puli dowodów, maksymalny możliwy poziom poparcia dla hipotezy nie przekracza połowy wszystkich poszlak. Wówczas większość przesłanek wskazuje na odrzucenie tej opcji.

Reguła kombinacji Dempstera

Matematyczny mechanizm fuzji niezależnych dowodów

Naturalną konsekwencją wcześniejszej teorii jest pytanie, czy jedno źródło informacji jest wystarczające, by wnioski były wiarygodne? Zazwyczaj jedna perspektywa, zwłaszcza subiektywna (jak w pierwszym przykładzie), to za mało. Aby zawęzić przedziały przekonania i uzyskać rzetelniejszy wynik, konieczne staje się dodanie kolejnych, niezależnych obserwacji. Teoria Dempstera-Shafera znajduje szerokie zastosowanie w takich procesach, oferując zaawansowany mechanizm fuzji informacji. Służy do tego reguła Dempstera, umożliwiającą łączenie ze sobą wartości funkcji podstawowego przyporządkowania prawdopodobieństwa m_i z różnych źródeł, a odnoszące się do tego samego zbioru hipotez. Dzięki temu, zamiast polegać na jednym pomiarze, budujemy spójny obraz oparty na wszystkich dostępnych dowodach.

Dla danych funkcji PPP m_1 oraz m_2 funkcję $m_1 \oplus m_2 \rightarrow [0,1]$, definiujemy za pomocą równości

$$(m_1 \oplus m_2)(C) = \frac{\sum_{A \cap B = C} m_1(A)m_2(B)}{1 - \sum_{A \cap B = \emptyset} m_1(A)m_2(B)}$$

dla $C \neq \emptyset$ oraz $(m_1 \oplus m_2)(\emptyset) = 0$.

Kluczowym elementem powyższego sformułowania jest mianownik, w którym występuje wyrażenie sumaryczne oznaczające tzw. współczynnik konfliktu (często oznaczany symbolem K). Wartość ta reprezentuje sumę iloczynów mas tych podzbiorów, których przecięcie jest zbiorem pustym, co w praktyce oznacza sytuację, w której analizowane źródła informacji podają całkowicie sprzeczne komunikaty. Poprzez odjęcie tej wartości od jedności, reguła Dempstera dokonuje operacji normalizacji. Dzięki temu procesowi sprzeczne dowody zostają odrzucone, a pozostałe wartości funkcji podstawowego przyporządkowania prawdopodobieństwa są proporcjonalnie rozdzielane między te hipotezy, co do których źródła wykazują choćby częściową zgodność. Taka konstrukcja zapewnia stabilność systemu decyzyjnego, uniemożliwiając paraliż decyzyjny w przypadku pozyskania częściowo zniekształconych danych.

Licznik reprezentuje nieznormalizowaną masę dowodową wspierającą hipotezę C . Powstaje on przez zsumowanie wszystkich iloczynów mas z poszczególnych źródeł, których przecięcie daje analizowany zbiór.

Zastosowanie w detekcji dezinformacji

Wykorzystajmy regułę kombinacji Dempstera do oceny wiarygodności rzekomego przemówienia polityka X. Rozważmy ramę rozróżnienia $\Theta = \{T, F, M\}$, gdzie T oznacza prawdziwość nagrania, F oznacza fałszywość, a M – lekką manipulację, czyli na przykład wyjęcie wypowiedzi z kontekstu. W procesie weryfikacji wykorzystamy niezależne źródła informacji.

Pierwszym źródłem będzie m_w oznaczające analizę wideo. Miało ono dobrą jakość, jednak zaobserwowano obecność podejrzanych artefaktów w tle, a ponadto wystąpiły lekkie przerwania. Kolejną obserwację stanowić będzie m_g , czyli badanie dźwięku. Tutaj głos w nagraniu zgadzał się z głosem polityka, jednak – podobnie jak poprzednio – zaobserwowano pewne nieciągłości.

Na podstawie gotowych rozkładów funkcji podstawowego przyporządkowania prawdopodobieństwa, wyznaczona zostanie wspólna wartość $m_1 = m_w \oplus m_g$, obliczona zgodnie z regułą kombinacji Dempstera. Wyniki obliczeń przedstawione zostały w tabeli 3.

Hipoteza	$m_w(\cdot)$	$m_g(\cdot)$	$m_1(\cdot)$
$\emptyset$	0	0	0
$\{T\}$	0,3	0,4	0,5088
$\{F\}$	0,15	0,1	0,0986
$\{M\}$	0,25	0,25	0,3251
$\{T, F\}$	0,07	0,05	0,0130
$\{T, M\}$	0,1	0,15	0,0431
$\{F, M\}$	0,08	0,03	0,0097
$\{T, F, M\}$	0,05	0,02	0,0017

Tab. 3. Użycie reguły kombinacji Dempstera przy dwóch pierwszych obserwacjach.

Wstępne obserwacje wskazują na dominację hipotezy T . W celu zilustrowania mechanizmu działania reguły kombinacji Dempstera, w tabeli 4 przedstawiono szczegółowe zestawienie par podzbiorów i obliczenia prowadzące do wyznaczenia nieznormalizowanej masy dla hipotezy T .

Podzbiór z rozkładu $m_w(\cdot)$	Podzbiór z rozkładu $m_g(\cdot)$	Iloczyn mas
$\{T\}$ (0,3)	$\{T\}$ (0,4)	0,12
$\{T\}$ (0,3)	$\{T, F\}$ (0,05)	0,015
$\{T\}$ (0,3)	$\{T, M\}$ (0,15)	0,045
$\{T\}$ (0,3)	$\{T, F, M\}$ (0,02)	0,006
$\{T, F\}$ (0,07)	$\{T\}$ (0,4)	0,028

$\{T, F\} (0,07)$	$\{T, M\} (0,15)$	0,0105
$\{T, M\} (0,1)$	$\{T\} (0,4)$	0,04
$\{T, M\} (0,1)$	$\{T, F\} (0,05)$	0,005
$\{T, F, M\} (0,05)$	$\{T\} (0,4)$	0,02

Tab. 4. Szczegółowe zestawienie iloczynów mas dla hipotezy T .

Suma iloczynów mas dla podzbiorów, których częścią wspólną jest rozważana hipoteza, wynosi 0,2895. Współczynnik konfliktu $K = 0,431$ reprezentuje sumę iloczynów mas wszystkich par podzbiorów, których przecięcie daje zbiór pusty. Wartość po normalizacji, uwzględniająca odjęcie tej wartości od jedności w mianowniku, wynosi ostateczne

$$m_1(T) = \frac{0,2895}{1-0,431} \approx 0,5088.$$

W dalszych badaniach uwzględnimy jednak kolejną obserwację m_t , która oznaczać będzie analizę mimiki twarzy. W nagraniu ruchu ust oraz mina polityka nie zgadzała się z jego głosem, co sugeruje fałszerstwo lub manipulację nagraniem. Pozostaje dokonać kolejnej kombinacji rozkładu m_1 z nową obserwacją. Niech $m = m_1 \oplus m_t$. W tabeli 5 przedstawione zostały wyniki obliczeń.

Hipoteza	$m_1 (\cdot)$	$m_t (\cdot)$	$m (\cdot)$
$\emptyset$	0	0	0
$\{T\}$	0,5088	0,05	0,3139
$\{F\}$	0,0986	0,4	0,199
$\{M\}$	0,3251	0,26	0,4659
$\{T, F\}$	0,0130	0,07	0,0045
$\{T, M\}$	0,0431	0,05	0,0116
$\{F, M\}$	0,0097	0,12	0,0049
$\{T, F, M\}$	0,0017	0,05	0,0002

Tab.5. Ostateczne użycie reguły kombinacji Dempstera.

Mając gotowy pełen rozkład funkcji podstawowego przyporządkowania prawdopodobieństwa m , ostatnim krokiem będzie wyznaczenie wartości funkcji przekonania i wiarygodności. Wyniki tych obliczeń zostały przedstawione w tabeli 6.

Hipoteza	$m(\cdot)$	$Bel(\cdot)$	$Pl(\cdot)$
$\emptyset$	0	0	0
$\{T\}$	0,3139	0,3139	0,3302
$\{F\}$	0,199	0,199	0,2086
$\{M\}$	0,4659	0,4659	0,4826
$\{T, F\}$	0,0045	0,5174	0,5341
$\{T, M\}$	0,0116	0,7914	0,801
$\{F, M\}$	0,0049	0,6698	0,6861
$\{T, F, M\}$	0,0002	1	1

Tab. 6. Obliczone przedziały przekonania dla hipotez.

Po uwzględnieniu trzeciej obserwacji możemy stwierdzić, że dominującą hipotezą stało się M z bardzo wąskim przedziałem przekonania wynoszącym około $[0,47; 0,48]$. Oznacza to niską niepewność we wnioskowaniu. Hipoteza F , mówiąca o fałszywości nagrania, urosła dwukrotnie, choć nadal jest słabo wspierana dowodowo. Z kolei hipoteza $\{T, M\}$ z przedziałem przekonania ok. $[0,79; 0,8]$ zostaje najbardziej wiarygodną. Wobec tego wnioskujemy, że o ile nagranie nie zostało sfalszowane, to najpewniej doszło do lekkiej manipulacji treścią, mimo iż samo nagranie jest prawdziwe.

Literatura

- [1] Lu Y.-J., He J.: Dempster-Shafer Evidence Theory and Study of Some Key Problems, Journal of Electronic Science and Technology, Vol. 15, No. 1, March 2017
- [2] Yadav V., Kaushik A., Verma R.: Dempster Shafer Theory, IJRDO – Journal of Mechanical and Civil Engineering, Vol. 1, No. 5, May 2015
- [3] Kohlas J., Monney P.-A.: A Mathematical Theory of Hints: An Approach to the Dempster-Shafer Theory of Evidence, 1995
- [4] Shafer G.: A Mathematical Theory of Evidence, Princeton University Press, 1976
- [5] Dempster A. P.: Upper and lower probabilities induced by a multivalued mapping, The Annals of Mathematical Statistics, 1967

TOPOLOGICZNA DETEKCJA BIFURKACJI W MODELU TERAPII ADAPTACYJNEJ NOWOTWORÓW

Hanna KLUS

Politechnika Gdańska, Gdańsk

Wstęp

Żyjemy w czasach niesamowitego postępu technologicznego i medycznego. Terapie nowotworowe nigdy wcześniej nie były skuteczniejsze i bardziej dostępne. Toczymy jednak nierówną walkę. Komórka rakowa, która wykazywała wrażliwość na leczenie może nagle się uodpornić, a nawet najnowocześniejsza nauka nie jest w stanie dokładnie przewidzieć tego zachowania. Konwencjonalnym sposobem dawkowania leku jest MTD (Maksymalna dawka tolerowana), metoda opierająca się na proporcjonalności podanego leku i uśmierconych komórek rakowych. Jej skuteczność natomiast drastycznie maleje w momencie, kiedy rak jest heterogeniczny, czyli składa się z kilku rodzajów komórek, a nie wszystkie wrażliwe są na lek. Jednym z rozwiązań jest terapia adaptacyjna zaproponowana w 2009 roku przez grupę w Moffitt Centre [1], pierwotnie dla raka prostaty. Podaż leku zależna jest od stężenia biomarkera PSA (prostate-specific antigen) w krwi i dawkowanie odbywa się cyklicznie. Efektem jest przekształcenie choroby nowotworowej w chorobę przewlekłą, która tylko minimalnie ogranicza pacjenta w codziennym życiu. Celem niniejszej pracy jest wykorzystanie topologicznej analizy danych do wczesnej detekcji utraty skuteczności terapii – nie tylko potwierdzenie faktu mutacji, ale sygnalizowanie zmiany zachowania układu zanim stanie się ona klinicznie widoczna.

Model matematyczny

Modelem matematycznym zastosowanym w niniejszej pracy jest układ czterech równań różniczkowych zaproponowany w publikacji [2]. Opisuje on zależności reprezentujących dynamikę trzech subpopulacji komórkowych (T^+ , T^- , T^P) oraz stężenie białka PSA będącego biomarkerem obecnym przy raku prostaty, który stanowi przybliżony wskaźnik wielkości guza:

$$\begin{aligned}\frac{dT^+}{dt} &= r_+ T^+ \left(1 - \frac{T^+ + \alpha_{12} T^P + \alpha_{13} T^-}{K_+(t)} \right), \\ \frac{dT^P}{dt} &= r_P T^P \left(1 - \frac{\alpha_{21} T^+ + T^P + \alpha_{23} T^-}{K_P} \right), \\ \frac{dT^-}{dt} &= r_- T^- \left(1 - \frac{\alpha_{31} T^+ + \alpha_{32} T^P + T^-}{K_-} \right), \\ \frac{d(PSA)}{dt} &= (T^+ + T^P + T^-) - \sigma \cdot PSA,\end{aligned}$$

gdzie poszczególne zmienne oznaczają:

T^+ , T^P , T^- – populacje komórek odpowiednio: zależnych od androgenów, produkujących testosteron oraz niezależnych (lekoopornych),

r_+, r_p, r_- – współczynniki tempa proliferacji dla poszczególnych typów komórek,
 $K_+(t), K_p, K_-$ – pojemność środowiska,
 α_{ij} – parametry macierzy Lotki-Volterra określające presję konkurencyjną, jaką populacja j wywiera na populację i ,
 σ – stała szybkości rozpadu biomarkera PSA.

Co istotne, $K_+(t)$ jest zmienną zależną od czasu – jej wartość ulega skokowej zmianie w zależności od tego, czy terapia jest włączona, czy wyłączona, co tworzy z układu system hybrydowy. Lek jest podawany, gdy stężenie PSA osiągnie 80% wartości wyjściowej i odstawiany w momencie spadku do 40% tej wartości, tworząc cykliczny protokół dawkowania sterowany biomarkerem. Zarówno parametry, jak i wartości początkowe zostały zaczerpnięte z oryginalnych materiałów udostępnionych wraz z badaniami klinicznymi w [2].

Metodologia

Twierdzenie Takensa-Starka

Niech M będzie zwartą, m wymiarową rozmaitością. Niech $D^r(M)$ oznacza przestrzeń dyfeomorfizmów klasy C^r na rozmaitości M , a $C^r(M, \mathbb{R})$ przestrzeń funkcji klasy C^r na M . Stan układu w czasie dyskretnym opisuje dyfeomorfizm $f \in D^r(M)$ oraz $\varphi \in C^r(M, \mathbb{R})$ jest funkcją obserwacji mierzalnej własności układu. Jeżeli wymiar zanurzenia $d \geq 2m + 1$, to dla $r \geq 2$ zbiór par (f, φ) , dla których odwzorowanie opóźnieniowe $\Phi_{f,\varphi}(x) = (\varphi(x), \varphi(f^\tau(x)), \dots, \varphi(f^{(d-1)\tau}(x)))$ jest zanurzeniem, stanowi podzbiór otwarty i gęsty $D^r(M) \times C^r(M, \mathbb{R})$.

Twierdzenie Takensa [3] gwarantuje, że zanurzenie szeregu czasowego w przestrzeni wymiaru d odtwarza topologię atraktora oryginalnego układu. Jednakże w analizowanym przypadku konieczne jest wykorzystanie stochastycznego rozszerzenia Starka [4]. Wynika to z faktu, że użyty model jest sterowany lekiem w sposób nieautonomiczny (skokowy). Zastosowany wektor zanurzenia konstruowany jest na podstawie szeregu czasowego stężenia PSA generowanego przez model. Umożliwia on odtworzenie topologii atraktora ukrytego układu komórkowego bez bezpośredniego dostępu do populacji:

$$x(t) = [PSA(t), PSA(t - \tau), \dots, PSA(t - (m - 1)\tau)],$$

gdzie τ to dobrane opóźnienie czasowe, m to wymiar przestrzeni fazowej.

Dwuwymiarowa projekcja portretu fazowego pozwala wizualnie potwierdzić obecność cykli w układzie, choć właściwa analiza prowadzona jest w wyższym wymiarze zanurzenia.

Homologia persystentna

Niech $X = \{x(t_1), \dots, x(t_N)\} \subseteq \mathbb{R}^m$ będzie chmurą punktów uzyskaną z zanurzenia Takensa. Dla parametru $\varepsilon \geq 0$ definiujemy kompleks Vietorisa-Rips $VR_\varepsilon(X)$ [5] jako kompleks symplecjalny, w którym sympleks $\{x_{\{i_0\}}, \dots, x_{\{i_k\}}\}$ należy do $VR_\varepsilon(X)$ wtedy i tylko wtedy, gdy $\|x_{\{i_p\}} - x_{\{i_q\}}\| \leq \varepsilon$ dla każdej pary p, q . Rodzina $\{VR_\varepsilon(X)\}_{\varepsilon \geq 0}$ tworzy filtrację: dla $\varepsilon_1 \leq \varepsilon_2$ zachodzi $VR_{\varepsilon_1}(X) \subseteq VR_{\varepsilon_2}(X)$. Dla każdego ε obliczamy pierwszą grupę homologii $H_1(VR_\varepsilon(X); \mathbb{Z}_2)$, liczonej ze współczynnikami w ciele $\mathbb{Z}_2$, której ranga $\beta_1(\varepsilon) = \text{rank } H_1(VR_\varepsilon(X))$ zlicza niezależne cykle obecne w kompleksie przy danym ε .

Każdej klasie homologii $\gamma \in H_1$ przypisujemy trwałość $p(\gamma) = d(\gamma) - b(\gamma)$, gdzie $b(\gamma)$ i $d(\gamma)$ oznaczają odpowiednio wartość ε , przy której klasa powstaje i znika w filtracji. W dalszej analizie

śledzimy trwałość najdłuższej żyjącej klasy $p^* = \max_{\gamma} p(\gamma)$, która stanowi wskaźnik obecności cyklu w zrekonstruowanej przestrzeni [6].

Sliding window

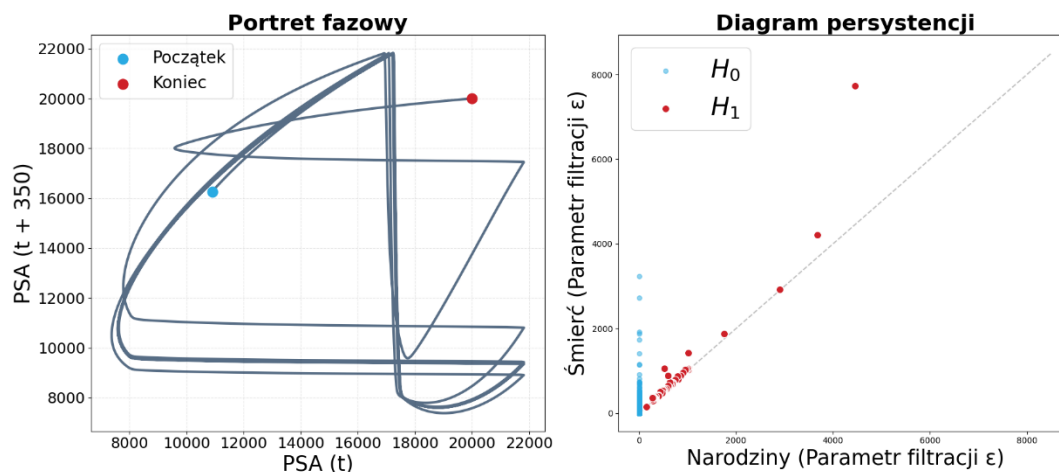
Obliczanie homologii z portretu fazowego całego układu nie dostarcza informacji o zmianach w czasie. Dlatego kluczowym narzędziem zastosowanym w tym celu jest technika okna przesuwającego (sliding window). Szereg czasowy dzielony jest na nakładające się okna o szerokości $W = 1800$ dni i kroku $s = 100$ dni. Wartości te dobrane zostały empirycznie, gwarantując dostateczną szerokość okna wykrywającą obecny cykl. Dla j -tego okna definiujemy chmurę punktów $X_j = \{x(t_i) : i \in [js, js + W]\}$ (przedział lewostronnie domknięty) i obliczamy $p_j^* = \max_{\gamma} p(\gamma)$ na kompleksie $VR_{\varepsilon}(X_j)$. Sygnał $p^*(t) = \{p_j^*\}$ interpretujemy jako wskaźnik bifurkacji. Gwałtowny spadek p^* do zera sygnalizuje zanik cyklu granicznego i utratę wrażliwości guza na terapię.

Obliczenia przeprowadzone zostały w języku Python za pomocą bibliotek Ripser [7] i Giotto-TDA [8], które posłużyły do wyznaczenia parametrów zanurzenia Takensa oraz obliczenia homologii persystentnej.

Wyniki

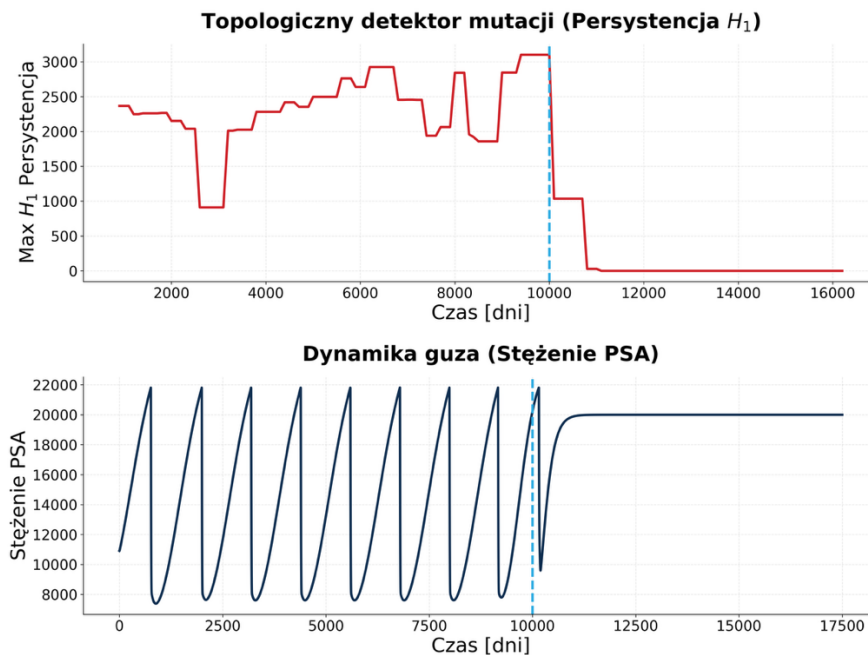
Analiza przeprowadzona została dla danych pacjenta nr 7 [2], wykazującego wielocykliczną odpowiedź na terapię, co umożliwia analizę długoterminowej dynamiki układu i detekcję bifurkacji przed utratą skuteczności leczenia.

Rysunek 1 przedstawia zrekonstruowany portret fazowy oraz diagram persystencji dowodzące obecność cykli w układzie.



Rys. 1. Portret fazowy zrekonstruowany metodą Takensa-Starka (z lewej strony) oraz odpowiadający mu diagram persystencji (z prawej strony). Punkt odległy od przekątnej wskazuje na obecność trwałego cyklu w odpowiednim wymiarze.

Rysunek 2 przedstawia wynik działania detektora topologicznego. Klasyczny wykres stężenia PSA pozostaje przez blisko 10000 dni pozornie stabilny – oscylacje nie wykazują żadnych anomalii dostrzegalnych metodami klinicznymi, a guz stabilizuje się na poziomie 20000, po nagłej ucieczce subpopulacji lekoopornej. Sygnał $p(t)$ rejestruje natomiast stopniową degradację trwałości najdłuższej żyjącej klasy H_1 już przed dniem 10000, a gwałtowny spadek p^* do zera w okolicach tego momentu wskazuje na zanik cyklu granicznego i bifurkację układu.



Rys. 2. Sygnał $p^*(t)$ maksymalnej trwałości H_1 (na górze) oraz stężenie PSA w czasie (na dole). Pionowa linia przerywana oznacza moment wykrytego zaniku cyklu – widoczny w sygnale topologicznym, nieobserwowalny w klasycznym sygnale $PSA(t)$.

Przedstawione wyniki wskazują, że sygnał $p(t)$ może stanowić czuły wskaźnik zbliżającej się bifurkacji, wykrywalnej w sygnale topologicznym wcześniej niż w klasycznym monitorowaniu PSA.

Wnioski

Zaproponowana metoda jest odporna na monotoniczną deformację sygnału i nie wymaga żmudnego doboru parametrów, co stanowi istotną zaletę w kontekście zaszumionych danych klinicznych. Głównym ograniczeniem podejścia jest zapotrzebowanie na gęste serie czasowe: pomiary PSA wykonywane co kilka tygodni nie pozwalają na pełną rekonstrukcję przestrzeni fazowej. Perspektywnym rozwiązaniem są metody interpolacji oraz nieinwazyjne urządzenia monitorujące biomarkery w sposób ciągły. Dalsze prace powinny objąć walidację metody na rzeczywistych danych klinicznych, porównanie z alternatywnymi metodami detekcji bifurkacji oraz rozszerzenie aparatu topologicznego o zigzag persistence – wariant homologii persystentnej pozwalający analizować sekwencje zmieniających się w czasie kompleksów symplecjalnych.

Literatura

- [1] Gatenby R.A., Silva A.S., Gillies R.J., Frieden B.R.: Adaptive therapy, *Cancer Res.*, vol. 69, no. 11, pp. 4894-903, 2009
- [2] Zhang J., Cunningham J.J., Brown J.S., et al.: Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer, *Nat. Commun.*, vol. 8, 1816, 2017
- [3] Takens F.: Detecting strange attractors in turbulence, *Lecture Notes in Mathematics*, vol. 898, pp. 366-381, Springer, 1981
- [4] Stark J., Broomhead D., Davies M., et al.: Delay Embeddings for Forced Systems. II. Stochastic Forcing, *J. Nonlinear Sci.*, vol. 13, pp. 519-577, 2003
- [5] Edelsbrunner H., Letscher D., Zomorodian A.: Topological persistence and simplification, *Discrete Comput. Geom.*, vol. 28, no. 4, pp. 511-533, 2002
- [6] Zomorodian A., Carlsson G.: Computing persistent homology, *Discrete Comput. Geom.*, vol. 33, pp. 249-274, 2005
- [7] Bauer U.: Ripser: efficient computation of Vietoris–Rips persistence barcodes, *J. Appl. Comput. Topology*, vol. 5, pp. 391-423, 2021
- [8] Tauzin G., Lupo U., Tunstall L., et al.: giotto-tda: A Topological Data Analysis Toolkit for Machine Learning and Data Exploration, *arXiv:2004.02551*, 2020

FROM EQUATION TO DOSE: THE MATHEMATICS BEHIND PERSONALIZED MEDICINE

Maria JĘDRZEJEWSKA

Politechnika Łódzka, Łódź

Introduction

This article presents the mathematical modelling behind the safety and effectiveness of drug use. These models provide a way to describe pharmacokinetic processes and predict how drug concentration changes over time. This makes them a useful tool for understanding the relationship between drug administration and its effects on the human body. The aim of this article is to show how such models can support appropriate dosing and contribute to the development of personalized medicine [1].

Basics of Pharmacokinetics

Pharmacokinetics is the branch of pharmacology that quantitatively describes the fate of a drug within a living organism following its administration. It encompasses four primary processes: absorption, distribution, metabolism, and excretion, collectively referred to as ADME.

Absorption: The process by which a therapeutic agent moves from its site of administration (e.g., oral, topical, or intravenous infusion) into the systemic circulation. The rate at which the drug enters the bloodstream is characterized by the absorption rate constant (K_a).

Distribution: The movement of drug molecules from the vascular space into various tissues and organs, which depends on cell membrane permeability and plasma protein binding affinity. The extent of distribution is quantified using the volume of distribution (V_d), a theoretical volume of fluid required to contain the total amount of drug at the same concentration as in the plasma.

Metabolism: The enzymatic biotransformation of the parent drug into active or inactive metabolites, primarily occurring in the liver.

Excretion: The irreversible elimination of the drug and its metabolites from the body, principally mediated via renal filtration into urine. The capacity of the body to eliminate the drug is defined by clearance (CL), which represents the volume of plasma completely cleared of the drug per unit of time [2].

Mathematical Models in Pharmacokinetics

Mathematical models are used in pharmacokinetics to simplify the complex processes that occur after a drug is administered. Since it is not possible to directly observe the movement of individual drug molecules throughout the body, mathematical equations can be used to describe changes in drug concentration over time. These models are based on assumptions about drug absorption, distribution, metabolism, and excretion and allow pharmacokinetic parameters to be estimated from experimental data [3].

One-compartment model

The one-compartment model is a simplified mathematical representation of the body, in which a drug is assumed to be instantaneously and uniformly distributed throughout a single compartment. This model does not represent the anatomical structure of the body but provides a simple way to describe changes in drug concentration over time. The amount of drug in the compartment can be expressed as:

$$A = V_d C,$$

where A is the amount of drug in the body, V_d is the apparent volume of distribution, and C is the drug concentration.

Assuming first-order elimination, the rate of drug elimination is proportional to the amount of drug present in the compartment. This relationship can be described by the following differential equation:

$$\frac{dA}{dt} = -kA,$$

where k is the elimination rate constant.

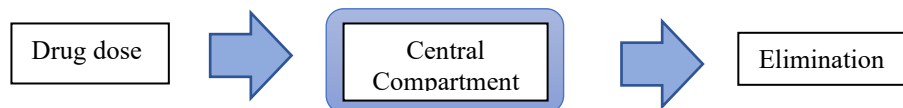


Fig.1. One-compartment model.

The differential equation can be solved by separating the variables:

$$\frac{1}{A} dA = -k dt.$$

Then, we integrate both sides which gives:

$$\ln A = -kt + C, \quad C \in \mathbb{R}.$$

Next, we apply the exponential function to both sides:

$$e^{\ln A} = e^{-kt+C}.$$

Therefore:

$$A = e^{-kt+C}.$$

Since e^C is a constant, we can write:

$$A = A_0 e^{-kt},$$

where A_0 is the initial amount of drug.

Using the relationship $A = V_d C$ and denoting $C_0 = \frac{A_0}{V_d}$ as the initial concentration, we can obtain the concentration as a function of time.

$$C(t) = C_0 e^{-kt}. \quad (1)$$

Equation (1) is the concentration-time equation for a one-compartment model with first-order elimination. It shows that the concentration decreases exponentially as time increases.

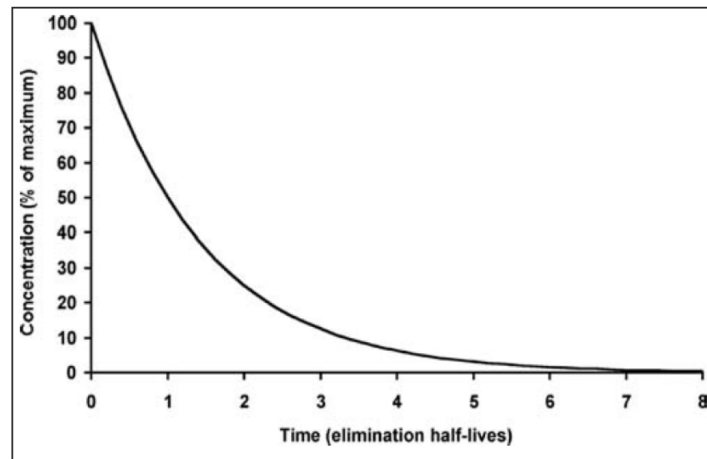


Fig.2. Drug concentration over time. Source: [4]

Half-Life

The half-life ($t_{1/2}$) is the time required for the concentration of a drug in the body to decrease by 50%. Setting $C(t_{1/2}) = \frac{1}{2} C_0$ in Equation (1) yields:

$$\frac{1}{2} C_0 = C_0 e^{-kt_{1/2}}.$$

Dividing both sides by C_0 :

$$\frac{1}{2} = e^{-kt_{1/2}}.$$

Taking the natural logarithm of both sides:

$$\ln\left(\frac{1}{2}\right) = -kt_{1/2} \Rightarrow -\ln 2 = -kt_{1/2}.$$

Solving for $t_{1/2}$ gives:

$$t_{1/2} = \frac{\ln 2}{k} \approx \frac{0.693}{k}.$$

The half-life is inversely proportional to the elimination rate constant. Therefore, a higher k results in a shorter half-life and faster drug elimination, while a lower k leads to a longer half-life. Half-life is an important pharmacokinetic parameter because it helps determine the appropriate dosing interval.

Therapeutic Window

The therapeutic window represents the range of drug concentrations between the minimum effective concentration (*MEC*) and the minimum toxic concentration *MTC*).

Below this range, the drug concentration may be insufficient to produce the desired therapeutic effect, whereas concentrations above it may increase the risk of toxicity [3]. Mathematical pharmacokinetic models can be used to predict changes in drug concentration over time and to determine an appropriate dose and dosing interval. This is particularly important for drugs with a narrow therapeutic window, for which relatively small changes in drug concentration can result in either insufficient therapeutic effect or toxicity.

For a drug undergoing first-order elimination, the time required for the concentration to decrease from the minimum toxic concentration to the minimum effective concentration, denoted by τ can be expressed as [3]:

$$\tau = \frac{1}{k} \ln \left(\frac{MTC}{MEC} \right),$$

where k is the elimination rate constant.

Repeated Dosing

When a drug is administered repeatedly, each new dose is added to the amount of drug remaining in the body from previous doses. This produces fluctuations between peak and trough concentrations. With regular dosing, the drug concentration gradually approaches a steady state.

The dosing interval is closely related to the drug's half-life, as it determines how much drug remains in the body when the next dose is administered. The aim of repeated dosing is to maintain drug concentrations within the therapeutic window, avoiding both ineffective concentrations and toxic levels.

Steady State

Steady state is reached when the average rate of drug administration is equal to the average rate of drug elimination. At this point, the average drug concentration remains constant over time, although peak and trough concentrations may still occur with repeated dosing.

For drugs following first-order elimination, steady state is typically reached after approximately 4–5 half-lives. The time required to reach steady state depends mainly on the drug's half-life and is generally independent of the dose.

At steady state, the average drug concentration can be described as:

$$C_{avg,ss} = \frac{F \cdot D}{CL \cdot \tau},$$

where F is bioavailability, D is the dose, CL is clearance, and τ is the dosing interval.

The maximum ($C_{max,ss}$) and minimum ($C_{min,ss}$) plasma concentrations at steady state are given by:

$$C_{max,ss} = \frac{C_0}{1 - e^{-k\tau}},$$

$$C_{min,ss} = C_{max,ss} e^{-k\tau}.$$

To ensure safety and efficacy, the dosing regimen should maintain drug concentrations within the therapeutic window:

$$MEC < C_{\min,ss} \leq C_{\max,ss} < MTC.$$

Maintaining these conditions helps achieve a consistent therapeutic effect while reducing the risk of toxicity.

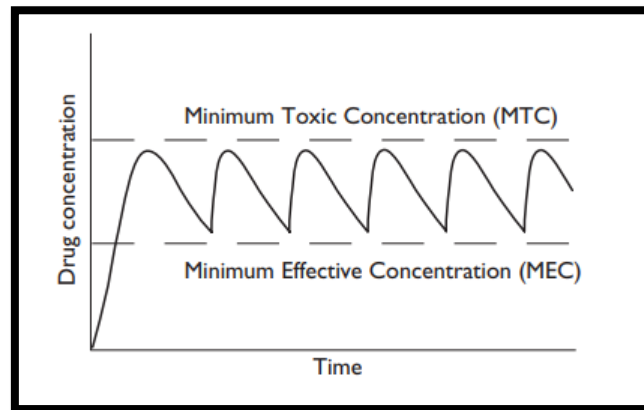


Fig.3. Therapeutic window and dosing interval. Source: [5]

Two-compartment model

The two-compartment model is used when drug distribution into tissues is not instantaneous and cannot be adequately described by a one-compartment model. It divides the body into a central compartment (blood and highly perfused tissues) and a peripheral compartment (slower distributing tissues).

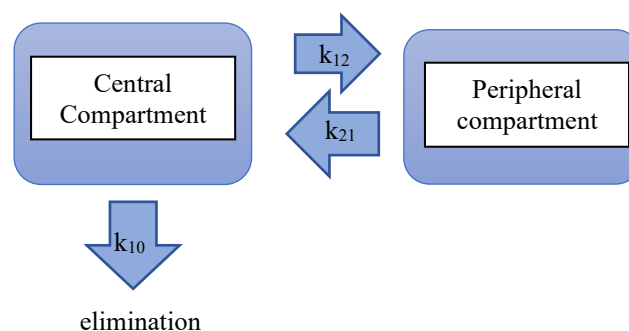


Fig.4. Two-compartment model.

For the Central Compartment:

$$\frac{dA_1}{dt} = k_{21}A_2 - k_{12}A_1 - k_{10}A_1.$$

For the Peripheral Compartment:

$$\frac{dA_2}{dt} = k_{12}A_1 - k_{21}A_2,$$

where

A_1, A_2 : amount of drug in the central and peripheral compartments, respectively,
 k_{12}, k_{21} : first-order transfer rate constants between central and peripheral compartments,
 k_{10} (or k_{el}): first-order elimination rate constant from the central compartment.

Solving this system for the amount of drug in the central compartment, $A_1(t)$, yields:

$$A(t) = Y_1 e^{-\alpha t} + Y_2 e^{-\beta t},$$

where

Y_1 and Y_2 are the mathematical intercepts for the distribution and elimination phases at $t = 0$.

The total amount of drug in the body at any given time is given by $A_{total}(t) = A_1(t) + A_2(t)$.

The constants α and β are hybrid rate constants defined as follows:

Alpha (α) describes the rapid initial decline in plasma concentration, governed primarily by drug transfer from the central to the peripheral compartment.

Beta (β) represents the slower decline corresponding to true drug elimination after a pseudo-equilibrium between compartments is established.

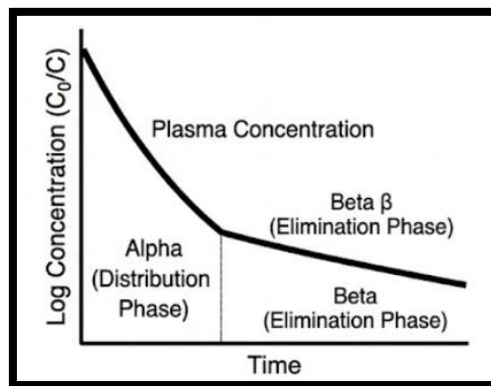


Fig.5. Two-compartment pharmacokinetic model showing plasma concentration over time after intravenous administration, featuring the alpha α distribution phase and beta β elimination phase. Source: [6]

Conclusions

Pharmacokinetic mathematical modelling plays a fundamental role in moving away from standard, "one-size-fits-all" dosing regimens towards the concept of personalized medicine. By replacing empirical dosage adjustments with deterministic equations, clinicians and pharmacologists can predict drug concentration profiles in plasma over time.

The single - and two-compartment models discussed in this article demonstrate how mathematical tools - such as ordinary differential equations, rate constants (k, k_{12}, k_{21}), and pharmacokinetic parameters ($t_{1/2}, CL, V_d$) - allow us to tailor therapy to individual physiological needs:

Optimizing the Therapeutic Window: Mathematical formulations can be used to determine the appropriate dosing interval (τ) and steady-state boundaries ($C_{avg,ss}$), minimizing the risk of adverse toxic effects while ensuring optimal drug efficacy.

Addressing Patient Variability: Individual patient characteristics (e.g., clearance impaired by renal function, variations in volume of distribution, or unique metabolic rates) can be directly integrated into these equations to adjust doses safely.

Model-Informed Precision Dosing (MIPD): Advanced mathematical frameworks serve as the foundation for modern MIPD computational tools, allowing clinicians to simulate drug accumulation under repeated dosing schedules prior to therapy onset.

In conclusion, mathematics bridges the gap between biological complexity and targeted pharmacological action. As personalized medicine continues to evolve, mathematical modelling remains an indispensable pillar for creating patient-specific dosing regimens that maximize therapeutic success and enhance patient safety.

Literature

- [1] Minichmayr I., et al.: Model-informed precision dosing: State of the art and future perspectives, *Advanced Drug Delivery Reviews*, vol. 215, 115421, 2024
- [2] Li Y., et al.: Current trends in drug metabolism and pharmacokinetics, *Acta Pharmaceutica Sinica B*, vol. 9, no. 6, pp. 1113-1144, 2019
- [3] Wijnand H. P.: Pharmacokinetic model equations for the one- and two-compartment models with first-order processes, *Journal of Pharmacokinetics and Biopharmaceutics*, vol. 16, no. 1, pp. 109–128, 1988
- [4] Thomson A.: Back to basics: pharmacokinetics, *The Pharmaceutical Journal*, vol. 272, pp. 769-771, 2004
- [5] Murphy J. E.: *Clinical Pharmacokinetics*, Blackwell Publishing, 2008
- [6] Rowland M., Tozer T. N.: *Clinical Pharmacokinetics and Pharmacodynamics: Concepts and Applications* (4th ed.), Lippincott Williams & Wilkins, 2011

JAK MATEMATYKA URATOWAŁA ZDJĘCIA JOWISZA?

Jan KOCIERZ

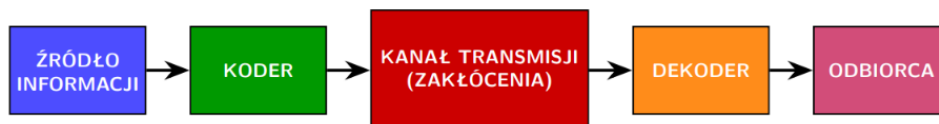
Politechnika Krakowska im. Tadeusza Kościuszki, Kraków

Wstęp

Algebraiczna teoria kodowania stanowi fundament niezawodnej komunikacji cyfrowej. W środowiskach o wysokim poziomie zakłóceń, takich jak przestrzeń kosmiczna, przesyłane sygnały narażone są na losowe błędy. Rozwiązaniem tego problemu są kody korekcyjne, które poprzez wprowadzenie precyzyjnej struktury algebraicznej i nadmiarowości pozwalają na odtworzenie pierwotnej informacji bez konieczności jej retransmisji. Sonda Voyager 1, która wyruszyła w przestrzeń kosmiczną 5 września 1977 roku, miała za zadanie przesłać pierwsze w historii precyzyjne zdjęcia Jowisza. Transmisja odbywała się przy użyciu nadajnika o mocy zaledwie 20 watów na dystansie setek milionów kilometrów. Niniejszy artykuł przedstawia, jak zastosowanie odpowiednich struktur algebraicznych uchroniło przesyłane dane przed bezpowrotnym uszkodzeniem.

Model kanału transmisyjnego

Zanim przejdziemy do bardziej szczegółowych rozważań dotyczących kodów korekcyjnych, należy zdefiniować środowisko w którym porusza się sygnał. Aby komputer mógł zinterpretować zdjęcie, musi oczywiście odebrać ciąg bitów. Proces przesyłania tej informacji z punktu A do punktu B można opisać za pomocą schematu blokowego kanału transmisji przedstawionego na rysunku 1.



Rys. 1. Schemat blokowy kanału transmisji. Opracowanie własne.

Sygnał generowany przez źródło informacji (w omawianym przez nas kontekście będzie to kamera sondy) trafia do kodera. Jego zadaniem jest wprowadzenie do wiadomości nadmiarowości, która zabezpieczy dane przed skutkami zakłóceń. Tak zakodowana informacja emitowana jest do kanału transmisyjnego, w którym podlega losowemu działaniu szumu.

Binarny kanał symetryczny

Do opisu zjawiska powstawania błędów w tego typu transmisji możemy wykorzystać model binarnego kanału symetrycznego. Kanał ten operuje na dwuelementowym alfabecie $\{0,1\}$. Zakładamy, że błędy dla poszczególnych bitów występują niezależnie od siebie z określonym

prawdopodobieństwem przekłamania p . Symetria kanału oznacza, że prawdopodobieństwo błędnej zamiany zera na jedynkę jest równe prawdopodobieństwu zamiany jedynki na zero, co możemy zapisać w postaci:

$$P(y = 1|x = 0) = P(y = 0|x = 1) = p,$$

gdzie x oznacza bit nadany, a y odebrany. Próba bezpośredniej transmisji surowych danych przez binarny kanał symetryczny bez zastosowania kodowania prowadziłaby do utraty informacji. Z tego względu zaszumiony sygnał z kanału trafia ostatecznie do dekodera, który wykorzystuje nałożoną w koderze strukturę algebraiczną w celu zrekonstruowania pierwotnej wiadomości.

Binarne kody liniowe i przestrzeń Hamminga

Rozważania nad algebraiczną teorią kodowania należy rozpocząć od zdefiniowania podstawowych struktur algebraicznych. Podstawową strukturą umożliwiającą korekcję błędów są binarne kody liniowe. Zakodowane wiadomości traktujemy jako wektory w n -wymiarowej przestrzeni wektorowej nad dwuelementowym ciałem $\mathbb{Z}_2 = \{0,1\}$.

Definicja 1. Binarnym kodem liniowym K o parametrach (n,k) nazywamy k -wymiarową podprzestrzeń liniową przestrzeni wektorowej $\mathbb{Z}_2^n$.

Elementy tej podprzestrzeni będziemy nazywać słowami kodowymi. Parametr n oznacza całkowitą długość słowa kodowego, natomiast k to długość właściwej wiadomości zakodowanej w tym słowie. Różnicę $r = n - k$ nazywamy nadmiarowością kodu.

Dowolny binarny kod (n,k) -liniowy można jednoznacznie opisać za pomocą jego macierzy generującej G o wymiarach $k \times n$. Wiersze macierzy G są liniowo niezależne i stanowią bazę podprzestrzeni K . Proces kodowania wiadomości $m \in \mathbb{Z}_2^k$ realizowany jest poprzez mnożenie $c = m \cdot G$, gdzie $c \in K$. Zastosowanie w tym miejscu podprzestrzeni liniowej zapewnia, że kombinacja liniowa dowolnych słów kodowych również jest słowem kodowym. W przypadku podprzestrzeni nad dwuelementowym ciałem $\mathbb{Z}_2$ jedynymi skalarami są 0 i 1. Oznacza to, że warunek ten sprowadza się do tego że jeśli $c_1, c_2 \in K$, to $c_1 + c_2 \in K$. Zobrazujemy to na przykładzie binarnego kodu $(3,2)$ -liniowego.

Przykład 1. Rozważmy liniową funkcję kodującą $f: \mathbb{Z}_2^2 \rightarrow \mathbb{Z}_2^3$, która do 2-bitowej wiadomości dopisuje jeden bit parzystości, tworząc 3-bitowe słowo kodowe:

$$f(00) = 000,$$

$$f(01) = 011,$$

$$f(10) = 101,$$

$$f(11) = 110.$$

Obraz tej funkcji jest naszym kodem liniowym K , czyli zbiorem wszystkich poprawnych słów kodowych. Mamy więc $K = \{000, 011, 101, 110\}$. Zgodnie z wcześniejszymi uwagami możemy łatwo zweryfikować, że K faktycznie stanowi podprzestrzeń liniową przestrzeni $\mathbb{Z}_2^3$. Zbiór ten zawiera wektor zerowy, a dodanie do siebie dowolnych dwóch niezerowych wektorów zawsze daje w wyniku trzeci niezerowy wektor należący do K :

$$101 + 110 = 011$$

$$011 + 110 = 101$$

$$011 + 101 = 110.$$

Ponieważ K to przestrzeń dwuwymiarowa, do jej rozpięcia potrzebujemy dwóch liniowo niezależnych wektorów, które utworzą jej bazę. Taką bazę mogą stanowić wektory 101 oraz 011, odpowiadające zakodowanym wiadomościom 10 oraz 01. Kod K możemy zapisać o wiele krócej, używając wyłącznie macierzy generującej G_K , w której wierszach umieszczamy wektory bazowe:

$$G_K = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}.$$

Mnożąc dowolny dwubitowy wektor przez macierz G_K , otrzymamy odpowiadające mu słowo kodowe. Przykładowo, jeśli chcemy zakodować wiadomość $m = [1,1]$ wykonujemy mnożenie wektora przez macierz w ciele $\mathbb{Z}_2$:

$$c = m \cdot G_K = [1,1] \cdot \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = [1,1,0]$$

Uzyskany wektor $c = [1,1,0]$ jest poprawnym słowem kodowym, co jednoznacznie pokrywa się z wartością zdefiniowanej funkcji $f(11) = 110$.

Aby jednak ocenić efektywność tak zdefiniowanej operacji kodowania w radzeniu sobie z zakłóceniami kanału, na przestrzeń $\mathbb{Z}_2^n$ należy nałożyć strukturę metryczną. W teorii kodowania kluczową rolę odgrywa w tym miejscu metryka Hamminga.

Definicja 2. Odległością Hamminga $d(x,y)$ dwóch wektorów $x, y \in \mathbb{Z}_2^n$ nazywamy liczbę współrzędnych, na których się one różnią, a więc

$$d(x,y) = |\{i : x_i \neq y_i\}|.$$

Definicja 3. Przestrzeń metryczną $(\mathbb{Z}_2^n, d)$ nazywamy przestrzenią Hamminga.

Dysponując tak określoną metryką, jesteśmy w stanie zdefiniować najważniejszy parametr każdego kodu korekcyjnego, który determinuje jego siłę i skuteczność w zaszumionym kanale.

Definicja 4. Rozstęp d^* kodu $K \subset \mathbb{Z}_2^n$ to najmniejsza odległość Hamminga między dwoma różnymi słowami tego kodu:

$$d^* = \min\{d(x,y) : x, y \in K, x \neq y\}.$$

Intuicyjnie, im większa jest wartość rozstępu d^* , tym kod jest bardziej odporny na zakłócenia. Większa odległość między poprawnymi słowami kodowymi oznacza bowiem, że potrzeba większej liczby przekłamań w kanale transmisyjnym aby jedno słowo kodowe zostało błędnie odebrane jako inne, wprowadzając dekodery w błąd.

Przykład 2. Obliczmy rozstęp dla analizowanego w przykładzie 1 kodu K . Sprawdzając odległość Hamminga między dowolnymi parami różnych słów kodowych (na przykład $d(000,011) = 2$, $d(011,101) = 2$, $d(000,110) = 2$), zauważamy, że zawsze różnią się one na dokładnie dwóch pozycjach. Zatem rozstęp kodu K wynosi $d^* = 2$.

W tym miejscu należy przywołać twierdzenie, które precyzuje zależność pomiędzy rozstępem kodu a jego zdolnością do naprawy uszkodzonych przy transmisji bitów.

Twierdzenie 1. Binarny kod liniowy o rozstępie d^* pozwala na skorygowanie co najwyżej t błędów powstałych w jednym słowie kodowym, jeśli spełniona jest nierówność:

$$d^* \geq 2t + 1 \quad [1].$$

Maksymalną wartość parametru t spełniającą tę nierówność nazywamy zdolnością korekcyjną danego kodu. Zależność z twierdzenia 1 najłatwiej objaśnić, odwołując się do geometrii przestrzeni Hamminga. Wprowadźmy jeszcze w tym kontekście pojęcie kuli.

Definicja 5. Kulą domkniętą w przestrzeni Hamminga, o środku $x \in \mathbb{Z}_2^n$ i promieniu $t \in \{0, \dots, n\}$ nazywamy zbiór:

$$\bar{B}_t(x) = \{y \in \mathbb{Z}_2^n : d(x, y) \leq t\}.$$

Jeśli $c \in K$ jest poprawnym słowem kodowym, to kula $\bar{B}_t(c)$ składa się ze wszystkich możliwych zaszumionych wariantów tego słowa, w których wystąpiło co najwyżej t błędów. Możemy zatem powiedzieć, że proces kodowania wiadomości opiera się na kryterium najbliższego sąsiada – odbiorca przyporządkowuje zaszumionemu sygnałowi to słowo kodowe, które różni się od niego na najmniejszej liczbie współrzędnych. Aby korekcja co najwyżej t błędów była jednoznaczna, kule domknięte o promieniu t i środkach będących elementami podprzestrzeni K muszą być parami rozłączne. Wynika z tego, że minimalna odległość między słowami kodowymi tych kul musi wynosić co najmniej $2t + 1$.

Przykład 3. Aby to lepiej zobrazować rozważmy trywialny kod z powtórzeniami o długości 3 składający się z dwóch poprawnych słów kodowych: $K = \{000, 111\}$. Rozstęp tego kodu wynosi $d^* = 3$. Zgodnie z twierdzeniem 1, kod ten gwarantuje korekcję $t = 1$ błędu. Skonstruujmy kule domknięte o promieniu 1 wokół obu poprawnych słów kodowych:

$$B_1(000) = \{000, 001, 010, 100\},$$

$$B_1(111) = \{111, 110, 101, 011\}$$

Zauważmy, że kule te są całkowicie rozłączne. Załóżmy teraz, że nadawca wysłał słowo 000, ale szum transmisyjny przekłamał środkowy bit. Odbiorca rejestruje zaszumiony ciąg $y = 010$. Dekoder sprawdza, do której kuli należy ten wektor. Ponieważ 010 leży wewnątrz $B_1(000)$, dekodek przyporządkuje zaszumionemu sygnałowi środek tej kuli. Wysłana wiadomość zostaje poprawnie odczytana jako 000. Rozłączne kule dają zatem gwarancję, że dopóki liczba przekłamań w kanale nie przekroczy t , zaszumiony sygnał nie zostanie błędnie zinterpretowany jako inne poprawne słowo kodowe.

Ograniczenia Hamminga i kody doskonałe

Wydłużanie słów kodowych poprawia korekcję, ale spowalnia transmisję, co chociażby przy słabym nadajniku sondy Voyager 1 było niedopuszczalne. Optymalnego rozwiązania należy szukać więc w geometrii przestrzeni Hamminga i idealnym upakowaniu kul. Skoro kule wokół poprawnych słów kodowych muszą być parami rozłączne, suma ich „objętości” (tzn. liczb elementów) nie może przekraczać liczby wszystkich możliwych wektorów. Stąd mamy następujące twierdzenie.

Twierdzenie 2. Dla każdego binarnego kodu o długości n , wymiarze k oraz zdolności korekcyjnej t zachodzi następująca nierówność:

$$2^k \sum_{i=0}^t \binom{n}{i} \leq 2^n \quad [3].$$

Okazuje się, że istnieją kody, które „szczelnie wypełniają” całą dostępną przestrzeń.

Definicja 6. Kod (n, k) -liniowy korygujący t błędów nazywamy doskonałym, jeśli spełnia równość

$$2^k \sum_{i=0}^t \binom{n}{i} = 2^n.$$

Kod Golaya

Jedynym kodem doskonałym, który koryguje więcej niż jeden błąd jest kod Golaya [4]. To on okazał się rozwiązaniem problemu ograniczonej przepustowości kanału transmisyjnego sondy Voyager. Parametry tego kodu to $(23, 12)$, co oznacza, że 12-bitowa wiadomość jest kodowana w postaci 23-bitowego bloku. Rozstęp tego kodu wynosi $d^* = 7$. Podstawiając tę wartość do warunku z twierdzenia 1 stwierdzamy, że kod $\mathcal{G}_{23}$ gwarantuje korekcję $t = 3$ błędów.

Fakt, że kod $\mathcal{G}_{23}$ jest kodem doskonałym, można łatwo udowodnić. Podstawiając parametry kodu do lewej strony równania z definicji 6, otrzymujemy:

$$2^{12} \left[\binom{23}{0} + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} \right] = 2^{12} (1 + 23 + 253 + 1771) = 2^{12} \cdot 2^{11} = 2^{23}$$

Będąc kodem doskonałym kod $\mathcal{G}_{23}$ rozbija przestrzeń $\mathbb{Z}_2^{23}$ na kule o promieniu 3. Wykorzystanie tego faktu zagwarantowało maksymalną efektywność przesyłu danych w ograniczonych warunkach misji Voyager 1.

Mimo że kod Golaya $\mathcal{G}_{23}$ jest kodem doskonałym, to w praktyce ograniczenia sprzętowe wymogły dodanie do niego jednego bitu parzystości. Było to spowodowane architekturą komputerów pokładowych sondy Voyager. Utworzony w ten sposób rozszerzony kod Golaya $\mathcal{G}_{24}$ ma parametry $(24, 12)$. Chociaż nowa struktura przestała spełniać ograniczenie Hamminga z równością, to lepiej dopasowała się do ograniczeń sprzętowych. Rozstęp wzrósł do $d^* = 8$.

Zgodnie z twierdzeniem 1, kod $\mathcal{G}_{24}$ zachował więc zdolność korekcji do co najwyżej 3 błędów

Cykliczność kodu Golaya

Zasadniczym problemem dotychczasowego podejścia była również jego złożoność obliczeniowa. Kodowanie wiadomości oparte na zwykłym mnożeniu przez macierz generującą przekraczało ówczesne możliwości sprzętowe. Rozwiązanie tego problemu znaleziono w kolejnej własności kodu Golaya, jaką jest cykliczność.

Definicja 7. Kod liniowy $K \subset \mathbb{Z}_2^n$ nazywamy cyklicznym, jeśli dla dowolnego słowa kodowego $(c_0, c_1, \dots, c_{n-1}) \in K$ również $(c_{n-1}, c_0, c_1, \dots, c_{n-2}) \in K$.

Przykład 4. Aby zilustrować tę własność, powróćmy do analizowanego w przykładzie 1 kodu $K = \{000, 011, 101, 110\}$. Weryfikacja przesunięć cyklicznych (w prawo) dla poszczególnych elementów tego zbioru daje nam następujące rezultaty:

$$000 \mapsto 000 \in K$$

$$\begin{aligned} 011 &\mapsto 101 \in K \\ 101 &\mapsto 110 \in K \\ 110 &\mapsto 011 \in K \end{aligned}$$

Ponieważ przesunięcie cykliczne dowolnego wektora zawsze generuje poprawne słowo kodowe należące do wyjściowego zbioru K , możemy stwierdzić, że jest to kod cykliczny.

Własność, jaką jest cykliczność pozwala na utożsamienie słów kodowych zapisanych jako wektory z pewnymi wielomianami w pierścieniu ilorazowym $\mathbb{Z}_2[X]/(X^n - 1)$ [3]. W przypadku kodu Golaya oznacza to, że cały kod $\mathcal{G}_{23}$ może być wówczas wygenerowany przez jeden wielomian generujący $g(X)$ stopnia 11. Wielomian ten przyjmuje postać:

$$g(X) = X^{11} + X^{10} + X^6 + X^5 + X^4 + X^2 + 1.$$

Zastosowanie tej reprezentacji pozwoliło na radykalną zmianę w implementacji inżynierskiej. Proces kodowania wiadomości $m(X)$ polega w tej metodzie wyłącznie na mnożeniu wielomianów: $c(X) = m(X) \cdot g(X)$. Sprzętowo, mnożenie wielomianów w takim układzie redukuje się do zastosowania prostego rejestru przesuwnego ze sprzężeniem zwrotnym, składającego się z bramek logicznych XOR. W praktyce system pokładowy najpierw błyskawicznie generował 23-bitowe słowo korzystając z cykliczności, a dopiero na końcu procesu doklejał do niego bit parzystości.

Podsumowanie

Bezpośrednia transmisja sygnału o niskiej mocy przez zaszumiony kanał prowadzi do utraty informacji. Niniejszy artykuł pokazał, że wprowadzenie odpowiednich struktur algebraicznych nad ciałem $\mathbb{Z}_2$ oraz wykorzystanie kodów korekcyjnych – w szczególności kodu Golaya – pozwala na skuteczną rekonstrukcję oryginalnej wiadomości. Sukces misji Voyager 1 potwierdza, że narzędzia z dziedziny algebry stanowią fundament budowy systemów telekomunikacyjnych.

Literatura

- [1] MacWilliams F. J., Sloane N. J. A.: The Theory of Error-Correcting Codes, North-Holland, 1977
- [2] Huffman W. C., Pless V.: Fundamentals of Error-Correcting Codes, Cambridge University Press, 2003
- [3] Jeż A.: Kody korekcyjne. Notatki do przedmiotu, Instytut Informatyki Uniwersytetu Wrocławskiego, 2019
<https://ii.uni.wroc.pl/~aje/KodKor2019/notatki.pdf> [dostęp: 06.2026]
- [4] Tietäväinen A.: On the nonexistence of perfect codes over finite fields, SIAM Journal on Applied Mathematics, 1973
- [5] Wicker S. B.: Error Control Systems for Digital Communication and Storage, Prentice Hall, 1995,
<https://archive.org/details/errorcontrolsyst0000wick/page/n9/mode/2up> [dostęp: 06.2026]
- [6] Yuen J. H. (red.): Deep Space Telecommunications Systems Engineering, Jet Propulsion Laboratory, California Institute of Technology, 1983, <https://descanso.jpl.nasa.gov/dstse/DSTSE.pdf> [dostęp: 06.2026]

WPŁYW GEOMETRII NA LUMINESCENCJĘ KROPEK KWANTOWYCH W UJĘCIU TEORETYCZNYM I APLIKACYJNYM

Alan MARCINIAK

Politechnika Łódzka, Łódź

Wstęp

Współczesna nanotechnologia oraz inżynieria materiałowa w obszarze struktur niskowymiarowych – a w szczególności zerowymiarowych – redefiniują klasyczne podejście do analizy właściwości fizykochemicznych materii. Głównym obiektem niniejszych rozważań są kropki kwantowe (ang. *quantum dots*, QDs) ze szczególnym uwzględnieniem grafenowych (ang. *graphene quantum dots*, GQDs) – nanometryczne kryształy półprzewodnikowe, których fizyczny rozmiar mieści się zazwyczaj w przedziale od 2 do 10 nm. Ograniczenie przestrzenne nośników ładunku we wszystkich trzech kierunkach przestrzennych prowadzi do ujawnienia się efektu uwięzienia kwantowego (ang. *quantum confinement effect*). Skutkuje to reorganizacją struktury pasmowej materiału, przejawiającą się dyskretyzacją gęstości stanów energetycznych. Makroskopowym efektem tego zjawiska jest silna zależność właściwości optoelektronicznych – w tym zdolności do luminescencji pod wpływem wzbudzenia promieniowaniem ultrafioletowym (UV) – od geometrycznego rozmiaru nanocząstki.

Fizyczny i matematyczny model sferycznej studni potencjału

W ujęciu teoretycznym kropkę kwantową modeluje się najczęściej jako cząstkę (elektron lub dziurę) uwięzioną w trójwymiarowej, sferycznie symetrycznej studni o nieskończonym potencjale i promieniu R . W klasycznym materiale litym pasma energetyczne (walencyjne i przewodnictwa) wykazują charakter ciągły. W momencie, gdy promień nanokryształu R staje się mniejszy lub porównywalny z efektywnym promieniem Bohra ekscytonu dla danego półprzewodnika, swoboda ruchu nośników zostaje drastycznie ograniczona. Z punktu widzenia mechaniki kwantowej oznacza to, że funkcje falowe cząstek muszą spełniać sztywne warunki brzegowe i zniknąć na granicy kropki.

Zmniejszenie wymiarów studni potencjału powoduje relatywny wzrost energii kinetycznej uwięzionych w niej cząstek. W efekcie, im mniejszy jest promień kropki kwantowej, tym szerokość efektywnej przerwy energetycznej staje się większa, co bezpośrednio determinuje większą energię (a więc mniejszą długość fali) emitowanego fotonu podczas rekombinacji. Matematyczny opis tego stanu uzyskuje się poprzez rozwiązanie stacjonarnego równania Schrödingera dla cząstki w sferycznej studni potencjału. Dozwolone stany energetyczne dla elektronu E_e oraz dziury E_h wyrażają się odpowiednio wzorami [1]:

dla elektronu:

$$E_{n,l} = \frac{\hbar^2 \rho_{n,l}^2}{2m_e R^2},$$

dla dziury:

$$E_{n,l} = \frac{\hbar^2 \rho_{n,l}^2}{2m_d R^2},$$

gdzie:

$\hbar$ – zredukowana stała Plancka ($\hbar = \frac{h}{2\pi}$),

m_e i m_d – efektywne masy odpowiednio elektronu i dziury w danym materiale,

n – główna liczba kwantowa ($n = 1, 2, 3, \dots$),

l – orbitalna liczba kwantowa ($l = 0, 1, 2, \dots, n - 1$),

$\rho_{n,l}$ – pierwiastki sferycznych funkcji Bessela, wynikające z warunków brzegowych.

Z powyższych zależności wynika kluczowa dla nanostruktur odwrotna proporcjonalność kwadratowa energii od promienia kropki. W najprostszym przybliżeniu całkowita energia przejścia E_{trans} stanowiłaby sumę energii przerwy wzbronionej materiału litego E_g oraz energii kwantyzacji obu nośników [1]:

$$E_{trans} \approx E_g + E_e + E_h.$$

Podejście to ignoruje jednak oddziaływania elektrostatyczne. W ograniczonej przestrzeni nanokryształu nie można pominąć energii przyciągania kulombowskiego pomiędzy elektronem a dziurą, tworzącymi powiązany stan ekscytonowy. Pełny model teoretyczny uwzględniający tę korektę sformułował profesor Louis Brus [2]:

$$\Delta E = \frac{\hbar^2 \pi^2}{2R^2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right) - \frac{1,8e^2}{4 \cdot \pi \cdot \epsilon_0 \cdot \epsilon_r \cdot R},$$

gdzie:

ΔE – energia przejścia kwantowego,

e – ładunek elementarny,

ϵ_0 – przenikalność elektryczna próżni,

ϵ_r – stała dielektryczna materiału kropki.

We wzorze Brusa wyrażenie $\frac{\hbar^2 \pi^2}{2R^2} \left(\frac{1}{m_e} + \frac{1}{m_h} \right)$, nazywane dalej członem kwantyzacyjnym i oznaczone literą T , reprezentuje opisaną wcześniej energię kinetyczną wynikającą z uwięzienia kwantowego, natomiast wyrażenie $\frac{1,8e^2}{4 \cdot \pi \cdot \epsilon_0 \cdot \epsilon_r \cdot R}$, nazywane dalej członem kulombowskim i oznaczone literą V , odpowiada za energię oddziaływań kulombowskich. Ogólny model kwantyzacji dla sferycznej kropki kwantowej wykorzystuje parametr ρ , oznaczający pierwiastki sferycznych funkcji Bessela, ponieważ pozwala on wyznaczyć energię dla dowolnego stanu wzbudzonego w nanokryształ. Wzór Brusa skupia się na podstawowym przejściu optycznym, dlatego wartość pierwszego pierwiastka sferycznej funkcji Bessela wynosi dokładnie π , co tłumaczy obecność tej stałej w klasycznej postaci równania Brusa [2].

Weryfikacja modelu Brusa dla klasycznych półprzewodników

Analityczną weryfikację wzoru Brusa przeprowadzono na przykładzie modelowych nanocząstek wykonanych z arsenku galu (GaAs), charakteryzujących się średnicami wynoszącymi odpowiednio $D = 5 \text{ nm}$ oraz $D = 7 \text{ nm}$. W tabeli 1 widnieją stałe zastosowane do obliczeń.

Nazwa stałej	Wartość
--------------	---------

Energia przerwy wzbronionej E_g	1,424 eV
Masa efektywna elektronu m_e	$0,063 \cdot m_0$
Masa efektywna dziury m_h	$0,51 \cdot m_0$
Masa spoczynkowa elektronu m_0	$9,11 \cdot 10^{-31} \text{ kg}$
Stała dielektryczna ϵ_r	12,9

Tab. 1. Zestawienie stałych fizycznych dla GaAs wykorzystanych w obliczeniach numerycznych.

Dla kropki o średnicy 5 nm ($R = 2,5 \text{ nm}$), po podstawieniu stałych materiałowych do równania, wartości dla członu kwantyzacyjnego i dla członu kulombowskiego wynoszą odpowiednio:

$$T_5 = \frac{\left(\frac{(6,626 \cdot 10^{-34} \text{ J} \cdot \text{s})}{2 \cdot \pi}\right)^2 \pi^2}{2 \cdot (2,5 \cdot 10^{-9} \text{ m})^2} \cdot \left(\frac{1}{0,063 \cdot 9,11 \cdot 10^{-31} \text{ kg}} + \frac{1}{0,51 \cdot 9,11 \cdot 10^{-31} \text{ kg}}\right) =$$

$$= 1,72 \cdot 10^{-19} \text{ J} = 1,073 \text{ eV}$$

$$V_5 = \frac{1,8 \cdot (1,602 \cdot 10^{-19} \text{ C})^2}{4 \cdot \pi \cdot 8,854 \cdot 10^{-12} \frac{\text{F}}{\text{m}} \cdot 12,9 \cdot 2,5 \cdot 10^{-9} \text{ m}} = 1,28 \cdot 10^{-20} \text{ J} = 0,080 \text{ eV}.$$

Całkowita wyznaczona energia przejścia optycznego dla kropki kwantowej o średnicy 5 nm E_5 wynosi:

$$E_5 = E_g + T_5 - V_5 = 1,424 \text{ eV} + 1,073 \text{ eV} - 0,080 \text{ eV} = 2,417 \text{ eV},$$

co odpowiada emisji fotonu o długości fali zgodnie ze wzorem $\lambda_5 = \frac{h \cdot c}{E_5}$:

$$\lambda_5 = \frac{(6,626 \cdot 10^{-34} \text{ J} \cdot \text{s}) \cdot 3 \cdot 10^8 \frac{\text{m}}{\text{s}}}{2,417 \text{ eV}} = 513 \text{ nm},$$

plasującej się w zielonym obszarze widma światła widzialnego.

Z kolei po zwiększeniu średnicy kryształu do 7 nm ($R = 3,5 \text{ nm}$) wartości członu kwantyzacyjnego i kulombowskiego wynoszą odpowiednio:

$$T_7 = \frac{\left(\frac{(6,626 \cdot 10^{-34} \text{ J} \cdot \text{s})}{2 \cdot \pi}\right)^2 \pi^2}{2 \cdot (3,5 \cdot 10^{-9} \text{ m})^2} \cdot \left(\frac{1}{0,063 \cdot 9,11 \cdot 10^{-31} \text{ kg}} + \frac{1}{0,51 \cdot 9,11 \cdot 10^{-31} \text{ kg}}\right) =$$

$$= 8,76 \cdot 10^{-20} \text{ J} = 0,547 \text{ eV}$$

$$V_7 = \frac{1,8 \cdot (1,602 \cdot 10^{-19} \text{ C})^2}{4 \cdot \pi \cdot 8,854 \cdot 10^{-12} \frac{\text{F}}{\text{m}} \cdot 12,9 \cdot 3,5 \cdot 10^{-9} \text{ m}} = 9,19 \cdot 10^{-21} \text{ J} = 0,057 \text{ eV}.$$

Całkowita wyznaczona energia przejścia optycznego dla kropki kwantowej o średnicy 7 nm E_7 wynosi:

$$E_7 = E_g + T_7 - V_7 = 1,424 \text{ eV} + 0,547 \text{ eV} - 0,057 \text{ eV} = 1,914 \text{ eV},$$

co odpowiada emisji fotonu o długości fali zgodnie ze wzorem $\lambda_7 = \frac{h \cdot c}{E_7}$:

$$\lambda_7 = \frac{1240 \cdot 10^{-9} \text{ eV} \cdot \text{m}}{1,914 \text{ eV}} = 647 \text{ nm},$$

odpowiadającej barwie czerwonej.

Uzyskane wyniki obliczeń wykazują pełną korelację z symulacjami numerycznymi układów GaAs i potwierdzają poprawność założeń teoretycznych uwięzienia kwantowego struktur półprzewodnikowych [3].

Grafenowe kropki kwantowe

Postęp w zakresie inżynierii materiałowej ukazał, iż nanokryształy luminescencyjne mogą być z powodzeniem wytwarzane z materiałów alternatywnych, w tym z alotropowych form węgla. Szczególne miejsce w tej grupie zajmują grafenowe kropki kwantowe stanowiące odizolowane, dwuwymiarowe fragmenty sieci grafenowej o średnicy $D < 10 \text{ nm}$.

Próba bezpośredniego zaadaptowania klasycznego wzoru Brusa do opisu struktury energetycznej GQDs napotyka jednak na barierę natury fizycznej. Specyficzna geometria sieci krystalicznej grafenu sprawia, iż zlokalizowane w niej nośniki ładunku wykazują liniową relację dyspersji w pobliżu punktów Diraca, zachowując się jak relatywistyczne cząstki o zerowej masie efektywnej. Podstawienie zerowych wartości mas do klasycznego równania Schrödingera lub wzoru Brusa – gdzie parametry te rezydują w mianowniku – generuje sprzeczność matematyczną. W konsekwencji do formalnego opisu teoretycznego GQDs niezbędne jest wdrożenie relatywistycznego równania Diraca [4]:

$$E = 2\hbar \cdot V_F \cdot k,$$

gdzie:

V_F – prędkość Fermiego dla grafenu,

k – wektor falowy.

Model ten operuje pojęciem prędkości Fermiego, która dla grafenu wynosi ok. $10^{15} \frac{\text{nm}}{\text{s}}$ oraz wektorem falowym elektronu, który z uwagi na ograniczenie przestrzenne interpretowany jest jako fala stojąca:

$$k = \frac{2\pi}{\lambda_n},$$

gdzie:

λ_n – długość n -tej harmonicznej.

Długość n -tej harmonicznej wyraża się wzorem:

$$\lambda_n = \frac{2L}{n},$$

gdzie:

L – długość ośrodka,

n – numer harmonicznej.

Na podstawie powyższych wzorów można stwierdzić, że dla małych wartości n , wektor falowy oraz odwrotność wymiaru ośrodka są wielkościami tego samego rzędu:

$$k = \frac{2\pi}{\frac{2L}{n}} \sim \frac{1}{L}.$$

Przyjmując geometryczne przybliżenie kropki do płaskiego okręgu o średnicy D , średnia długość ośrodka wynosi:

$$L_{Gdot} = \pi \frac{P}{Ob} = \pi \frac{\pi \frac{D^2}{4}}{\pi D} = \frac{\pi D}{4},$$

gdzie:

L_{Gdot} – średnia długość ośrodka jakim jest grafenowa kropka kwantowa,

P – pole powierzchni grafenowej kropki kwantowej,

Ob – obwód grafenowej kropki kwantowej,

D – średnica grafenowej kropki kwantowej.

Wektor falowy przyjmuje ostatecznie postać:

$$k = \frac{4}{\pi D}.$$

Po podstawieniu stałych fizycznych, relatywistyczna energia przejścia przyjmuje postać, w której pozostaje odwrotnie proporcjonalna do średnicy kropki:

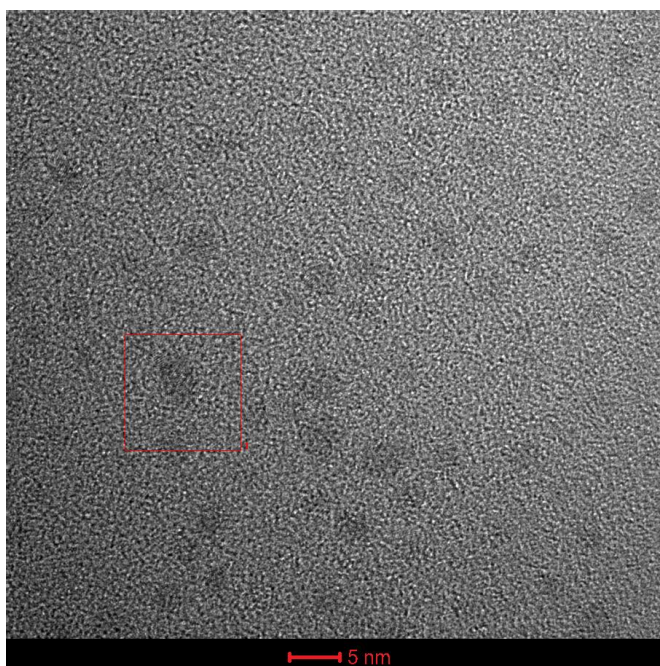
$$E = 2 \cdot \frac{6,626 \cdot 10^{-34} \text{ J} \cdot \text{s}}{2 \cdot \pi} \cdot 10^6 \frac{\text{m}}{\text{s}} \cdot \frac{4}{\pi \cdot D} = \frac{1,68 \text{ eV} \cdot \text{nm}}{D}.$$

Dla idealnego, pozbawionego defektów skrawka grafenu o średnicy $D = 5 \text{ nm}$, model ten implikuje energię przejścia rzędu $0,336 \text{ eV}$, co odpowiada długości fali $\lambda = 3690 \text{ nm}$. Oznacza to, iż teoretyczna luminescencja idealnej cząstki GQD przypada na pasmo głębokiej podczerwieni, pozostając niewidoczną dla ludzkiego oka [4].

Zielona synteza grafenowych kropek kwantowych

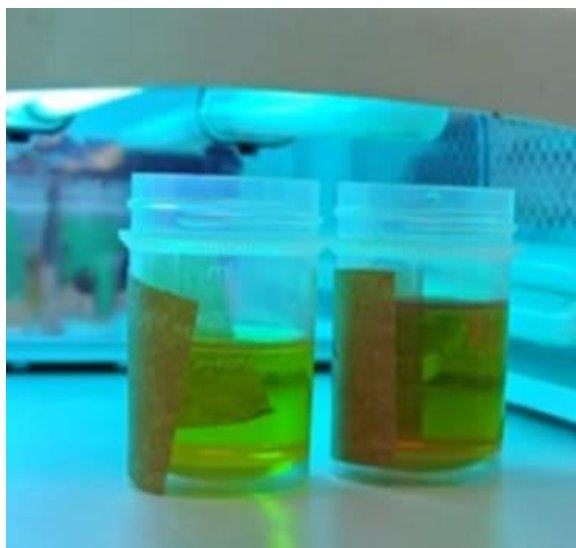
Izolowanie geometrycznie idealnych płatów grafenu metodami odgórnymi *top-down* wiąże się z wysokimi kosztami i skomplikowanymi procedurami laboratoryjnymi. Z tego względu paradygmat współczesnej chemii nanostruktur faworyzuje syntezę oddolną *bottom-up*, wykorzystującą łatwo dostępne, bogate w węgiel prekursor organiczne. W nurt ten wpisuje się tzw. zielona synteza, w której jako surowce węglowe stosuje się naturalne ekstrakty roślinne i owocowe, obfitujące w węglowodany [5], [6].

W ramach badań prowadzonych m.in. na Politechnice Łódzkiej zrealizowano proces syntezy GQDs aplikując sok owocowy jako prekursor. Za pomocą transmisyjnej mikroskopii elektronowej (TEM) potwierdzono pomyślnie wytworzenie nanostruktur o średnicy oscylującej wokół 5 nm , co widać na rysunku 1. Zaobserwowano jednak wyraźną anomalię eksperymentalną – uzyskany koloid poddany naświetlaniu UV wykazywał intensywną fotoluminescencję w paśmie zielonym (rysunek 2), co stoi w bezpośredniej sprzeczności z relatywistycznym modelem teoretycznym przewidującym emisję w podczerwieni.



Rys. 1. Zdjęcie grafenowych kropek kwantowych z mikroskopu TEM. Źródło: materiały własne

Wyjaśnieniem tego fenomenu jest zaawansowana chemia powierzchni nanocząstek. Naturalny ekstrakt owocowy stanowi wieloskładnikową mieszaninę, zawierającą oprócz węgla także znaczne ilości tlenu, azotu i wodoru. W trakcie procesu syntezy pierwiastki te ulegają wbudowaniu w strukturę krawędziową siatki węglowej, tworząc liczne grupy funkcyjne (m.in. hydroksylowe, karboksylowe). Powstałe w ten sposób powierzchniowe stany defektowe generują nowe, stabilne podpoziomy energetyczne wewnątrz natywnej przerwy wzbronionej. To właśnie te stany powierzchniowe, a nie uwięzienie kwantowe samego rdzenia węglowego, dominują w mechanizmie rekombinacji ekscytonów, powodując drastyczne przesunięcie widma emisji w obszar światła widzialnego [7], [8].



Rys. 2. Dyspersja koloidalna grafenowych kropek kwantowych wzbudzona światłem ultrafioletowym.
Źródło: materiały własne

Potencjał aplikacyjny GQDs

Unikalna struktura energetyczna, połączona z uniwersalną matrycą węglową, determinuje szerokie spektrum aplikacyjne grafenowych kropek kwantowych. Cechują się one wysoką stabilnością koloidalną, doskonałą biokompatybilnością i brakiem toksyczności w porównaniu do klasycznych kropek na bazie metali ciężkich [5], [9], [10], [11].

W sferze zaawansowanej biomedycyny, GQDs są intensywnie badane pod kątem zastosowań w teranostyce. Dzięki modyfikacjom powierzchniowym możliwe jest funkcjonalizowanie kropek przeciwciałami, co wykazano m.in. w badaniach nad celowanym wykrywaniem komórek nowotworowych raka piersi (HER2+). Ponadto, wysoki stosunek powierzchni do objętości predestynuje je do roli nanośników w systemach kontrolowanego dostarczania leków, gdzie wbudowana fotoluminescencja umożliwia jednocześnie śledzenie pozycji nanośnika w komórce w czasie rzeczywistym. Wykazują one również potencjał jako czułe biosensory optyczne – wprowadzenie roztworu sody oczyszczonej o wysokiej zasadowości skutkuje natychmiastowym wygaszeniem luminescencji, co potwierdza ich funkcjonalność jako sensorów pH [10], [12].

W inżynierii odnawialnych źródeł energii, GQDs stanowią obiecujący komponent optoelektroniczny. W ramach prac badawczych zrealizowano struktury barwnikowych ogniw fotowoltaicznych (DSSC), w których owocowe kropki kwantowe pomyślnie zaimplementowano w charakterze sensybilizatora (barwnika) powłoki mezoporowatego ditlenku tytanu (TiO_2), tworząc ekologiczną i tanią alternatywę dla ogniw krzemowych. Dodatkowym atutem GQDs jest zdolność części struktur do luminescencyjnej konwersji, co pozwala na transformację niskoenergetycznych fotonów z pasma bliskiej podczerwieni do pasma widzialnego, optymalizując stopień wykorzystania spektrum słonecznego [13], [14].

Podsumowanie

Analiza właściwości optycznych grafenowych kropek kwantowych, syntetyzowanych metodami chemii ekologicznej z biomasy, stanowi doskonały dowód na interdyscyplinarny charakter współczesnej nauki o materiałach. Wykazano, iż o ile fizyka kwantowa i modele relatywistyczne precyzyjnie definiują zachowanie idealnych układów niskowymiarowych, o tyle w warunkach eksperymentalnych ostateczna struktura energetyczna jest w równym stopniu determinowana przez inżynierię geometrii rdzenia, co przez chemię stanów powierzchniowych. W ujęciu fundamentalnym to właśnie aparat matematyczny tłumaczy zjawisko uwięzienia kwantowego, sprowadzając je do problemu wartości własnych przy założonych warunkach brzegowych. Otrzymane rozwiązania analityczne, dostarczają precyzyjnego, ilościowego opisu relacji między fizyczną geometrią nanokryształu a jego strukturą energetyczną. Osiągnięte rezultaty w zakresie biosensorów pH oraz ogniw fotowoltaicznych DSSC potwierdzają, że „zielone” nanostruktury węglowe stanowią kluczowy filar rozwoju nowoczesnej optoelektroniki i nanomedycyny.

Literatura

- [1] Griffiths D. J., Schroeter D. F.: Introduction to Quantum Mechanics, Cambridge University Press, 2019
- [2] Brus L. E.: Electron–electron and electron-hole interactions in small semiconductor crystallites: The size dependence of the lowest excited electronic state, J. Chem. Phys., vol. 80, no. 9, pp. 4403–4409, May 1984, doi: 10.1063/1.447218
- [3] Lee H., Kim D.: Quantum Mechanical Analysis Based on Perturbation Theory of CdSe/ZnS Quantum-Dot Light-Emission Properties, Nanomaterials, vol. 12, no. 20, pp. 3590, Oct. 2022, doi: 10.3390/nano12203590

- [4] Ritter K. A., Lyding J. W.: The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons, *Nat. Mater.*, vol. 8, no. 3, pp. 235–242, Mar. 2009, doi: 10.1038/nmat2378
- [5] Kadyan P., et al.: Comprehensive Review on Synthesis, Applications, and Challenges of Graphene Quantum Dots (GQDs), *J. Nanomater.*, vol. 2023, no. 1, pp. 2832964, 2023, doi: 10.1155/2023/2832964
- [6] Abbas A., et al.: One-step green synthesis of biomass-derived graphene quantum dots as a highly selective optical sensing probe, *Mater. Today Chem.*, vol. 30, pp. 101555, Jun. 2023, doi: 10.1016/j.mtchem.2023.101555
- [7] Zhu S., Song Y., Zhao X., Shao J., Zhang J., Yang B.: The photoluminescence mechanism in carbon dots (graphene quantum dots, carbon nanodots, and polymer dots): current state and future perspective, *Nano Res.*, vol. 8, no. 2, pp. 355–381, Feb. 2015, doi: 10.1007/s12274-014-0644-3
- [8] Huang P., et al.: Anomalous Light Emission and Wide Photoluminescence Spectra in Graphene Quantum Dot: Quantum Confinement from Edge Microstructure, *J. Phys. Chem. Lett.*, vol. 7, no. 15, pp. 2888–2892, Aug. 2016, doi: 10.1021/acs.jpclett.6b01309
- [9] Rasheed P. A., Ankitha M., Pillai V. K., Alwarappan S.: Graphene quantum dots for biosensing and bioimaging, *RSC Adv.*, vol. 14, no. 23, pp. 16001–16023, May 2024, doi: 10.1039/D4RA01431F
- [10] Cui Y., et al.: A Review of Advances in Graphene Quantum Dots: From Preparation and Modification Methods to Application, *C*, vol. 10, no. 1, pp. 7, Jan. 2024, doi: 10.3390/c10010007
- [11] Ghaffarkhah A., et al.: Synthesis, Applications, and Prospects of Graphene Quantum Dots: A Comprehensive Review, *Small*, vol. 18, no. 2, pp. 2102683, 2022, doi: 10.1002/sml.202102683
- [12] Osorio H. M., Castillo-Solís F., Barragán S. Y., Rodríguez-Pólit C., Gonzalez-Pastor R.: Graphene Quantum Dots from Natural Carbon Sources for Drug and Gene Delivery in Cancer Treatment, *Int. J. Mol. Sci.*, vol. 25, no. 19, pp. 10539, Jan. 2024, doi: 10.3390/ijms251910539
- [13] Jahantigh F., Ghorashi S. M. B., Bayat A.: Hybrid dye sensitized solar cell based on single layer graphene quantum dots, *Dyes Pigments*, vol. 175, pp. 108118, Apr. 2020, doi: 10.1016/j.dyepig.2019.108118
- [14] Vercelli B.: The Role of Carbon Quantum Dots in Organic Photovoltaics: A Short Overview, *Coatings*, vol. 11, no. 2, pp. 232, Feb. 2021, doi: 10.3390/coatings11020232

KODOWANIE KSZTAŁTU: MATEMATYKA W PROJEKTOWANIU PARAMETRYCZNYM

Szymon WOŹNIAK, Łukasz ZYCHOWICZ

Politechnika Łódzka, Łódź

Wstęp

Zastosowanie zaawansowanych algorytmów oraz sztucznej inteligencji (AI) we współczesnej inżynierii lądowej i architekturze wymaga uświadomienia sobie fundamentalnej prawdy: proces projektowania obiektów budowlanych od zawsze opierał się na ścisłych zasadach matematycznych. Historycznie, dyscyplina ta była silnie zdeterminowana przez paradygmaty geometrii euklidesowej. Twórcy i inżynierowie pozostawali ograniczeni do operowania liniami prostymi, kątami prostymi oraz powtarzalnymi, elementarnymi bryłami.

Pojawienie się projektowania parametrycznego stanowi radykalne odejście od tego klasycznego schematu, całkowicie transformując proces twórczy i inżynierski. Architektura staje się w nim matematyczną funkcją wielu zmiennych, a nie zbiorem statycznych linii prostych na płaszczyźnie.

Powierzchnie matematyczne jako obiekty budowlane

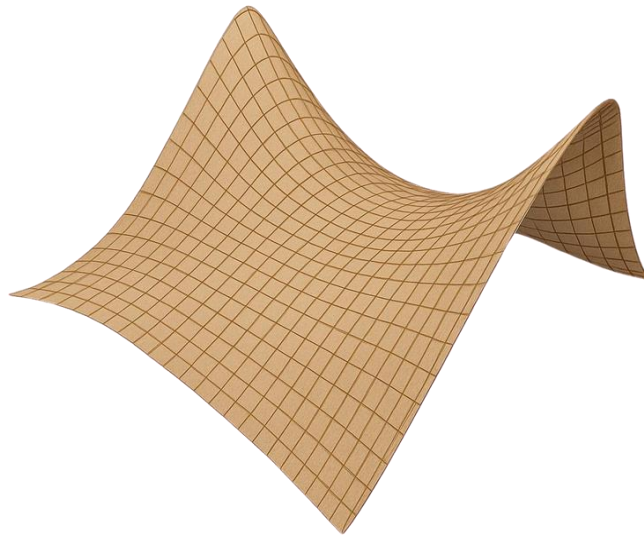
Zrozumienie tej rewolucji wymaga jednak powrotu do absolutnych podstaw – mianowicie do analizy powierzchni matematycznych w przestrzeni trójwymiarowej (R^3). To właśnie te struktury geometryczne stanowią fundament i punkt wyjścia dla najbardziej ikonicznych i skomplikowanych budowli współczesnego świata.

Paraboloida hiperboliczna – L'Oceanogràfic (Walencja)

Jedną z najbardziej efektywnych form geometrycznych wykorzystywanych we współczesnym budownictwie cienkościennym jest paraboloida hiperboliczna, w literaturze technicznej klasyfikowana często jako powierzchnia siodłowa. Z matematycznego punktu widzenia jest to powierzchnia drugiego stopnia, opisana w układzie kartezjańskim równaniem:

$$\frac{z}{c} = \frac{x^2}{a^2} - \frac{y^2}{b^2},$$

gdzie: a, b, c – stałe rzeczywiste oraz $a, b > 0$, $c \neq 0$.



Rys. 1. Paraboloida hiperboliczna. Źródło: [1]

Wybitnym przykładem praktycznej implementacji tej geometrii w architekturze monumentalnej jest dach kompleksu L'Oceanogràfic w Walencji – największego oceanarium w Europie, zaprojektowanego przez Félixą Candelę. Przekrycie to stanowią żelbetowe powłoki, których grubość w strefach centralnych wynosi zaledwie kilka centymetrów.

Z punktu widzenia mechaniki budowli, zdolność tak cienkiej struktury do przenoszenia znacznych obciążeń zewnętrznych (w tym parcia i ssania wiatru) wynika bezpośrednio z jej geometrii. Podwójna, przeciwnokierunkowa krzywizna sprawia, że w powłoce dominują naprężenia błonowe (ściskanie i rozciąganie w płaszczyźnie powłoki), podczas gdy momenty zginające zostają zredukowane do minimum.



Rys. 2. Budynek L'Oceanogràfic. Źródło: [2]

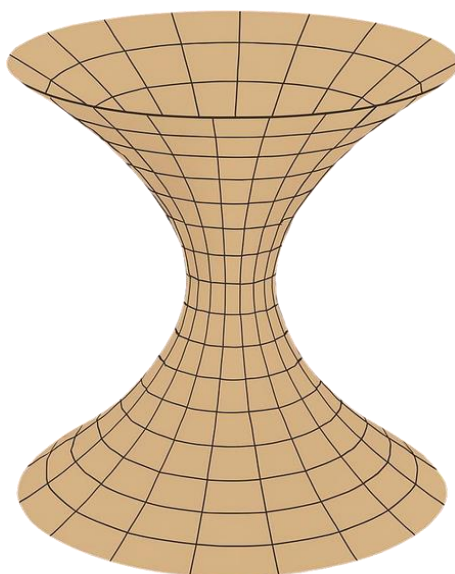
Hiperboloida jednopowłokowa - Canton Tower (Chiny)

Kolejną geometrią o kluczowym znaczeniu dla współczesnej architektury megastruktur i obiektów wieżowych jest hiperboloida jednopowłokowa.

Jej równanie kanoniczne w trójwymiarowym układzie współrzędnych kartezjańskich przyjmuje postać:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2} = 1,$$

gdzie: a, b, c – stałe rzeczywiste oraz $a, b, c > 0$.



Rys. 3. Hiperboloida jednopowłokowa. Źródło: [1]

Wizualnie struktura ta charakteryzuje się płynnym zwężeniem w swojej środkowej części, przypominając kształtem symetryczną klepsydę.

Modelowym przykładem pełnej implementacji tej geometrii w skali architektonicznej jest wieża telewizyjna Canton Tower w Kantonie (Chiny) o całkowitej wysokości 604 metry. Zewnętrzny szkielet wieży został zaprojektowany bezpośrednio w oparciu o geometrię hiperboloidy jednopowłokowej. Efekt wizualnego „skręcenia” i zwężenia talii obiektu osiągnięto nie poprzez gięcie elementów, lecz poprzez odpowiednie nachylenie i przesunięcie względem siebie prostoliniowych kolumn strukturalnych.

Wybór tej formy geometrycznej podyktowany był przede wszystkim rygorystycznymi wymaganiami aerodynamicznymi oraz mechanicznymi. Smukła, hiperboliczna sylwetka wykazuje wybitne właściwości w zakresie minimalizacji oddziaływania wiatru. Dzięki zmiennemu przekrojowi poprzecznemu i obłemu kształtowi, struktura nie tylko stawia minimalny opór czołowy, ale przede wszystkim w istotny sposób redukuje zjawisko regularnego odrywania się wirów (tzw. *vortex shedding*). Zjawisko to jest główną przyczyną powstawania niebezpiecznych drgań poprzecznych w obiektach smukłych. Ponadto, rozszerzająca się ku dołowi podstawa geometryczna zapewnia optymalne ramię sił do przenoszenia ogromnych momentów wywracających, generowanych przez parcie wiatru oraz siły sejsmiczne, gwarantując maksymalną sztywność struktury przy jednoczesnej optymalizacji zużycia materiałów budowlanych.



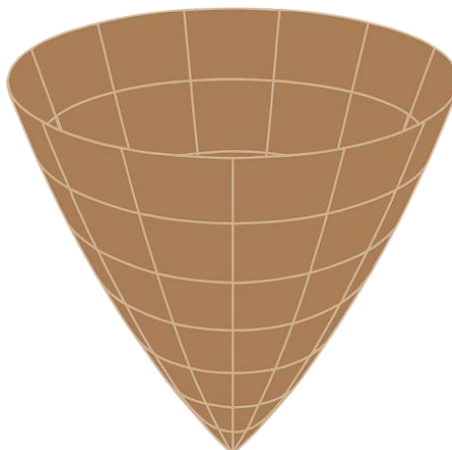
Rys. 4. Budynek Canton Tower. Źródło: [3]

Paraboloida eliptyczna – Kopuła Reichstagu (Berlin)

Ostatnią z prezentowanych klas powierzchni drugiego stopnia, odgrywającą istotną rolę w kształtowaniu nowoczesnych przekryć, jest paraboloida eliptyczna. W ortogonalnym układzie współrzędnych kartezjańskich powierzchnia ta reprezentowana jest przez równanie:

$$z = \frac{x^2}{a^2} + \frac{y^2}{b^2},$$

gdzie: a, b – stałe rzeczywiste oraz $a, b > 0$.



Rys. 5. Paraboloida eliptyczna. Źródło: [1]

Doskonałym przykładem zaawansowanej implementacji tej geometrii jest nowa kopuła gmachu Reichstagu w Berlinie, zaprojektowana przez pracownię Foster + Partners. Ta transparentna, stalowo-szklana struktura nie jest jedynie reinterpretacją tradycyjnego zwieńczenia budynku państwowego, lecz zaawansowanym technologicznie komponentem inżynieryjnym. Wybór geometrii zbliżonej do paraboloidy eliptycznej pozwolił na płynne zintegrowanie trzech kluczowych funkcji w ramach jednej bryły.

Pierwszą z nich jest optymalizacja strukturalna, w której lekki, siatkowy szkielet stalowy, stanowiący ortogonalną siatkę odwzorowującą linie krzywizny powierzchni, optymalnie przenosi obciążenia własne oraz środowiskowe, takie jak wiatr i śnieg, głównie poprzez siły osiowe, co pozwoliło zminimalizować przekroje elementów konstrukcyjnych. Jednocześnie struktura ta pełni funkcję zaawansowanego systemu termodynamicznego i wentylacyjnego, ponieważ w centralnej osi paraboloidy umieszczono odwrócony stożek, działający jako naturalny kanał wywiewny. Ostatnim elementem tej integralnej synergii jest ustrój optyczny, zrealizowany poprzez pokrycie wspomnianego stożka centralnego systemem trzystu sześćdziesięciu sterowanych komputerowo pojedynczych zwierciadeł.



Rys. 6. Budynek Reichstagu. Źródło: [4]

Architektura bez granic geometrycznych

Omówione dotychczas przykłady opierają się na relatywnie prostej geometrii – pojedynczym równaniu definiującym konkretną powierzchnię. Współczesne projektowanie parametryczne inicjuje jednak fundamentalną zmianę. Klasyczne powierzchnie matematyczne przestają być ostatecznym rezultatem, a stają się jedynie bazowymi modułami składowymi. Środowisko algorytmiczne pozwala na ich swobodne łączenie, przenikanie i transformowanie, przekształcając architekturę w płynny, wysoce złożony system przestrzenny, który można niemal dowolnie kształtować.

Morpheus Hotel (Makau)

Zmiana podejścia do środowiska cyfrowego – polegająca na odejściu od traktowania go jako pasywnego narzędzia kreślarskiego na rzecz aktywnego generatora form przestrzennych – otwiera przed inżynierami i architektami niemal nieograniczone możliwości projektowe. Doskonałym

przykładem, w którym algorytmy w pełni determinują ostateczny kształt konstrukcji, jest kompleks Morpheus w Makau.

Obiekt ten wyróżnia się zastosowaniem, swobodnego egzoszkieletu stalowego o wysoce organicznej geometrii. Przeniesienie głównego układu nośnego na zewnętrzną powłokę budynku przekształciło elewację w samonośną, trójwymiarową strukturę. Taki zabieg inżynierski pozwolił na całkowite uwolnienie przestrzeni wewnętrznej od tradycyjnych, wertykalnych elementów nośnych oraz ścian konstrukcyjnych.

Zewnętrzny szkielet składa się z tysięcy stalowych węzłów, z których każdy charakteryzuje się całkowicie unikalną geometrią i układem przestrzennym. Osiągnięcie stabilności tak nietypowego ustroju wymagało implementacji zaawansowanych skryptów algorytmicznych, które w środowisku parametrycznym przeprowadziły globalną analizę rozkładu naprężeń. Dzięki precyzyjnym obliczeniom numerycznym zagwarantowano stateczność całego układu, co byłoby nieosiągalne przy zastosowaniu tradycyjnych, manualnych metod wymiarowania.



Rys. 7. Budynek Morpheus Hotel. Źródło: [5]

Museum of the Future (Dubaj)

Kolejnym obiektem redefiniującym granice współczesnej inżynierii i architektury parametrycznej jest Muzeum Przyszłości (Museum of the Future) w Dubaju. Bryła tego budynku przyjmuje formę asymetrycznego, eliptycznego torusa, w którym całkowicie zrezygnowano z zastosowania jakichkolwiek linii prostych i klasycznych płaszczyzn ortogonalnych. Funkcję głównego ustroju nośnego powierzono tu przestrzennej, stalowej strukturze siatkowej typu diagrid. Jest to zaawansowany system konstrukcyjny oparty na diagonalnie krzyżujących się elementach, które tworzą sztywną sieć modułów trójkątnych. Dzięki rygorystycznej optymalizacji matematycznej algorytmów projektowych, ta samonośna powłoka efektywnie przejmuje i dystrybuje naprężenia, kreując sztywną „skorupę”. Rozwiązanie to pozwoliło na całkowite wyeliminowanie wewnętrznych słupów nośnych, uwalniając płynną i niezakłóconą przestrzeń użytkową.

Spektakularność i złożoność tego obiektu uwidaczniają się również w zewnętrznej warstwie elewacyjnej, którą stanowi 1024 paneli kompozytowych, z których każdy charakteryzuje się całkowicie unikalną geometrią zintegrowaną ze wzorem arabskiej kaligrafii. Realizacja tak bezprecedensowego poziomu skomplikowania wymusiła wdrożenie bezpośredniego paradygmatu produkcji cyfrowej BIM-to-Fabrication. W procesie tym zintegrowany model parametryczny przestał pełnić jedynie funkcję reprezentacyjną, a stał się bezpośrednim nośnikiem danych produkcyjnych, sterującym obrabiarkami numerycznymi (CNC) w zakładach prefabrykacji. Wykorzystanie algorytmów na każdym etapie projektowania i produkcji było jedynym sposobem na poprawne zdefiniowanie geometrii węzłów oraz bezkolizyjny montaż unikalnych paneli z trójwymiarową podkonstrukcją przy zachowaniu precyzji wykonawczej.



Rys. 8. Budynek Museum of the Future. Źródło: [6]

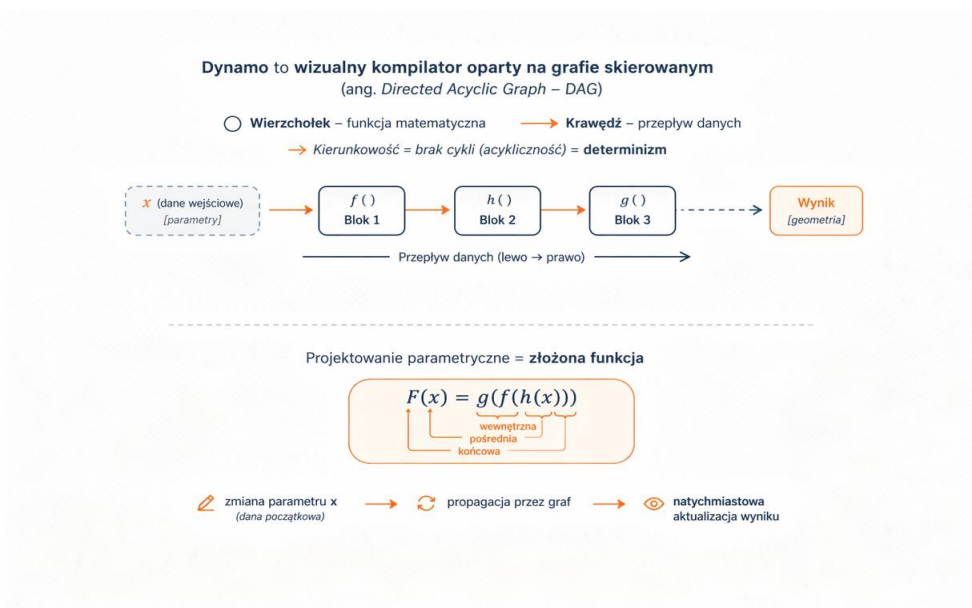
Budowanie logiki zamiast kształtów – idea projektowania parametrycznego

Projektowanie parametryczne stanowi radykalną zmianę paradygmatu inżynierskiego – to przejście od manualnego odwzorowywania form do konstruowania systemów opartych na logice i ścisłych regułach matematycznych. Proces ten polega na zdefiniowaniu parametrów wejściowych (reprezentujących wymiary lub metadane elementów) oraz nałożeniu ograniczeń i relacji, które wspólnie kształtują ostateczną formę obiektu. W efekcie powstaje dynamiczny model przestrzenny o niemal dowolnym stopniu złożoności. Jednym ze środowisk programistycznych umożliwiających tego typu modelowanie jest Dynamo, którego architektura opiera się na rygorystycznych zasadach matematycznych.

Skierowane grafy acykliczne (DAG) jako środowisko obliczeniowe

W ujęciu teoretycznym, oprogramowanie do programowania wizualnego nie jest narzędziem kreślarskim, lecz kompilatorem opartym na strukturze skierowanych grafów acyklicznych (ang. *Directed Acyclic Graph* – DAG). Każdy element logiczny skryptu (węzeł) stanowi wierzchołek grafu i reprezentuje konkretną funkcję matematyczną, natomiast połączenia między węzłami (krawędzie) definiują kierunek przepływu danych.

Kierunkowość tego przepływu zapobiega powstawaniu pętli logicznych, co gwarantuje pełen determinizm systemu. Z perspektywy analizy matematycznej, projektowanie algorytmiczne sprowadza się do budowy złożonej funkcji wielu zmiennych.



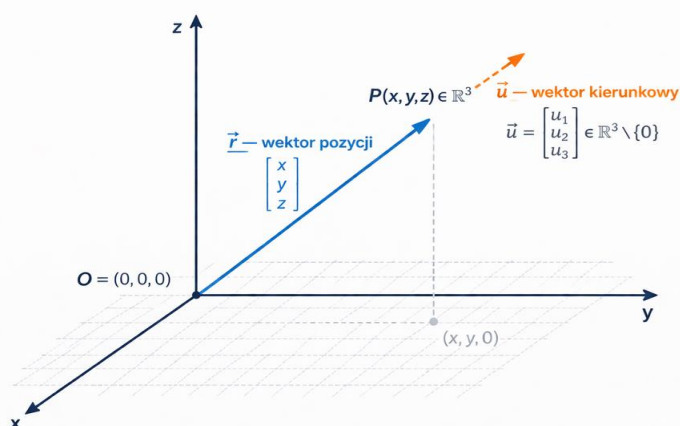
Rys. 9. Dynamo – skierowane grafy acykliczne. Źródło: [1]

W tak zdefiniowanym systemie każda modyfikacja zmiennej niezależnej na wejściu propaguje się wzdłuż krawędzi grafu od lewej do prawej strony, automatycznie i w czasie rzeczywistym aktualizując stan całej geometrii końcowej.

Algebra wektorowa w przestrzeni euklidesowej R^3

Fundamentem generowania i manipulacji formami przestrzennymi w systemach parametrycznych jest algebra wektorowa w trójwymiarowej przestrzeni euklidesowej R^3 . Każdy komponent konstrukcyjny definiowany jest operacyjnie poprzez wektory wodzące (określające jego pozycję w układzie współrzędnych) oraz wektory kierunkowe (opisujące wektory przesunięć i orientację). Kluczowe operacje analityczne opierają się na fundamentalnych działaniach wektorowych: weryfikacja prostopadłości czy rzutowanie wektorów wykorzystuje iloczyn skalarny (generujący wartość liczbową), natomiast wyznaczanie wektorów normalnych płaszczyzn bazuje na iloczynie wektorowym.

Punkt w $\mathbb{R}^3$ – wektor pozycji i wektor kierunkowy



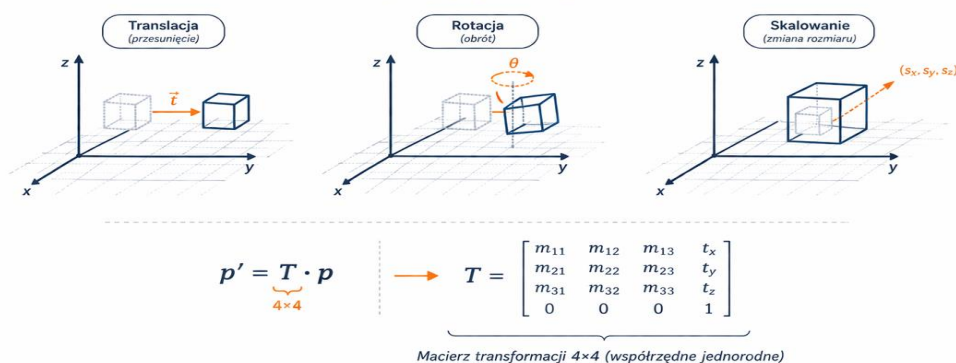
Rys. 10. Dynamo – algebra wektorowa. Źródło: [1]

Macierze transformacji i współrzędne jednorodne

Wszelkie przekształcenia afiniczne obiektów w przestrzeni, takie jak translacja (przesunięcie), rotacja (obróć) czy skalowanie (zmiana skali), realizowane są za pomocą macierzy transformacji o wymiarach 4×4 . Operacje te przeprowadzane są w układzie współrzędnych jednorodnych, co pozwala na reprezentację punktów w wielowymiarowej przestrzeni rzutowej. Zastosowanie takiego aparatu matematycznego umożliwia złożenie szeregu skomplikowanych przekształceń w jedną, wydajną obliczeniowo operację poprzez mnożenie macierzy. Skomplikowane operacje przestrzenne na tysiącach elementów sprowadzają się do zagadnień algebry liniowej.

TRANSFORMACJE OBIEKTÓW W DYNAMO

Translacja (przesunięcie) • Rotacja (obróć) • Skalowanie (zmiana rozmiaru)
realizowane są za pomocą macierzy transformacji 4×4
w układzie współrzędnych jednorodnych.

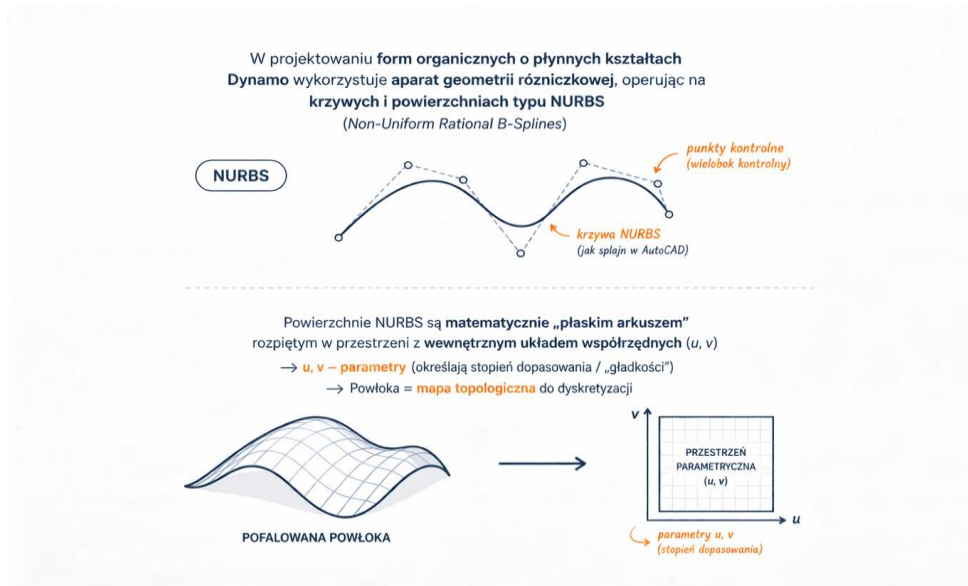


Rys. 11. Dynamo – macierze transformacji. Źródło: [1]

Geometria różniczkowa i reprezentacja NURBS

Do generowania form organicznych o płynnej krzywiźnie środowiska algorytmiczne wykorzystują aparat geometrii różniczkowej, operując na krzywych i powierzchniach typu NURBS (ang. *Non-Uniform Rational B-Spline*). Kształt tych struktur jest definiowany i modyfikowany za pomocą siatki punktów kontrolnych (wieloboku kontrolnego).

Z perspektywy algorytmu, każda, nawet najbardziej skomplikowana topologicznie powłoka o podwójnej krzywiźnie, traktowana jest jako matematycznie zdefiniowana, płaska dziedziną rozpiętą w trójwymiarowej przestrzeni. Powierzchnia ta posiada własny, lokalny układ współrzędnych krzywoliniowych, określany parametrami u oraz v , które sterują stopniem funkcji dopasowania. Dzięki takiemu odwzorowaniu formy powłokowe stają się mapami topologicznymi, co pozwala na precyzyjną, numeryczną dyskretyzację (podział ciągłej powierzchni na skończone elementy, takie jak panele elewacyjne).



Rys. 12. Dynamo – powierzchnie NURBS. Źródło: [1]

Teoria zbiorów i algorytmiczne mapowanie danych

Zarządzanie zaawansowanymi strukturami danych w modelowaniu parametrycznym opiera się na paradygmatach programowania funkcyjnego oraz teorii zbiorów. Informacje o dziesiątkach tysięcy elementów grupowane są w wielowymiarowe listy. Do nawiązywania relacji między rozbudowanymi zbiorami zmiennych stosuje się precyzyjną logikę mapowania, zwaną zbieżnością (ang. *Lacing*), która w klasycznym ujęciu przyjmuje trzy główne strategie obliczeniowe.

Najkrótsze dopasowanie (ang. *Shortest Lacing*) polega na łączeniu elementów obu list w pary relacją jeden do jednego (1:1); operacja zostaje wstrzymana w momencie wyczerpania elementów w krótszym zbiorze. Najdłuższe dopasowanie (ang. *Longest Lacing*) działa analogicznie, jednak po wyczerpaniu krótszej listy jej ostatni element jest replikowany i używany do obsługi pozostałych zmiennych z najdłuższej listy. Najbardziej zasobochłonnym działaniem jest iloczyn kartezjański (ang. *Cross Product*), polegający na skojarzeniu każdego elementu z pierwszej listy z każdym elementem listy drugiej, co prowadzi do wygenerowania pełnej macierzy kombinacji układu zbiorów.

Jednym z najbardziej zaawansowanych matematycznie aspektów Dynamo jest zarządzanie strukturami danych w oparciu o teorię zbiorów i programowanie funkcyjne (sposób działań na listach danych). Przy wielu elementach zbiera się je w listy, a do zarządzania nimi korzysta się z tzw. **Lacingu**.



Rys. 13. Dynamo – działania na listach. Źródło: [1]

Podsumowanie

Zastosowanie Dynamo na styku matematyki i inżynierii ma największą wartość w procesach optymalizacji. Wykorzystując algorytmy oraz metody numeryczne, Dynamo potrafi przeszukiwać wielowymiarową przestrzeń rozwiązań w celu znalezienia ekstremum funkcji celu (np. minimalizacja masy stali przy zachowaniu sztywności). W tym ujęciu proces projektowy staje się zagadnieniem optymalizacyjnym, gdzie parametry wejściowe są zmiennymi decyzyjnymi, a algorytm matematyczny iteracyjnie dąży do rozwiązania optymalnego. Podejście to zmienia schemat pracy z „rysowania rozwiązania” na „definiowanie problemu”, w którym ostateczny kształt obiektu jest wynikiem precyzyjnego, matematycznego dowodu.

Literatura

- [1] Grafiki wygenerowane z wykorzystaniem narzędzi AI
- [2] Gabaldón F., Wikipedia, https://pl.wikipedia.org/wiki/Plik:L%27Oceanografic_%28Valencia,_Spain%29_01.jpg [dostęp: 04.2026]
- [3] Nelson A., Veryrealandrew, <https://veryrealandrew.com/site-visits/2019/12/16> [dostęp: 04.2026]
- [4] Heller R., Rachelsruminations, <https://rachelsruminations.com/reichstag-dome-berlin/> [dostęp: 04.2026]
- [5] Romain C., Cromain, <https://cromain.com/morpheus-hotel-macau/> [dostęp: 04.2026]
- [6] Sahib K., Gettyimages, <https://www.gettyimages.com/detail/news-photo/this-picture-taken-on-june-30-2021-shows-a-view-of-the-news-photo/1233733073> [dostęp: 04.2026]



MathUp

Konferencja Zastosowań
Matematyki

Pod redakcją: dr inż. Gertruda Gwóźdź-Łukawska, prof. uczelni
Recenzja naukowa:

dr inż. Gertruda Gwóźdź-Łukawska, prof. uczelni
dr Elżbieta Galewska

dr inż. Elżbieta Kotlicka-Dwurzniak
dr Monika Potyrała

Skład i edycja: dr hab. inż. Ewa Korzeniewska, prof. uczelni

Wydział Elektrotechniki, Elektroniki, Informatyki i Automatyki
Centrum Nauczania Matematyki i Fizyki
Politechnika Łódzka 30.09.2026

ISBN 978-83-938538-6-1

