Badania operacyjne i systemy wspomagania decyzji
05 Indeks Kendalla - objaśnienie

Macierze turniejowe i uogólnione macierze turniejowe

Do tej pory przy badaniu spójności macierzy porównań parami zajmowaliśmy się sprawdzaniem na ile proporcje między porównaniami trójek obiektów są zachowane. Innymi słowy, jeżeli obiekt A był dwa razy lepszy od obiektu B, a B dwa razy lepszy od obiektu C, to oczekiwaliśmy ,że obiekt A będzie cztery razy lepszy od obiektu C, a wszelkie odchylenie od takiego wyniku używaliśmy do mierzenia niespójności. Dziś skupimy się tylko na sprawdzaniu spójności porządkowej, to jest będziemy sprawdzać, czy jeżeli A jest lepszy od B, a B jest lepszy od C, to A jest lepszy od C.

Widać, że w takim przypadku nie potrzebujemy wielu różnych wartości w naszej macierzy porządkowej, wystarczy że będzie ona przyjmować wartość 1 w komórce m_{ij}, jeżeli i-ty obiekt jest lepszy od j-tego, 0 gdy są porównywalne i -1 gdy j-ty jest lepszy od i-tego. Tym samym, jeżeli mamy normalną macierz porównań parami wystarczy zamienić wszystkie wartości większe od 1 na 1, wszystkie 1 na 0 i wszystkie wartości mniejsze od 1 na -1.

Macierz turniejowa przyjmuje swoją nazwę od tego, że jest to macierz sąsiedztwa dla grafu turniejowego, to jest takiego grafu skierowanego, że istnieje krawędź skierowana między dowolnymi dwoma wierzchołkami. Taki graf może być użyty by reprezentować wyniki turnieju w którym każdy rozegrał jeden mecz z każdym innym przeciwnikiem, a krawędź od wierzchołka A do B oznacza, że B wygrał mecz z A. Klasycznie turniejem nazywa się sytuację w której mecze nie dopuszczają remisów, innymi słowy 0 pojawiają się tylko na przekątnej. Jeżeli 0 pojawia się także poza przekątną, mówimy o uogólnionym turnieju.

Nie jest jednak tak, że każda macierz wypełniona wartościami 1,0 i -1 jest macierzą (uogólnioną) turniejową. Żeby tak było każda komórka w macierzy musi spełniać równość $m_{ij} = -m_{ji}$.

Wyznaczanie indeksu Kendalla i uogólnionego indeksu Kendalla

Żeby wyznaczyć indeks Kendalla dla macierzy turniejowej potrzebujemy sprawdzić ile trójek obiektów nie jest w naszym pojęciu spójna. Jeżeli mamy macierz turniejową M to liczbę takich trójek oznaczamy przez $|T_M|_i$. Ponieważ wszystkie krawędzie są skierowane, trójka obiektów będzie problematyczna tylko, jeżeli A jest lepsze od B, B lepsze od C i C lepsze od A. Łatwo taką sytuację odczytać z naszej macierzy, gdyż jeżeli mamy trzy obiekty o indeksach (różnych) i, j, k to spełniona jest równość $m_{ij} = m_{jk} = m_{ki}$ (kolejność w indeksach jest ważna!). Kiedy zliczymy wszystkie takie trójki musimy porównać ich ilość do maksymalnej możliwej ilości niespójnych trójek dla danego rozmiaru macierzy, o wzorze postaci (podstawy wyprowadzania wzoru były na wykładzie, są to żmudne obliczenia i nie będziemy na nie się tu powoływać):

$$\mathcal{I}(n) = \begin{cases} \frac{n^3-n}{24} & \text{, gdy n jest nieparzyste} \\ \frac{n^3-4n}{24} & \text{, gdy n jest parzyste} \end{cases}.$$

Teraz wystarczy podstawić do wzoru:

$$\zeta(M) = 1 - \frac{|T_M|_i}{\mathcal{I}(n)},$$

żeby uzyskać indeks Kendalla.

Przypadek dla uogólnionego turnieju jest trochę bardziej skomplikowany, gdyż jest więcej sytuacji w których trójka obiektów może mieć niespójne porównanie. Żeby poprawnie wyznaczyć $|T_M|_i$ dla turnieju uogólnionego rozważymy które trójki należy liczyć w zależności od tego ile krawędzi mają skierowanych (we wszystkich przypadkach indeksy muszą być parami różne a ich kolejność ma znaczenie, ponadto interesują nas trójki w jakiś sposób niespójne, czyli zliczamy problemy):

- trójka bez krawędzi skierowanych: wszystkie trzy elementy są tak samo dobre, zatem nie ma niespójności. W naszej macierzy rozpoznajemy to po równości $m_{ij} = m_{jk} = m_{ki} = 0$.
- trójka z jedną krawędzią skierowaną: tu zawsze będzie problem, gdyż sytuacja sprowadza się do powiedzenia: A jest porównywalne do B, B jest porównywalne do C, zaś C jest lepsze od A. Te sytuacje można rozpoznać w macierzy po tym, że dwa spośród m_{ij}, m_{jk}, m_{ki} są równe 0, a jeden jest od 0 różny.
- trójka z dwiema krawędziami skierowanymi: w takim przypadku problemu nie ma jeżeli jeden element jest lepszy od dwóch pozostałych porównywalnych, lub gorszy od dwóch pozostałych porównywalnych. Problem pojawia się, gdy A jest lepsze od B, B lepsze od C, zaś C porównywalne do A. Takie sytuacje znajdujemy w naszej macierzy, gdy dwa spośród m_{ij}, m_{jk}, m_{ki} są równe i różne od 0, zaś trzeci jest równy 0.

- trójka z trzema krawędziami skierowanymi: omówiona została przy zwykłych turniejach. Liczymy, gdy $m_{ij} = m_{jk} = m_{ki} \neq 0$.

Teraz musimy znaleźć maksymalną możliwą liczbę niespójnych trójek dla danego rozmiaru grafu, którą możemy wyliczyć z poniższego (odrobinę bardziej skomplikowanego) wzoru:

$$y(n) = \begin{cases} \frac{13n^3-24n^2-16n}{96} & \text{, gdy } n = 4q, q \in \mathbb{Z} \\ \frac{13n^3-24n^2-19n+30}{96} & \text{, gdy } n = 4q+1, q \in \mathbb{Z} \\ \frac{13n^3-24n^2-4n}{96} & \text{, gdy } n = 4q+2, q \in \mathbb{Z} \\ \frac{13n^3-24n^2-19n+18}{96} & \text{, gdy } n = 4q+3, q \in \mathbb{Z} \end{cases}$$

Teraz wystarczy otrzymane wartości podstawić do wzoru:

$$\zeta_g(M) = 1 - \frac{|T_M|_i}{y(n)},$$

by otrzymać uogólniony indeks Kendalla.