ANALIZA RUCHU PASAŻERSKIEGO W KRAKOWIE, PRZY POMOCY MODELU DYSKRETNEJ SYMULACJI ZDARZEŃ

Jakub Dyrcz

Studenckie Koło Naukowe Matematyków Politechniki Krakowskiej

W referacie przedstawiony zostanie proces tworzenia tytułowej symulacji. Wyjaśnimy, czym jest model dyskretnej symulacji zdarzeń (DES). Przedstawimy rozwiązania, jakie znalazły zastosowanie w opartym na DES modelu krakowskiej komunikacji miejskiej, oraz pochodzenie i sposób przetwarzania danych. Pokażemy związki tego modelu z topologią i rachunkiem prawdopodobieństwa. Na koniec omówimy wyniki otrzymane za pomocą symulacji i opowiemy o możliwościach dalszego rozwoju projektu.

Komunikacja miejska w dużych miastach, to ogromny system opierający się na samym sobie. Pojedyncza zmiana w rozkładzie jazdy jednej linii, spowoduje lawinowe skutki wpływając na zachowania pasażerów korzystających z tej linii (dojeżdżających do pracy, lub tych którzy będą się przesiadać). W referacie przedstawimy proces twórczy takiego powstawania modelu.

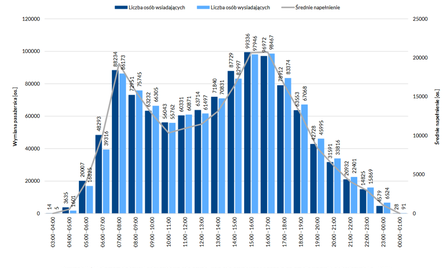
Zacznijmy od zbudowania podstawowej intuicji. Załóżmy, że pracujemy w rejestracji. W trakcie dnia pacjenci umawiają się na wizyty, my tych zapisujemy w takiej kolejności w jakiej się zapisali.

L.p.	Nazwisko	Godzina
1	Kowalski	10:00
3	Dudek	10:30
4	Wojtyła	11:00
2	Nowak	12:00
4	~~Wojtyła~~	~~11:00~~

Jeśli posortujemy tę listę względem czasu, kiedy przyjdą kolejni pacjenci będziemy wiedzieć w których momentach coś się wydarzy. Od teraz dodając nowe zdarzenie będziemy od razu sortować tabelę. Taka czynność pozwoli nam oszczędzać czas i moc obliczeniową ponieważ interesuje nas tylko moment w którym dane zdarzenie się dzieje.

Takie podejście do sortowania zdarzeń względem czasu jest podstawą do stworzenia modelu DES (*ang. Discrete Event Simulation*). W naszym modelu użyjemy m.in. takich zdarzeń jak to, że pasażer wsiadł/wysiadł z autobusu, pasażer pojawił się/znikł z systemu, pojazd przyjechał/zatrzymał się na przystanku. Jest to wystarczająca liczba zdarzeń, ponieważ każde zdarzenie może rejestrować kolejne zdarzenia w liście np. zdarzenie, że autobus przyjechał na przystanek doda do listy zdarzenie, że za kilkanaście sekund autobus skończy obsługiwać ten przystanek.

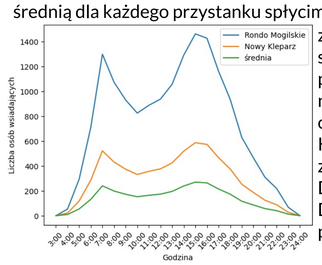
Aby model działał prawidłowo, należy wyposażyć się w odpowiednie dane, wszystkie dane statyczne pobierzemy z plików GTFS (*ang. General Transit Feed Specification*) , jest to standard opracowany przez inżynierów Google, do ujednolicenia danych transportowych. Dane te są używane przez wiele popularnych aplikacji takich jak Google Maps, jakdojade.pl, czy czynaczas.pl do śledzenia położenia pojazdów w czasie rzeczywistym, a także rozkładów jazdy. My z tych plików pobierzemy przede wszystkim listę przystanków, a także rozkład jazdy Komunikacji Miejskiej w Krakowie. Ponieważ przeprowadzenie takiej symulacji dla wszystkich linii byłoby bardzo ciężkie obliczeniowo, my ograniczymy się tylko do linii (co za tym idzie przystanków też) zawartych granicach miasta Kraków (wszystkie linie 5XX, 1XX i tramwajowe). Pozostałą część danych znajdziemy w Kompleksowym Badaniu Ruchu w Krakowskim Obszarze Metropolitalnym 2023. Jest to cykliczne badanie przeprowadzane na zlecenie UM Kraków, gdzie bada się m.in. liczbę pasażerów przewożonych przez KMK. My zwrócimy uwagę jedynie na wybrane dane z tego badania. Jest to wykres opisujący wymianę pasażerską na przedziale godziny na wszystkich badanych przystankach. Dane te zostały zmierzone przez człowieka w związku z tym zakładamy ich poprawność i na podstawie nich będziemy bazować dalej.



Rysunek 3.5 Histogram dobowy liczby osób wsiadających i wysiadających wraz z napełnieniem dla wszystkich badanych linii KMK

Ponieważ te dane są zbiorczo podane, można w ten sposób wyznaczyć średnią wymianę pasażerską dla danej grupy przystanków. Od tego momentu przystankiem będziemy nazywać grupę miejsc, gdzie zatrzymują się pojazdy KMK. W ten sposób mówiąc przystanek „Politechnika” będziemy mieli na myśli miejsce zatrzymania przy ul. Pawia, ul. Warszawskiej itd.

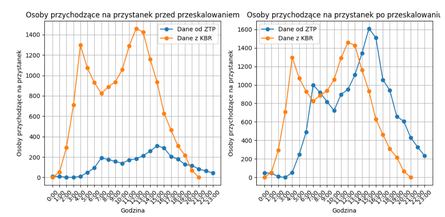
We wspomnianym badaniu jesteśmy w stanie znaleźć informację o sumarycznej liczbie obsłużonych pasażerów dla kilku przystanków. Między innymi dla przystanku „Rondo Mogilskie” jest to liczba 64 081. Ponieważ, jest to liczba wymiany pasażerskiej (wsiadających i wysiadających z pojazdu) podzielimy ją przez dwa, zakładając równość między tymi liczbami. Dane więc w czasie będą wyglądać w następujący sposób. Zwróćmy uwagę, że kiedy zastosujemy średnią dla każdego przystanku spłycimy nasze badania. Wartości takie jak przystanki



z dużą wymianą pasażerską będą niedoszacowane, tak samo jak przystanki, które „otrzymają” dodatkowych pasażerów. Postaramy się zatem znaleźć jakiś inny sposób na dotarcie do tych danych. W tej kwestii najłatwiej odezwać się do Zarządu Transportu Publicznego w Krakowie (ZTP). ZTP udostępniło nam dane liczbowe, zbierane przez kamery/bramki liczące w pojazdach KMK. Dane te zostały udostępnione za cały listopad 2025 roku. Dokładne szczegóły tych danych można znaleźć w poniższej tabeli.

Data	Linia	Kierunek	Przystanek	Godzina	Wejścia	Wyjścia
2025-11-24	152	Cmentarz Olszanica	Pilotów	09:29:10	42	1
2025-11-24	152	Czyżyny Dworzec	Teatr Bagatela	15:09:10	37	20

Ponieważ dane te nie są w żaden sposób powiązane z tymi w plikach GTFS, musimy je ręcznie przyporządkować. Do tego celu użyjemy biblioteki *pandas*. Zwróćmy uwagę, że po nałożeniu tych danych na siebie wykresy nie pokrywają się. Dlaczego? Okazuje się, że dane które otrzymaliśmy z urzędu są wybrakowane, otrzymujemy tylko część informacji z rozkładu jazdy. Skorzystamy teraz z wcześniejszego założenia, że dane z KBR są prawdziwe. Ponieważ dane z KBR i od ZTM opisują dokładnie to samo zjawisko, znajdziemy taką stałą k, że suma danych ZTP pomnożonych przez tą stałą będzie równa sumie z KBR. Robiąc proste obliczenia wychodzi, że stała $k = 5.21$.



Ponieważ przystanek „Rondo Mogilskie” jest największym centrem przesiadkowym, zakładamy, że wystarczy te dane dopasować tylko do jednego przystanku. A każdy następny przystanek będzie tak samo niedoszacowany. Następnym krokiem jest wyznaczenie odpowiednich wartości dla każdej grupy przystanków.

Dla każdego przystanku przypiszemy teraz średnią liczbę (x) pasażerów (odrzucimy wartości z sobót, niedziel i świąt) przychodzących na ten przystanek w oknie 5 minutowym. Korzystając z fizycznej interpretacji rozkładu wykładniczego (jako czasu do następnego zdarzenia) zbudujemy teraz dla każdego okna czasowego rozkład wykładniczy o parametrze $\lambda = 5/x$. Kolejnym krokiem w procesie przetwarzania tych danych będzie sprawienie, żeby były one ciągłe. Do tego celu skorzystamy z homotopii. Niech będą dwa ciągłe odwzorowania $f, g: X \to Y$ Homotopią nazwiemy takie ciągłe odwzorowanie $H: X \times [0,1] \to Y$, że $H(x,0) = f(x)$ i $H(x,1) = g(x)$. Przykładowe homotopie dla dwóch rozkładów wykładniczych f i g o parametrach λ_i i λ_j:

- *homotopia liniowa* $H(x,y) = \lambda_t e^{-\lambda_t x}$ *, gdzie* $\lambda_t = t(\lambda_j - \lambda_i) + \lambda_i$,
- *mieszanina rozkładów* $H(x,t) = (1-t) \cdot f(x) + \mathrm{t} \cdot \mathrm{g}(x)$.

Jesteśmy teraz gotowi przejść do modelowania ruchu pasażerów.

Summary in English. This project presents a Discrete Event Simulation (DES) model designed to analyze passenger behavior within the public transport network of Kraków, Poland. The foundation of the model lies in event-driven logic: rather than processing time continuously, the simulation tracks only the moments when meaningful events occur — such as a passenger boarding or alighting a vehicle, or a bus arriving at or departing from a stop. This approach significantly reduces computational overhead. Static data is sourced from GTFS files — a standard developed by Google engineers to unify transit data — and includes stop locations and timetables for Kraków's Municipal Transport (KMK). The simulation is scoped to lines operating within city limits. Passenger flow data is drawn from the 2023 Comprehensive Traffic Study of the Kraków Metropolitan Area (KBR), a cyclical survey commissioned by the City of Kraków. Since KBR and ZTM timetable data describe the same phenomenon but differ in magnitude, a scaling constant (k = 5.21) was derived to align them. Passenger arrival rates at each stop are modeled using an exponential distribution with parameter λ = 5/x, where x represents the average number of passengers per 5-minute window. Homotopy is applied to ensure continuity between time windows. Passengers navigate the network via a dynamic (time-dependent) graph, with routes computed using Dijkstra's algorithm. The model is implemented in Python using SimPy, pandas, and NumPy. All events are logged to a database, enabling analysis in tools such as Microsoft Power BI. Planned extensions include live delay tracking, passenger habit modeling, route editing, and an analysis of a potential metro line's impact on the existing network.

Nasi pasażerowie powinni teraz potrafić dobrać dla siebie odpowiednią trasę. Możemy więc zbudować graf, w którym wierzchołkami będą przystanki, z kolei krawędziami będą połączenia komunikacją miejską. Zwróćmy uwagę, że nie jest to taki zwykły graf skierowany, ponieważ pasażer szukający najszybszej trasy dla siebie o godzinie 10:00 nie powinien przeszukiwać wcześniejszych połączeń. W związku z tym jest to graf, który w zależności od czasu będzie te krawędzie usuwać. Taki graf nazwiemy grafem dynamicznym, chociaż w transporcie często mówi się o takim grafie jako o grafie transportowym. Jeśli mamy już działający graf zaimplementujemy algorytm Dijkstry, który będzie przeszukiwać najkrótszą trasę dla pasażerów.

Cały projekt został napisany w języku *Python*, przy pomocy takich bibliotek jak *simpy*, *pandas i numpy*. Na dzień dzisiejszy model zawiera wszystkie wspomniane elementy i jest gotowy jako bardzo podstawowe narzędzie do analizy zachowań pasażerów. Wszystkie wydarzenia zapisywane są w bazie danych, co pozwala już teraz analizować je w zewnętrznych programach takich jak *Microsoft PowerBI*, *Tableau*, czy *Google Looker*.

Model DES jest bardzo elastyczny, można go rozbudowywać o kolejne elementy, które będą zwiększać jego dokładność. W omówionym dzisiaj modelu, mam zamiar wdrożyć:

1. dokładniejszą analizę danych o pasażerach, aby dane o średniej liczbie pasażerów wsiadających i wysiadających na przystankach były bardziej miarodajne,
2. zbieranie „na żywo” danych na temat opóźnień, aby model był bardziej dokładny - dzięki temu istniałaby możliwość prognozowania przyszłości, a nie tylko opisywania przeszłości,
3. zwrócenie uwagi na przyzwyczajenia pasażerów (np. część pasażerów woli podjechać troszkę dłużej, zamiast przejść się kawałek pieszo),
4. większą możliwość edytowania tras, dzięki czemu zaistniałaby możliwość prognozowania zmian w rozkładzie jazdy i ich wpływu na pozostałe linie,
5. przeprowadzenie przy pomocy tego modelu, analizy wpływu potencjalnego metra na komunikację miejską w Krakowie.

Bibliografia:

- Dan Chalk, Sammi Rosser and Amy Heather, “HSMA - the little book of DES”, 2024.

Koło Naukowe Matematyków, Katedra Matematyki Stosowanej,
ul. Warszawska 24, 31-155 Kraków
kontakt.sknmpk@gmail.com, dr Marcin Skrzyński