

MACIEJ HELBIN

mgr, Zarząd Komunikacji Miejskiej
w Gdyni, ul. Zakręt do Oksywi 10,
81-244 Gdynia, e-mail: m.helbin@
zkmgdynia.pl

OLGIERD WYSZOMIRSKI

prof. dr hab., Uniwersytet Gdański,
Katedra Rynku Transportowego, ul.
Armii Krajowej 119/121, 81-824
Sopot, e-mail: o.wyszomirski@wp.pl

Możliwości wykorzystania Big Data w badaniach popytu i podaży w transporcie miejskim¹

Streszczenie: W badaniach popytu i podaży w transporcie miejskim powinno stosować się zarówno metody tradycyjne, jak i nowoczesne. Te ostatnie są związane z wykorzystywaniem Big Data. W artykule przedstawiono badania popytu i podaży jako podstawę kształtowania oferty przewozowej transportu miejskiego. Następnie omówiono tradycyjne metody badania popytu potencjalnego i czasów jazdy pojazdów tego transportu. Na ich tle zwrócono uwagę na coraz większe możliwości wykorzystywania Big Data w tych badaniach. Określono ich specyfikę, posługując się m.in. doświadczeniami praktycznymi Zarządu Komunikacji Miejskiej w Gdyni jako organizatora transportu miejskiego. Jednostka ta dysponuje zbiorami danych z systemu nadzoru miejskiego ruchu drogowego Tristar powiązanymi z programem Busman służącym do opracowywania i udostępniania pasażerom rozkładów jazdy oraz współuczestniczy z opracowywaniem koncepcji nowoczesnego systemu biletowego, który ma zapewnić istotny zbiór danych dla badań popytu i podaży.

Słowa kluczowe: transport miejski, badania popytu i podaży, Big Data.

Wprowadzenie

W celu racjonalnego organizowania, zarządzania i finansowania usług transportu miejskiego niezbędne jest posiadanie i wykorzystywanie informacji z badań popytu i podaży. Informacje te można uzyskać, prowadząc badania metodami tradycyjnymi opartymi na zbieraniu danych bezpośrednio przez człowieka. Stosowanie tych metod jest obciążone jednak licznymi wadami. Od pewnego czasu dzięki rozwojowi technologii możliwe stało się stosowanie nowoczesnych technik. Umożliwiają one gromadzenie i analizowanie dużej liczby danych. Zjawisko to nazwano Big Data. Pozwala ono na lepsze zrozumienie otaczającej rzeczywistości. W związku z tym dotychczasowe tradycyjne metody badawcze zastępowane są w coraz większym zakresie nowymi. Celem opracowania jest przedstawienie możliwości wykorzystania Big Data w transporcie miejskim w badaniach popytu i podaży. Dążąc do realizacji tego celu, porównano metody badań na przykładzie wybranych systemów umożliwiających automatyczne zbieranie danych.

Badania popytu i podaży jako podstawa kształtowania oferty przewozowej

Podstawowym zadaniem transportu miejskiego jest oferowanie odpowiedniej jakości usług przewozowych wewnątrz obszarów zurbanizowanych. Usługi te powinny charakteryzować się określonymi cechami, których zbiór tworzy ofertę przewozową [1].

Oferta przewozowa transportu powinna być stale weryfikowana przez zachowania transportowe klientów-pasażerów i nieustannie ulepszana [2,4]. Kształtowanie oferty przewozowej transportu miejskiego polega na oddziaływaniu na ilość i jakość podaży na rynku, uwarunkowanych dostępnością do określonych czynników produkcji [3,4]. Wszystkie elementy oferty przewozowej powinny być do siebie dopasowane. Właściwym rozwiązaniem jest kompleksowe kształtowanie oferty przewozowej, przy uwzględnieniu nadrzędnej roli rozkładu jazdy w stosunku do parametrów pojazdów, systemu taryfowego i systemu dystrybucji [4].

Rozkład jazdy, jako podstawowy element oferty przewozowej, musi być opracowany według określonych zasad, zapewniających jego dostosowanie do potrzeb przewozowych, popytu i warunków ruchu [5]. W związku z tym jego konstrukcja powinna być oparta na podstawie informacji z badań:

- potrzeb przewozowych i popytu potencjalnego,
- popytu efektywnego,
- czasów jazdy.

Badania potrzeb przewozowych i popytu potencjalnego pozwalają na określenie potencjalnej wielkości popytu na planowanych liniach transportu miejskiego i powinny przesądzać o celowości ich uruchamiania. Badania popytu efektywnego dają natomiast obraz wykorzystania oferty przewozowej i wyznaczają obszary jej niezbędnej modyfikacji [4]. Jednym z istotnych etapów w procesie konstruowania rozkładu jazdy jest obliczanie międzyprzystankowych czasów jazdy. W celu dostosowania czasów rozkładowych do faktycznego ruchu drogowego i pasażerskiego obliczenia te powinny być oparte na pomiarach czasowych przejazdu pojazdów pomiędzy poszczególnymi przystankami [6].

Obecnie potrzeby przewozowe i popyt potencjalny bada się głównie tradycyjnymi metodami, z których najlepiej sprawdza się w praktyce wywiad indywidualny przeprowadzany w gospodarstwie domowym respondentą. W przypadku popytu efektywnego oraz czasów jazdy, obok metod tradycyjnych, upowszechniają się nowoczesne metody oparte na zautomatyzowanym zbieraniu i analizowaniu danych.

Tradycyjne badania popytu efektywnego

Badanie popytu efektywnego służy przede wszystkim do określenia liczby i struktury pasażerów oraz do analizy na tej podstawie poziomu rentowności poszczególnych linii transportu miejskiego z dokładnością nawet do pojedynczych

¹ ©Transport Miejski i Regionalny, 2019. Wkład autorów w publikację: M. Helbin 50%, O. Wyszomirski 50%.

kursów. Podstawową tradycyjną metodą wykorzystywaną podczas badań popytu efektywnego jest obserwacja. W zależności od zakresu obserwatorzy mogą prowadzić badania, przebywając wewnątrz lub na zewnątrz pojazdu. W celu standaryzacji obserwacji dane powinny być rejestrowane na specjalnie przygotowanych instrumentach pomiarowych [1]. Obserwacje prowadzone wewnątrz pojazdów przez doświadczonych obserwatorów pozwalają określić liczbę pasażerów z błędem nie przekraczającym 3–5%, przy wypełnieniu sięgającym 120 osób w pojeździe [1,7]. Obserwacje na zewnątrz pojazdów są znacznie mniej dokładne, ale pozwalają na jednoczesne zbadanie większej liczby kursów. W przypadku badań prowadzonych na zewnątrz pojazdów zasadnicze znaczenia ma ustalenie punktów obserwacji, które powinny znajdować się w miejscach umożliwiających określenie przeciętnego napelnienia pojazdów na danym odcinku, przy czym ważne jest, aby obserwacje w poszczególnych punktach były prowadzone w tym samym czasie [1].

Zbieranie danych dotyczących popytu efektywnego tradycyjnymi metodami wiąże się z następującymi problemami metodologicznymi i organizacyjnymi:

- próba poddawana badaniom jest niewielka;
- dane zbierane podczas badań organizowanych 1–2 razy do roku mogą być niereprezentatywne, np. zakłócone przez warunki pogodowe;
- proces zbierania i obróbki danych jest czasochłonny;
- koszty zbierania i obróbki danych są wysokie;
- człowiek, realizujący badania wykorzystuje swoje zmysły, które są zawodne, a ponadto wykazuje skłonność do niedokładnej, a nawet nieuczciwej pracy, co powoduje konieczność pracochłonnego i kosztownego nadzoru przez człowieka o tych samych niedoskonałościach.

Tradycyjne badania czasów jazdy

Pomiary czasów jazdy pojazdów transportu zbiorowego mogą być realizowane zarówno w punktach sieci (zazwyczaj na kolejnych przystankach), jak i w pojazdach transportu zbiorowego na poszczególnych liniach. Pomiary w punktach sieci mają zastosowanie głównie w badaniach o charakterze ogólnym lub w badaniach korytarzowych. Umożliwiają pozyskanie danych odnoszących się do wszystkich pojazdów przejeżdżających przez punkt pomiarowy. Pomiary tego rodzaju charakteryzują się podatnością na błędy obserwacji, gdyż obserwator (na ogół wyposażony w długopis i formularz) ma niewiele czasu na rejestrację danych, zwłaszcza w sytuacji pojawienia się w punkcie pomiarowym więcej niż jednego pojazdu. Z kolei pomiary prowadzone w pojazdach umożliwiają pozyskiwanie dokładnych i wiarygodnych danych dotyczących zmienności położenia konkretnego pojazdu w czasie i przestrzeni. Pomiary wykonywane przez obserwatorów wymagają zazwyczaj zaangażowania wielu osób, jednak zwłaszcza w przypadku badań szczegółowych, nastawionych na precyzyjne odwzorowanie procesów ruchu (np. czas wymiany, czas oczekiwania), są bardziej efektywne od pomiarów w pełni zautomatyzowanych (rejestrujących z góry ustalony zakres wyników) [8].

Pomiary czasów jazdy pojazdów transportu zbiorowego prowadzi na co dzień w ramach swojej działalności Zarząd Komunikacji Miejskiej w Gdyni. Od momentu powstania ZKM w 1992 roku pomiary te wykonywane były przez obserwatorów wyposażonych w standaryzowane karty pomiarowe i zegarki. W zależności od potrzeb były one przeprowadzane na przystankach lub w pojazdach transportu miejskiego. Zbierane dane, poza dopasowywaniem rozkładowych czasów przejazdu do rzeczywistych, służyły także do generowania statystyk regularności i punktualności świadczonych usług oraz kontroli jakości usług świadczonych przez operatorów. Dużą wadą tzw. ręcznej metody zbierania danych był fakt, iż nawet przy stałym zatrudnieniu na stanowiskach obserwatorów 5 osób pozwalała ona na zbadanie w sieci ZKM w Gdyni jedynie próby około 2% wszystkich odjazdów z przystanków w skali roku [6].

Big Data w badaniach popytu efektywnego

Tradycyjne metody badań popytu i podaży w transporcie miejskim mają jedną cechę wspólną – przy dużych nakładach pracy i środków umożliwiają zebranie jedynie danych dotyczących mniejszej lub większej próby.

Obecnie badanie próby to pozostałość ery ograniczeń w zbieraniu i przetwarzaniu danych. W wielu obszarach dochodzi do zmiany – zamiast zbierania części danych gromadzi się możliwie wiele, a jeśli to wykonalne wszystkie, tj. $N = \text{całość}$. Ograniczenia w gromadzeniu i analizowaniu nie są już takim problemem. Czujniki, telefony wyposażone w GPS, Internet czy Twitter, pasywnie zbierają dane, a komputery radzą sobie z liczbami z rosnącą łatwością [9].

Pół wieku po wejściu komputerów do powszechnego użytku informacje tak się zaczęły kumulować, że powstaje coś nowego i wyjątkowego. Świat nie tylko jest zalewany niespotykaną wcześniej liczbą danych, ale ich ilość rośnie coraz szybciej. Zmiana ilości doprowadziła do zmiany jakości. W naukach, takich jak astronomia czy genetyka, które pierwsze w XXI wieku zmierzyły się z olbrzymią ilością danych, sformułowano termin „Big Data”. Termin ten przenosi się obecnie na wszelkie obszary ludzkiej działalności [9]. Jednym z nich jest transport.

Geolokalizacja elementów przyrody, obiektów i tras, jakimi poruszają się poszczególne osoby, generuje dane. Do danej sytuacji położenia konieczne jest spełnienie kilku warunków:

- pomiar każdego centymetra kwadratowego powierzchni Ziemi,
- standaryzacja sposobu zapisania danych z pomiarów,
- instrumentalizacja do monitorowania i zapisania danych [9].

W 1978 roku wystrzelono pierwszy z 24 satelitów tworzących system GPS (ang. Global Positioning System). Osoby na Ziemi mogą za pomocą triangulacji określić swoją pozycję dzięki różnicom czasowym w sygnale docierającym do nich z różnych satelitów zawieszonych 20 300 kilometrów nad ich głowami. System stworzony przez Departament Obrony Stanów Zjednoczonych po raz pierwszy można było używać do niemilitarnych celów w latach 80.

ubiegłego wieku, a w kolejnej dekadzie stał się powszechnie użyteczny. Wprowadzenie tego systemu, za pomocą którego można określić pozycję z dokładnością do jednego metra, wyznaczyło moment, w którym pomiar położenia stał się dzięki środkom technicznym szybki, tani i łatwy do wykonania dla osób bez specjalistycznej wiedzy [9]. Coraz powszechniejsze urządzenia GPS montowane w pojazdach transportu miejskiego, samochodach i telefonach komórkowych, w połączeniu z systemami informatycznymi i aplikacjami mobilnymi pozwalają na automatyzację zbierania danych dotyczących popytu i podaży w transporcie miejskim.

Wraz z wprowadzeniem nowego typu taboru dostęp do danych na temat jego aktualnej lokalizacji, liczby pasażerów przebywających w pojazdach czy punktualności odjazdu staje się coraz łatwiejszy. Każdy nowo kupowany pojazd przez przedsiębiorstwa transportowe, obsługujące duże miasta lub aglomeracje miejskie, wyposażony jest standardowo w urządzenie lokalizacyjne GPS, autokomputer oraz systemy łączności z dyspozytornią. Do podstawowych systemów stosowanych w pojazdach transportu miejskiego coraz częściej dołączają automatyczne systemy zliczania pasażerów. Rozwiązania tego typu dostępne są już u wielu producentów, a ich montaż jest możliwy na etapie produkcji pojazdu [10].

Sposób przedstawiania wyników pomiarów w formie macierzy w arkuszu kalkulacyjnym ułatwia i skraca proces opracowywania i wgrywania danych wejściowych. Macierze te już na etapie pobierania danych z systemu są odpowiednio przygotowane i sformatowane. Dzięki temu możliwe jest proste wykorzystanie ich w celu wykonania prognozy popytu na usługi transportowe. System dostarcza dane o popycie obejmujące:

- lokalizację przystanku (współrzędne GPS);
- dokładny czas, w którym pojazd zatrzymał się na przystanku;
- liczbę pasażerów, którzy wsiadli do pojazdu;
- liczbę pasażerów, którzy wysiedli z pojazdu;
- liczbę pasażerów w pojeździe po odjeździe z przystanku [10].

Duża różnorodność automatycznie zbieranych danych umożliwia stworzenie modelu prognozy nie tylko dla popytu,

ale także punktualności odjazdów w przekroju całej sieci transportowej oraz dla wybranych linii lub ciągów transportowych. Prognozy te przydatne są szczególnie przy optymalizowaniu częstotliwości kursowania pojazdów oraz optymalizacji przesiadek [10].

Systemy automatycznego zliczania pasażerów znoszą ograniczenia dotyczące próby – umożliwiają zebranie dowolnej ilości danych. Problemy jednak mogą występować w kwestii dokładności pomiarów. W literaturze przedstawiono analizę zbieżności danych z systemu DILAX z danymi z pomiarów przeprowadzonymi przez obserwatorów przy zachowaniu największej ich dokładności. Wyniki przeprowadzonych pomiarów pogrupowano zgodnie z wcześniej przyjętym podziałem taboru ze względu na typ pojazdu (standardowe – 12 metrowe i długie – 18 metrowe). Odrębnie analizowano wejścia i wyjścia do pojazdów. Następnym krokiem było uszeregowanie wyników pod względem liczby osób wsiadających i wysiadających od wartości minimalnej do maksymalnej. Na koniec porównano dane z klasycznych pomiarów wykonanych przez obserwatorów w pojazdach oraz dane z automatycznego systemu zliczania pasażerów. Przy uwzględnieniu przedziału zgodności z dopuszczeniem różnicy pięciu pasażerów między danymi z obu źródeł pomiarów uzyskano prawie 100% zgodność zgromadzonych danych dla autobusu standardowego i 60% zgodność dla autobusu przegubowego. Wykonane analizy porównawcze wykazały, że dokładność danych nie tylko zależy od liczby pasażerów wsiadających i wysiadających, ale także od wielkości pojazdu. Dodatkowo badania wykazały przydatność danych z systemu automatycznego w procesie prognozowania popytu, jednakże należy rozszerzyć i powtórzyć w niektórych przypadkach pomiary porównawcze oraz wspierać działania mające na celu podwyższenie dokładności zbierania danych [10].

System DILAX testowany był także przez Zarząd Komunikacji Miejskiej w Gdyni. Na rysunku 1 przedstawiono przykładowy raport z systemu, przedstawiający dane zebrane na jednej z linii trolejbusowych. Zapisane w ustrukturyzowanej postaci dane pozwalają na analizę napełnienia pojazdów w przekrojach linii, tras, przystanków, przedziałów czasowych, gmin czy też dla całej sieci.

Data	Zadanie	Pojazd	Pojemność pojazdu	Nr przystanku	Przystanek	Czas odjazdu (plan.)	Czas przyjazdu (fakt.)	Czas odjazdu (fakt.)	Wejścia	Wyjścia	Napełnienie przy odjeździe	Wskaźnik napełnienia pojazdu
09.07.2018	27-01	3076	70	1	7.20-Cisowa SKM	14:52:00	14:39:21	14:52:37	7.00		7.00	10.00%
09.07.2018	27-01	3076	70	2	7.21-Piaskowa	14:53:00	14:53:59	14:54:22	3.00	0.00	10.00	14.29%
09.07.2018	27-01	3076	70	3	7.22-Kocynska	14:54:00	14:55:01	14:55:20	4.00	0.00	14.00	20.00%
09.07.2018	27-01	3076	70	4	7.23-Chylonia Dworzec PKP	14:56:00	14:56:33	14:57:00	8.00	2.00	20.00	28.57%
09.07.2018	27-01	3076	70	5	7.24-Pucka	14:57:00	14:58:03	14:58:32	7.00	3.00	24.00	34.29%
09.07.2018	27-01	3076	70	6	7.25-Leszczynki SKM	14:58:00	14:59:45	15:00:07	1.00	2.00	23.00	32.86%
09.07.2018	27-01	3076	70	7	7.26-Zamenhofa	14:59:00	15:00:49	15:01:11	0.00	1.00	22.00	31.43%
09.07.2018	27-01	3076	70	8	7.40-Wezeł Żołnierzy Wykolejonych	15:01:00	15:02:17	15:02:42	2.00	4.00	20.00	28.57%
09.07.2018	27-01	3076	70	9	7.41-Morska - Estakada	15:03:00	15:04:28	15:04:52	1.00	6.00	15.00	21.43%
09.07.2018	27-01	3076	70	10	7.42-Kalksztajnow	15:04:00	15:05:41	15:06:01	1.00	1.00	15.00	21.43%
09.07.2018	27-01	3076	70	11	7.44-Mireckiego	15:06:00	15:07:21	15:07:49	2.00	0.00	17.00	24.29%
09.07.2018	27-01	3076	70	12	7.45-Akademia Morska	15:08:00	15:08:48	15:09:09	2.00	1.00	18.00	25.71%
09.07.2018	27-01	3076	70	13	7.46-Stocznia SKM - Morska	15:09:00	15:09:50	15:10:10	3.00	0.00	21.00	30.00%
09.07.2018	27-01	3076	70	14	7.47-Bpa Okoniewskiego	15:10:00	15:10:51	15:11:11	0.00	1.00	20.00	28.57%

Rys. 1. Dane zebrane przez automatyczny system zliczania pasażerów na linii 27 w Gdyni

Źródło: materiały Zarządu Komunikacji Miejskiej w Gdyni

Wykorzystanie automatycznych systemów liczenia pasażerów ma określone zalety, a mianowicie:

- zamiast niewielkiej próby uzyskuje się całą zbiorowość pasażerów w wybranym przedziale czasu;
- zbieranie danych następuje automatycznie bez udziału człowieka;
- automatycznie opracowywane są raporty z badań;
- wyeliminowanie człowieka wykorzystującego jako instrument pomiarowy zmysł wzroku i skłonnego do odstępstw od nałożonych standardów może zwiększyć dokładność danych;
- dane przetwarzane są na bieżąco w czasie rzeczywistym, a nie z mniejszymi lub większymi opóźnieniami w miarę posiadanych możliwości;
- bieżące zbieranie danych umożliwia wykrycie czasu i miejsc przepełnień pojazdów, które trudno jest zidentyfikować podczas rutynowych badań (np. w piątki wieczorem podczas imprez masowych);
- zapewnia dane dotyczące popytu (bez ponoszenia dodatkowych kosztów) w dni nietypowe, w które z reguły nie prowadzi się badań, np. święta, długie weekendy;
- pozwala, dzięki posiadaniu kompleksowych danych, na znalezienie w nich nowych wzorców i prawidłowości.

Systemy zliczania pasażerów stają się nie tylko dokładniejsze, ale często umożliwiają pozyskiwanie danych w czasie rzeczywistym [10, 14]. Podstawowe systemy automatyczne umożliwiają jednak wyłącznie określenie wielkości potoków pasażerskich między poszczególnymi przystankami. Jest to przydatne w takich działaniach jak:

- planowanie przydziału taboru do linii,
- określanie częstotliwości kursowania taboru,
- ocenianie warunków podróży i poziomu komfortu w pojazdach [10].

W przypadku potrzeby zbudowania macierzy podróży uwzględniającej źródła i miejsce docelowe podróży tego typu systemy mogą okazać się niewystarczające. Wówczas mogą być wykorzystane dane z automatycznych systemów pobierania opłat, które umożliwiają także określenie źródła i miejsca docelowego każdej podróży transportem zbiorowym [10, 11, 12].

Systemy biletowe, w których pasażer posiada kartę uprawniającą do przejazdu wykrywaną w momencie wejścia do i wyjścia z pojazdu, mogą zostać wykorzystane jako bogate źródło danych. Pożądane jest automatyczne pobieranie danych w strukturze: ID biletu, pojazd, czas, wejście/wyjście. Systemy identyfikacji biletu mogą być bezobsługowe i bezdotykowe. Pasażer nie musi wówczas wykonywać żadnych dodatkowych czynności. Systemy automatycznego pobierania opłat zazwyczaj powstają z inicjatywy organizatora transportu w celu poprawienia jakości usług, zmniejszania kosztów dystrybucji i uzyskania danych dla analiz ekonomicznych. Z punktu widzenia funkcjonowania transportu systemy takie są bogatym źródłem danych o popycie, takich jak pasażerokilometry, napełnienie pojazdów, trasy przejazdów, częstość przejazdów,

czas podróży. Za pomocą tych danych można zbudować model popytu [11].

Systemy biletowe pobierają dane o określonym zdarzeniu w określonym miejscu, w określonym czasie, od określonego użytkownika. Użytkownik identyfikowany jest za pomocą rodzaju biletu (jednorazowy, sieciowy) lub za pomocą dodatkowych danych takich jak wiek i płeć. Baza danych z systemu biletowego stanowi ogromny potencjał do zmiany modelowania popytu. Popyt dotychczas aproksymowany za pomocą modeli, opartych na podstawie niepełnej bazy danych, może zostać skalibrowany do pełnej bazy danych o realizacji popytu. Pozwala to zwiększyć dopasowanie podaży do popytu, a tym samym efektywność ekonomiczną realizacji przewozów. Lepiej rozumiane mogą być przyczyny wyboru środka transportu i wpływ jakości podróży na wybór trasy [11].

Technologie analityczne mogą przetwarzać surowe dane o lokalizacji pojazdów w użyteczne dane dotyczące:

- grupowania się pojazdów na trasie,
- realizacji priorytetu pojazdów transportu zbiorowego w ruchu drogowym,
- czasu wysiadania i wsiadania pasażerów na przystankach,
- zależności czasu przejazdu od dnia tygodnia [11].

Big Data w badaniach czasów jazdy

W oparciu o GPS działa, wdrożony w 2015 roku przez miasta Gdańsk, Gdynię i Sopot, system zarządzania ruchem TRISTAR. Jednym z użytkowników systemu jest Zarząd Komunikacji Miejskiej w Gdyni. Podstawowym zadaniem tej inwestycji jest zwiększenie przepustowości układu drogowego i płynności ruchu. System TRISTAR wykorzystuje dane pochodzące z detektorów ruchu zlokalizowanych na wlotach skrzyżowań. Dzięki złożonym algorytmom obliczeniowym zapewnia automatyczne sterowanie sygnalizacjami świetlnymi, maksymalizuje przepustowość skrzyżowań, redukuje opóźnienia i zmniejsza liczbę zatrzymań środków transportu. Jednym z ważniejszych elementów zrealizowanej inwestycji jest moduł zarządzania pojazdami miejskiego transportu zbiorowego. System TRISTAR, dzięki usłudze pozycjonowania, zapewnia wizualizację na cyfrowej mapie wszystkich eksploatowanych środków transportu, z zaznaczeniem pojazdów opóźnionych i przyspieszonych w stosunku do rozkładu jazdy, co czyni go użytecznym narzędziem kontroli punktualności. Trójmiejski system zarządzania ruchem zapewnia także dokładne dane, które znajdują zastosowanie w ramach dynamicznej informacji pasażerskiej prezentowanej na przystankach, w Internecie i urządzeniach mobilnych. Dzięki możliwości rejestracji godzin odjazdów z poszczególnych przystanków organizator jest w stanie skutecznie optymalizować ofertę przewozową i korygować rozkłady jazdy [13].

Wraz z wdrożeniem w 2015 roku systemu TRISTAR prowadzone przez Zarząd Komunikacji Miejskiej w Gdyni pomiary czasów jazdy pojazdów komunikacji miejskiej za pomocą obserwatorów zostały znacznie ograniczone. Obecnie są one stosowane tylko w sytuacjach, w których

konieczne jest uzyskanie danych dotyczących zdarzeń nie rejestrowanych przez system (np. przyczyn opóźnień w danym miejscu). W codziennej praktyce zaczęto wykorzystywać zapisane w bazie danych czasy jazdy zarejestrowane przez system TRISTAR. W system ten wyposażone są wszystkie pojazdy obsługujące zadania w sieci ZKM w Gdyni. Urządzenia zamontowane w pojazdach rejestrują czas przyjazdu i odjazdu z każdego przystanku na trasie pojazdu. Wdrożenie systemu spowodowało zatem, iż niewielkie próby badane przez obserwatorów zastąpił zbiór danych zbliżający się do opisanego wcześniej, pożądanego $N = \text{całość}$ (braki danych na poziomie nie przekraczającym 5% występują jedynie w przypadku awarii urządzeń czy zakłóceń sygnału GPS) [6].

Wyposażone w system GPS urządzenia zainstalowane we wszystkich pojazdach w sieci ZKM w Gdyni wysyłają do systemu TRISTAR telegramy zawierające dane identyfikacyjne pojazdu, numer linii wraz z numerem zadania przewoźowego oraz powiązane z rozkładem jazdy dane o faktycznych czasach przyjazdów i odjazdów z przystan-

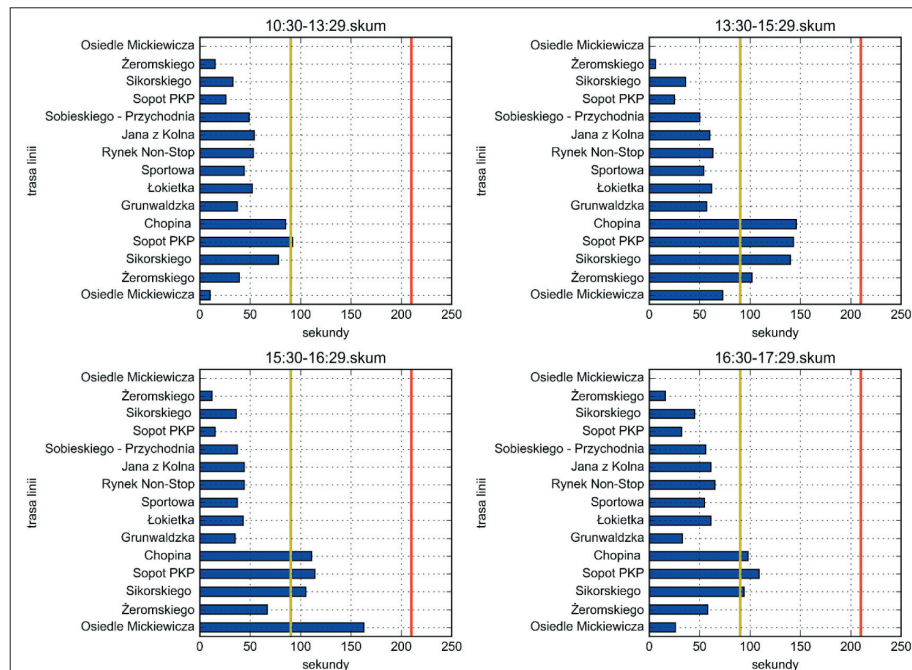
ków. Dane te rejestrowane są w formie znaczników czasowych (rok, miesiąc, dzień, godzina, minuta, sekunda) przy każdym wjeździe i wyjeździe pojazdów ze zdefiniowanej w metrach strefy przystankowej i na bieżąco przekazywane są do Centrum Zarządzania Ruchem. Służą one dyspozytorom do nadzoru i regulacji ruchu. Wszystkie dane zapisywane są w relacyjnej bazie danych [6].

Relacyjna baza danych TRISTAR umożliwia przeprowadzenie analizy faktycznych czasów jazdy w odniesieniu do rozkładowych w istotnych z punktu widzenia układania rozkładów jazdy przekrojach. Dane mogą być analizowane w przekroju linii i przystanków oraz zdefiniowanych przez organizatora transportu przedziałów czasowych i rodzajów dni (powszedni, sobota, niedziela). Na rysunku 2 zaprezentowano przykładową kwerendę będącą wynikiem zapytania do bazy danych TRISTAR, natomiast na rysunku 3 przykładową wizualizację opóźnień na wybranej linii w przekroju wyszczególnionych przez organizatora transportu przedziałów czasowych, obowiązujących dla danej linii [6].

data	linia	zadanie	pojazd	przystanek	nr_przyst	rozkładowy_przyjazd	faktyczny_przyjazd	rozkładowy_odjazd	faktyczny_odjazd
2018-09-16	144	144-01	7860	Osiedle Mickiewicza	0	09:21:00.0000000	08:57:52.0000000	09:21:00.0000000	09:21:31.0000000
2018-09-16	144	144-01	7860	Zeromskiego	1	09:22:00.0000000	09:22:19.0000000	09:22:00.0000000	09:22:44.0000000
2018-09-16	144	144-01	7860	Sikorskiego	2	09:23:00.0000000	09:23:40.0000000	09:23:00.0000000	09:24:19.0000000
2018-09-16	144	144-01	7860	Sopot PKP	3	09:26:00.0000000	09:25:51.0000000	09:26:00.0000000	09:26:31.0000000
2018-09-16	144	144-01	7860	Sopot PKP - Marynarzy	4	09:27:00.0000000	09:26:56.0000000	09:27:00.0000000	09:27:32.0000000
2018-09-16	144	144-01	7860	Władysława IV	5	09:28:00.0000000	09:28:31.0000000	09:28:00.0000000	09:29:00.0000000
2018-09-16	144	144-01	7860	Jana z Kolna	6	09:29:00.0000000	09:29:22.0000000	09:29:00.0000000	09:29:50.0000000
2018-09-16	144	144-01	7860	Rynek Non-Stop	7	09:30:00.0000000	09:30:16.0000000	09:30:00.0000000	09:30:48.0000000
2018-09-16	144	144-01	7860	Sportowa	8	09:31:00.0000000	09:31:22.0000000	09:31:00.0000000	09:31:38.0000000
2018-09-16	144	144-01	7860	Lokietka	9	09:32:00.0000000	09:32:18.0000000	09:32:00.0000000	09:32:58.0000000
2018-09-16	144	144-01	7860	Grunwaldzka	10	09:35:00.0000000	09:35:14.0000000	09:35:00.0000000	09:35:38.0000000
2018-09-16	144	144-01	7860	Chopina	11	09:36:00.0000000	09:36:23.0000000	09:36:00.0000000	09:36:57.0000000
2018-09-16	144	144-01	7860	Sopot PKP	12	09:37:00.0000000	09:37:41.0000000	09:37:00.0000000	09:38:15.0000000
2018-09-16	144	144-01	7860	Sikorskiego	13	09:39:00.0000000	09:40:05.0000000	09:39:00.0000000	09:40:27.0000000
2018-09-16	144	144-01	7860	Zeromskiego	14	09:41:00.0000000	09:41:22.0000000	09:41:00.0000000	09:41:41.0000000
2018-09-16	144	144-01	7860	Osiedle Mickiewicza	15	09:42:00.0000000	09:42:14.0000000	09:42:00.0000000	09:43:25.0000000

Rys. 2. Czasy jazdy zarejestrowane przez system TRISTAR na linii 144 ZKM w Gdyni

Źródło: materiały Zarządu Komunikacji Miejskiej w Gdyni

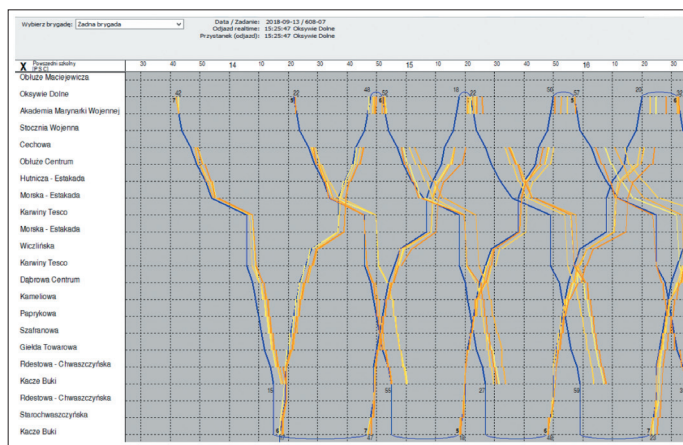


Rys. 3. Przedziały czasowe, w których na linii 144 ZKM w Gdyni wystąpiło co najmniej na jednym przystanku opóźnienie powyżej 1 minuty.

Źródło: opracowanie własne na podstawie danych z systemu TRISTAR

Zautomatyzowane zbieranie i przetwarzanie danych dotyczących czasów jazdy umożliwia bieżące analizowanie dużych ich zbiorów, co znacznie poprawia jakość i aktualność uzyskiwanych informacji. Duże zbiory danych umożliwiają zatem lepsze zarządzanie ofertą transportu miejskiego przez organizatora. Dzięki nim można ułożyć rozkłady jazdy lepiej dopasowane do warunków ruchu na drodze. Języki programowania automatycznie doprowadzają dane do pożądanej postaci, wyciągają czasy jazdy i wizualizują dane na wykresach. Narzędzia te umożliwiają analizę kompleksową i stanowią podstawę korekt rozkładów jazdy [6].

Rozwój technologii oraz standaryzacja zapisu danych umożliwiają integrację różnych systemów, np. nadzoru ruchu i oprogramowania do układania rozkładów jazdy. Przykładem jest wykorzystywane przez Zarząd Komunikacji Miejskiej w Gdyni w codziennej praktyce oprogramowanie Busman z modułem do analizy realizacji rozkładu jazdy pod kątem międzyprzystankowych czasów jazdy i urealniania rozkładów jazdy. Oprogramowanie to sprzężone jest z danymi z systemu TRISTAR. W graficzny sposób przedstawia ono realizację wykonanych kursów w zestawieniu z kursami zaplanowanymi w rozkładzie jazdy. Przykładową wizualizację rozkładu jazdy wybranej linii przedstawiono na rysunku 4.



Rys. 4. Wizualizacja rozkładu jazdy linii X ZKM w Gdyni w programie Busman

Źródło: materiały Zarządu Komunikacji Miejskiej w Gdyni

Pomarańczowe kursy systemu Busman to graficzny sposób prezentacji danych lokalizacyjnych (prezentowany od cienia mówi o czasie pozyskania danych – jasnożółte to kursy najstarsze, a ciemnopomarańczowe – najnowsze) zgromadzonych w bazie danych, pochodzących z urządzeń pokładowych pojazdów. Jednoczesna prezentacja założeń projektowych (linie niebieskie) oraz danych o czasach międzyprzystankowych zrealizowanych kursów pozwala w łatwy sposób zweryfikować założenia konstrukcji rozkładów i wprowadzić dokładne korekty planu kursów.

Podsumowanie

Dla prawidłowego funkcjonowania transportu miejskiego niezbędne są badania popytu i podaży. W badaniach tych coraz większego znaczenia nabiera wykorzystywanie Big Data. O ile w badaniach potrzeb przewozowych i popytu potencjalnego badacze posługują się głównie tradycyjnymi metodami, o tyle w badaniach popytu efektywnego i czasów jazdy metody tradycyjne są coraz częściej zastępowane wykorzystaniem Big Data i uzyskaniem próby $N=całość$. Nowoczesny sposób gromadzenia i analizy danych oparty na wielkich zbiorach danych nie ma wad metod tradycyjnych. Przede wszystkim zapewnia on możliwość zrezygnowania z próby na rzecz dowolnej wybranej liczby danych. Pozwala na oszczędność czasu i ograniczenie kosztów, wymaga jednak poniesienia nakładów inwestycyjnych. Jego walorem jest posługiwanie się w określonym zakresie danymi, które powstają przy okazji funkcjonowania systemów skonstruowanych z myślą o zapewnieniu wyższej jakości usług, czy to przez elektroniczną regulację i nadzór ruchu, czy też przez elektroniczny pobór opłat za przejazdy. Korzystając z danych wykorzystujących Big Data w badaniach popytu i podaży w transporcie miejskim, należy być jednak świadomym pewnych ograniczeń związanych z gromadzeniem i analizą danych, wynikających przede wszystkim z zawodności urządzeń i procesów.

Zbiory danych Big Data nie muszą mieć wielkości liczonej w terabajtach. Czasem $N=całość$ może zawierać mniej bitów, niż obecnie składa się na standardowe zdjęcie cyfrowe. Analiza z zakresu Big Data uwzględnia więcej danych, niż wynika to z użycia próby losowej. Kiedy

mówi się o Big Data, w mniejszym stopniu ma się na myśli wielkość w wartościach bezwzględnych, a w większym we względnych (względem całego zbioru danych). Przez długi czas metoda próby losowej była drogą na skróty. Przed erą cyfrową umożliwiała analizę problemów obejmujących wielką ilość danych. Jednak w takim przypadku traci się część danych. Posiadanie całego (lub prawie całego) zbioru danych zapewnia dużo większą swobodę jego eksplorowania, umożliwia przyglądanie się danym z różnych perspektyw czy też bliższe przyjrzenie się pewnym ich aspektom [9]. W przypadku badania popytu efektywnego oraz badania czasów jazdy w transporcie miejskim obecnie już można badać pożądane $N=całość$. Wiąże się to z wyposażeniem taboru w systemy zliczania pasażerów oraz systemy nadzoru ruchu.

Literatura

1. *Transport miejski. Ekonomia i organizacja*, red. O. Wyszomirski, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk 2008.
2. Bąkowski W., *Konkurencja w przewozach pasażerskich na poziomie regionalnym*, „Zeszyty Naukowo-Techniczne Stowarzyszenia Inżynierów i Techników Komunikacji Rzeczpospolitej Polskiej Oddział w Krakowie” Seria: Materiały Konferencyjne, nr 63, 2004.
3. Szałucki K., Wyszomirski O., *Efektywne zarządzanie komunikacją miejską w miastach i przedsiębiorstwami komunikacji miejskiej na przykładzie Gdyni*, [w:] *Efektywność transportu zbiorowego w miastach*, IV Ogólnopolska Konferencja „Komunikacja Miejska”, Materiały Konferencyjne, Izba Gospodarcza Komunikacji Miejskiej, Łódź 2000.
4. Hebel K., *Zachowania transportowe mieszkańców w kształtowaniu transportu miejskiego*, Fundacja Rozwoju Uniwersytetu Gdańskiego, Gdańsk 2013.
5. Gromadzki M., Hebel K., *Kształtowanie rozkładów jazdy jako elementu oferty przewozowej w transporcie miejskim. Aspekty teoretyczne i praktyczne*, „Logistyka”, 2008, nr 6, CD.
6. Helbin M., *Wykorzystanie bazy danych czasów jazdy do optymalizacji rozkładów jazdy transportu miejskiego w Sopocie*, Szkoła Główna Handlowa w Warszawie, Studia podyplomowe „Inżynieria danych – Big Data”, Warszawa 2016, praca dyplomowa niepublikowana.
7. Grzelec K., *Metody liczenia pasażerów w komunikacji miejskiej*, „Biuletyn Komunikacji Miejskiej”, 2006, nr 89.
8. Bauer M., *Porównanie technik pomiaru czasu przejazdu pojazdów transportu zbiorowego*, „Transport Miejski i Regionalny”, 2013, nr 7.
9. Mayer-Schönberger V., Cukier K., *BIG DATA – rewolucja, która zmieni nasze myślenie, pracę i życie*, MT Biznes sp. z o.o., Warszawa 2014.
10. Aleksandrowicz J., *Przydatność automatycznych systemów zliczania pasażerów w celach predykcji popytu na usługi transportowe*, „Transport Miejski i Regionalny”, 2018, nr 4.
11. Kucharski R., *Przegląd możliwych technologii ITS dla transportu zbiorowego*, IV Polski Kongres ITS, Warszawa 2017.
12. Wyraz E., Lorenc A.K., *System pomiaru napełnienia pojazdów komunikacji miejskiej – przegląd współczesnych rozwiązań*, „Logistyka”, 2015, nr 3.
13. Bujak K., *Znaczenie i uwarunkowania kontroli jakości usług przewozowych prowadzonej przez Zarząd Komunikacji Miejskiej w Gdyni*, „Transport Miejski i Regionalny”, 2014, nr 3.
14. <https://www.dilax.com/pl/> dostęp 10.09.2018.