

Metoda analizy ruchu drogowego na podstawie zachowania użytkowników sieci telefonii komórkowej

G.SABAK

e-mail:gsabak@wat.edu.pl

Wydział Cybernetyki WAT
ul. S. Kaliskiego 2, 00-908 Warszawa

Artykuł ten poświęcony jest możliwości wykorzystania informacji o przemieszczaniu się użytkowników sieci telefonii komórkowej do analizy sytuacji na drogach objętych zasięgiem tej sieci.

W pracy przedstawiony został formalny model danych lokalizacyjnych dostępnych w sieci, model drogi jako linii łamanej oraz model pokrycia drogi zasięgiem telefonii komórkowej. Przy użyciu wymienionych modeli sformułowany został problem uzyskiwania *informacji o ruchu drogowym*.

Artykuł zawiera propozycję metody wyznaczania informacji o ruchu na podstawie strumienia danych lokalizacyjnych. W celu wstępnej weryfikacji zaproponowanej metody, wykonane zostały obliczenia dla trasy Kraków-Zakopane, dla danych zebranych z sieci jednego z operatorów telefonii komórkowej. Wyliczona informacja drogowa została porównana z, uzyskaną na podstawie przejazdu testowego, informacją referencyjną.

Słowa kluczowe: lokalizacja CellId, transport, zatłoczenia

1. Wprowadzenie

Transport i telekomunikacja mobilna, ze względu na ich powszechne wykorzystanie, mają dziś olbrzymie znaczenie zarówno dla gospodarek państw i regionów, jak i dla poszczególnych obywateli. Można wymienić następujące cechy wspólne dla obu tych dziedzin:

- zajmują się zapewnianiem dostępności dóbr (transport) i informacji (telekomunikacja) w określonym miejscu i czasie
- w obu występuje konieczność zapewnienia oczekiwanej efektywności procesów przy jednoczesnym istnieniu ograniczeń na koszty ich działania
- problemy występujące w obu dziedzinach są podobne. Mianowicie są to m.in.: znajdowanie efektywnych dróg transportu, wykrywanie i unikanie blokad oraz zatłoczeń.

Te podobieństwa oraz fakt, że bardzo wiele osób jednocześnie korzysta z systemu transportu i z sieci telefonii komórkowej, skłaniają do przeanalizowania możliwości wykorzystania informacji o zmieniającej się lokalizacji użytkowników komórek do wnioskowania o stanie systemu drogowego.

Na rys. 1. przedstawione zostały wybrane ślady stacji mobilnych należących do sieci telefonii komórkowej jednego z operatorów w Polsce. Uwzględnione zostało 200 stacji

mobilnych, które w dniu 6.08.2009 r. przemieściły się na największe odległości.



Rys. 1. Ślady przemieszczeń wybranych stacji mobilnych

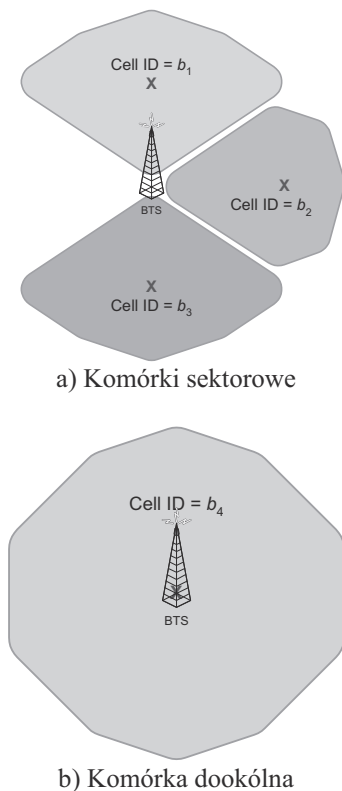
Na rysunku widoczne jest, że ślady te pokrywają się z przebiegiem głównych szlaków transportowych w naszym kraju. Potwierdza to przypuszczenie, że istnieje grupa użytkowników telefonów komórkowych, którzy w tym samym czasie korzystają z transportu samochodowego.

1.1. Metoda lokalizacji w sieci GSM na podstawie CellId

Jedną z najczęściej wykorzystywanych przez operatorów sieci komórkowych metod

lokalizacji użytkowników jest wykorzystanie informacji o identyfikatorze komórki (CellId), która w danej chwili obsługuje stację mobilną. Obsługa ta polega na realizacji komunikacji (zarówno przez kanał głosowy jak i sygnalizacyjny) pomiędzy stacją mobilną, a częścią stałą sieci telefonii komórkowej. W tej metodzie każdemu z identyfikatorów komórki przypisywane są współrzędne geograficzne wskazujące miejsce, które przyjmuje się za lokalizację stacji mobilnej.

Podstawowym ograniczeniem tej metody jest mała dokładność lokalizacji, która wynosi ok. 300-500 metrów dla obszaru gęstej zabudowy miejskiej i 35-50 km dla obszarów pozamiejskich [1]. Dodatkowo, ponieważ na ogół obszary obsługiwane przez komórki pokrywają się, często występuje sytuacja, w której dla pewnego położenia stacji mobilnej, lokalizacja uzyskiwana metodą CellId jest skokowo różna w sąsiednich chwilach.



Rys. 2. Różne typy komórek

Na rys. 2. przedstawione zostały schematycznie dwa rodzaje komórek występujące w sieci: sektorowa i dookólna. Typowo, dla komórki dookólnej lokalizacja zwracana przez metodę CellId (zaznaczona na rysunku znakiem X) jest tożsama z położeniem wieży stacji bazowej (BTS). Dla komórek sektorowych zwracana lokalizacja jest z reguły

bliska środka obszaru, jaki obsługuje dana komórka.

Inne metody lokalizacji użytkowników w sieci GSM korzystają z: parametrów czasowych komunikacji stacji mobilnej z siecią, informacji o sile sygnału radiowego, danych z trzech i więcej stacji bazowych, informacji z systemu GPS [2, 3]. Zapewniają one większą dokładność lokalizacji, jednak bardzo wysokie koszty związane z ich implementacją powodują, że niewielu operatorów decyduje się na ich realizację i uruchomienie.

1.2. Wstępna analiza problemu

Trudności związane z praktyczną realizacją pomysłu wnioskowania o stanie systemu transportowego na podstawie przemieszczania się użytkowników sieci komórkowej są następujące:

- informacja o lokalizacji jest bardzo niedokładna, z praktycznego punktu widzenia użyteczne informacje o sytuacji na drodze wymagają dokładności przynajmniej o rząd wielkości większej
- brak jest informacji podanej wprost o tym, które stacje mobilne są użytkownikami systemu transportowego
- w praktyce, dostęp do informacji o wszystkich użytkownikach jest niemożliwy ze względu na koszty, które musiałby ponieść operator. Koszty te wynikają ze znacznego rozproszenia zarządzania obsługą użytkowników i tego, że są to podsystemy specjalizowane, zamknięte, bez interfejsów pozwalających na łatwy dostęp do przetwarzanych przez nie danych.

Celem opisywanej pracy badawczej było zaproponowanie i weryfikacja metody, przy pomocy której, pomimo powyżej przedstawionych trudności, operator sieci komórkowej będzie mógł uzyskać informację o systemie transportowym na obszarze, który jest obsługiwany przez jego sieć. Informacja ta powinna być na tyle dokładna i aktualna, by na jej podstawie opłacalne było przygotowanie usług, które mogłyby być następnie zaoferowane klientom instytucjonalnym i korporacyjnym.

Na podstawie uzyskiwanych informacji, w ogólnym przypadku, możliwe byłoby rozwiązywanie następujących problemów praktycznych:

- analiza strumienia ruchu na drodze w zależności od dnia tygodnia, pory dnia itp.
- wykrywanie blokad i zagęszczeń ruchu

- znajdowanie fragmentów dróg o niezadawalającej przepustowości.

Te potencjalne zastosowania są dodatkową motywacją do zajęcia się opisanym powyżej zagadnieniem.

2. Model danych i definicja problemu

2.1. Definicje i oznaczenia

Obiektem mobilnym nazywany będzie obiekt świata rzeczywistego poruszający się po powierzchni Ziemi. Przykładem obiektu mobilnego będącego w zakresie zainteresowania może być np. pojazd, telefon komórkowy, moduł GSM lub odbiornik GPS. Komórką nazywamy obszar zasięgu sygnału radiowego emitowanego przez jedną ze stacji przekaźnikowych. Stacją mobilną nazywany będzie zestaw składający się z urządzenia technicznego użytkownika (wraz z oprogramowaniem) wykorzystywany do komunikacji w sieci telefonii komórkowej.

Wprowadźmy następujące oznaczenia:

- N – zbiór liczb naturalnych
- R – zbiór liczb rzeczywistych,
- $A \subset R^2$ – obszar, na którym „działa” sieć telefonii komórkowej
- $t \in T \subset R$ – czas, znacznik czasu
- $p \in A$ – punkt na obszarze (płaszczyźnie), lokalizacja
- $d(p_1, p_2)$ – odległość (euklidesowa) pomiędzy punktami $p_1 = \langle x_1, y_1 \rangle$ i $p_2 = \langle x_2, y_2 \rangle$, obliczana wg wzoru:

$$d(p_1, p_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (1)$$

- $O = \{1, 2, \dots, N_o\}$ – zbiór identyfikatorów obiektów mobilnych o
- $\lambda : O \times T \mapsto A$ – funkcja, która przypisuje obiektowi jego lokalizację w chwili t
- $\lambda^{GPS} : O \times T \mapsto A$ – funkcja, która przypisuje obiektowi jego lokalizację w chwili t wyznaczoną przez system GPS.

2.2. Model sieci telefonii komórkowej

Wprowadźmy następujące oznaczenia:

- B – zbiór identyfikatorów b komórek wchodzących w skład sieci komórkowej. Niech 0 oznacza „brak komórki”

$$B = \{0, 1, 2, \dots, N_B\}$$

- M – zbiór identyfikatorów stacji mobilnych m

$$M = \{1, 2, \dots, N_M\}$$

- $\theta : M \times T \mapsto B$ – funkcja przypisująca stacji mobilnej identyfikator komórki, przez którą, w chwili t , jest ona obsługiwana. Jeśli $\theta(m, t) = 0$, to oznacza to, że w chwili t stacja mobilna m nie jest widoczna w sieci komórkowej

- $\beta : B \mapsto A$ – funkcja przypisująca każdej z komórek lokalizację interpretowaną jako miejsce, w którym znajduje się stacja mobilna obsługiwana przez tę komórkę. Punkt $p^b = \beta(b)$ nazywaliśmy będziemy środkiem komórki b .

Metodą lokalizacji w sieci komórkowej będziemy nazywać funkcję $M \times T \mapsto A$, która stacji mobilnej m przypisuje dla danej chwili t pewne położenie geograficzne.

Metodą lokalizacji opartą o CellId będziemy nazywać funkcję określoną wzorem:

$$\lambda^{CellId}(b, t) = \beta(\theta(b, t)) \quad (2)$$

Niech piątka $\Lambda = \langle A, B, M, \theta, \beta \rangle$ będzie modelem sieci telefonii komórkowej działającej na obszarze A , ze zdefiniowaną metodą lokalizacji w oparciu o CellId.

2.3. Model danych lokalizacyjnych

Zdarzeniem nazywana będzie para $e = \langle t, P \rangle$, gdzie t jest znacznikiem czasu określającym chwilę, w której miało miejsce zdarzenie, a P to lista parametrów definiujących typ zdarzenia. Strumieniem zdarzeń nazywamy ciąg zdarzeń

$$E = \{e_1, e_2, \dots\}$$

uporządkowanych niemalejąco względem znacznika czasu t .

Funkcję $\phi : e \rightarrow \{0, 1\}$ będziemy określać mianem filtra strumienia zdarzeń. Wartość 1 przypisana zdarzeniu będzie oznaczała, że zostało ono przepuszczone przez filtr. Wartość 0 oznaczała będzie, że filtr odrzucił zdarzenie e . Operację, która z ciągu E tworzy nowy ciąg przez usunięcie tych zdarzeń, dla których $\phi(e) = 0$, nazywaliśmy będziemy operacją filtrowania.

Zdarzeniem lokalizacyjnym l nazywane będzie zdarzenie, w którym $P = \langle o, p \rangle$. Zdarzenie lokalizacyjne zawiera informację o fakcie znajdowania się obiektu o w chwili t w lokalizacji p .

Zdarzeniem lokalizacyjnym CellId l^b nazywamy zdarzenie, w którym $P = \langle m, b \rangle$, gdzie $b = \theta(t, m)$. Strumieniem zdarzeń

lokalizacyjnych CellId nazywamy ciąg zdarzeń lokalizacyjnych

$$L^b = \{l_1^b, l_2^b, \dots\}$$

uporządkowanych niemalejąco względem znacznika czasu t .

2.4. Model drogi

Niech modelem drogi będzie linia łamana otwarta, zbudowana z odcinków, których końcami są kolejne punkty ciągu

$$\hat{Q} = \{\hat{q}_1, \hat{q}_2, \dots, \hat{q}_{N_Q}\}$$

Zbiór punktów Q należących do linii łamanej oznaczany będzie przez $Q \subset A$

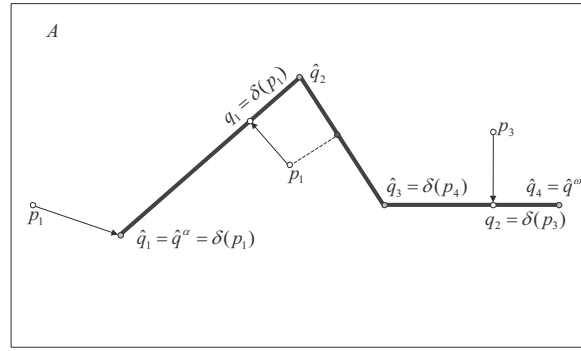
Punkty będące wyrazami ciągu \hat{Q} będą nazywane *punktami wierzchołkowymi drogi*, a odcinki $\hat{q}_k \hat{q}_{k+1}$ *segmentami drogi*. Punkt $\hat{q}^\alpha = \hat{q}_1$ nazywany będzie *punktem początkowym drogi*, a punkt $\hat{q}^\omega = \hat{q}_{N_Q}$ jej *punktem końcowym*.

Pikietażem nazywać będziemy funkcję $\pi: Q \mapsto R_{\geq 0}$ przypisującą wszystkim punktom należącym do drogi ich odległość wzdłuż linii łamanej licząc od punktu początkowego drogi. Można zauważyć, że $\pi(\hat{q}^\alpha) = 0$. Wartość $\pi(\hat{q}^\omega)$ jest długością całkowitą drogi, a różnica $\pi(\hat{q}_{k+1}) - \pi(\hat{q}_k)$ jest długością segmentu $\hat{q}_k \hat{q}_{k+1}$.

Dla przedziału $\langle 0, \pi(\hat{q}^\omega) \rangle$ określona jest

funkcja $\pi^{-1}(x): R \mapsto Q$ odwrotna do funkcji pikietażu, której wartościami są punkty na drodze leżące w odległości x od początku drogi.

Dla drogi Q , zdefiniujemy funkcję $\delta(p)$ jako funkcję $A \mapsto Q$ przypisującą każdemu z punktów należących do A punkt drogi najbliższy w sensie odległości euklidesowej. Jeśli istnieje wiele takich punktów, to wybierany jest ten o najmniejszym pikietażu. Operację znajdowania $q = \delta(p)$ będziemy nazywać przyciąganiem punktu do drogi. Zasady przyciągania punktów do modelu zilustrowane zostały na rys. 3.

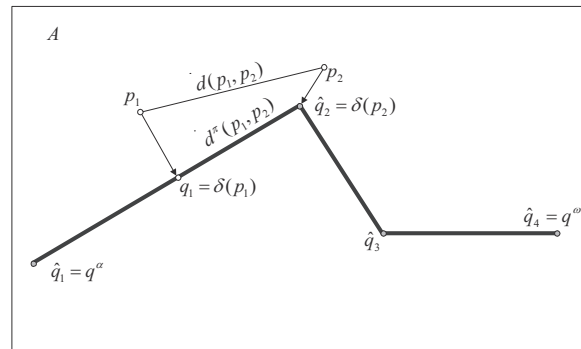


Rys. 3. Przyciąganie punktów do drogi

Odległością pikietażową punktów $p_1, p_2 \in A$ będziemy nazywali wielkość wyliczoną wg następującej definicji:

$$d^\pi(p_1, p_2) = |\pi(\delta(p_1)) - \pi(\delta(p_2))|. \quad (3)$$

Przykład pokazujący różnicę pomiędzy obliczaniem odległości euklidesowej, a odległości pikietażowej przedstawiony został na rys. 4. Zauważmy, że odległość pikietażowa dwóch punktów może być zarówno większa jak i mniejsza od ich odległości euklidesowej.



Rys. 4. Odległość euklidesowa, a odległość pikietażowa

2.5. Model pokrycia drogi zasięgiem sieci telefonii komórkowej

Zdefiniujemy *funkcję zasięgu* $\gamma: A \times B \mapsto \{0,1\}$ jako funkcję przyporządkowującą wartość 1 tym punktom obszaru A , które są w zasięgu komórki, a wartość 0 tym punktom, które są poza jej zasięgiem. Poprzez bycie punktu p w zasięgu komórki rozumiemy fakt, że stacja mobilna znajdująca się w punkcie p może być w pełni obsługiwana przez tę komórkę.

Funkcją pokrycia drogi przez komórkę będziemy nazywali taką funkcję

$$\gamma^\pi: \langle 0, \pi(\hat{q}^\omega) \rangle \times B \mapsto \{0,1\},$$

która przyporządkowuje wartość 1 tym wartościom pikietażu, które odpowiadają

punktem drogi będącym w zasięgu komórki. W pozostałych punktach funkcja przyjmuje wartość 0.

Pokryciem drogi przez komórkę b nazywamy zbiór $c_b^Q \subseteq \langle 0, \pi(\hat{q}^\omega) \rangle$ zdefiniowany jak poniżej:

$$c_b^Q = \left\{ x : x \in \langle 0, \pi(\hat{q}^\omega) \rangle \wedge \gamma^\pi(x, b) = 1 \right\} \quad (4)$$

Jeżeli $c_b^Q = \emptyset$, to będziemy mówili, że droga Q nie jest w zasięgu komórki o identyfikatorze b .

Pokryciem drogi Q przez sieć telefonii komórkowej będziemy nazywali zbiór zasięgów wszystkich komórek, w których zasięgu jest droga Q :

$$C^Q = \{c_1^Q, c_2^Q, \dots, c_{N_C}^Q\}$$

Dla danej drogi Q i pokrycia C^Q definiujemy filtr zdarzeń ϕ^c wzorem:

$$\phi^c(l^b) = \begin{cases} 1 & \text{gdy } l^b = \langle t, m, b \rangle \wedge c_b^Q \neq \emptyset, \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (5)$$

Początkiem pokrycia drogi przez komórkę b będziemy nazywali wartość $c_b^<$ określoną przez poniższy wzór:

$$c_b^< = \min c_b^Q \quad (6)$$

Końcem pokrycia drogi przez komórkę b będziemy nazywali wartość $c_b^>$ określoną jak poniżej:

$$c_b^> = \max c_b^Q \quad (7)$$

Z powyższych definicji wynika, że $c_b^< \leq c_b^>$.

Przez C^q oznaczany będzie zbiór tych komórek, w których zasięgu znajduje się punkt $q \in Q$.

Długością zasięgu komórki b będziemy nazywali wielkość $\bar{c}_b \in R_{\geq 0}$ obliczaną zgodnie z poniższym wzorem:

$$\bar{c}_b = \int_0^{\pi(\hat{q}^\omega)} \gamma^\pi(x, b) dx \quad (8)$$

Środkiem ciężkości zasięgu komórki b , będziemy nazywali liczbę

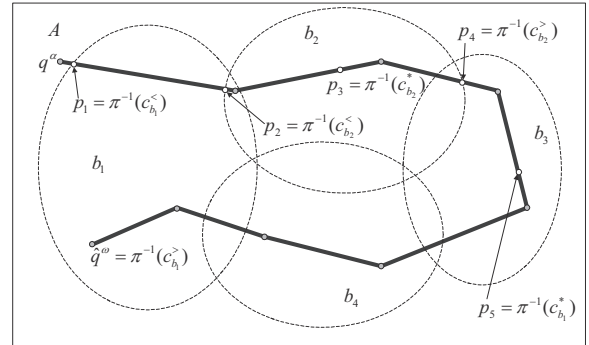
$$c_b^* = \frac{1}{\bar{c}_b} \int_0^{\pi(\hat{q}^\omega)} x \gamma^\pi(x, b) dx \quad (9)$$

Przy założeniu, że prawdopodobieństwo znajdowania się stacji mobilnej w każdym miejscu drogi jest takie samo, punkt $\pi^{-1}(c_b^*)$ może być uznawany za oczekiwane miejsce

znajdowania się stacji mobilnej będącej w zasięgu komórki b . Dodatkowo zauważmy, że nie musi być spełnione $c_b^* \in c_b^Q$, czyli że środek ciężkości zasięgu nie musi być wartością należącą do pokrycia drogi przez komórkę.

Można zauważyć, że jeśli pokrycie drogi przez komórkę jest zbiorem spójnym (przedziałem), to $c_b^* = \frac{c_b^> + c_b^<}{2}$.

Na rys. 5. pokazane zostało przykładowe pokrycie drogi zasięgiem sieci telefonii komórkowej.



Rys. 5. Pokrycie drogi przez sieć komórkową

Zaznaczone zostały punkty odpowiadające początkom, końcom i środkom ciężkości zasięgów komórek b_1 i b_2 . W przypadku komórki b_1 , środek ciężkości jej zasięgu (punkt p_5 na Rys. 5) nie należy do jej zasięgu.

2.5.1. Odległość dwóch komórek

Odległością dwóch komórek $b_i, b_j \in B$ będziemy nazywali funkcję $d: B \times B \mapsto R_{\geq 0}$ spełniającą warunki:

$$\begin{aligned} d(b_i, b_i) &= 0 \wedge \\ d(b_i, b_j) &= d(b_j, b_i) \wedge \\ d(b_i, b_k) &\leq d(b_i, b_j) + d(b_j, b_k) \end{aligned} \quad (10)$$

Na potrzeby dalszych rozważań wprowadzmy następujące definicje:

- odległością zwykłą dwóch komórek będziemy nazywać funkcję: $\bar{d}(b_i, b_j) = d(\lambda^{\text{CellId}}(b_i), \lambda^{\text{CellId}}(b_j))$ (11) czyli odległość euklidesową środków komórek.

- odległością pikietażową dwóch komórek będziemy nazywać funkcję $\hat{d}(b_i, b_j) = d^\pi(\lambda^{\text{CellId}}(b_i), \lambda^{\text{CellId}}(b_j))$ (12)

Inaczej mówiąc jest to odległość pikietażowa ich środków

- odległością *pokryciową* dwóch komórek o identyfikatorach b_i i b_j będziemy nazywać funkcję określoną następującym wzorem:

$$\tilde{d}(b_i, b_j) = |c_{b_j}^* - c_{b_i}^*| \quad (13)$$

Można zauważyć, że ponieważ wielkości te są definiowane jako odległość euklidesowa dwóch punktów na płaszczyźnie (dla odległości zwykłej) bądź jako odległość dwóch punktów na linii (odległości pikietażowa i pokryciowa), spełniają warunki (10).

2.5.2. Charakterystyki pokrycia drogi siecią komórkową

Zdefiniujemy *współczynnik krotności pokrycia* jako wielkość określoną poniższym wzorem:

$$c_x = \frac{\sum_{i=1}^{N_B} \tilde{c}_i}{\pi(\hat{q}^\omega)} \quad (14)$$

Funkcją *krotności pokrycia* $\tilde{c}(q)$ będziemy nazywali funkcję $\langle 0, \pi(\hat{q}^\omega) \rangle \mapsto N$, zdefiniowaną wzorem:

$$\tilde{c}(x) = C^q, \quad (15)$$

która pikietażowi, każdego z punktów należących do drogi przypisuje liczbę komórek w których zasięgu znajduje się ten punkt.

2.6. Model informacji o stanie systemu transportowego

Informacją o ruchu na drodze będziemy nazywali ciąg skończony

$$S = \{s_1, s_2, \dots, s_{N_s}\},$$

którego elementami są trójki $s = \langle q, u^+, u^- \rangle$, gdzie

- $q \in Q$ - punkt na drodze, którego dotyczy informacja
- $u^+, u^- \in \{0, 1, 2, 3\}$ - stan systemu drogowego dla ruchu odpowiednio: w kierunku rosnącego pikietażu i kierunku malejącego pikietażu. Wartościom u^+, u^- nadajemy następującą interpretację:
 - 0 - 'unknown' - brak danych, sytuacja nieokreślona,
 - 1 - 'low' - zatłoczenie, ruch powolny,
 - 2 - 'medium' - ruch utrudniony,
 - 3 - 'high' - ruch swobodny.

W celu ocenienia jak wybrana informacja o ruchu różni się od innej informacji o ruchu

(przy założeniu, że obie dotyczą tego samego zbioru punktów) wprowadźmy następującą funkcję oceny różnicy w informacji dla ustalonego punktu drogi i dla ustalonego kierunku:

$$u_d(u^*, u) = \begin{cases} 0 & \text{gdy } u^* = 0 \\ |u - u^*| & \text{gdy } u^* \neq 0 \text{ i } u \neq 0 \\ 2 & \text{gdy } u^* \neq 0 \text{ i } u = 0 \end{cases} \quad (16)$$

W powyższej definicji informację u^* nazywamy informacją referencyjną. Pod pojęciem informacji referencyjnej rozumiana jest informacja, która jest traktowana jako „pewna”. Może ona pochodzić z wykonanych pomiarów, lub być informacją zadaną w eksperymencie symulacyjnym.

Funkcję porównującą dwie informacje o ruchu (dla ciągu punktów) będziemy oznaczali jako s_d i definiowali wg poniższego wzoru:

$$s_d(s^*, s) = \frac{\sum_{i=1}^{N_s} u_d(u_i^{*+}, u_i^+) + \sum_{i=1}^{N_s} u_d(u_i^{*-}, u_i^-)}{2N_s} \quad (17)$$

Wartością różnicy pomiędzy dwoma informacjami o ruchu jest wartość średnia różnicy informacji obliczona dla poszczególnych punktów wchodzących w skład informacji o ruchu przy uwzględnieniu obu kierunków. Funkcja s_d przyjmuje wartość 0, kiedy to uznajemy, że informacje o ruchu są identyczne, wtedy, kiedy wartości u są takie same dla wszystkich punktów, w których informacja referencyjna ma wartości inne niż 'unknown'.

2.7. Sformułowanie problemu

W oparciu o wprowadzone powyżej definicje i oznaczenia, problem może być sformułowany następująco: Dla danych Λ , \hat{Q} , C^Q i L^b oraz dla określonego przedziału czasu $\langle t_1, t_2 \rangle$ wyznaczyć *zadowalającej jakości* informację o stanie ruchu drogowego.

Na obecnym etapie prac pojęcie *zadowalającej jakości* nie jest jeszcze formalnie definiowane. Jakość rozwiązań jest oceniana poprzez subiektywną ocenę użyteczności uzyskiwanych informacji.

3. Rozwiązanie problemu

3.1. Uzupełniające definicje i oznaczenia

Zdefiniujmy funkcję $v(q)$ jako funkcję $Q \mapsto R_{\geq 0}$ przyporządkowującą punktom drogi

pewną wielkość, która jest proporcjonalna do średniej prędkości pojazdów w tym miejscu. Wielkość tą będziemy nazywali *szacowaną prędkością*.

Profilom prędkości W dla drogi Q będziemy nazywać ciąg

$$W = \{w_1, w_2, \dots, w_{N_W}\}$$

krotek $w = \langle q, v^+, n^+, v^-, n^- \rangle$, gdzie:

- q – punkt drogi, którego dotyczy informacja
- v^+ – szacowana prędkość pojazdów poruszających się w kierunku rosnącego pikietażu
- n^+ – liczba zdarzeń, dla których została obliczona wartość v^+
- v^- – szacowana prędkość pojazdów poruszających się w kierunku malejącego pikietażu
- n^- – liczba zdarzeń, dla których została obliczona wartość v^- .

3.2. Wyznaczanie profilu prędkości

Dla danych: A , Λ , i Q , profil prędkości będzie wyznaczany na podstawie L^b wg następującego ogólnego algorytmu VelProfile (t_{move}, d_{min}), który przedstawiony został w tabeli 1. Algorytm ten przyjmuje dwa następujące parametry:

- t_{move} – długość odcinka czasu w przeszłości, jaka jest analizowana
- d_{min} – minimalna wymagana odległość, jaką musiała pokonać stacja mobilna w ciągu czasu t_{move} poprzedzającego zdarzenie, by zostało ono uwzględnione w wyliczeniu szacowanej prędkości.

Tabela 1. Algorytm wyznaczania profilu prędkości

VelProfile (t_{move}, d_{min})
1. Zainicjuj W jako ciąg pusty.
2. Dla każdego zdarzenie lokalizacyjnego $l^b = \langle t, m, b \rangle$ należącego do L^b :
a) Znajdź $q' := \pi^{-1}(c_b^*)$.
b) Jeśli istnieje $w \in W$, dla którego $q = q'$ to podstaw $w' := w$. W przeciwnym wypadku do W dodaj $w' = \langle q', 0, 0, 0, 0 \rangle$.

c) Znajdź $\hat{l}^b = \langle \hat{t}, \hat{m}, \hat{b} \rangle$ takie, że $\hat{m} = m$ i $\tilde{d}(b, \hat{b})$ jest maksymalne spośród zdarzeń, dla których $t - \hat{t} \leq t_{move}$.

d) Na podstawie c_b^* i $c_{\hat{b}}$ wyznacz kierunek ruchu związany ze zdarzeniem l^b .

e) Jeśli $\tilde{d}(b, \hat{b}) < d_{min}$ przejdź do analizy następnego zdarzenia i wróć do punktu 2a).

f) Oblicz szacowaną prędkość związaną ze zdarzeniem wg wzoru

$$v' = \frac{\tilde{d}(b, \hat{b})}{t - \hat{t}} \quad (18)$$

g) Jeśli zdarzenie jest związane z ruchem w kierunku *rosnącego* pikietażu, to zmodyfikuj w' następująco:

$$w' := \left\langle q', \frac{n^+ v^+ + v'}{n^+ + 1}, n^+ + 1, v^-, n^- \right\rangle \quad (19)$$

h) Jeśli zdarzenie jest związane z ruchem w kierunku *malejącego* pikietażu, to zmodyfikuj w' następująco:

$$w' := \left\langle q', v^+, n^+, \frac{n^- v^- + v'}{n^- + 1}, n^- + 1 \right\rangle \quad (20)$$

3.3. Wyznaczanie informacji drogowej

Wyznaczanie S na podstawie W odbywa się zgodnie z algorytmem TraffInfo ($itval, n_{min}$), który przedstawiony został w tabeli 2. Parametrami algorytmu są:

- $itval$ – odległość od punktu q , dla której zliczane są elementy ciągu W
- n_{min} – minimalna liczba zdarzeń potrzebna żeby uznać wyliczoną informację o ruchu drogowym za istotną.

Tabela 2. Algorytm wyznaczania informacji o ruchu

TraffInfo ($itval, n_{min}$)
1. Zainicjuj S jako ciąg elementów $\langle q, 0, 0 \rangle$, dla punktów q , dla których ma zostać wyznaczona informacja o statusie.
2. Dla każdego elementu $s = \langle q, u^+, u^- \rangle$ ciągu S :
a) Wyznacz przedział pikietażu, dla którego zliczane są szacowane prędkości. Dla kierunku rosnącego pikietażu będzie to:
$\langle \max\{0, \pi(q) - itval\}, \pi(q) \rangle$,

a) dla kierunku malejącego pikietażu:

$$\langle \pi(q), \min\{\pi(q) + itval, \pi(\hat{q}^o)\} \rangle$$

b) Dla obu kierunków oblicz wartości v jako średnią ważoną szacowanych prędkości dla tych $w \in W$, które należą do,

odpowiedniego dla kierunku, przedziału pikietażu.

c) Dla obu kierunków oblicz wartości n jako sumę liczby zdarzeń dla tych $w \in W$, które należą do, odpowiedniego dla kierunku, przedziału pikietażu.

d) Dla obu kierunków oblicz informację o ruchu drogowym wg poniższego ogólnego wzoru:

$$u = \begin{cases} 0 & \text{gdy } n < n_{\min} \\ 1 & \text{gdy } n \geq n_{\min} \wedge v < 30 \\ 2 & \text{gdy } n \geq n_{\min} \wedge 30 \leq v < 60 \\ 3 & \text{gdy } n \geq n_{\min} \wedge v \geq 60 \end{cases} \quad (21)$$

Progi dla wartości v w równaniu (21) zostały wybrane arbitralnie, jako reprezentujące subiektywną ocenę warunków na drodze. Ich dobór powinien być przedmiotem dokładniejszej analizy w przyszłości.

4. Wyniki obliczeniowe

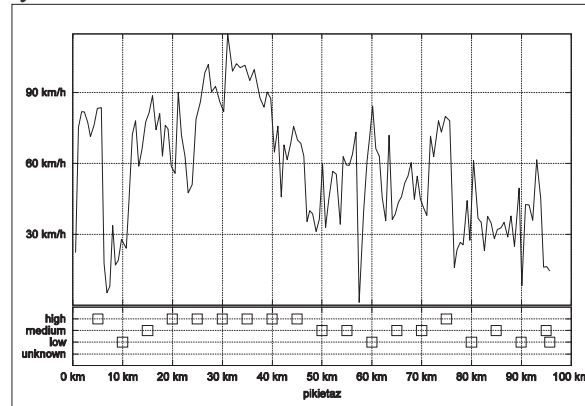
W celu weryfikacji zaproponowanej metody przeprowadzone zostały obliczenia dla trasy prowadzącej z Krakowa do Zakopanego (zwanej potocznie Zakopianką). Długość analizowanej trasy wynosi 95.71 km. Podczas obliczeń porównane zostały uzyskiwane różnice pomiędzy wyliczonymi informacjami drogowymi a informacją referencyjną, którą traktujemy jako wiarygodny opis rzeczywistej sytuacji na drodze.

W przedstawionych poniżej obliczeniach ograniczono się do analizy ruchu tylko w jednym kierunku (rosnącego pikietażu).

4.1. Referencyjna informacja drogowa

Ponieważ szczegółowe dane o rzeczywistej sytuacji na drodze nie są dostępne, do obliczeń wykorzystana została informacja drogowa przygotowana na podstawie pojedynczego przejazdu testowego na analizowanej trasie. Przejazd ten miał miejsce 6 sierpnia 2009 roku pomiędzy godziną 13-tą a 16-tą w kierunku od Krakowa do Zakopanego. Założono, że zarejestrowana prędkość pojedynczego pojazdu odpowiada szacowanej średniej prędkości pojazdów poruszających się drogą i że sytuacja

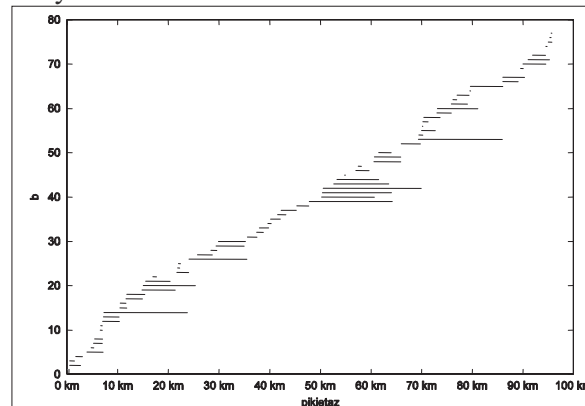
na drodze w tym czasie odpowiada typowej sytuacji na drodze w ciągu dnia w tym okresie. Na jej podstawie została wygenerowana referencyjna informacja drogowa dla zbioru punktów rozłożonych na drodze równomiernie, co 5 km. Prędkość przejazdu oraz uzyskana informacja drogowa zostały przedstawione na rys. 6.



Rys. 6. Referencyjna informacja o ruchu na trasie Kraków-Zakopane

4.2. Pokrycie trasy zasięgiem sieci komórkowej

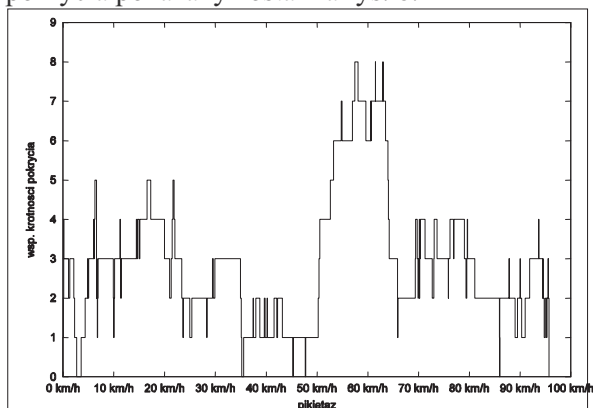
W celu określenia pokrycia analizowanej drogi zasięgiem sieci komórkowej wykonano jazdę testową, podczas której monitorowano i logowano identyfikatory komórek, które obsługiwały w danym miejscu każdą z sześciu testowych stacji mobilnych. Monitorowanie wartości CellId odbywało się co 20 sekund. Zasięgi poszczególnych komórek zostały przedstawione na rys. 7.



Rys. 7. Zasięgi i położenie komórek

Trasa Zakopianka pokrywana jest przez zbiór 78 komórek, o zasięgu od kilkuset metrów do kilkunastu kilometrów. Średnia długość zasięgu komórki to 3.60 km. Komórka o maksymalnym zasięgu pokrywa odcinek drogi o długości 19.31 km.

Przebieg wykresu funkcji krotności pokrycia pokazany został na rys. 8.



Rys. 8. Indeks pokrycia drogi siecią komórkową

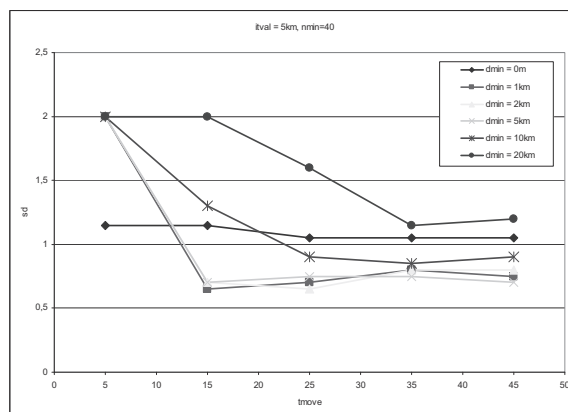
Widać na nim, że na większej części droga pokryta jest przez 2-3 komórki a maksymalna wartość funkcji krotności pokrycia wynosi 8. Obliczony dla analizowanej trasy współczynnik krotności pokrycia wynosi $c_x = 2.93$.

Jako strumień danych lokalizacyjnych wykorzystane zostały logi platformy Gateway Mobile Location Centre (GMLC) działającej w sieci jednego z operatorów komórkowych. Platforma GMLC jest odpowiedzialna za przekazywanie danych o położeniu abonentów serwerom realizującym lokalizacyjne usługi dodane. Zaletą takiego podejścia jest to, że poprzez logi, dane lokalizacyjne są dostępne w bardzo prosty sposób i niskim kosztem (znajdują się w plikach binarnych, które mogą być zgrane z platformy protokołem ftp). Wadą jest to, że informacja dotyczy jedynie niewielkiej części abonentów - tylko tych, którzy korzystali z usług dodanych oferowanych przez operatora.

Analizowane były dane z okresu od 1. do 21. sierpnia 2009 z godzin 8:00÷20:00. Przed dalszymi obliczeniami na strumieniu danych lokalizacyjnych wykonana została operacja filtrowania przy użyciu filtra ϕ^c , co spowodowało, że uwzględniane były tylko te zdarzenia, dla których b było w zbiorze komórek pokrywających Zakopiankę.

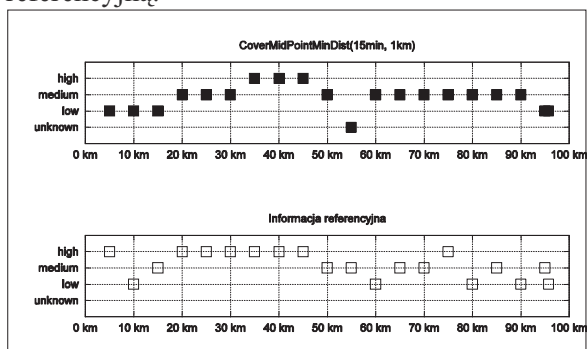
4.3. Rezultaty obliczeń

Uzyskane wartości różnicy pomiędzy informacją wyznaczoną a referencyjną, dla różnych wartości parametrów zostały przedstawione na wykresie na rys. 9.



Rys. 9. Wartości s_d dla uzyskanych informacji drogowych

Z obliczeń wynika, że informację najbardziej zbliżoną do referencyjnej uzyskano dla wartości parametrów d_{min} i t_{move} wynoszących odpowiednio 1 km i 15 minut oraz 2 km i 25 minut. Na rys. 10. zestawiona została jedna z uzyskanych informacji z informacją referencyjną.



Rys. 10. Zestawienie informacji o ruchu drogowym

Na podstawie analizy wyników można wyciągnąć wniosek, że zaproponowana metoda daje możliwość uzyskiwania informacji o stanie systemu drogowego. Jednak dla przyjętych założeń i wykorzystanego zestaw danych wejściowych nie było możliwe uzyskanie informacji równej informacji referencyjnej. Z jednej strony może to być spowodowane niedoskonałością metody, ale z drugiej tym, że informacja referencyjna mogła nie do końca odzwierciedlać standardowej sytuacji na drodze w tym okresie.

5. Podsumowanie

W artykule przedstawiona została metoda analizy danych lokalizacyjnych generowanych w elementach sieci telefonii komórkowej. Celem takiej analizy jest wnioskowanie o sytuacji na wybranej trasie, która jest pokryta zasięgiem sieci.

Przedstawione wyniki obliczeniowe pokazują na przykładzie trasy z Krakowa do Zakopanego, że przy odpowiednim doborze parametrów, możliwe jest uzyskanie informacji o ruchu, który jest bliski informacji referencyjnej, która z założenia opisuje typowy stan na drodze. Dalsze badania mogą obejmować:

- weryfikację dla jakich wartości parametrów zaproponowanej metody analizy zdarzeń lokalizacyjnych uzyskiwane są najlepsze rezultaty
- rozszerzenie badania o statystyczną analizę istotności wyników, np. potraktowanie profilu prędkości jako ciągu statystyk wyznaczających wartość oczekiwaną prędkości w danym punkcie drogi i obliczenie przedziałów ufności dla poszczególnych wartości
- weryfikację poprzez symulację jak zmienia się informacja o ruchu w zależności od napływających danych lokalizacyjnych, jeśli zmienia się realna sytuacja na drodze. Pozwoli to odpowiedzieć na pytanie, przy jakich warunkach (okres monitorowania i liczba monitorowanych stacji mobilnych) możliwe jest wykrywanie zatłoczeń w zadanym czasie
- zwiększenie ilości danych wejściowych przez dołączenie informacji lokalizacyjnej z innych elementów sieci komórkowej niż tylko platformy GMLC
- analizę, jaki wpływ na wyniki działania metody ma przyjęta procedura wyznaczania pokrycia drogi przez sieć i czy np. wykorzystanie informacji o pokryciu, która znajduje się w systemach wspomagających projektowanie sieci u operatora mogłoby poprawić jakość uzyskiwanych wyników.

ekonomicznych i prawnych. Etap 1: Charakterystyka metod służących do określania lokalizacji abonentów w sieciach GSM i UMTS., Instytut Łączności, Warszawa, 2007.

6. Bibliografia

- [1] *CGALIES final report*, Report on implementation issues related to access to location information by emergency services (E112) in the European Union, 2002.
- [2] 3G TS 23.171 3rd Generation Partnership Project; Technical Specification *Group Services and System Aspects; Functional stage 2 description of location services in UMTS*, (Release 1999).
- [3] W. Michalski, *Przegląd metod określania lokalizacji abonentów w ruchomych publicznych sieciach komórkowych GSM/UMTS z uwzględnieniem dokładności dostarczanej informacji, technicznych możliwości wdrożenia oraz czynników*