



Pohyb obyvatel v Brně dle dat mobilního operátora

Realizováno na základě objednávky POHYB OBYVATEL V BRNĚ DLE DAT MOBILNÍHO OPERÁTORA s č. j. MMB/0482638/2019.

Vypracoval: CE-Traffic, a.s.
Prvního pluku 8a
Praha 8
186 00
IČ: 28082656

Kontaktní osoba: Hynek Mašata
+ 420 776 283 206
hynek.masata@ce-traffic.com

Obsah

| | | |
|-----|------------------------------------------------------------------------------------------|---|
| 1 | Zadavatel..... | 3 |
| 2 | Popis vstupních dat..... | 3 |
| 2.1 | Struktura streamu signalizačních dat..... | 3 |
| 2.2 | Cell mapa - v případě identifikace polohy nepřímo přes Cell-ID..... | 3 |
| 3 | Výpočet matice zdrojů a cílů | 3 |
| 3.1 | Identifikace míst pobytů a cest..... | 3 |
| 3.2 | Přepočítání prostorové identifikace z Cell-ID na jednotky administrativního členění..... | 3 |
| 3.3 | Finalizace matice pro export..... | 4 |
| 3.4 | Kalibrace počtu cest..... | 4 |
| 3.5 | Časy cest | 5 |
| 3.6 | Rozsahy hodnot zdrojů a cílů použitých v matici | 5 |
| 4 | Výstupy | 5 |
| 4.1 | Specifikace | 5 |

1 Zadavatel

Statutární město Brno, Dominikánské nám. 196/1, 602 00 Brno, IČ 44992785

2 Popis vstupních dat

2.1 Struktura streamu signalizačních dat

Data vstupující do výpočtu jsou na straně mobilního operátora před samotným zpracováním anonymizována. Data jsou průběžně v reálném čase přepočítávána na data o mobilitě.

Signalizační data ve streamu jsou souborem v reálném čase předávaných elementárních signalizačních transakcí s následující strukturou:

- Anonymní ID – anonymní identifikátor
- Časová známka – okamžik vzniku transakce
- Místo vzniku transakce - identifikace polohy přes Cell-ID - poloha vzniku transakce je dána pomocí tzv. Cell-ID, přičemž interpretace polohy Cell-ID je předávána dávkově v podobě GIS vrstvy pro potřeby dekódování polohy ve streamu – viz níže v části „Cell mapa“.
- Typ transakce (různé transakce jsou generovány pro různé situace v síti)

Datové transakce ve streamu signalizačních dat jsou generovány v reálném čase ze všech druhů sítí (2G, 3G, LTE800, LTE1800, LTE2100 a LTE2600). Těchto 6 druhů sítí má plošně pokrytí území různého rozsahu blízké plnému pokrytí republiky, má rozdílnou velikost buněk a navzájem se překrývají. V závislosti na druhu telefonu (například nepodporuje LTE) a s ohledem na druh aktivity (hlas nebo datové přenosy) se přihlašuje k některé z dostupných sítí mobilního operátora. V případě českého zákazníka T-Mobile se telefon přihlašuje k nevhodnější síti tohoto operátora, v případě zahraničního zákazníka je tento proces ještě komplexnější, protože telefon si ještě volí mezi sítěmi všech mobilních operátorů.

2.2 Cell mapa - v případě identifikace polohy nepřímo přes Cell-ID

Poloha vzniku transakce je udána pomocí Cell-ID. Pro potřebu dekódování skutečné geografické polohy se využívá aktuální cell mapy s přesnou prostorovou informací o poloze buněk v jednotlivých sítích (mapy 2G, 3G, LTE800, LTE1800, LTE2100 a LTE2600). Cell mapa je pravidelně aktualizována. Přesnou identifikaci polohy a rozsahu buněk v území poskytuje polygonová vrstva jednotlivých buněk v podobě – tzv. „best service“ oblastí – oblastí, ve kterých je dle modelu pokrytí nejlepší signál z dané buňky.

3 Výpočet matice zdrojů a cílů

3.1 Identifikace míst pobytů a cest

Matice zdrojů a cílů se počítá z předzpracovaných signalizačních dat s prostorovou identifikací na úrovni Cell-ID. Algoritmus v sekvencích transakcí obsahujících časoprostorové stopy pohybu identifikuje místa pobytu a následně propojuje takto identifikovaná místa pobytu cestami. Parametry pro nastavení místa pobytu (minimální doba pobytu v lokalitě) jsou dány tak, aby při nastavené frekvenci vzorkování (každý mobil fixuje v jedné ze sítí minimálně jednou časoprostorovou známku za každých 30 minut) byly výsledky optimální. Dále je nasazen algoritmus, který se vyrovnává s přesakováním signálu mezi buňkami, které se děje i v případě, kdy se mobil nepohybuje.

3.2 Přepočet prostorové identifikace z Cell-ID na jednotky administrativního členění

Vzhledem k tomu, že jednotky administrativního členění se neshodují s oblastmi jednotlivých buněk mobilní sítě (tři různá pokrytí buňkami mnoha druhů sítí) je třeba provést přepočet zdrojů a cílů, respektive jejich kombinací do jednotek administrativního členění. Je třeba mít na paměti, že jednotlivé mobilní buňky různých sítí pokrývají více administrativních územních jednotek a to přitom každou buňku jen z části nebo někdy zcela, a analogicky, jedna administrativní územní jednotka je vykryta často více než jednou buňkou (dle best service area) daného typu sítě (tj.

často více než jednou buňkou 2G, dále často více než jednou buňkou 3G a současně jednou nebo více buňkou LTE), přičemž každá taková buňka často obsluhuje i území jiných administrativních jednotek.

S ohledem na minimalizaci deformací vzniklých při přepočtu je aplikován systém kvótního rozdělení dle následujících principů:

- Zasahuje-li místo pobytu dle Cell-ID do více administrativních jednotek, je cesta přiřazena té jednotce, kde je nejvyšší nenaplněná relativní kvóta umístěných pobytů, každé alokované místo následně mění poměr již alokovaných míst a hledá se další jednota s nejvíce nenaplněnou relativní kvótou umístěných ve srovnání s kvótou bydlících.
- Kvóty bydlících jsou počítány následujícím způsobem:
 - dle dat ze sčítání jsou každé administrativní jednotce přiřazeny počty bydlících
 - počty bydlících dle sčítání v jednotce jsou rozděleny dle podílu plochy mezi část administrativní jednotky, která spadá pod Cell-ID, k němuž je alokováno místo pobytu, a část administrativní jednotky, která pod pokrytí daného Cell-ID již nespadá.
- Cílem je místa pobytu alokovat v rámci jednoho Cell-ID tak, aby místa veškerá místa pobytu v daný den byl rozdělena proporcionálně tak, aby odpovídala podílům bydlící populace přiřazené částem administrativních jednotek, které jsou pokrývané buňkou, jejíž místa pobytů se rozdělují.

3.3 Rozsah dodávky

Maticе cest se dodává v rozsahu 14 dnů. 7.10.2019 až 20.10.2019, tj. 14 matic po jednotlivých dnech. K tomu 7 matic, kde jsou zprůměrované dny v týdnu a k tomu ještě zprůměrovaný pracovní a víkendový den za celých 14 dní.

3.4 Finalizace matice pro export

Ve vypočtené matici zdrojů a cílů se mohou vyskytovat fragmenty údajů spočívající ve výskytech specifických kombinací zdrojů a cílů, které vyžadují další dodatečnou anonymizaci při exportu matice zdrojů a cílů.

Všechny hodnoty větší než 0 a nižší než 20 zprůměrujeme v krocích po 5, aby součet nových hodnot dal stejný počet jako součet původních hodnot. Hodnoty větší než 20 (včetně) uvádíme neprůměrované.

Před anonymizačním průměrováním malých hodnot:

- trip1: 1.3 osob
- trip2: 1.7 osob
- trip3: 2.4 osob

Po anonymizačním průměrování malých hodnot:

- $(1.3 + 1.7 + 2.4) / 3 = 1.8$
- trip1: 1.8 osob
- trip2: 1.8 osob
- trip3: 1.8 osob

Tato operace se aplikuje jak na nekalibrovanou matici, tak na matici kalibrovanou. U průměrovaných matic (viz 3.3) se anonymizovalo až po zprůměrování.

3.5 Kalibrace počtu cest

Po vytvoření nekalibrované matice (v interním mezivýsledku před dodatečnou anonymizací) je provedena kalibrace. Kalibrací se rozumí přepočtení uskutečněných jednotlivých cest ze vzorku populace T-Mobile, představující obecně 40 % populace, na celkovou populaci. Kalibrace se provádí vynásobením počtu cest dle příslušných koeficientů místa bydliště. Koeficienty pro jednotlivé oblasti bydliště jsou počítány jako podíl počtu obyvatel v dané jednotce dle posledního sčítání obyvatel (trvalá bydliště) k počtu nalezených unikátních ID ve streamu signalizačních dat v nočních hodinách. Každá cesta je přepočítána na očekávaný počet cest v celkové populaci přes porovnání počtu unikátních ID

v signalačním streamu vůči referenčním datům ze sčítání. Kalibrační koeficienty jsou počítány pro všechny administrativní jednotky.

Pokud existuje administrativní jednotka, kde se převážně pohybují zaměstnanci např. jedné firmy, která má rámcovou smlouvu s jedním operátorem, může to kalibraci vychýlit.

V případě přepočtených dat na celou populaci (kalibrované matice zdrojů a cílů) je provedena finalizace matice pro export v podobě přepočítání počtů kombinací zdrojů a cílů s nižšími četnostmi než 5 na průměrné hodnoty.

3.6 Časy cest

V předaném výstupu jsou časy, kdy cesty začaly a skončily. Nejnižší čas je 0 a nejvyšší 23.

Cesty začínající v čase 0 v datech jsou, ale tyto cesty mohly začít předchozí den. Podobně i cesty končící v čase 23 mohou končit až den následující. Pokud je třeba eliminovat tripy, které začínají a končí jiné dny (např. 22:00 – 2:00), z dat se musí skrýt tripy, které začínají v čase 0 (`start_cas = 0`) a ty, které končí v 23 (`cil_cas = 23`).

3.7 Rozsahy hodnot zdrojů a cílů použitých v matici

Matice uskutečněných cest nabývá následujících maximálních rozsahů možných zdrojů a cílů:

1. Území města Brna – úroveň katastrálního území (celkem maximálně 48 jednotek)
2. Okres Brno-venkov – úroveň obce (celkem maximálně 187 jednotek)
3. Zbytek Jihomoravského kraje – úroveň okresů (celkem maximálně 6 jednotek)
4. Jiné kraje ČR – úroveň krajů (celkem maximálně 13 jednotek)

4 Výstupy

4.1 Specifikace

Zpracovaná data jsou předávána v surové podobě v 7 csv tabulkách za každý monitorovaný den s následující strukturou:

`start_cas,cil_cas,start_level,start_kod,cil_level,cil_kod,pocet,pocet_kalibrovano`

- **`start_cas`**: začátek cest příslušného zprava otevřeného intervalu (2 znamená čas od 2:00:00 do 2:59:59)
- **`cil_cas`**: konec cest příslušného zprava otevřeného intervalu (2 znamená čas od 2:00:00 do 2:59:59)
- **`start_level`**: úroveň administrativní jednotky, viz kapitola 3.6
- **`start_kod`**: kód jednotky dle číselníku ČSÚ
- **`cil_level`**: úroveň administrativní jednotky, viz kapitola 3.6
- **`cil_kod`**: kód jednotky dle číselníku ČSÚ
- **`cz`**: rozlišení, zda je trip uskutečněn tuzemskou SIM (=1), nebo zahraniční (=0)
- **`stay_before`**: interval, jak dlouho dané SIM pobýly na místě v rámci své jednotky (`start_kod`), než se vydali na danou cestu. Intervaly jsou po 4 hodinách (8 znamená, že SIM pobyla na místě v zprava otevřeném intervalu 8 až 16 hodin)
- **`pocet`**: nekalibrovaný naměřený počet cest (anonymizace pod hodnotu 5)
- **`pocet_kalibrovano`**: naměřený počet cest vynásobený koeficientem místa bydliště každé jednotlivé SIM