

Výpočet matice zdrojů a cílů

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 přeskokováním signálu mezi buňkami, které se děje i v případě, kdy se mobil nepohybuje.

2 Přepočítání 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 přepočítání 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 buď 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 jednotka s nejvíce nenaplněnou relativní kvótotou umístěných ve srovnání s kvótotou 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 částí administrativní jednotky, která spadá pod Cell-ID, k němuž je alokováno místo pobytu, a částí administrativní jednotky, která pod pokrytí daného Cell-ID již nespádá.
- Cílem je místa pobytu alokovat v rámci jednoho Cell-ID tak, aby místa veškerá místa pobytu v daný den byla rozdělena proporcionálně tak, aby odpovídala podílům bydlící populace přiřazené částem administrativních jednotek, které jsou pokrývané buňkou, jejíž místa pobytů se rozdělují.

3 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 5 zprůměrujeme, aby součet nových hodnot dal stejný počet jako součet původních hodnot. Hodnoty větší než 5 (včetně) uvádíme zaokrouhlené na celá čísla.

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

Poznámka: Tato operace se aplikuje jak na nekalibrovanou matici, tak na matici kalibrovanou.

4 Kalibrace počtu cest

Po vytvoření nekalibrované matice (v interním mezivýsledku před dodatečnou anonymizací) je provedena kalibrace. Kalibrací se rozumí přepočítání skutečnějších 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 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 signalizačním 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.

5 Časy cest

V předaném výstupu jsou časy, kdy cesty začaly a skončily. Nejnižší čas je 1 a nejvyšší 22. Časy 0 a 23 se v matici cest nevyskytují – je tomu tak z toho důvodu, že minimální doba pobytu je nastavena na 60 minut a tudíž, aby začala cesta, musí mobil být nejdříve vygenerováno místo pobytu. Výpočet úvodních míst pobytu začíná pro daný den v 0:00 a výpočet posledních míst pobytu je ukončen k 23:59.

6 Rozsahy hodnot zdrojů a cílů použitých v matici

Matice skutečnějších 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)

Výstupy

1 Specifikace

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

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

- ***start_cas***: začátek cest
- ***cil_cas***: konec cest
- ***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
- ***pocet***: nekalibrovaný naměřený počet cest (anonymizace pod hodnotu 5)
- ***pocet_kalibrovano***: naměřený počet cest vynásobený koeficientem místa bydliště každé jednotlivé SIM