

TRAVEL TIME PREDICTION

Andrej Mudroch

Master Degree Programme (2.), FEEC BUT

E-mail: xmudro01@stud.feec.vutbr.cz

Supervised by: Peter Honec

E-mail: honecp@feec.vutbr.cz

Abstract: This paper discusses the methods of travel time prediction based on the usage of machine learning and historical data. The developed prediction models are described as well as the data sources which were used as the input of the prediction models. Finally, the comparison of the models' performance is shown, providing proof the developed models have ability to outperform the widely used model based on instantaneous travel time that is not using statistical learning.

Keywords: travel time, prediction, regression

1 ÚVOD

Presné stanovenie dojazdovej doby vozidiel na cestných komunikáciách má v doprave veľký význam. Umožňuje lepšie plánovanie trás pre jednotlivcov či priemysel. Presná predikcia dojazdových dôb má taktiež nezanedbateľný psychologický efekt. Táto práca sa venuje metóde na spresnenie stanovenia dojazdovej doby, pričom na určenie dojazdových dôb využíva prediktívne štatistické modely postavené na historických dátach.

2 STANOVENIE DOJAZDOVEJ DOBY

Dojazdová doba vozidla $\Delta t = t_2 - t_1$ je definovaná ako doba potrebná na prekonanie vzdialenosti medzi bodmi A a B, keď vozidlo vyráža z bodu A v čase t_1 a prichádza do bodu B následne v budúcnosti v čase $t_2 = t_1 + \Delta t$.

Dojazdová doba sa dá stanoviť viacerými spôsobmi. Jedným z nich je sledovanie prejazdov vozidiel medzi dvomi bodmi, pričom vozidlá môžu byť identifikované napríklad opticky (zosnímanie a rozpoznanie ŠPZ) alebo na základe MAC adres bluetooth/wifi zariadení, ktoré sú v nich umiestnené. Dojazdová doba, ktorú takto získame v reálnom čase, sa nazýva okamžitá dojazdová doba. Hodnota okamžitej dojazdovej doby je ale k dispozícii až vtedy, keď vozidlo dosiahne cieľovú destináciu. Nie je teda k dispozícii v reálnom čase odjazdu, ale až s oneskorením, čo má za následok, že vodičovi je poskytnutý údaj platný pre vozidlo, ktoré vyrazilo v minulosti v čase $-\Delta t$. Rozdiel medzi reálnou dojazdovou dobou vozidla a okamžitou dojazdovou dobou je viditeľný na obrázku 1.

Z tohto dôvodu boli na predikovanie dojazdovej doby vyvinuté nižšie popísané modely, ktoré v reálnom čase dávajú presnejšiu informáciu o dojazdovej dobe.

3 PREDIKČNÉ MODEL Y

Pri tvorbe štatistických modelov boli využité historické dáta z rôznych zdrojov. Jedná sa o hodnoty okamžitej dojazdovej doby, dáta z radarových senzorov umiestnené na rôznych miestach sledovanej komunikácie a časové údaje týchto dát. Tieto dáta boli následne zoskupené podľa času a boli z nich vytvorené príznaky zahŕňajúce priemerné rýchlosti a počty vozidiel, obsadenie komunikácie a časové príznaky ako deň v týždni a víkend.

3.1 TIME DOMAIN MODEL

TimeDomainModel je model založený na časovom pôvode dát, ktorý je inšpirovaný algoritmom LOKRR [1]. Model využíva predpoklad, že dojazdová doba je závislá od času dňa. Dáta - príznaky zo všetkých dní z celého obdobia použitého na tréning algoritmu, sú teda rozdelené do N časových okien w_i $i = 0 \dots N - 1$ reprezentujúcich 10 minútové časové úseky.

Pre každé časové okno w_i je vytvorený prediktor, ktorý zaisťuje predikciu iba pre toto okno. Na tréning prediktoru sú použité dáta z časového okna w_i a z prilahlých okien w_{i-1} a w_{i+1} . Tak je vytvorené tolerančné pásmo pre nepravidelnosti v doprave, keď v niektoré dni môže očakávaný vzor nastať s istým posunom a už by nespadal do očakávaného časového okna a teda oblasti pôsobnosti prediktoru. Ako prediktory pre jednotlivé časové okná je možné použiť ľubovoľný regresný algoritmus.

3.2 CLUSTER MODEL

Tento model je založený na tom, že existujú vektory príznakov (body), ktoré sú si natoľko podobné, že tvoria clustre - zhľuky. Pri predpoklade, že takéto zhľuky dátových bodov existujú, je výhodnejšie vytvoriť prediktor pre každý cluster zvlášť, čo umožňuje lepšie zachytenie štruktúry dát.

V prípade tohto modelu je potrebné vybrať clusterizačný algoritmus, ktorým budú vytvorené clustre (fáza tréningu) a ktorý dokáže priradiť príslušnosť nových dát ku clusterom (fáza predikcie). Je potrebné podotknúť, že výsledný model je vo veľkej miere ovplyvnený výstupom clusterizačného algoritmu a teda jeho počiatočnými podmienkami. Tiež je potrebné stanoviť regresný algoritmus, ktorý bude využitý na vytvorenie prediktorov pre jednotlivé clustre.

3.3 COMBINED MODEL

Tento model využíva kombináciu N plnohodnotných modelov (TimeDomainModel alebo ClusterModel) a spracováva hodnoty dojazdovej doby predikované týmito modelmi. Na kombináciu výsledkov jednotlivých modelov môže byť použitá ľubovoľná agregáčna funkcia, v tejto práci bol použitý aritmetický priemer, čiže výstupná hodnota \hat{y} je stanovená ako

$$\hat{y} = \frac{1}{N} \sum_{i=0}^N f_i(X) \quad (1)$$

4 POROVNANIE VÝSLEDKOV MODELOV

Modely boli otestované a porovnané s reálnou hodnotou dojazdovej doby. Pre účely testu boli využité dáta z mestskej komunikácie s dĺžkou 6.6 km. Na vytvorenie modelov boli využité dáta z obdobia 10/10/2016 - 17/1/2017, predikcia bola overená na dátach 17/1/2017 - 14/2/2017. Ako porovnávacie kritériá boli využité rozšírené metriky pre vyhodnocovanie regresných modelov - MAPE [2], MAE a RMSLE [3], ktoré sú definované nasledovne:

$$MAPE = 100 \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \hat{y}_i|}{y_i} \quad [\%] \quad (2) \quad MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad [s] \quad (3)$$

$$RMSLE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\ln(\hat{y}_i + 1) - \ln(y_i + 1))^2} \quad [-] \quad (4)$$

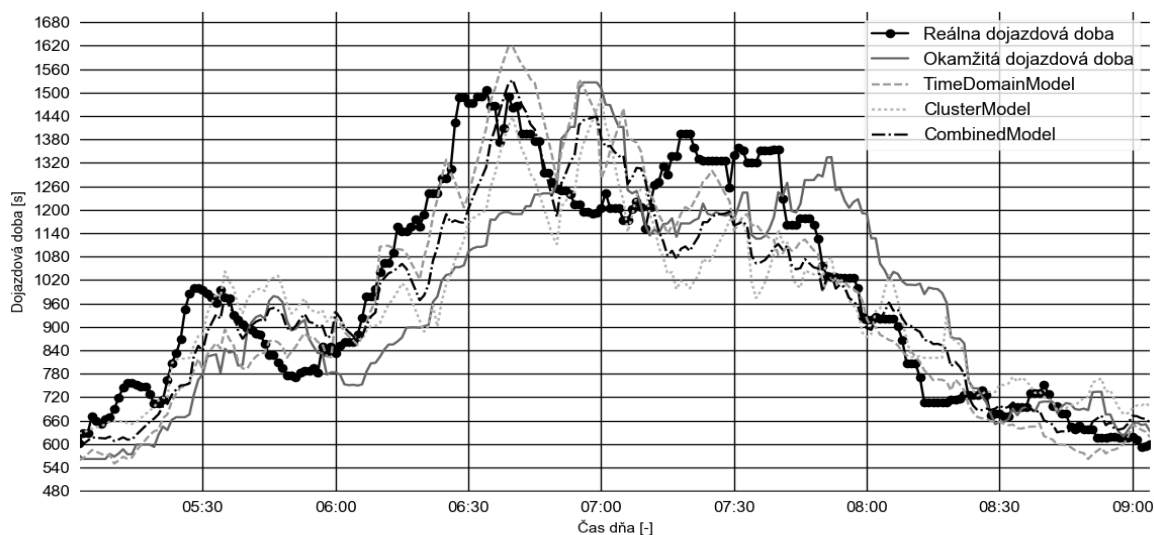
kde y_i je referenčná (reálna) hodnota dojazdovej doby a \hat{y}_i je hodnota estimovaná modelmi.

Porovnanie výsledkov predikcií je uvedené v tabuľke 1. Parametre pre použité modely boli určené prostredníctvom metódy grid-search s využitím 80% dát pre tréning a 20% pre testovanie (dáta boli

brané v súlade s ich časovou postupnosťou, čiže pre fázu testovania boli použité najnovšie dáta).

Metóda predikcie	MAPE [%]	MAE [s]	RMSLE [-]
Okamžitá dojazdová doba	10.9	78	0.16
TimeDomainModel	10.6	76	0.15
ClusterModel	8.8	65	0.13
CombinedModel	9.2	66	0.14

Tabuľka 1: Porovnanie chýb predikcií modelov pre obdobie 17/1/2017 - 14/2/2017



Obr. 1: Znáročenie predikcií modelov počas ranej dopravnej špičky

5 ZÁVER

V tejto práci boli predstavené štatistické modely pre predikciu dojazdových dôb. Bolo ukázané, že tieto modely produkujú pri stanovovaní dojazdovej doby lepšie výsledky ako metóda okamžitej dojazdovej doby. Zlepšenie spočíva v zmenšení chyby MAPE v porovnaní s metódou okamžitej dojazdovej doby o 3 % pre TimeDomainModel až 19 % pre ClusterModel, chyby MAE o 3 - 17% a RMSLE o 6 - 19%, čo je badateľné z tabuľky 1. Tieto modely budú použité v predikčnom systéme na predikovanie dojazdových dôb, pričom budú extenzívne otestované v praxi.

LITERATÚRA

- [1] LUNDE, Emil Øien a Thomas WOLFF. *Time Prediction: A Comparison Study on a common Data Set*. Trondheim, 2015. Dostupné také z: https://brage.bibsys.no/xmlui/bitstream/id/401494/12828_FULLTEXT.pdf. Master's thesis. Norwegian University of Science and Technology. Vedoucí práce Anders Kofod-Petersen.
- [2] Mean absolute percentage error. In: *Wikipedia: the free encyclopedia* [online]. San Francisco (CA): Wikimedia Foundation, 2001- [cit. 2017-03-15]. Dostupné z: https://en.wikipedia.org/wiki/Mean_absolute_percentage_error
- [3] CUKIERSKI, William. Metrics. *Kaggle* [online]. Kaggle Inc. [cit. 2016-10-24]. Dostupné z: <https://www.kaggle.com/wiki/Metrics>