

تخمین زمان سفر در شبکه راه‌های شهری



دانشجو: محمد صادق جزایری
استاد راهنما: دکتر فتانه تقی‌پاره
دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران



نتایج

با توجه به مقادیر معیار ارزیابی که در جدول زیر آمده‌اند می‌توان مشاهده کرد که در روش ارائه شده ترکیب تمام مدل‌ها و مجموعه خصیصه‌ها عملکرد قابل قبولی برای مسئله مورد نظر دارند. مورد جالب توجه کم‌تر بودن دقت مدل‌های رگرسیون درخت تصمیم و رگرسیون جنگل تصادفی نسبت به رگرسیون خطی است. با توجه به تعداد خصیصه‌ها و اینکه دقت رگرسیون جنگل تصادفی از درخت تصمیم بیشتر است این کم‌تر بودن دقت را می‌توان به فرابرازش نسبت داد.

Model	Dataset Index	RMSE
LR	1	۵/۴۸
DTR	1	۶/۶۷
RFR	1	۶/۳۴
LR	2	۵/۴۸
DTR	2	۸/۲
RFR	2	۶/۰۲
LR	3	۵/۴۶
DTR	3	۸/۱۹
RFR	3	۶/۰۴
LR	4	۵/۹
DTR	4	۶/۹۹
RFR	4	۶/۲۷

جمع بندی

در این تحقیق روشی برای ادغام داده‌های تاریخی سرعت بخش راه‌ها با اطلاعات مربوط به ساختار گراف ارائه شده که می‌تواند به بهبود دقت پیش‌بینی‌های سرعت بخش راه‌ها در سامانه‌های پیش‌بینی زمان سفر در شبکه‌های راه‌های شهری کمک کند. محدودیت‌های کار انجام شده شامل اما نه محدود به، در نظر گرفتن یال‌های دقیقاً همسایه یک یال، استفاده نکردن از محلیت زمانی، و سادگی مدل و خصیصه‌های استفاده شده هستند. مهندسی دقیق‌تر خصیصه‌ها، تلاش برای استفاده از مدل‌های پیچیده‌تر، استفاده بهتر از اطلاعات موجود در ساختار گراف، و تلاش برای استفاده از محلیت زمانی زمینه‌هایی هستند که کار بر روی آن‌ها می‌تواند نتایج این تحقیق را بهبود بخشد.

کاربرد های صنعتی:

تحقیق انجام شده می‌تواند به بهبود کیفیت مسیرهای پیشنهادی در نرم‌افزارهای مسیریابی و همچنین دقت قیمت‌گذاری در سرویس‌های تاکسی آنلاین کمک کند.

مراجع اصلی

1. D. Freedman, *Statistical models: theory and practice*. Cambridge: Cambridge University Press, 2009.
2. L. Rokach and O. Maimon, *Data mining with decision trees: theory and applications*. New Jersey: World Scientific, 2010.
3. F. Zheng and H. V. Zuylen, "Urban link travel time estimation based on sparse probe vehicle data," *Transportation Research Part C: Emerging Technologies*, vol. 31, pp. 145–157, 2013.

مقدمه / خلاصه

در این تحقیق با استفاده از داده‌های مکانی جمع‌آوری شده توسط گوشی‌های همراه هوشمند مجهز به سامانه مکان‌یاب جهانی متعلق به رانندگان یک سرویس تاکسی اینترنتی سرعت جابجایی ترافیک در بخش‌های مختلف شبکه‌های راه‌های شهری محاسبه شده و سپس با استفاده از سه مدل رگرسیون خطی، رگرسیون درخت تصمیم، و رگرسیون جنگل تصادفی روی خصیصه‌های استخراج شده از روی این داده‌های سرعت جابجایی ترافیک و ساختار گرافی شبکه مدلی برای پیش‌بینی سرعت جابجایی ترافیک در راه‌ها و زمان‌هایی که اطلاعات مربوط به سرعت جابجایی ترافیک در آن‌ها موجود نبوده توسعه داده شده.

مدل و روش پیشنهادی

شبکه راه‌های شهری توسط یک گراف جهت‌دار $D = (V, E, f)$ مدل شده. هر خیابان به یک یا چند بخش راه تقسیم می‌شود و هر جهت هر بخش راه توسط یک یال مدل می‌شود. رئوس نماینده محل‌هایی هستند که از یک بخش راه می‌توان به یک بخش راه دیگر رفت. با استفاده از داده‌های تاریخی سرعت حرکت ترافیک، تابع $v: E \times W \rightarrow \mathbb{R}$ ساخته شد که در آن $W = \{0, 1, \dots, 7 \times 24 - 1\}$ مجموعه ساعات موجود در هفته است. مقدار $v(e, h)$ نمایانگر سرعت معمول در جهت بخش راه متناظر با e در ساعت h هفته است. این مقادیر با پردازش اطلاعات به دست آمده از تلفن‌های همراه هوشمند رانندگان در طول یک بازه زمانی چند هفته‌ای محاسبه گردیده‌اند.

هر یال (u, v) را همراه با همسایگی آن مانند شکل زیر در نظر گرفته شده. به ازای هر چنین یال (u, v) و هر ساعت h که $v((u, v), h)$ از طریق داده‌های تاریخی موجود باشد یک مدخل در هر مجموعه داده خواهیم داشت.

در طول این تحقیق در مجموع چهار مجموعه داده مختلف به روش مذکور تولید شده که هر مدخل هر مجموعه داده متشکل از یک زیر مجموعه از خصیصه به دست آمده از طریق پردازش همسایگی هر یال از نظر سرعت در ساعت مورد نظر، حداکثر سرعت، درجه ورودی و خروجی رئوس و ... است.

هر مجموعه داده به دو بخش یادگیری و تست شامل ۹۰ درصد مدخل‌ها و درستی سنجی شامل ۱۰ درصد مدخل‌ها تقسیم شده و یک مدل رگرسیون خطی، یک مدل رگرسیون درخت تصمیم و یک مدل رگرسیون جنگل تصادفی به آن برآزش شدند. برای ساختن و برآزش مدل‌ها با توجه به حجم داده از Apache Spark استفاده شده.

معیار ارزیابی در این تحقیق به مانند معمول مسائل رگرسیون $RMSE$ می‌باشد.

