

# توسعه ابزار آدرسیابی آدرس های پستی به کمک مدل پنهان مارکوف

دانشجو: محمدرضا بخشایش  
استاد راهنما: دکتر احمد کلهر  
دانشکده مهندسی برق و کامپیوتر، دانشگاه تهران



## مقدمه

با گسترش خدمات مکان‌محور در حوزه‌هایی مانند حمل‌ونقل هوشمند، مدیریت شهری و مسیریابی، نیاز به پردازش داده‌های مکانی افزایش یافته است. بسیاری از این داده‌ها به صورت توصیفات متنی، مانند آدرس‌های پستی، ارائه می‌شوند که برای استفاده در سامانه‌های دیجیتال باید به مختصات جغرافیایی تبدیل شوند. این فرایند که به آدرسیابی (Geocoding) معروف است، در کشورهایی با آدرس‌های استاندارد به سادگی انجام می‌شود، اما در ایران به دلیل پیچیدگی ساختار آدرس‌ها، خلأ جدی برای ابزارهای کارآمد وجود دارد.

هدف از این پروژه، توسعه ابزاری مستقل و کارآمد برای تبدیل آدرس‌های پستی فارسی به یک محدوده جغرافیایی بوده که علاوه بر پوشش شهرهای بزرگ، در شهرهای کوچک و مناطق روستایی نیز قابل استفاده باشد. این ابزار با استفاده از مدل پنهان مارکوف و یا کمک داده‌های OpenStreetMap طراحی شده است.

## مدل پیشنهادی

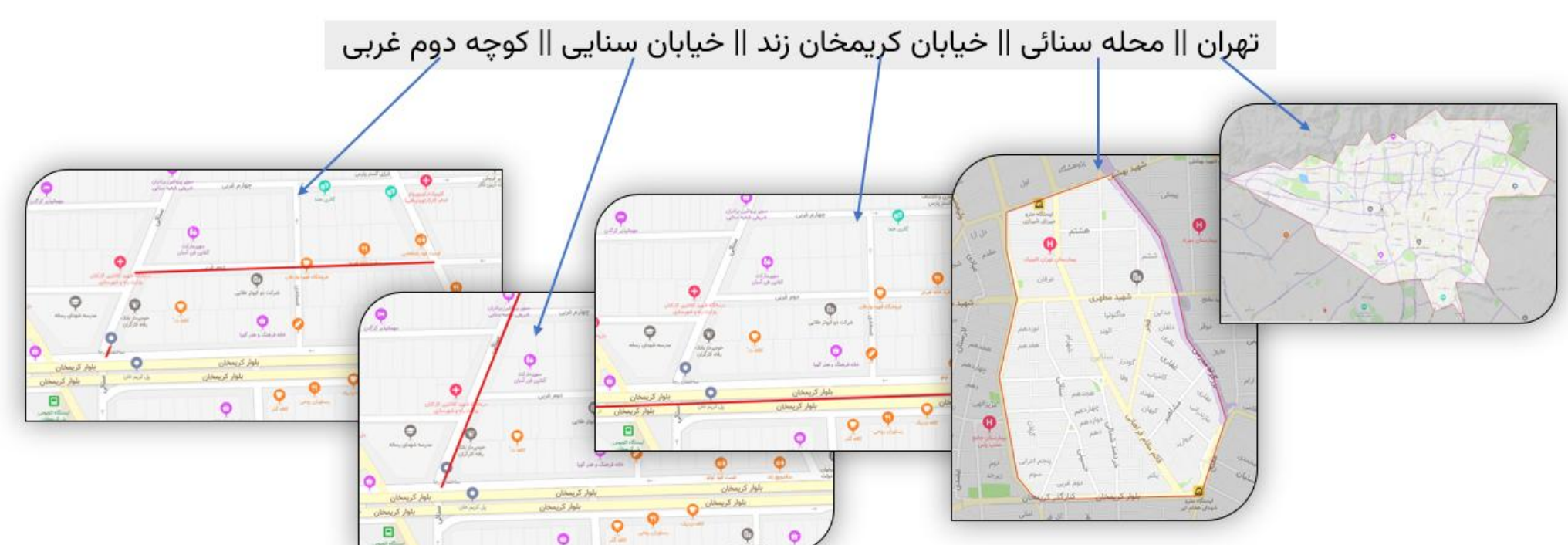
این ابزار، بر مبنای مدل پنهان مارکوف (HMM) طراحی شده تا مسئله آدرسیابی را به شیوه‌ای دقیق و انعطاف‌پذیر حل کند. در این ساختار، حالت‌های پنهان به عنوان مختصات جغرافیایی در نظر گرفته می‌شوند و بخش‌های مختلف آدرس به عنوان مشاهدات عمل می‌کنند.

ابتدا، آدرس ورودی برای اصلاح مشکلاتی مثل فاصله گذاری، حذف کاراکترهای عربی، حذف علائم نگارشی و ... نرمال‌سازی شده و به قسمت‌های سازنده تقسیم می‌شود. سپس، هر بخش از آدرس با رکوردهای موجود در نقشه مرجع (مأخوذ از OpenStreetMap) تطبیق داده می‌شود. در این مرحله، تابع امتیاز Emission با استفاده از فاصله ویرایش (Levenshtein Distance) میزان شباهت بین بخش آدرس و نام کانیدها را محاسبه می‌کند.

برای تضمین پیوستگی منطقی بین بخش‌های آدرس، تابع امتیاز Transition با بررسی فاصله اقلیدسی بین مختصات جغرافیایی مورد استفاده قرار می‌گیرد. به کمک الگوریتمی مشابه ویتربی که با نرمال‌سازی گوسی و SoftMax بهینه شده، احتمال تطبیق هر کاندید در هر مرحله به صورت تجمعی محاسبه می‌شود تا خطاهای محلی اصلاح شوند.

همچنین، با اعمال استراتژی‌هایی نظیر حذف کاندیدهای با امتیاز پایین و محدودسازی تعداد آن‌ها، زمان اجرای الگوریتم به طور قابل‌توجهی بهبود یافته است. استفاده از پایگاه داده Elasticsearch نیز امکان ذخیره و بازیابی سریع داده‌های نقشه مرجع را فراهم آورده و کارایی کلی سیستم را افزایش می‌دهد.

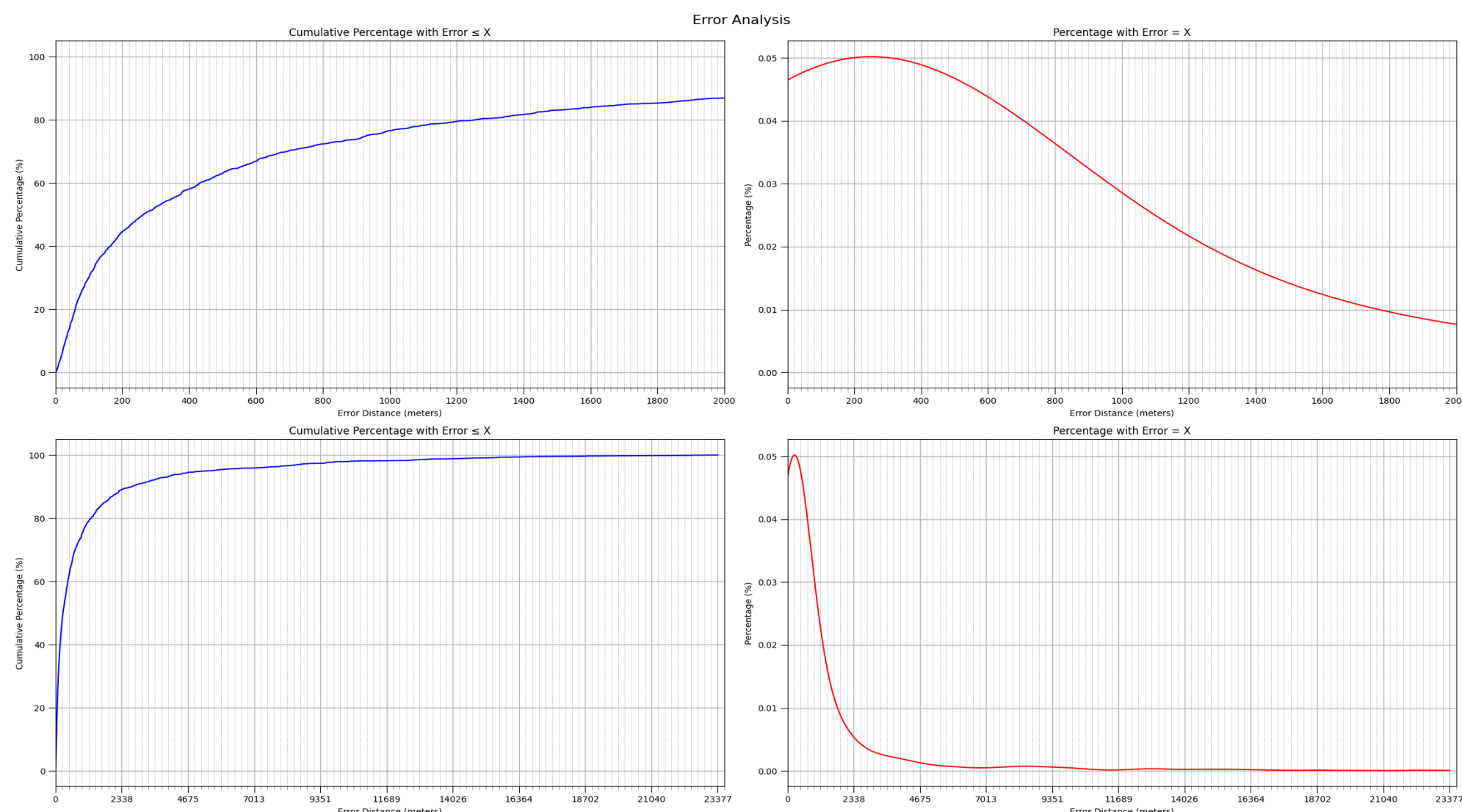
این ساختار مدولار، قابلیت تطبیق آدرس‌ها در شهرهای بزرگ و مناطق روستایی را دارا بوده و انعطاف‌پذیری لازم برای مدیریت پیچیدگی‌های آدرسیابی در ایران را به خوبی ارائه می‌دهد. شکل ۱ مثالی از نتیجه آدرسیابی این ابزار را نمایش می‌دهد.



شکل ۱. مثالی از عملکرد ابزار آدرسیاب

## نتایج

برای ارزیابی دقت ابزار، ۲۰۰۰ آدرس پستی آزمایش شد. نتایج نشان می‌دهد که ابزار توسعه‌یافته توانست حدود ۸۷٪ از آدرس‌ها را با حداکثر خطای ۲ کیلومتر مکان‌یابی کند که در مقایسه با نرخ موفقیت ۸۵٪ ابزار رقیب (TehranGeocode)، عملکرد بهتری دارد. این نتایج، کارایی ابزار را در تبدیل آدرس‌های پستی به مختصات جغرافیایی برای کاربردهای پژوهشی و تحلیل‌های مکانی تأیید می‌کند. شکل ۲ توزیع خطای هندسی پاسخ‌های سامانه را نمایش می‌دهد.



شکل ۲. توزیع خطای هندسی ابزار: در نمودار بالا، محور افقی خطای هندسی پاسخ‌های سامانه را بر حسب متر و محور عمودی درصد پاسخ‌ها را مشخص می‌کند. خط آبی رنگ، به صورت تجمعی حجم آدرس‌های آدرسیابی شده با خطای کمتر مساوی هر آستانه خطا را نشان می‌دهد در حالی که خط قرمز رنگ، توزیع خطا را برای آستانه‌های مختلف خطا به صورت نقطه‌ای نشان می‌دهد. برای رسیدن به درک بهتر از توزیع خطا، در دو نمودار بالاتر محور افقی به خطای ۰ تا ۲ کیلومتر محدود شده تا اثر داده‌های پرت از نمودار حذف شود

## جمع بندی

هدف اصلی این پژوهش، توسعه ابزاری برای تبدیل آدرس‌های پستی فارسی به مختصات جغرافیایی بود. در نهایت ابزار توسعه‌یافته توانست با دقت حدود ۹۵٪ آدرس‌های پستی را با حداکثر خطای ۵ کیلومتر به مختصات جغرافیایی تبدیل کند. این ابزار به دلیل استفاده از یک روش کلاسیک نیاز به حجم بالای داده آموزشی لیبیل گذاری شده ندارد و بروزرسانی تغییرات معابر شهری بدون نیاز به آموزش مجدد مدل امکان‌پذیر است. مجموعه این ویژگی‌ها، مدل توسعه‌یافته را به ابزاری کارآمد برای تسریع فرایند تبدیل آدرس‌های محاوره‌ای به مختصات در مراکز خدمات مشتریان، پژوهش‌های علمی و نرم‌افزارهای مسیریاب تبدیل کرده است.

در کنار این ویژگی‌های مثبت، عدم برخورداری از امکان آدرسیابی با دقت نقطه‌ای (Rooftop Accuracy) و عدم امکان تجزیه دستورات پیچیده مسیریابی در آدرس‌های فارسی از بزرگ‌ترین چالش‌های ابزار توسعه‌یافته بوده هستند که می‌توان در پژوهش‌های بعدی به کمک استفاده از مدل‌های پیچیده‌تر بر مبنای یادگیری عمیق، آنها را برطرف کرد.

## مراجع اصلی

1. R. Mazochi, s. Bourbour, M. R. Ghofrani and S. Momtazi, "Persian Address Geocoding: an LALR Parsing and Dynamic Programming Approach," *Journal of AI and Data Mining*, vol. 11, pp. 291-302, 2023.
2. Javidaneh, F. Karimipour and N. Alinaghi, "How Much Do We Learn from Addresses? On the Syntax, Semantics and Pragmatics of Addressing Systems," *ISPRS International Journal of Geo-Information*, vol. 9, no. 5, p. 317, 2020.
3. Koumarelas, A. Kroschk, C. Mosley and F. Naumann, "Experience: Enhancing Address Matching with Geocoding and Similarity Measure Selection," *J. Data and Information Quality*, vol. 10, no. 2, pp. 1-16, 2018.