BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023

http://excel.fit.vutbr.cz

Model of Cycling Traffic Intensity in Brno

Bc. Radoslav Eliáš

Abstract

The Brno Data Department has access to multiple datasets regarding cycling traffic numbers. They approached the faculty with a goal of developing a model integrating these sources for the Transportation Department of the city planning office to gain insight about how the infrastructure is used daily. Each dataset is aggregated to a different basemap with a slightly different street network. This thesis introduces an algorithmic approach to street matching based on similarity, overlap percentage and other parameters. Two algorithms for matching point-based and polyline-based geometries are presented, as well as a dashboard visualizing values from different datasets side-by-side. The robustness of the algorithms enables usage in any geographical application using spatial data. The dashboard provides useful information about cycling transport for both casual users and professionals designing the infrastructure of Brno.

*xelias18@vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

[Motivation] Car-centric cities are bad for the environment, space-consuming, and exclusive to car owners, disregarding children, the elderly, disabled people, and many others. Bicycles are a great alternative way of transport, but integrating them into the existing infrastructure comes with challenges. Providing more information to the Transportation Department enables them to make better decisions and make cycling more appealing and popular.

[Problem definition] There are multiple datasets available, however, each has only a partial temporal or spatial coverage of the city. Integrating them would improve both. Basemap inconsistencies deny the possibility of a simple join operation, so an algorithmic approach is required to match streets among different networks.

[Existing solutions] Most cities use a single dataset or multiple ones but separately. The University of Salzburg built an agent-based model of cycling traffic intensity based on pre-defined rules of route selection and origin-destination combinations.

[Our solution] The algorithm for integrating different street networks uses coordinates rounding to transform a pseudo-continuous space into a discrete one. It weights polyline similarity by angles and uses a bounding-box overlap percentage approach to matching. The final data model utilizes indirect encoding and mapping to endpoint datasets. Data are visualized by an interactive dashboard.

[Contributions] The matching algorithm is a unique approach and is built robustly, so it can be used in a multitude of applications, any time basemap inconsistencies happen. The integrated model of cycling traffic in Brno will help both the general public and the Transportation Department to better understand the behavior of cyclists.

2. Poster commentary

2.1 Problem Definition

The first Section titled "Problem definition" introduces the problem this thesis is trying to solve. Available datasets include automatic bike sensors that detect cyclists at various places around the city, data from the "BikeToWork" campaign that encourages sustainable ways of transport by creating a sort of competition where groups of people or companies can contest who can ride the most kilometers in a month, data from cycling census statistics conducted by the Brno Communications Company, and data from the Strava sports application used to collect and share GPS routes.

Each dataset aggregates the gathered number of cyclists to a different street network. These are incompatible and have multiple inconsistencies listed in the poster, like different partitioning or an occurrence where the same street in different datasets is a few meters apart. The Figure in this section tries to show an example of all the inconsistencies in one place. The same street is split into smaller segments differently in each basemap, they're not exactly parallel, and the BikeToWork instance is offset significantly.

2.2 Algorithm for Street Matching

The goal of the algorithm is to find the same street in different basemap networks and create a mapping that enables comparing values among sources of the data. To eliminate some of the offset, the coordinates of all the points of a street are rounded to a specific number of decimal places. This transforms the pseudo-continuous space into a discrete one, as can be seen in the bottom Figure of this section. Rounding to 5 decimal places means points of the discrete space are spaced out by around 2 meters.
Next, it's assumed that streets with a big angle between them are less likely to represent the same road. However, to calculate an angle between two polylines, it's required to take the bounding box of the street and extract the diagonal line of the box. Then, an angle can be calculated by the arc-cosine of the dot product of unit vectors. Streets with smaller angles are preferred by the algorithm.
Finally, the best match score is evaluated by a percentage of the overlap between buffers around both streets. Thus, the lines may not align or intersect at all and still be matched. Both the angle and the overlap principles are displayed in the top two figures in the section.

2.3 Segmentation Optimalization

The selected area of Brno for the case study has around 65,000 streets. Comparing them naively N to N between just two datasets would sum up to 4,225,000,000 comparisons alone. To bring this number down, an optimization technique was implemented, where it's assumed that significantly distant streets can't represent the same section of the road, thus it's not needed to compare them.
The selected window of the map is broken down into smaller, equally sized segments. These are generated as a read-only collection and each dataset is annotated with a new column with IDs of the segments, where the street in question lies. Then, streets are attempted to match with only ones in the same segment. The number of segments is a hyper-parameter of the method, but for example, 1024 segments in case of an even distribution would have 63 streets on average per each segment, therefore the final number of comparisons would be $63^2 * 1024 = 4,064,256$. A visual representation of this technique is shown in the figure on the poster.

2.4 Data Model

The proposed data model (Figure in Section Data Model of the poster) uses the OpenStreetMap[1] street network as a source of truth. It provides useful information about the road type, state, etc. It's then matched by the aforementioned algorithm to the end-point datasets of Brno cycling traffic. They are mapped by their corresponding IDs, so their internal values aren't duplicated but just indirectly encoded and mapped to one another. This way, appending new data is handled by their owners and doesn't need to be addressed in this thesis. Some values are omitted in the Figure for simplicity.

2.5 Dashboard

A screenshot of the Dashboard used to present these integrated data to users is displayed at the bottom part of the poster. It's created using the ArcGIS portal, as the Brno Data Department already uses this solution, and the data owned by the city are already hosted there. This means that the datasets don't have to be transferred or hosted independently, as the ArcGIS dashboard can natively query these layers that are public.
The main view of the dashboard uses the integration model. Picking a street in the map viewer displays the user a selection of tables of charts on the right-hand side, built from the data regarding the selected street. The left sidebar holds a date picker that serves only to query data from the automatic sensors, as these are generated on an hourly basis. The map viewer can be switched to a variety of heatmaps showing some common usages of the datasets in the below section.
Please note that the Strava data couldn't be displayed here, as they are subject to a very strict set of GDPR rules and can't be shared with anyone without explicit access from the Strava Metro team. Another dashboard will be made including Strava data, which will be **private** for the Brno Transportation Department. Anyone is welcome to try the dashboard, which can be found using the QR code next to this section. ArcGIS portal currently has a bug where it might prompt you to log in to an account, but closing this window a few times will provide you with access to the dataset.

References

[1] OpenStreetMap contributors. Openstreetmap.