A Road Segment Attribute Completion System

Razvan-Gabriel Cirstea, Hilmar Gústafsson, Rasmus Riis Grønbæk Pedersen,
Rolf Hakon Verder Sehested, Tamas Imre Winkler, Bin Yang
Department of Computer Science, Aalborg University, Denmark
razvan@cs.aau.dk; {hgusta16, rrgp16, rsehes16, twinkl18}@student.aau.dk; byang@cs.aau.dk

***Abstract*—High-quality location based services rely on complete and accurate information of road segments. However, the attributes of road segments in online maps are often incomplete. For example, to compute fastest routes, a navigation system requires information, such as speed limits and road categories, of all road segments. While in OpenStreeMap, such attributes are often missing for many road segments. To contend with incomplete attributes, we propose a system that is able to utilize different machine learning techniques, including both non-deep learning and deep learning algorithms, to fill in the missing attributes. The system is developed and integrated into *aSTEP*, a spatio-temporal data analytic platform developed by Aalborg University, and is tested using data collected from four major Danish cities.**

I. Introduction

Recent transportation innovations, e.g., transportation-as-a-service and autonomous driving, call for high-resolution routing, such as stochastic routing [1]–[3], personalized routing [4]–[6], and eco-routing [7]. A prerequisite of any kind of routing is that all road segments are associated with relevant attributes, such as speed limits, road categories, travel time, and fuel consumption, etc. However, it is often the case that only some, but not all, road segments, are associated with such attributes. We call this the *data sparseness challenge*. For example, in OpenStreetMap, a significant portion of the Danish road segments are missing speed limits and road categories. Travel time and fuel consumption are often derived from GPS data, but GPS data is often skewed, which cannot cover all road segments, even when using large GPS data sets [8]. Thus, the travel time and fuel consumption information is also often incomplete. The data sparseness challenge adversely affects routing quality.

We demonstrate a system that contends with the data sparseness challenge—it provides practical solutions to fill in missing road segment attributes such that a road network with complete road segment attributes can be offered to various routing algorithms. Specifically, the system models the problem as a classification problem. The road segments with relevant attributes are employed as training data. The available attribute of a training road segment is employed as its label, indicating, e.g., the road category or the speed limit of the road segment. The system offers a wide variety of machine learning algorithms to solve the classification problem, including deep learning vs non-deep learning and inductive vs transductive learning. Finally, the system is able to return estimated attributes for the road segments that are originally without attributes.

We test the system on the four largest cities in Denmark: Copenhagen, Aarhus, Odense, and Aalborg. The system fills in eight types of attributes, e.g., speed limit and road categories, using relevant information collected from OpenStreetMap. The system is developed based on *aSTEP*, a spatio-temporal data analytics platform developed at Aalborg University [9].

II. System overview

Figure 1 gives a system overview. The input is the OpenStreetMap data from a city or a rectangular region.

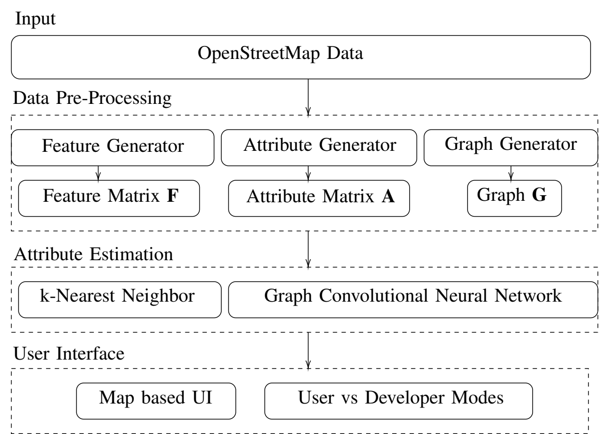

Fig. 1. System Overview

The input data is pre-processed through a Segment Feature Generator, a Segment Attribute Generator, and a Graph Generator. The Segment Feature Generator extracts relevant features of each segment, e.g., length, number of lanes, coordinates of the endpoints, whether it is connected with roundabouts or junctions. The output is a feature matrix $\mathbf{F} \in \mathcal{R}^{n \times m}$, where n is the number of road segments and m is the number of features. Segment Attribute Generator associates attributes, e.g., speed limits or road categories, with the edges which have relevant information. The output is an attribute matrix $\mathbf{A} \in \mathcal{R}^{n' \times x}$, where $n' < n$ is the number of road segments that are already associated with relevant attributes, which are used as training data; and x is the number of attributes. The Graph Generator returns a graph G that models the topology of the road segments in the underlying road network.

The Attribute Estimation module uses two algorithms to estimate road attributes. First, the k-nearest-neighbor algorithm, a non-deep learning and inductive learning algorithm, builds a classifier using $\mathbf{F}$ and $\mathbf{A}$ to estimate the missing road attributes. Second, the graph convolution neural network [10],

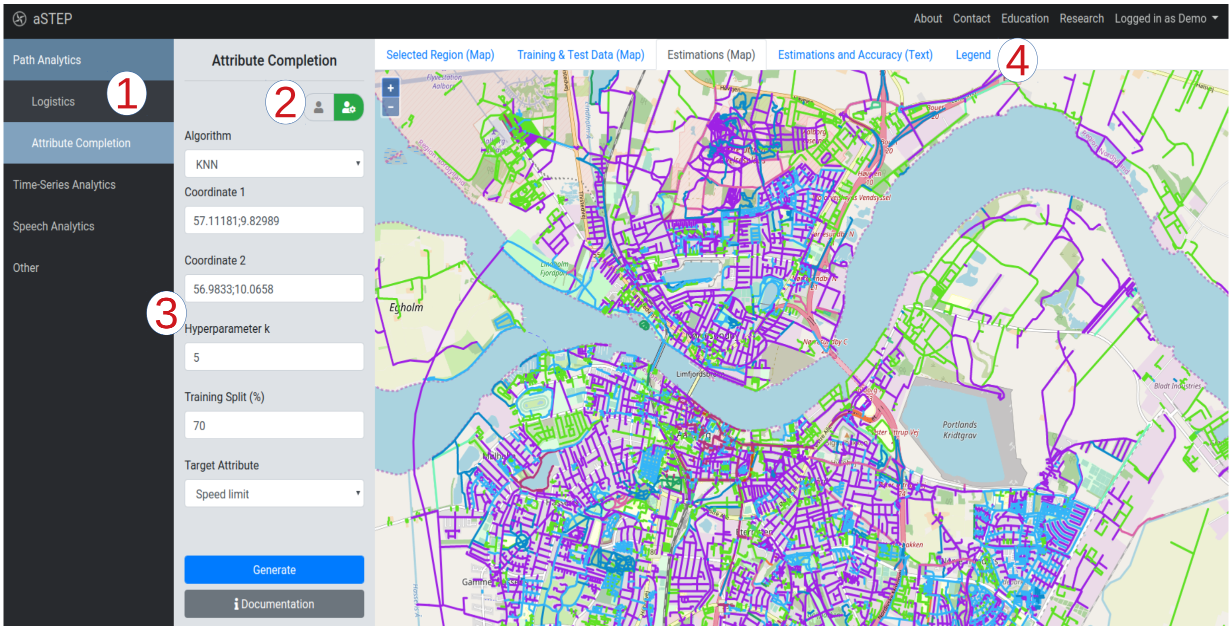

Fig. 2. Demonstration Outline

[11], a deep-learning and transductive learning algorithm, estimates the missing road attributes using features in $\mathbf{F}$, available road attributes in $\mathbf{A}$, and the road segment topology captured by graph G, which facilitates to propagate attributes to the segments without attributes. As future work, ensemble learning [12] and distributed computing [13] may be used to improve accuracy and efficiency, respectively.

Finally, we provide a map based UI with two modes, a user mode and a developer mode. In the user mode, a user is able to see the original road attributes vs. the estimated road attributes along with the accuracy of the estimations. The developer mode provides means to change different parameter settings of the learning algorithms which facilitate developers to fine tune the algorithms to achieve the best accuracy.

III. Demonstration Outline

We proceed to describe how users may interact with the system, which is integrated into aSTEP (https://astep.cs.aau.dk), a spatio-temporal data analytics platform developed by Aalborg University [9]. To try the segment attribute completion system, click "Path Analytics" and then "Attribute Completion" (Label 1 in Fig. 2). Users may switch between the user vs. the developer modes by using the toggle button (Label 2). Fig. 2 shows the developer mode. Label 3 allows users to select various settings in the developer mode—selecting a specific learning algorithm, a specific region defined by the coordinates of the region corners, hyper-parameters, training-testing split, and the target attribute to be estimated, e.g., speed limits. After filling in all relevant information, clicking the "Generate" button enables the system to estimate the missing attributes.

The system provides multiple views to show the estimated results (Label 4). The "Estimations (Map)" view allows users to see the results on a map where different roads are associated with different colors based on their estimated attributes. Clicking the "Legend" button shows an explanation of different colors. By clicking the "Estimations and Accuracy (Text)" button, the estimated attributes are shown in text. In addition, the average accuracy of the system is also shown.

Acknowledgements: This work was supported by Independent Research Fund Denmark under agreements 8022-00246B and 8048-00038B.

References

[1] J. Hu, B. Yang, C. Guo, and C. S. Jensen, "Risk-aware path selection with time-varying, uncertain travel costs: a time series approach," *VLDB J.*, vol. 27, no. 2, pp. 179–200, 2018.
[2] J. Hu, B. Yang, C. S. Jensen, and Y. Ma, "Enabling time-dependent uncertain eco-weights for road networks," *GeoInformatica*, vol. 21, no. 1, pp. 57–88, 2017.
[3] S. A. Pedersen, B. Yang, and C. S. Jensen, "Fast stochastic routing under time-varying uncertainty." *VLDB J.*, online first, 2019.
[4] C. Guo, B. Yang, J. Hu, C. S. Jensen, and L. Chen, "Context-aware, preference-based vehicle routing." *VLDB J.*, online first, 2020.
[5] H. Liu, C. Jin, B. Yang, and A. Zhou, "Finding top-k shortest paths with diversity," *IEEE TKDE*, vol. 30, no. 3, pp. 488–502, 2018.
[6] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Distinguishing trajectories from different drivers using incompletely labeled trajectories," in *CIKM*, 2018, pp. 863–872.
[7] C. Guo, B. Yang, O. Andersen, C. S. Jensen, and K. Torp, "Ecosky: Reducing vehicular environmental impact through eco-routing," in *ICDE*, 2015, pp. 1412–1415.
[8] C. Guo, B. Yang, J. Hu, and C. S. Jensen, "Learning to route with sparse trajectory sets," in *ICDE*, 2018, pp. 1073–1084.
[9] M. Beuchert, S. H. Jensen, O. A. Sheikh-Omar, M. B. Svendsen, and B. Yang, "aSTEP: aau's spatio-temporal data analytics platform," in *MDM*, 2018, pp. 278–279.
[10] J. Hu, C. Guo, B. Yang, and C. S. Jensen, "Stochastic weight completion for road networks using graph convolutional networks," in *ICDE*, 2019, pp. 1274–1285.
[11] J. Hu, C. Guo, B. Yang, C. S. Jensen, and H. Xiong, "Stochastic origin-destination matrix forecasting using dual-stage graph convolutional, recurrent neural networks," in *ICDE*, 2020, pp. 1417–1428.
[12] T. Kieu, B. Yang, C. Guo, and C. S. Jensen, "Outlier detection for time series with recurrent autoencoder ensembles," in *IJCAI*, 2019, pp. 2725–2732.
[13] P. Yuan, C. Sha, X. Wang, B. Yang, and A. Zhou, "XML structural similarity search using mapreduce," in *WAIM*, 2010, pp. 169–181.