Extended Abstract submitted for TransitData 2019

Extraction of public transportation networks from Openstreetmap data and estimation of OD patterns through a graph convolutional approach

Koji Fukuda[1], Saeed Maadi[2], Kai Shen[3] and Jan-Dirk Schmöcker[3]

[1]Hitachi Kyoto University Laboratory; [2]School of Computing Science, University of Glasgow; [3]Department of Urban Management Kyoto University

Introduction: OSM data

OpenStreetMap (OSM) is the prime crowed-sourced open access platform used by a range of applications. Though quality of data varies, it can provide information on transportation networks in most parts of the world. Recently, also the quantity and quality of uploaded information regarding the public transport networks has significantly improved though issues remain. In this research we first describe the development of an algorithm that allows simple extraction of Points of interests, road networks and public transportation networks in terms of stops and routes. This allows us to easily compare and analyse network quality in different cities. With a few examples we discuss first potential problems of the OSM data. We then propose that in particular the extracted PT network can be used to understand passenger origin and destination patterns of the network due to the endogeneous relationship between demand and supply.

OD estimation

Recently, with the advancement of deep learning, graph convolutional network (GCN) analysis has attracted much attention. As an application of the extracted PT graph from OSM, we estimate origin and destination (OD) patterns. We suggest that the OD patterns can be estimated from the map features including learning land-use patterns from the map. For example, from a residential area there will be commuting traffic to a nearby industrial area, and shopping traffic to commercial areas. We assume that the traffic demand T_{ij} from point i to the point j is expressed as following generalised gravity model using d-dimensional vector value functions f and g:

$$T_{ij} = f\left(a_i, a_j\right)^T g\left(d_{ij}\right)$$

where a_i is a feature vector representing the "*atmosphere*" *(land-use)* of point i, *e.g.* a residential area or an industrial area, and d_{ij} is the "*closeness*" from point i to point j considering generalised travel costs. We deduct the functions f and g as well as a_i and d_{ij} by supervised learning with GCN. With this we estimate the OD patterns.

First, the atmosphere of the point i is derived. Let $\mathrm{G} = (\mathcal{V}, \mathcal{E}, \mathcal{R})$ denote a PT graph with nodes $v_{\mathrm{i}} \in \mathcal{V}$ and labelled edges $\left(v_{\mathrm{I}}, r, v_{\mathrm{j}}\right) \in \mathcal{E}$ with a relation r, where r represents the type of transportation, i.e., road, bus, and railway. Each node v_{I} has a one-dimensional vector x_i that represents the POIs (station, school, house, factory, commercial facility, etc.) located at the point.

By applying a graph convolution layer defined below L times with an input value $h_i^{(0)} = x_i$, the atmosphere of the node is derived as $a_i = h_i^{(L)}$,

$$h_i^{(l+1)} = \sigma\left(\sum_{r\in\mathcal{R}} \sum_{j\in\mathcal{N}_i^r} W_r^{(l)} h_j^{(l)} + W_o^{(l)} h_i^{(l)}\right)$$

where $\mathcal{N}_i^r = \{j \mid (v_\mathrm{I}, r, v_\mathrm{j}) \in \mathcal{E}\}$ denotes the set of neighboring nodes of i with relation r, σ is the ReLU function and $W_r^{(l)}$ and $W_o^{(l)}$ are learned parameters representing convolutional weights. We use a ComplEx model [1] to represent the function f to consider asymmetry of the origin and the destination.

$$f(a_i, a_j) = a_i^T\ Q\ a_j$$

where Q is a set of learned parameters representing a $d \times d$ real normal matrix.

Next, the function $g(d_{ij})$ is derived. Given a path (chain of edges), we use a long short time memory (LSTM) structure to deduce its closeness. $g(d_{ij})$ is the sum of the LSTM output applied to all the reasonable paths $p \in P_{ij}$ from point i to point j.

$$g(d_{ij}) = \sum_{p\in P_{ij}} \mathrm{LSTM}(p)$$

Validation with Smartcard data

In order to validate the proposed method, we estimated the OD pattern using trip data from networks available by *Transportation Networks for Research Core Team* [2]. First, all the nodes are divided into 70% and 30%, and using the OD pattern between 70% of the nodes as teacher data, the parameters were trained so as to minimize the square error between the estimated T_{ij} and the true value. Then, the OD flow value with the remaining 30% nodes as the origin or destination was estimated and compared with the ground truth value. In addition, in order to confirm the effectiveness of the actual data, we estimated the OD pattern using bus smart card data in a mid-sized city in Japan. Using the parameters trained with the boarding/alighting data between 70% of the bus stops, the OD patterns among the remaining 30% of the bus stops was estimated. Based on these results, we discuss the effectiveness and robustness of the proposed method.

[1] Théo Trouillon, Johannes Welbl, Sebastian Riedel,Éric Gaussier, and Guillaume Bouchard. *Complex embeddings for simple link prediction.* In Proceedings of the 33rd ICML, pp. 2071–2080, 2016.
[2] Transportation Networks for Research Core Team. Transportation Networks for Research. https://github.com/bstabler/TransportationNetworks. Accessed March 2019.