New Techniques in Road Network Comparison

Mahmuda Ahmed*, Brittany Terese Fasy†, Carola Wenk‡

14 March 2014

Abstract

Road networks are always changing: new streets are built, accidents and floods close roads, etc. Detecting when, and if, a change has occurred is an important question. In this presentation, we highlight recent progress in computing the distance between two road networks.

Audience

This talk is intended for a general computer science audience, who need not be familiar with road network reconstruction and comparison. We will highlight several algorithms, but will avoid the tedious technical details.

1 Introduction

The task of comparing road networks has received a lot of attention lately with the emergence of algorithms to reconstruct road networks from GPS trajectory data. For example, the Open Street Map project[1] provides crowd-sourced data freely to the public. Additionally, several automatic road network reconstruction algorithms have been proposed; see e.g. [1, 3, 7, 8, ?]. However, it remains a challenge to evaluate the quality of the reconstructed networks, even in the presence of the true road network. In this short abstract, we highlight several distance measures between road networks.

We model a road network as an embedded planar graph, $G = (V, E) \subset \mathbb{R}^2$. We assume that V is a set of vertices with degree $\neq 2$ and each edge in E is represented as a polygonal curve. That is, intersections in the road networks are vertices and (piecewise linear) road segments connecting consecutive intersections make up the edges.

*Department of Computer Science, University of Texas at San Antonio, mahmed@cs.utsa.edu

†Department of Computer Science, Tulane University, britttany.fasy@alumni.duke.edu

‡Department of Computer Science, Tulane University, cwenk@tulane.edu

[1]http://www.openstreetmap.org

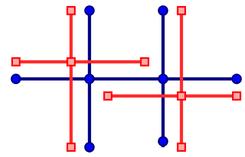

Figure 1: Road networks can have a small Hausdorff distance, but different topological structures.

2 Approaches

We now assume that we have two embedded graphs: $G_0 = (V_0, E_0)$ and $G_1 = (V_1, E_1)$. For example, G_0 could represent the true road network, and G_1 could be the approximation of the true network, computed from GPS trajectory data. We discuss several techniques for measuring the distance between G_0 and G_1.

One of the natural ways to measure distance given two embedded objects is the Hausdorff distance. However, one could then allow networks with disconnected travel paths to be very similar, even though driving routes on the two would necessarily be very different; consider the maps in Figure 1. Below, we outline three distance measures between embedded graphs that are useful for the application of road network comparison.

2.1 Hänsel and Gretel Distance

As we mentioned above, the Hausdorff distance does not take the local topology into account. In [5], a sampling-based distance is proposed that incorporates the topology of the graphs.

Fix parameters $r > 0$ (locality radius), $d > 0$ (jump distance), and $\delta > 0$ (neighborhood threshold). We choose a random point in G_0. This is our seed s. We then place a red bread crumb at s, as well as at all points in G_0 at distance kd from s for k an integer and $kd < r$. Here, we are measuring the distance within G_0. We repeat this process with graph G_1, placing blue bread crumbs. We now find a maximum matching between the red bread crumbs and the blue bread crumbs, where we can match a red bred crumb and a blue bread crumb if their distance is at most δ. We now have n_s red bread crumbs and m_s blue ones, of which k_s are matched. Repeating this process for a large number of seeds taken i.d.d., let $n = \sum_s n_s$, $m = \sum_s m_s$, and $k = \sum_s k_s$. We then compute the statistical precision and recall:

$$\texttt{pre}_{0,1} = \frac{k}{m} \quad \text{and} \quad \texttt{rec}_{0,1} = \frac{k}{n}.$$

The F-score is the distance measure:

Definition 2.1 (HG-Distance) *The* Hänsel and Gretel (HG) distance *is the statistical F-score, with respect to G_0, given by*

$$F(G_0, G_1) = 2\frac{pre_{0,1}\,rec_{0,1}}{pre_{0,1} + rec_{0,1}}.$$

2.2 Path-Based Distance

Another approach to computing distances between maps is to quantify how similar or different is to travel within a road network. We consider the set of paths between two vertices u and v in G_i. A *path* between u and v is the image of a continuous map $\alpha\colon [0,1] \to G_i$ such that $\alpha(0) = u$ and $\alpha(1) = v$. We denote the set of all paths in G_i by Π_i. Our distance measure between two embedded graphs is based on the Fréchet distances between paths in Π_0 and Π_1.

Definition 2.2 (Fréchet Distance) *For two planar curves $f, g : [0,1] \to \mathbb{R}^2$, the Fréchet distance δ_F between them is defined as*

$$\delta_F(f,g) = \inf_{\alpha} \max_{t\in[0,1]} \|f(t) - g(\alpha(t))\|, \qquad (1)$$

where $\alpha : [0,1] \to [0,1]$ ranges over all continuous, surjective, non-decreasing re-parameterizations.

The Fréchet distance is a well-suited distance measure for comparing curves, or paths, because it takes continuity and monotonicity of the curves into account. The Fréchet distance between two polygonal curves with m and n vertices, respectively, can be computed in $O(mn \log mn)$ time [4].

For each path in one graph, we find the closest path in the other, with respect to the Fréchet distance. This leads to a directed distance measure:

Definition 2.3 (Path-Based Distance) *The directed Path-Based Distance between G_0 and G_1 is defined as:*

$$\overrightarrow{d}(G_0, G_1) = \max_{p_0\in\Pi_0} \min_{p_1\in\Pi_1} \delta_F(p_0, p_1). \qquad (2)$$

The undirected distance, $d(G_0, G_1)$ is defined as $\max(\overrightarrow{d}(G_0,G_1), \overrightarrow{d}(G_1,G_0))$, similar to the undirected Hausdorff distance. Like the Hausdorff distance, the path-based distance is not symmetric, i.e., $\overrightarrow{d}(G_0,G_1) \neq \overrightarrow{d}(G_1,G_0)$. This anti-symmetry is desirable in our setting. For example, G_1 can be the reconstructed road network from bus route data. In this case, the bus routes correspond to a subgraph of the complete road network G_0.

Our proposed distance measure is defined so that if the directed distance $\overrightarrow{d}(G_0,G_1)$ is small, then for any path p in G_0 there exists a corresponding path q in G_1 such that the Fréchet distance between p and q is small. From this property, we are able to show that, under reasonable assumptions on G_0 and for $\overrightarrow{d}(G_0,G_1)$ small enough, every vertex in G_0 has at least one naturally corresponding vertex in G_1 and every travel route defined in G_0 has a similar travel route in G_1. Moreover, we are able to give theoretical quality guarantees, and that can approximate the distance in polynomial time.

2.3 Local Homology-Based Distance

Recently ([2]), we have defined a distance measure between road networks that uses a concept called local homology (this paper is currently in progress and will be submitted to a conference in the near future). For the sake of brevity in the current exposition, we refer the reader to [10] for the formal definition of local homology, and to [6] for the definition of persistent homology.

Letting $r > 0$ be fixed, we create a local signature for every $x \in D$ as follows. Let B_x be the ball of radius r centered at x. We consider the distance function f_0 to G_0. Let $F_0(t)$ denote the set of all points in D that are distance at most t from G_0. The persistence module is the following sequence of homology groups:

$$H_1(F_0(0), D - B_x) \to H_1(F_0(t_1), D - B_x) \to \dots$$

$$\dots \to H_1(F_0(t_n), D - B_x) \to H_1(F_0(r), D - B_x),$$

where t_i are chosen to interleave the homological critical values. In words, we are thickening G_1 and monitoring the homology of the thickened graph restricted to B_r, taken relative to $\partial(B_r)$. Let $\mathcal{P}_{0,x}$ be the resulting persistence diagram. We define $\mathcal{P}_{1,x}$ similarly.

We now use $\mathcal{P}_{1,x}$ and $\mathcal{P}_{2,x}$ to compare the local topologies of G_0 and G_1 near x. We use the bottleneck distance $d_\infty(\mathcal{P}_{0,x}, \mathcal{P}_{1,x})$ (see [6, Ch. VIII]) as a local distance between G_0 and G_1 at x, for every $x \in D$.

This point-based distance can be used to visualize the distance between two graphs, as demonstrated in Figure 2, where we plot one map in gray and the second map colored by the local distance, with yellow indicating a small distance and red indicating a large distance. We integrate this local distance over $\mathbb{X}$ to obtain the fixed-radius local homology distance:

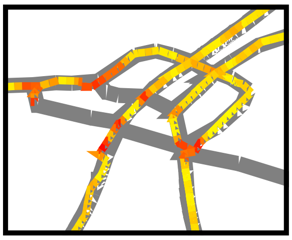

Figure 2: We see two road networks for an intersection in Athens, Greece. We plot one map in gray and the second map is colored by the local distance observed along that graph.

Definition 2.4 (LH Distance) *The fixed-radius local homology distance is:*

$$d^{LH}(G_0, G_1) = \frac{1}{|\mathbb{X}|} \int_{\mathbb{X}} d_\infty(\mathcal{P}_{0,x}, \mathcal{P}_{1,x})\, dx,$$

where $|\mathbb{X}|$ *denotes the Lebesgue measure of* $\mathbb{X}$.

3 Conclusion

This abstract does not present an exhaustive list of road network distance measures; see e.g. [9] for another one. We have, however, briefly described three distance measures that use the local connectivity of the road networks, which we believe to be necessary for the purpose of road network comparison.

Since there are various reconstruction algorithms, we may wish to determine which algorithm is better than the others. To do so, we can choose one of the distance measures above and evaluate $d(G_0, G_i)$ for each reconstruction G_i. Perhaps we could see a trend demonstrating a ranking of the reconstruction algorithms if we do so. We are also interested in understanding when one distance measure would be preferred over another.

Bio

Mahmuda Ahmed is a PhD candidate at University of Texas at San Antonio. Her research focuses on geospatial algorithms.

Brittany Terese Fasy is a postdoctoral researcher at Tulane University. She received her PhD from Duke University in 2012. Her research is in the intersection of topological data analysis and statistics.

Carola Wenk is an associate professor in the Department of Computer Science at Tulane University, in the School of Science and Engineering. She also holds an adjunct appointment in the Department of Mathematics.

References

[1] Aanjaneya, M., Chazal, F., Chen, D., Glisse, M., Guibas, L. J., and Morozov, D. Metric graph reconstruction from noisy data. In *Proc. ACM SoCG* (2011).

[2] Ahmed, M., Fasy, B. T., and Wenk, C. Local homology based distance between maps, 2014. In submission.

[3] Ahmed, M., and Wenk, C. Constructing street networks from GPS trajectories. In *Proc. European Symposium on Algorithms (ESA)* (2012), pp. 60–71.

[4] Alt, H., and Godau, M. Computing the Fréchet distance between two polygonal curves. *Int. J. of Computational Geometry and Applications 5* (1995), 75–91.

[5] Biagioni, J., and Eriksson, J. Inferring road maps from global positioning system traces. *Transportation Research Record: Journal of the Transportation Research Board 2291*, 1 (2012), 61–71.

[6] Edelsbrunner, H., and Harer, J. *Computational Topology. An Introduction.* Amer. Math. Soc., Providence, RI, 2010.

[7] Ge, X., Safa, I., Belkin, M., and Wang, Y. Data skeletonization via Reeb graphs. In *25th Annual Conf. on Neural Info. Proc. Sys.* (2011), pp. 837–845.

[8] Karagiorgou, S., and Pfoser, D. On vehicle tracking data-based road network generation. SIGSPATIAL '12, ACM, pp. 89–98.

[9] Karagiorgou, S., and Pfoser, D. On vehicle tracking data-based road network generation. SIGSPATIAL '12, ACM, pp. 89–98.

[10] Munkres, J. R. *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, CA, 1984.