A Hybrid Learning Approach to Stochastic Routing

Pedersen, Simon Aagaard; Yang, Bin; Jensen, Christian S.

Publication date:
2019

Citation for published version (APA):
Pedersen, S. A., Yang, B., & Jensen, C. S. (2019). *A Hybrid Learning Approach to Stochastic Routing*. Poster presented at Kick-off: AI for the people, Aalborg, Denmark.

A Hybrid Learning Approach to Stochastic Routing

Simon Aagaard Pedersen, Bin Yang, Christian S. Jensen
{sape, byang, csj}cs.aau.dk
Department of Computer Science, Aalborg University, Denmark

daisy
Center for Data-intensive Systems

Introduction

- Emerging disruptive innovations in transportation, e.g., autonomous vehicles and transportation-as-a-service, will benefit from high-resolution routing, where travel-time uncertainty is captured accurately.
- For example, when an autonomous taxi needs to arrive at an airport within a deadline, having accurate travel-time distributions of candidate paths enables the taxi to choose the best path.
  - Consider the path costs below. If the deadline is 60 minutes, path P_1 is better than path P_2, since P_1 gives a 0.9 probability of arriving within 60 minutes, which exceeds P_2's probability of 0.8.
  - If using average travel times, the taxi will choose P_2 that has an average travel time of 51 minutes vs. 53 minutes for P_1. Thus, the taxi has a higher risk of being late.

Travel Time Distributions of Two Paths to the Airport

Travel time (mins)	[40, 50)	[50, 60)	[60, 70)
P_1	0.3	0.6	0.1
P_2	0.6	0.2	0.2

- Traditional road network models often assume spatial independence for adjacent roads.
- This leads to the use of convolution for computing the cost of traversing a path in stochastic road networks.
- We propose a Hybrid Model: a model that combines machine learning and convolution to construct stochastic traversal costs in spatially dependent road networks.

Motivating Example

Observed trajectories

Observations	e_1	e_2	Total cost
T_1	10	20	30
T_2	15	25	40

Distributions for e_1 and e_2

H_1		H_2	
Travel Time	Probability	Travel Time	Probability
10	0.5	20	0.5
15	0.5	25	0.5

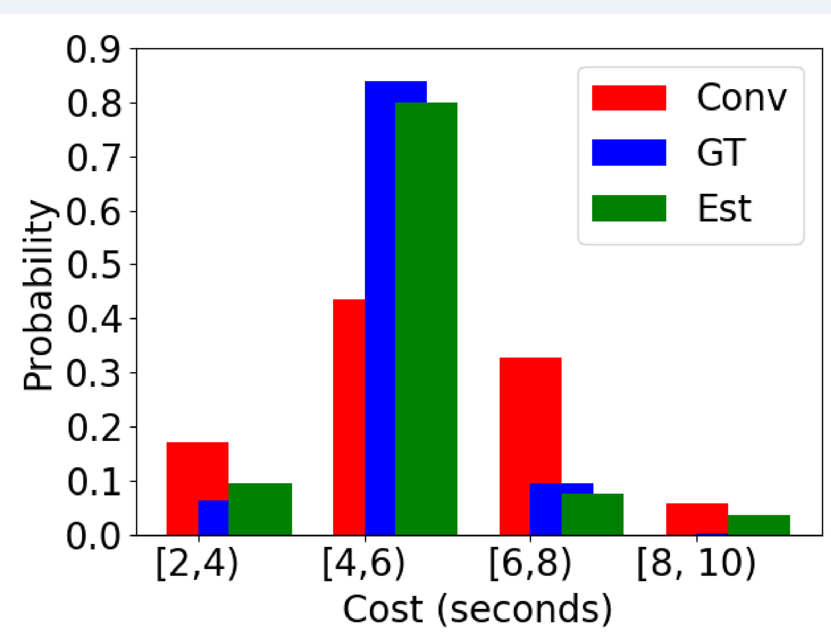
Result of convolving distributions for e_1 and e_2

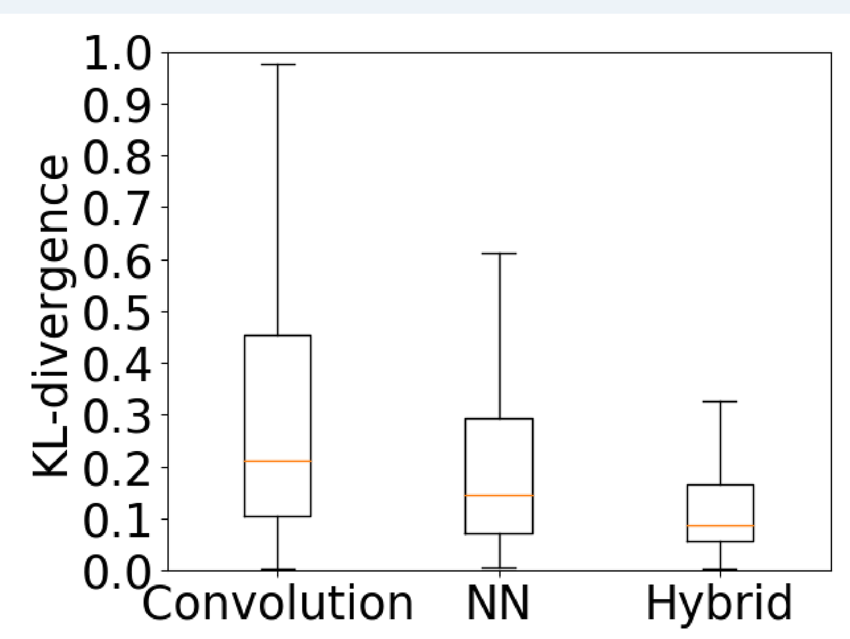
Travel Time	Probability
30	0.25
35	0.50
40	0.25

Ground Truth based on observed trajectories

Travel Time	Probability
30	0.50
40	0.50

Convolution vs. Estimation



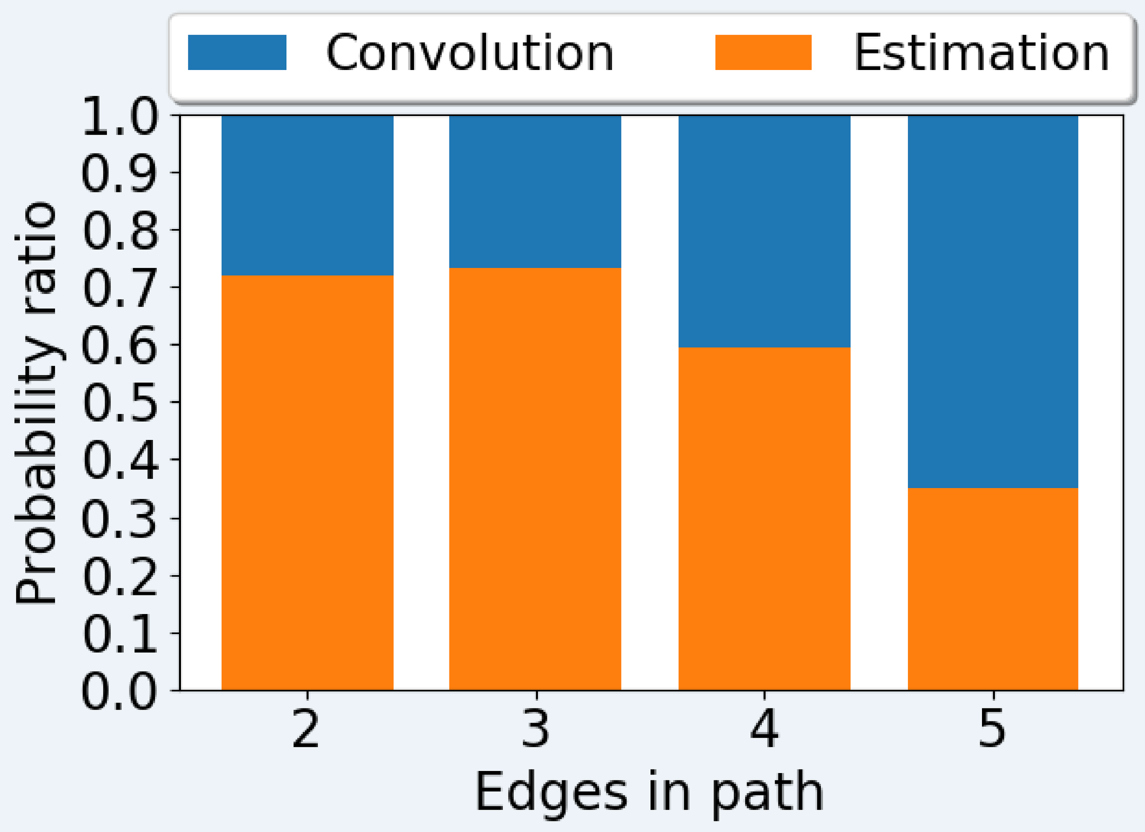


Hybrid Model Overview

- The Hybrid Model consists of two machine learning models:(i) a **distribution estimation model** that estimates the dependent uncertain cost of traversing two edges, and (ii) a **binary classifier** that determines if we should use convolution or estimation at a specific intersection.
- The estimation model is trained on 4000 edge pairs with sufficient data. An instance of the classifier is initialized for each estimation model.
- Following training, we test the model with a set of 1000 edge pairs, measuring the KL-divergence between the output and ground truth trajectories.

Path Cost Computation

- Path cost computation is an iterative process, as the cost of a path is computed by repeatedly combining the cost of the path so far with the cost of the next edge until the last edge is reached.
- We can use the distribution estimation model built for short paths to estimate the costs of longer paths by treating the path so far (pre-path) as a "virtual" edge.



Routing

- **Probabilistic Budget Routing**: Given a source, destination, and time budget t, find the path that maximizes the probability of arrival within t.
- A base algorithm uses pruning to run faster, including using (a) an A* inspired optimistic cost of reaching the destination for each vertex, (b) a pivot path representing the most promising return candidate, (c) distribution cost shifting, and (d) stochastic dominance pruning.
- To control the run-time, we also propose an *anytime* extension that limits the total run-time. With this approach, we give an acceptable maximum run-time x as an additional input, and the algorithm returns the pivot path if search has not terminated after x time units.

Empirical Studies

- Experiments done on the Danish road network using OpenStreetMap.
- The graph consists of 667,950 vertices and 1,647,724 edges.
- Approximately 75% of all edge pairs with data are dependent.
- We look at queries in distance categories: $[0, 1)$, $[1, 5)$, $[5, 10)$ km.

Quality

Dist (km)	P_∞	P_1	P_5	P_{10}
$[0, 1)$	13%	13%	13%	13%
$[1, 5)$	53%	51%	53%	53%
$[5, 10)$	60%	54%	59%	60%

Efficiency

Dist (km)	Mean (sec)
$[0, 1)$	0.06
$[1, 5)$	3.37
$[5, 10)$	9.73

AALBORG UNIVERSITET
INDEPENDENT RESEARCH FUND DENMARK