CENTER FOR DATA-INTENSIVE CYBER-PHYSICAL SYSTEMS

Data-Intensive Routing

Christian S. Jensen

www.cs.aau.dk/~csj

With numerous collaborators from

Daisy and beyond.

Roadmap

- Big data
- Take 1: Data-intensive conventional routing
  - Time-varying, uncertain weights
  - Complete weight annotation using incomplete data
  - Advanced routing
- Take 2: Trajectory-based routing
  - Routing based on local-driver behavior
- Closing

Digitization

- Instrumentation of reality
  - E.g., smartphones
- Digitization of processes
  - E.g., e-commerce, public services, payment, communications, social interactions, transportation

Big Data

Every day, we create 2.5 quintillion bytes of data — so much that 90% of the data in the world today has been created in the last two years alone. This data comes from everywhere: sensors used to gather climate information, posts to social media sites, digital pictures and videos, purchase transaction records, and cell phone GPS signals to name a few. This data is **big data**.

http://www-01.ibm.com/software/data/bigdata/

Big Data – Synthesis

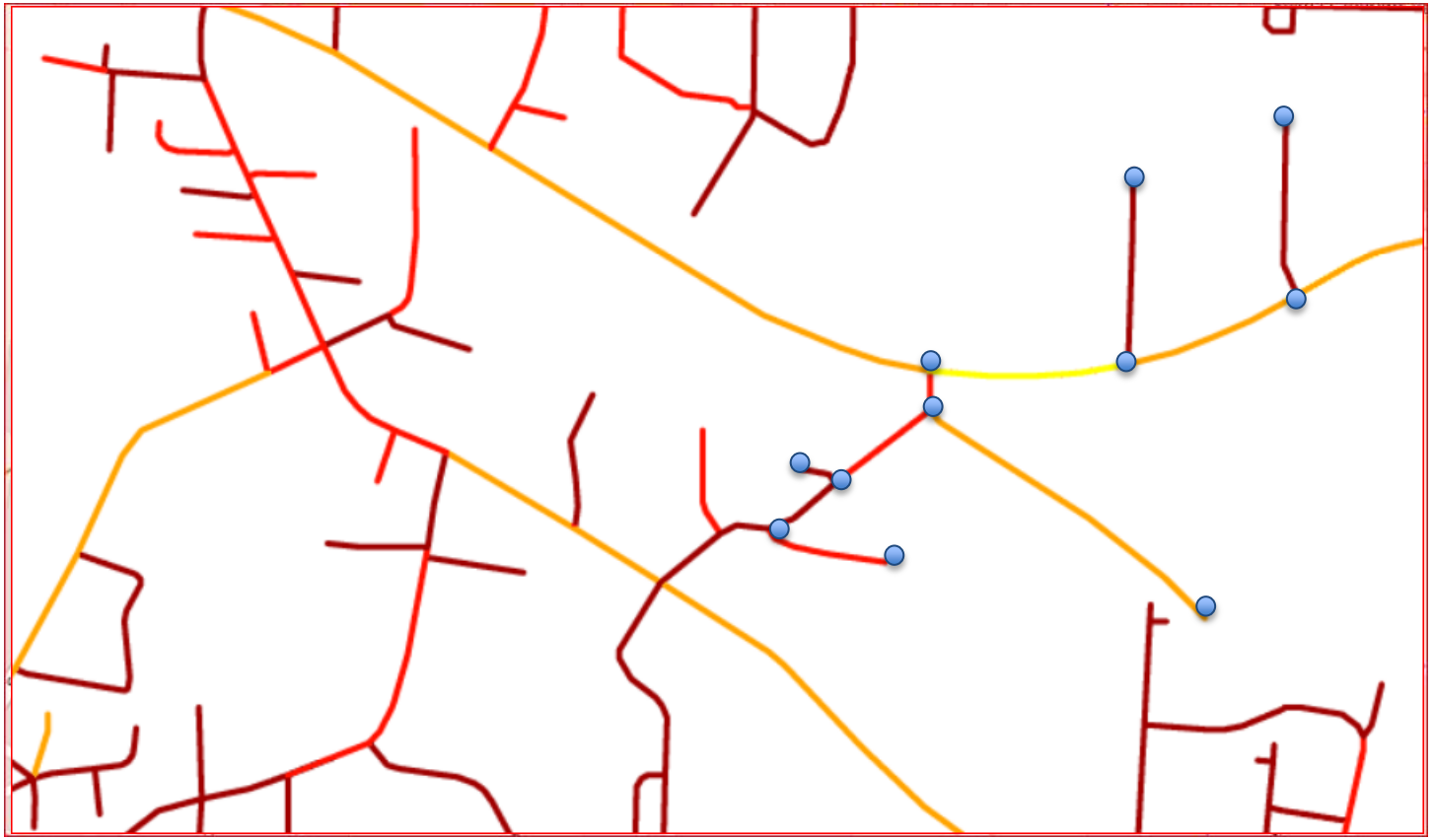
- Lots of data and unprecedented computing infrastructure combine to offer potentials for value creation from data.

- To be competitive, society and businesses must be able to create value from data.
- Data-based decisions and data-driven processes
  - Decisions based on good data beat decisions based on feelings or opinions.

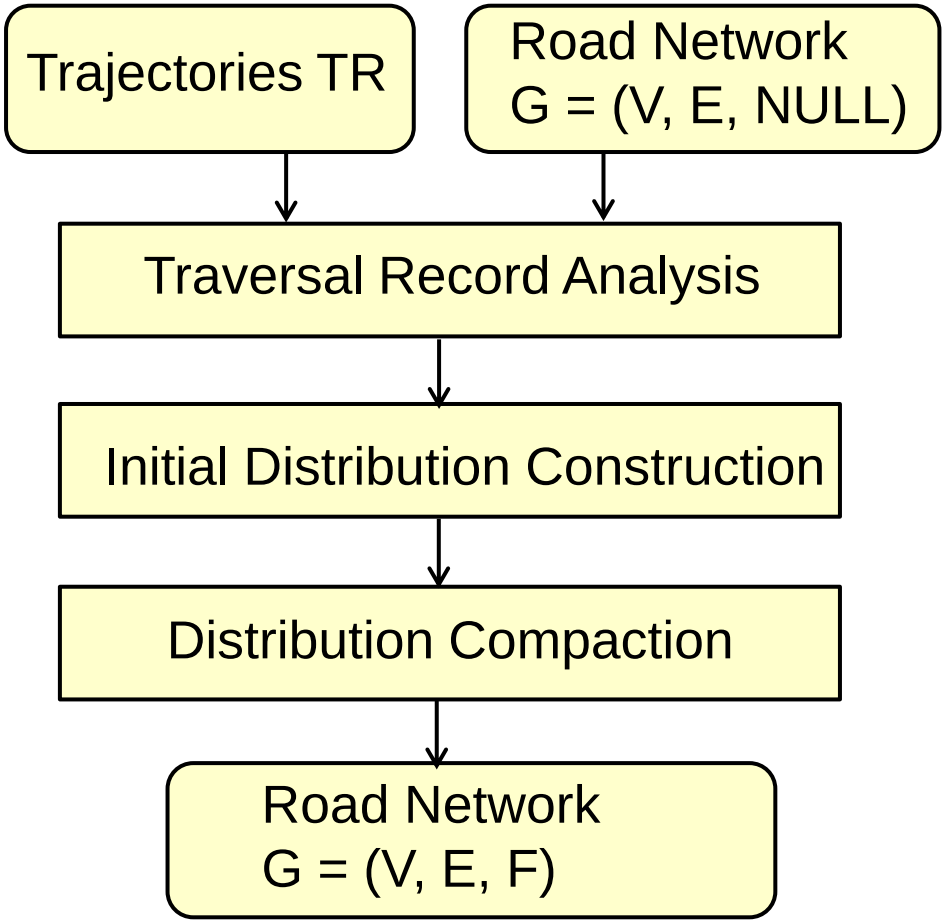
- A finer granularity of services
- Entirely new services

Take 1:
Extended Conventional Routing

Road Networks

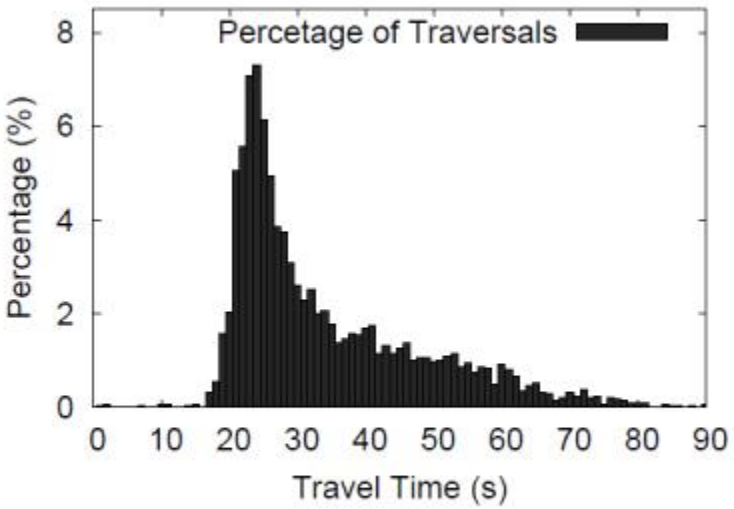


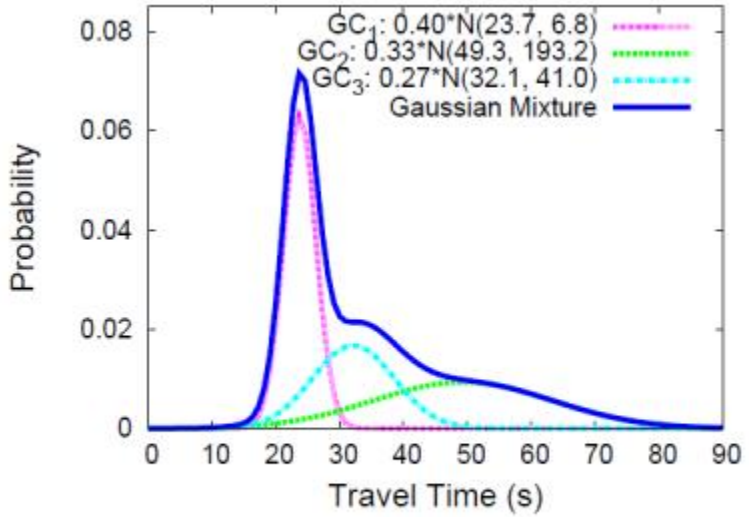
Framework


Trajectories TR
Road Network
G = (V, E, NULL)
Traversal Record Analysis
Initial Distribution Construction
Distribution Compaction
Road Network
G = (V, E, F)

Example Details

- Partition a week into 672 15-min intervals.
- For each (edge, interval) pair, obtain a multi-set containing the costs on the edge and interval.
  - {10 s, 20 s, 25 s, 30 s, 40 s, …}
- Estimate a distribution based on the multi-set.
  - Use a Gaussian Mixture Model or a histogram

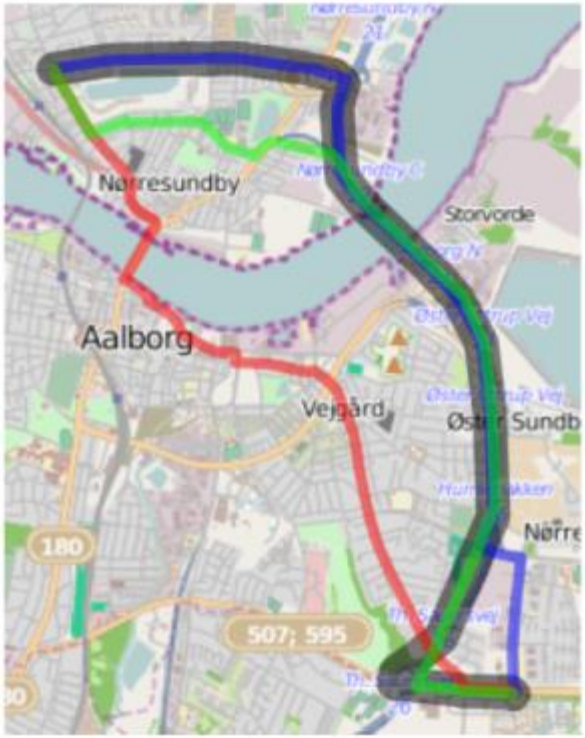




Research Topics

- Obtain weights for all road network edges
  - Incomplete data
  - Use data from one place in another place
- Stochastic skyline routing
  - Consider distance, travel time, fuel consumption
  - Return multiple routes that are not dominated by other routes
- Personalized routing
  - Different drivers may prefer different routes
  - Learn driver preferences and use these in routing
  - Compute personal weights

Example Results



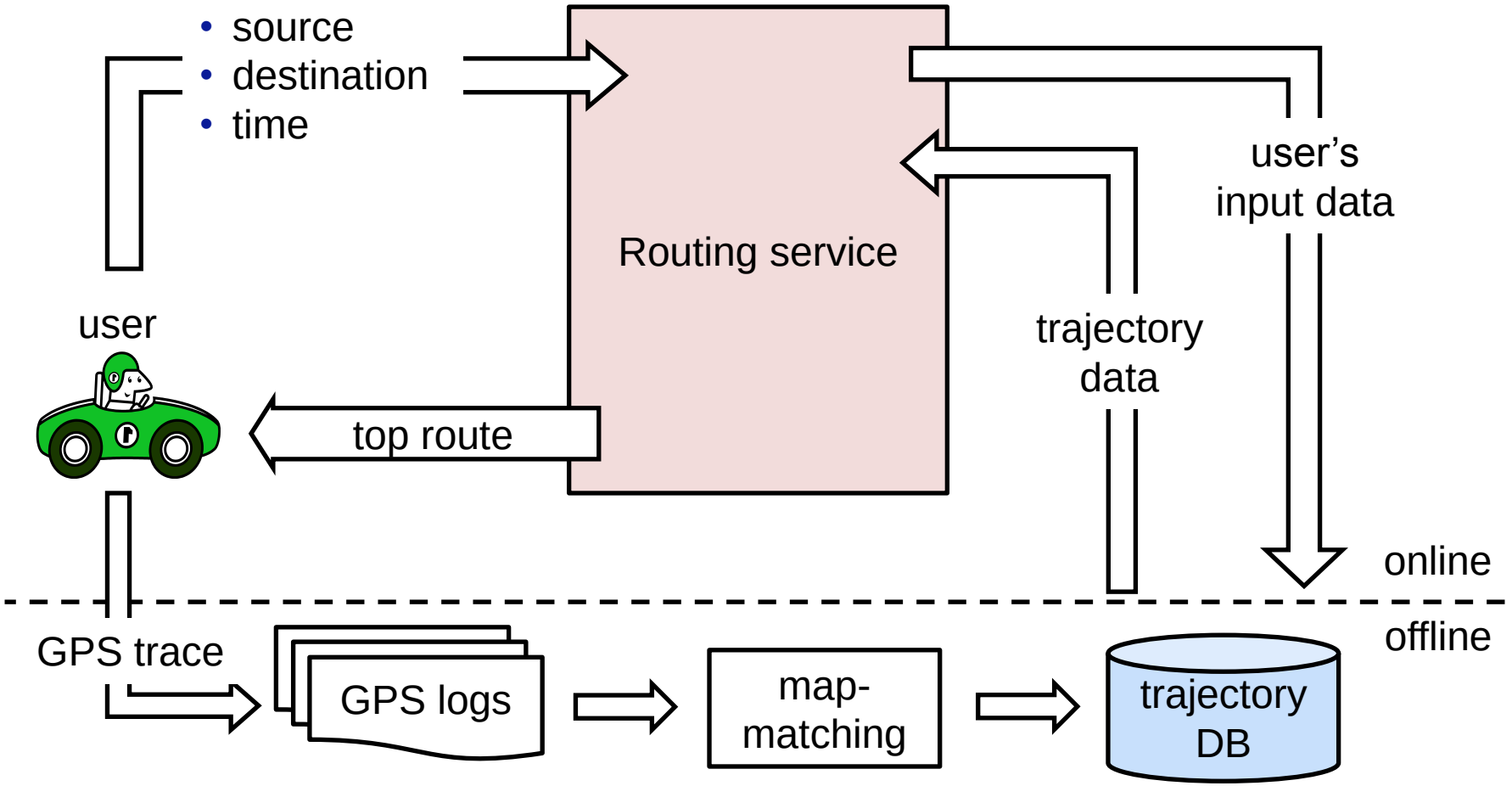
- Dark, bold routes: actual routes used by drivers.
- Red routes: shortest
- Green routes: fastest
- Blue routes: predicted routes.

Take 2:
Trajectory-Based Routing

Intuition

- Local travel
  - Knowledge of the surroundings
  - Follow familiar routes
- Travel in unfamiliar surroundings to unknown destinations
  - Depend on available routing services
  - Expect that the provided route is the best
- Idea: Use GPS data to let those who travel in unfamiliar surroundings benefit from the insights of local travelers

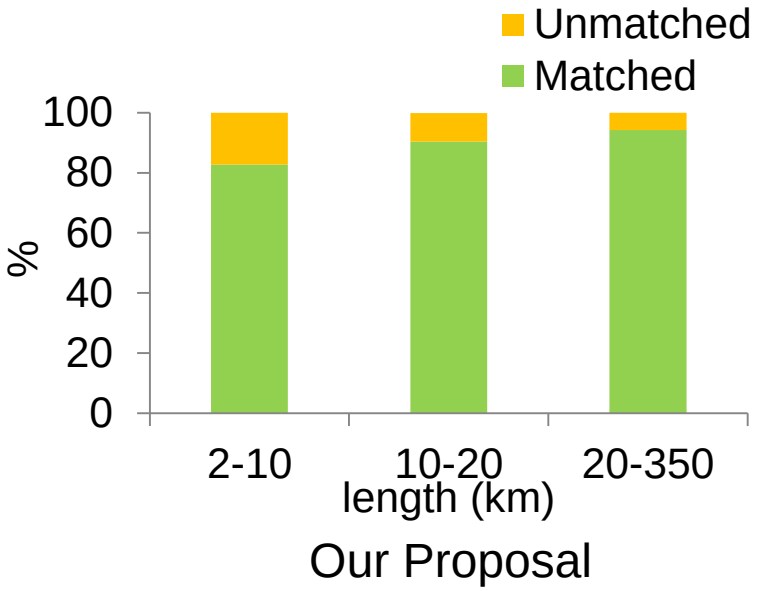
Framework


source
destination
time
Routing service
user's input data
user
trajectory data
top route
online
offline
GPS trace
GPS logs
map-matching
trajectory DB

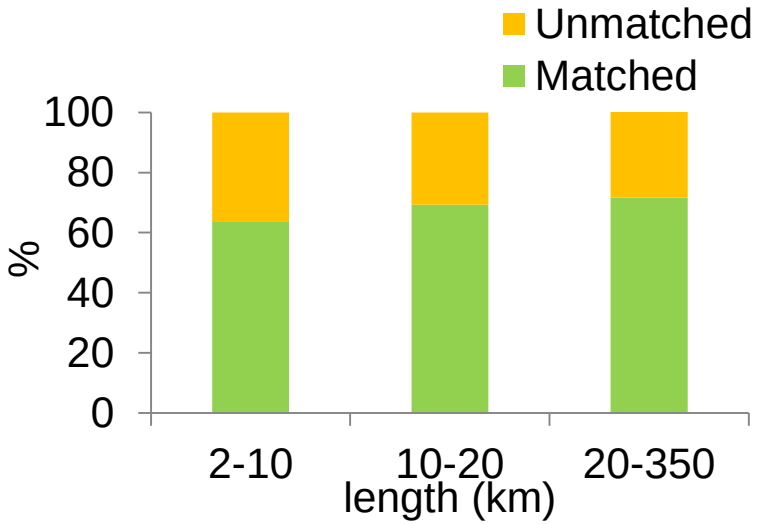
Empirical Study: Data

- Monitoring period: 2 years
- Number of drivers: 285
- Number of GPS points (raw data): ~182,700,000
- Number of trips: ~275,000
- “Pay as You Speed” project (http://www.sparpaafarten.dk)

Routing Quality Evaluation: Some Results



Our Proposal



Google Directions API (top-1)

Closing

The Future

- Much more travel data
  - GPS data from vehicles
  - Inductive loop detectors, Wi-Fi/Bluetooth
  - Collective transport data, e.g., bus data,
  - Multimodal collective transport data, e.g., “Rejsekortet”
- Much more connected vehicles
- New services
  - Routing
  - Safety and warnings
  - Parking, fees, insurance, road pricing
  - Car sharing, multi-modality
- Self-driving vehicles