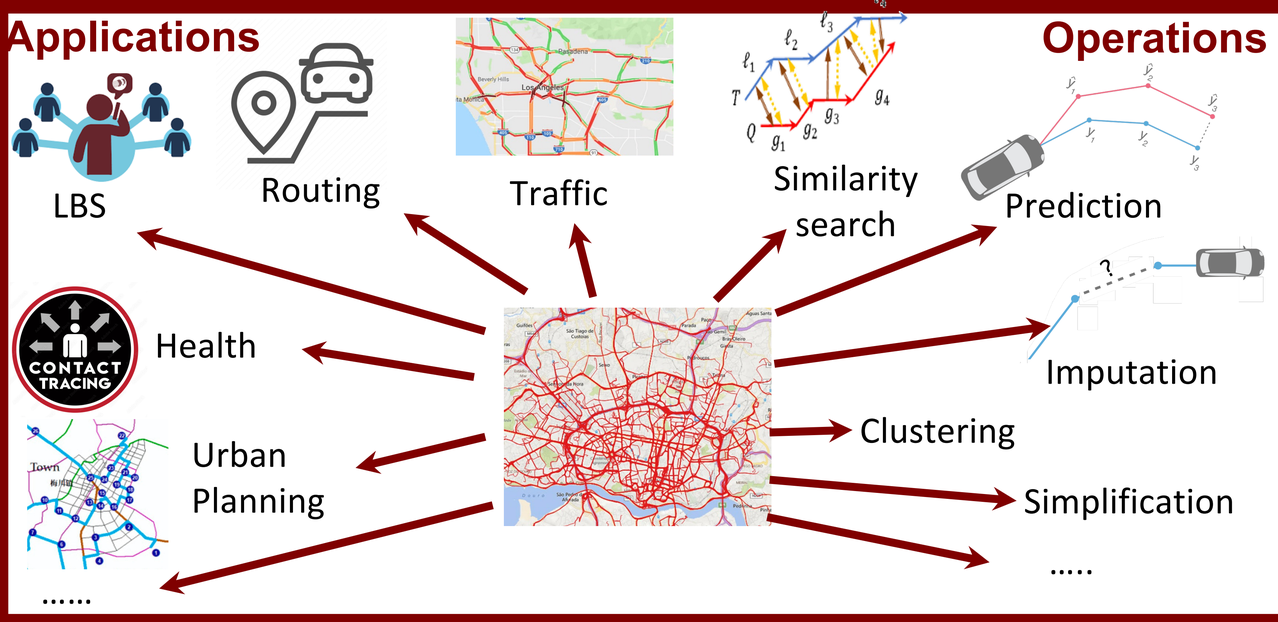
TOWARDS A UNIFIED DEEP MODEL FOR TRAJECTORY ANALYSIS

Paper ID: 2553

Mashaal Musleh

University of Minnesota

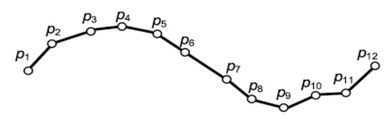
Trajectories are everywhere



Trajectories are statements

My husky dog was barking loudly

- Ordered set of ***words***.
- Drawn from a defined ***pool of words*** (underlying language)
- Words in a statement are ***semantically related***
- Constrained by ***language rules***
- Depends on the user and topic ***writing style***



- Ordered set of ***GPS points***
- Drawn from a defined ***pool of points*** (the underlying space)
- Points in a trajectory are ***spatio-temporally related***
- Constrained by ***physical rules*** (road topology and speed)
- Depends on the user and ***driving modality***

NLP Problems

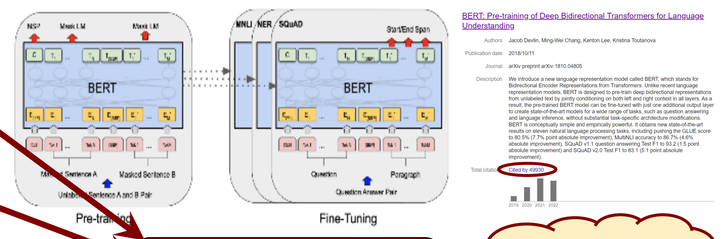
Predicting the missing word

Next sentence prediction

Text classification

Text summarization

Text similarity



BERT

Solution

In 2018, Google introduced the BERT deep learning model as a unified solution infrastructure for a wide variety of NLP tasks.

Trajectory Problems

Trajectory imputation

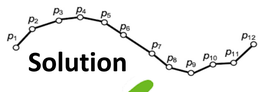
Trajectory prediction

Trajectory classification

Trajectory summarization

Trajectory similarity

???

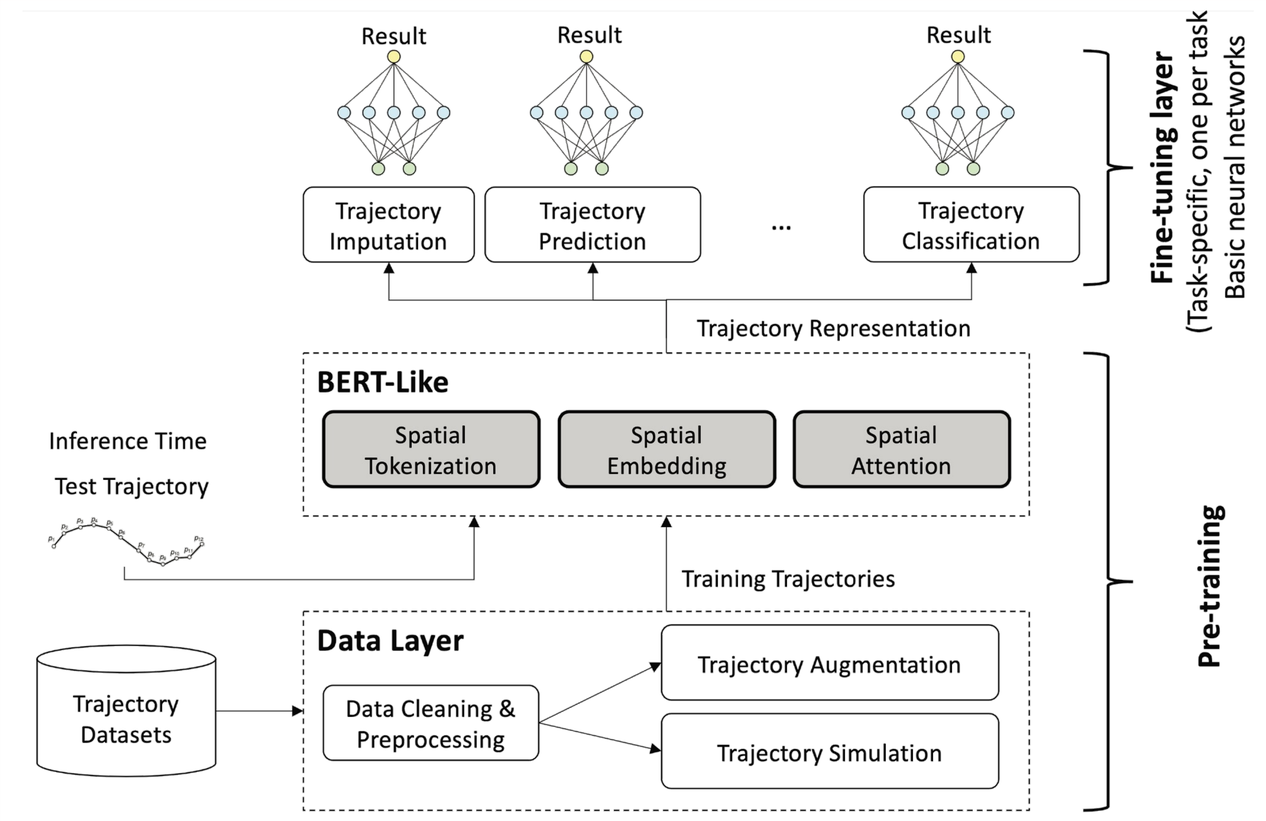


How can we have unified deep BERT-like model for trajectory problems?

Meet TrajBERT

No one needs to worry again about each specific trajectory operation.

A step towards the "Let's Speak Trajectories" vision (SIGSPATIAL 2022)



TrajBERT changes the core of the BERT system itself to make it deal with spatial data in general and trajectory data in particular as first-class citizens.

TrajBERT components understand that spatial data is special and support its unique characteristics.

1- Data Layer:

Address the data quality and availability issues:

- **Data Cleaning and Processing**
- **Trajectory Augmentation**
- **Trajectory Simulation**

Challenges:

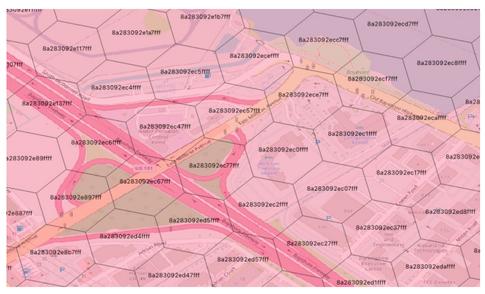
- Limited trajectory data.
- Ratio of available training trajectories to possible GPS points.
- Noisy trajectory data.
- Spatial and temporal constraints.
- Long and unrelated consecutive trajectories.

TrajBERT components address these challenges

2- BERT-Like Layer:

Address spatio-temporal constraints and relationships:

- **Spatial Tokenization:**
  - Minimizes the number of possible tokens (points)
  - Accurately represents any original GPS point to token and vice versa.
  - E.g. Uber's H3 Hexagonal Hierarchical Spatial Index



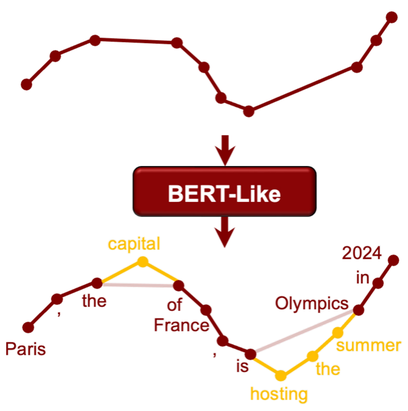
- **Spatial Embedding:**
  - Based on the trajectories they appear in (similar to statements), and
  - their spatial attributes such as proximity to each other/roads/and other geospatial features.
- **Spatial Attention :**
  - Identifies other key points in the trajectories for a give point P.

3- Fine-Tuning Layer:

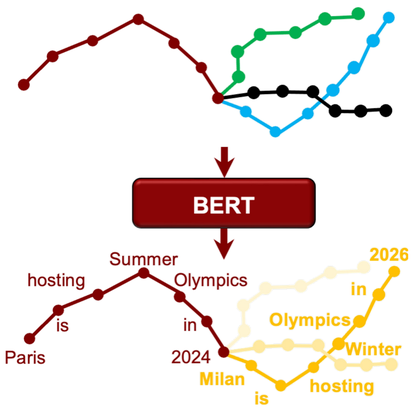
- Trains one additional neural network per trajectory analysis task, such as classification, imputation, and prediction.

Example Operations

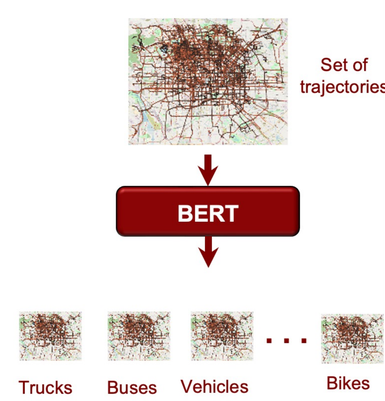
Trajectory Imputation



Trajectory Prediction

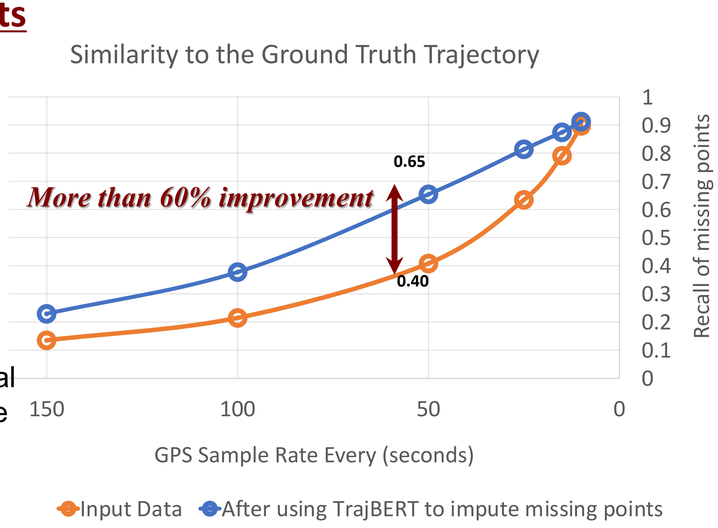


Trajectory Classification



Deployment & Early Experiments

- **Trajectory Imputation as a use case:**
  - GIS CUP 2017 dataset
  - 20K trajectories
  - 16k/4k train/test split
  - Assuming the ground truth GPS points were sampled @ 5 seconds
  - TrajBERT gives **more than 60%** improvement (similarity to the original trajectory) if GPS points were sparse (samples @ 50 seconds).



This work is supported by the National Science Foundation (NSF), USA, under grants IIS-1907855 and IIS-2203553