Transportation Means Analysis

Abdelrahman Mohamed Aly Sobh* 30061A

Geospatial Data Management
A.Y. 2023/2024

Abstract

The project aims to gather real-world trajectory through location tracking app (GPX Tracker in our case) and analyzing it, exhibiting some possibile statistics and eventually estimate the transportation means used in such a trajectory. So the process's steps can be summerized as {Trajectory Data Collection, Trajectory Segmentation based on Change Points, Producing Length and Time Interval Statistics per segments, Producing Speed statistic per segments, Producing Transportation Means Estimation per segment compared to clustering methods}.

1. Index

*https://github.com/Abudo-S/GDM_Project.

2. Trajectory Data Collection

We have scheduled a 50-minute trip between Paderno d'adda (LC) and Treviolo(BG) with marked points on transportation mean change as the following:

- Train [18:09:13,18:33:24]
- On foot [18:33:24, 18:42:50]
- Car [18:42:50,18:56:36]

Then we have exported 2 GPX files {tracks, way points}; for which, the first contains only one lineString record for the whole trajectory and the second contains all timestamped points added on mean changes.

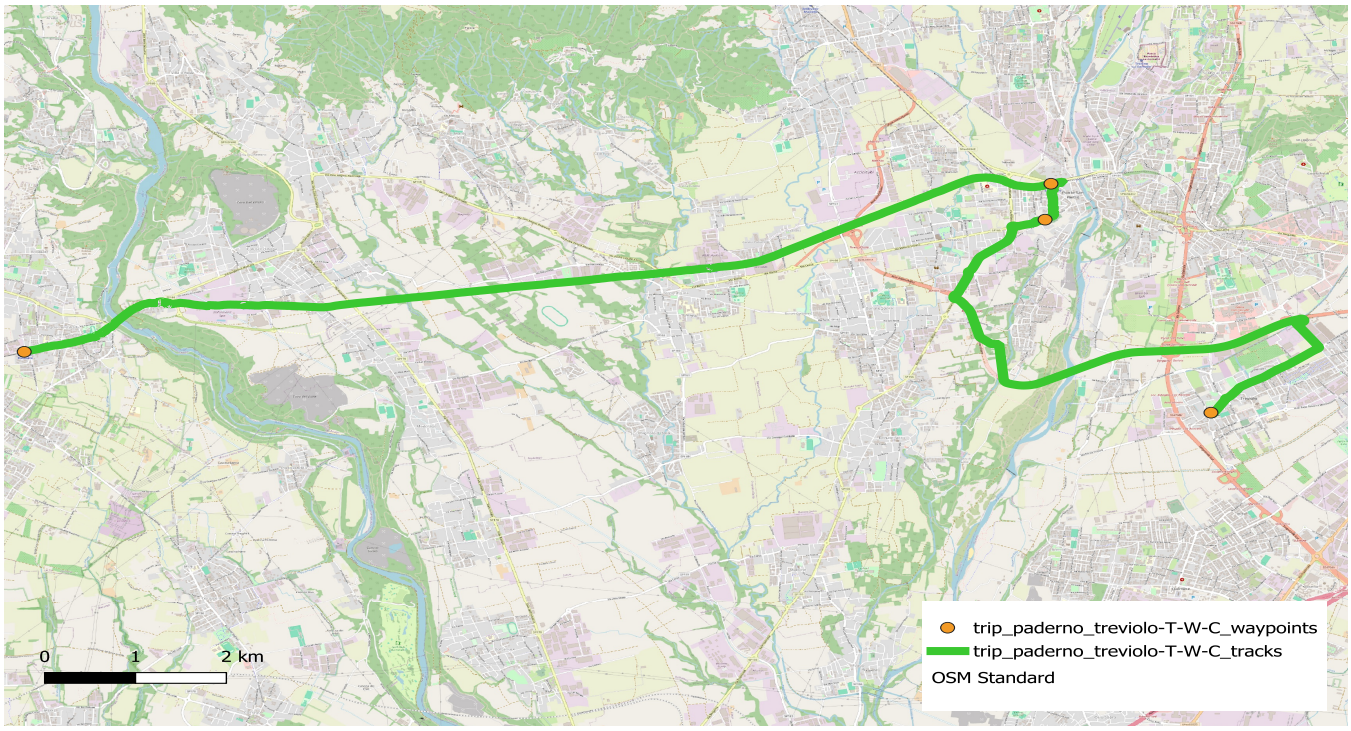

Figure 2.1: Captured trajectory with change points

3. Trajectory Segmentation Based on Change Points

The idea consists in splitting the whole trajectoy in segments based on change points, so we can determine the length and time interval for each segment.

```
--comment: create a table for trip's segments, since ST_Split() returns a GEOMETRYCOLLECTION
    , so we'll need to extract each linestring from the collection through ST_Dump
//https://postgis.net/docs/ST_Dump.html --

CREATE Table trip_paderno_treviolo_segments as
(SELECT row_number() over () as id, geom from
    (
        SELECT (ST_Dump(
                ST_Split(lineString.geom,
             (
                SELECT ST_Collect(changePoint.geom) AS geom FROM public."
                    trip_paderno_treviolo-T-W-C_waypoints" AS changePoint
                      WHERE ST_Intersects(lineString.geom, changePoint.geom)
                 )
             ))).geom as geom
        FROM public."trip_paderno_treviolo-T-W-C_tracks" AS lineString
    )
)
```

4. Length and Time Interval Statistics

In order to produce the required statistics, we need to create and populate these columns {start_time, end_time, duration, length} for each segment.

```
Alter table trip_paderno_treviolo_segments
ADD column start_time timestamp,
ADD column end_time timestamp,
ADD column duration interval,
ADD column "length" double precision

--comment: update each segment through points'data intersected with such a segment. Note
    that srid 3857 uses meters as unit of measurement.--
UPDATE trip_paderno_treviolo_segments segment
SET start_time = (
                  SELECT min(time) FROM
                  public."trip_paderno_treviolo-T-W-C_waypoints" as changePoint
                                  WHERE ST_Intersects(changePoint.geom, segment.geom)
                 ),
end_time = (
             SELECT max(time) FROM
             public."trip_paderno_treviolo-T-W-C_waypoints" as changePoint
                         WHERE ST_Intersects(changePoint.geom, segment.geom)),
"length" = ST_length(ST_Transform(geom, 3857))

UPDATE trip_paderno_treviolo_segments
SET duration = end_time - start_time
```

Having all the columns valorized, we can obtain the segmented lengths and the segmented time intervals as shown in the next page.

Figure 4.1: Trajectory segmented lengths (in meter)

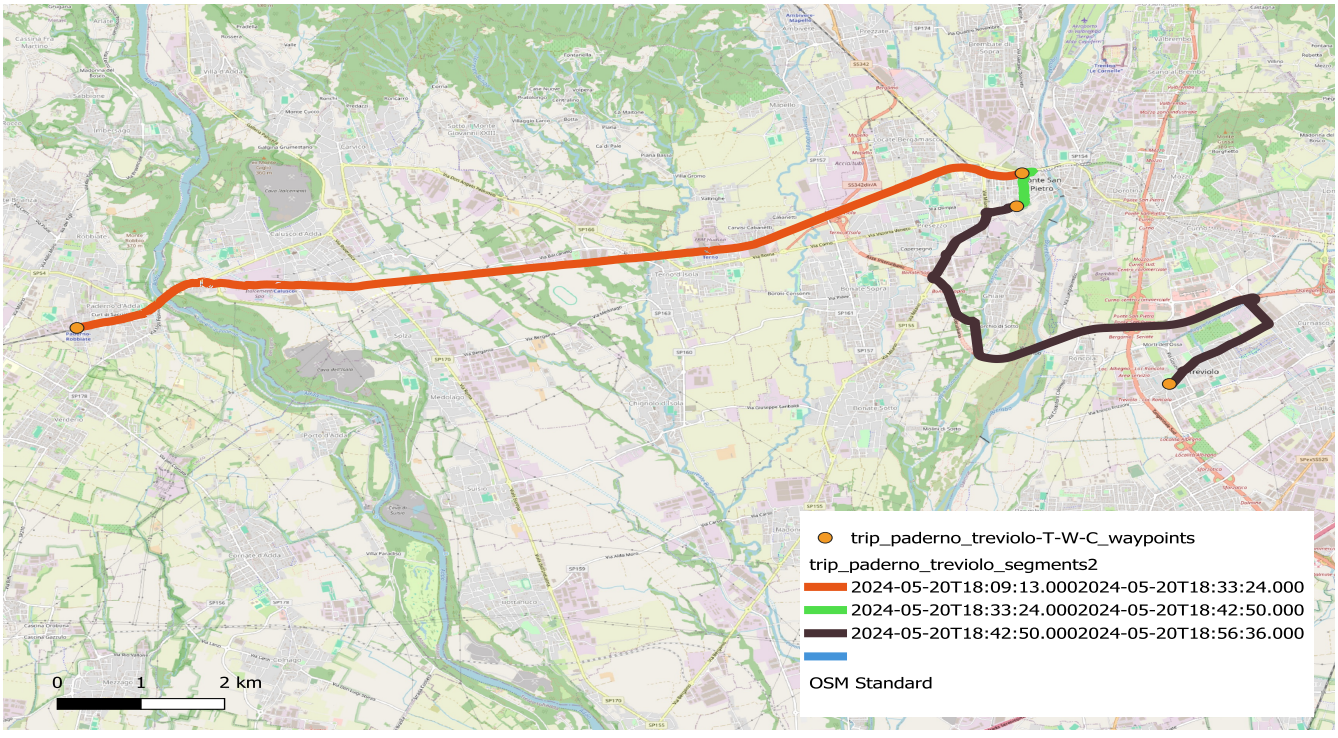

Figure 4.2: Trajectory segmented time intervals

5. Speed Statistic

Given that we have already calculated the length and time duration per segment, so we can just calculate the speed as length/time.

```
UPDATE trip_paderno_treviolo_segments
SET speed = "length"/ Extract(epoch from duration)
```

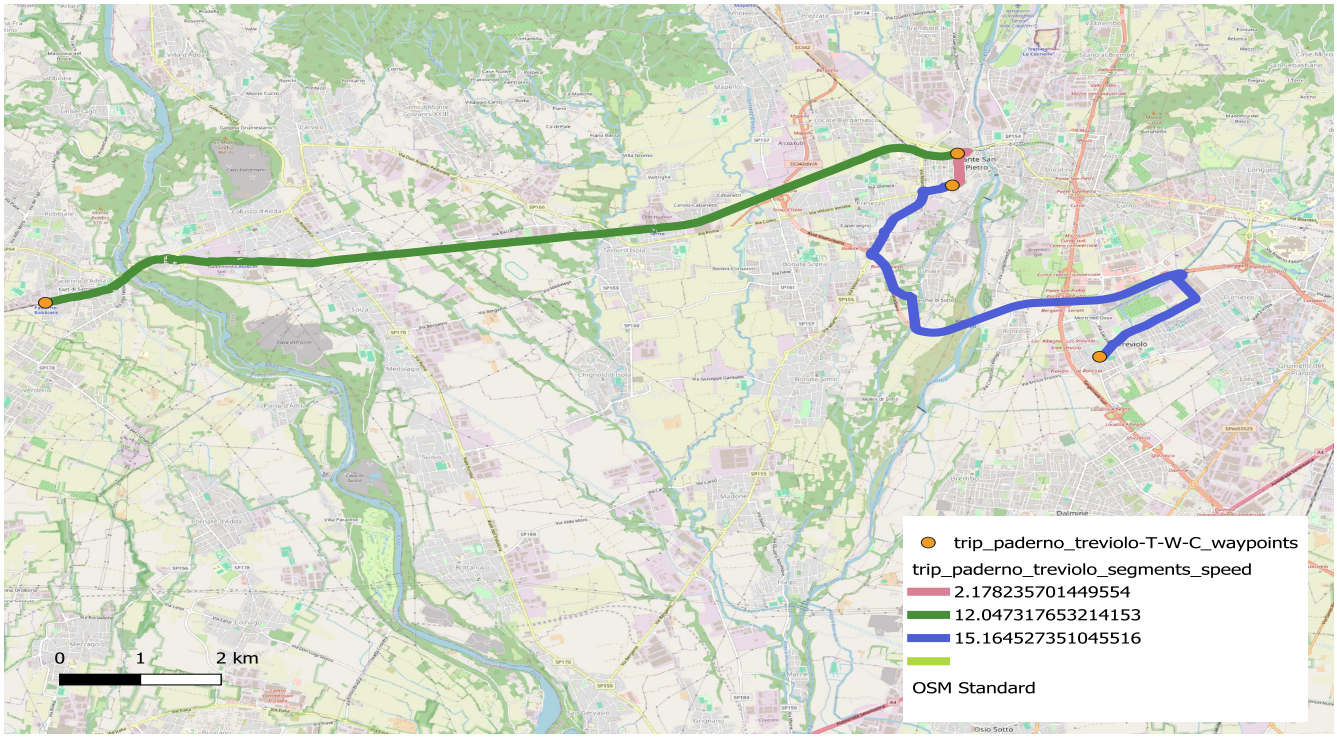

Figure 5.1: Trajectory segmented speeds (meter/second)

6. Transportation Means Estimation

Since we have already calculated the speed for the previous point, so we can determine the mean of transporation based on speed limits.

```
SELECT id, geom,
    (
     CASE
        WHEN speed * 3.6 >= 1 and speed * 3.6 <= 20 then 'bicycle or on foot'
                  WHEN speed * 3.6 >= 20 and speed * 3.6 <= 50 then 'car on urban streets'
                WHEN speed * 3.6 > 50 and speed * 3.6 <= 100 then 'train or extra urban car'
            WHEN speed * 3.6 > 100 and speed * 3.6 <= 160 then 'fast train or highway car'
            ELSE 'unkown mean'
        END
    ) as mean_transportation
FROM trip_paderno_treviolo_segments
```

So we can obtain the following statistic based on the value present in "mean_transportation" per segment.

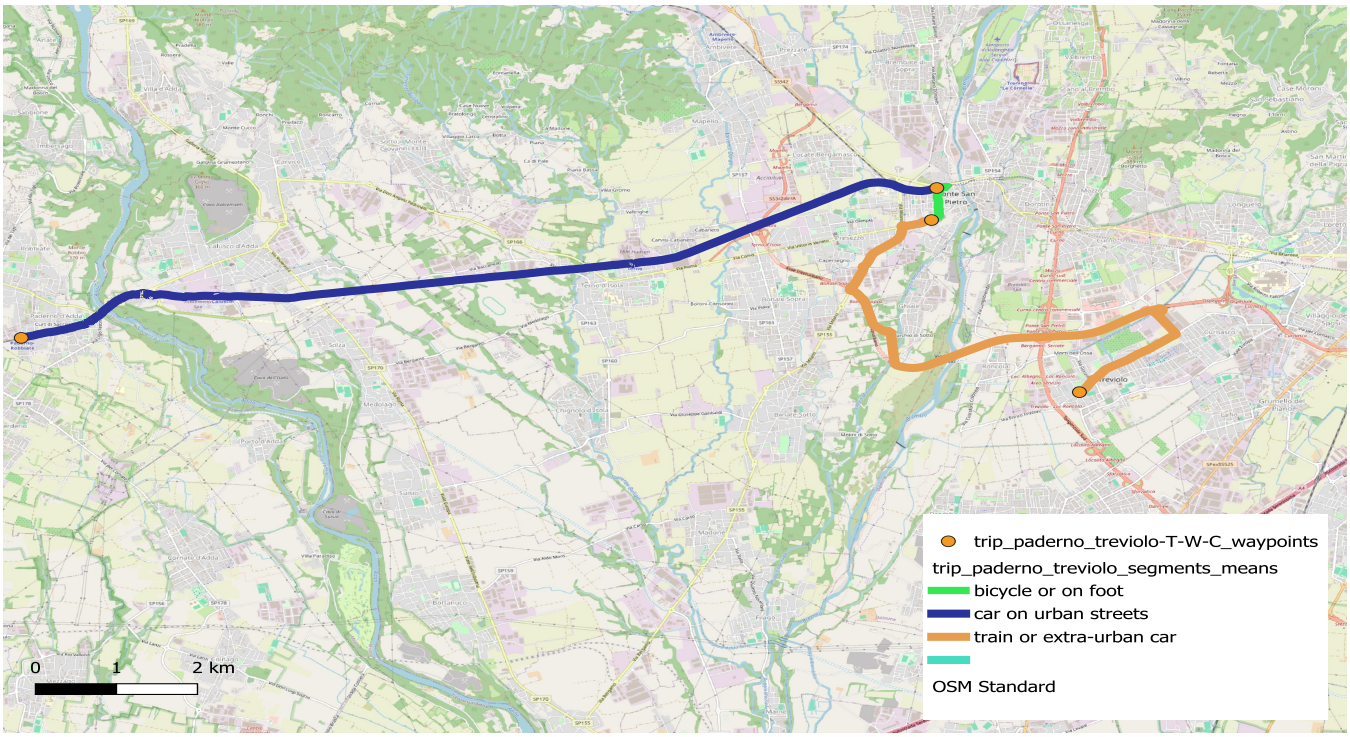

Figure 6.1: Trajectory segmented means

7. Transportation Means Estimation Based on Clustering

In our scenario clustering methods are useful to determine all similar points in different segments; thus, we can determine the transporation mean based on cluster's speed. First of all, in order to apply clustering on geometry data types, they should be of type `point`; so starting from the segments table, we can extract all timestamped point by applying the following query:

```
CREATE Table trip_paderno_treviolo_segments_points as
(
SELECT row_number() over () as id, geom,
        to_timestamp(((Extract(epoch from "segment_start_time") * 1000) + (((ST_Distance("
            segment_start_point", geom)/"segment_speed") * 1000)::integer))/1000)::timestamp
             without time zone as time
        FROM
    (
        SELECT seg.start_time as segment_start_time, ST_StartPoint(ST_Transform(seg.geom,
            3857)) as segment_start_point,
        seg.speed as segment_speed,
          (ST_DumpPoints(ST_Transform(seg.geom, 3857))).geom as geom
        FROM trip_paderno_treviolo_segments as seg
   ) ORDER BY time
)
```

7.1. Transportation Means Using K-means

By applying K-means on the points' table as the following:

```
CREATE Table trip_paderno_treviolo_kmeans_clusters as
(
    SELECT *,
    ST_ClusterKMeans(geom, 3 ) over() as cluster_id
    FROM (select * from trip_paderno_treviolo_segments_points)
)
```

Then calculating the speed of each cluster considering the `planer` distance between the lowest and greatest timestamped-points. `(Note that the planer distance makes the calculation less accurate since it doesn't take into consideration the real sequence of points in trajectory).`

```
UPDATE trip_paderno_treviolo_kmeans_clusters as clustered_point
SET cluster_speed = cluster.cluster_speed FROM
(
    SELECT cluster.cluster_id, ST_Distance(
    (
        SELECT geom FROM trip_paderno_treviolo_kmeans_clusters where cluster_id = cluster.
            cluster_id ORDER time desc LIMIT 1), (
        SELECT geom FROM trip_paderno_treviolo_kmeans_clusters where cluster_id = cluster.
            cluster_id order by time asc LIMIT 1)
    )
    /Extract(epoch from (max(cluster.time) - min(cluster.time))) as cluster_speed
    FROM trip_paderno_treviolo_kmeans_clusters as cluster
    GROUP BY cluster.cluster_id
) as cluster
WHERE cluster.cluster_id = clustered_point.cluster_id
```

So the resulting k-means clusters speed calculated with 3 clusters are shown in fig.7.1; thus, the transportation means are determined as fig.7.2 using the same query of sec.6 for clustered points.

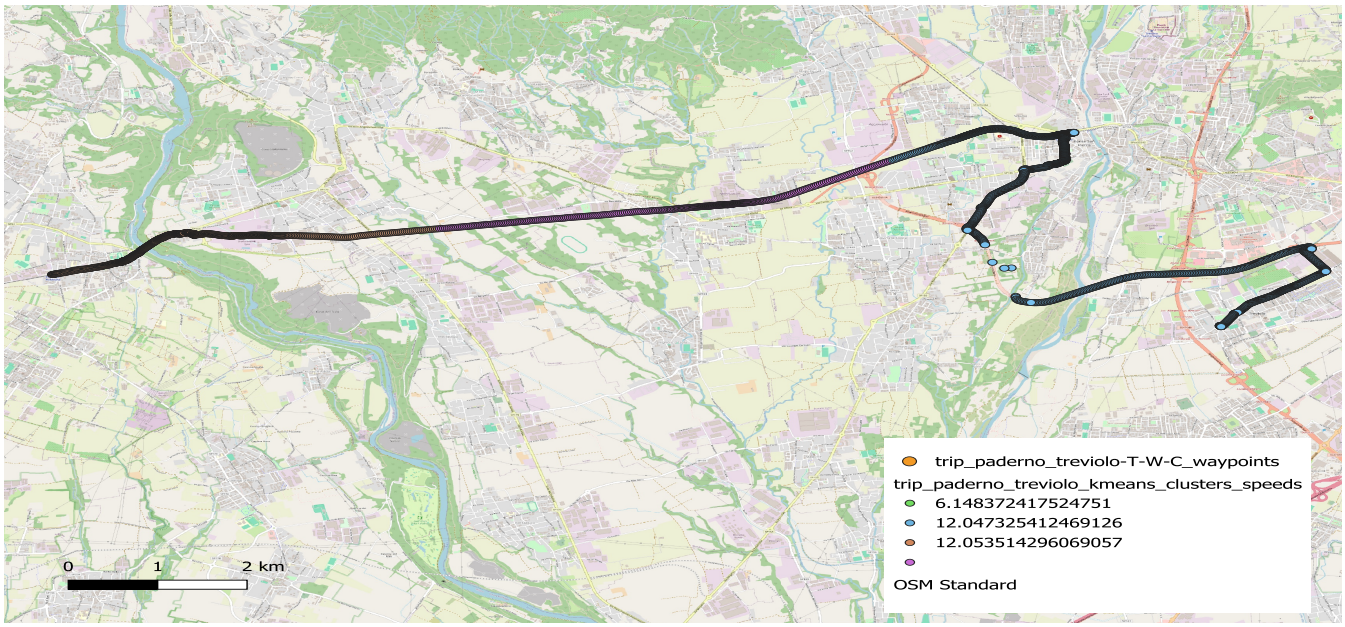

Figure 7.1: Trajectory kMeans clusters speeds (meter/second)

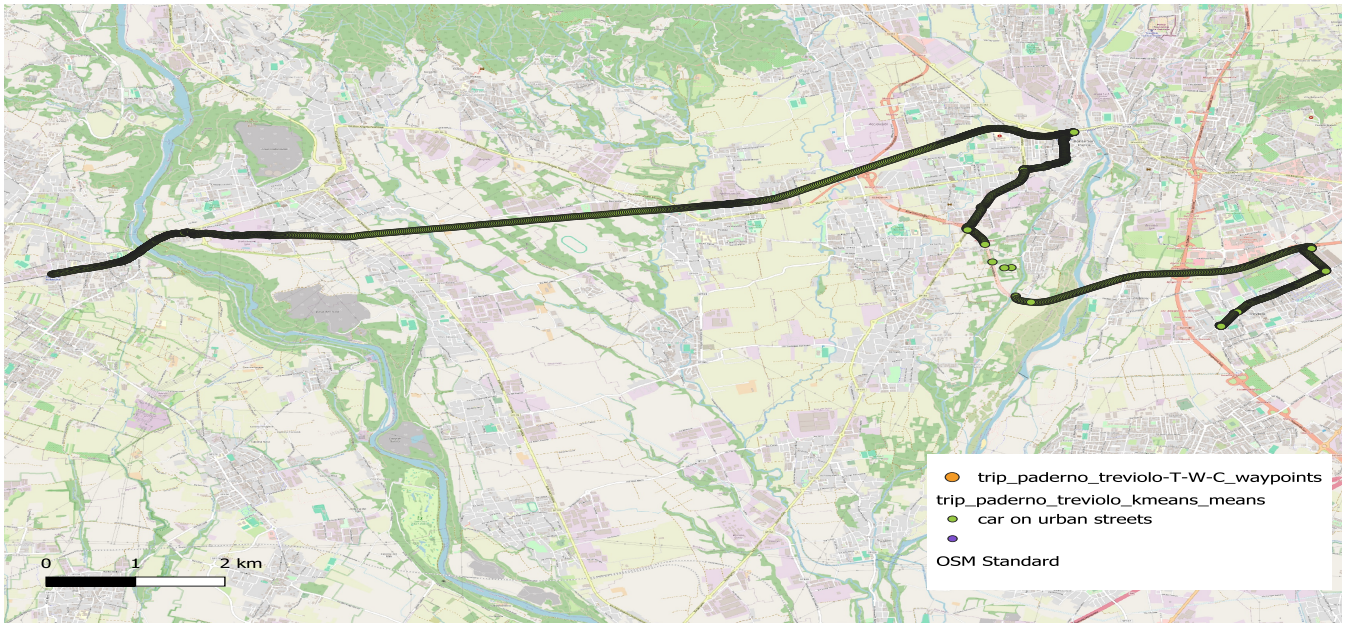

Figure 7.2: Trajectory transportation means - kMeans

8. Comparison Between Knowlege-based Estimation and Clustering

In the first estimation in sec.6, we have used the change points (points created on map when the transportation mean changes during the data acquisition) which have helped us to know the start and end points of a transportation mean. Meanwhile, the kMeans has started with only timestamped points (regardless the existance of change points) then by adding "cluster_speed", we have got an estimation.
Nevertheless, the knowledge-based estimation has given us more accurate estimation "Not 100% accurate" over the clustering method with respect to the ground truth.

9. Envisage a New Method for Transportation Mean Prediction

In real-world scenarios, we don't know the change points of transportation means, also as we have seen that the knowledge-based estimation gives us more accurate results. But we still have confusion between car or train as transporation mean, even between bicycle or on foot as transportation mean!

9.1. Infrastruture knowledge-based Method

A more accurate solution consists in considering also the infrastructure of cities like {train railways and stations, car highways, bicycle paths}, such information can help us to give more accurate estimation by checking the continous intersection (ST_Intersects) or even if the device is completely inside the geomtry of infrastructure (ST_Contains). Have such useful information helps us to exclude improbable estimations like:

- the tracked points are contained along the railway, so the transportation mean for these points must be a train.
- the tracked points are contained along the highway, so the transportation for these points must be a car.
- etc...

9.2. GPS Points feature-based Method

GPS devices typically capture more points when the transportation mean is at higher speeds, meanwhile they capture fewer points when moving slowly. So for such a hypothesis, we can consider the temporal and spatial distances between `each two consecutive points` to give an overall estimation for a similar subset or segment of points.

We can even mix between the infrastruture knowledge-based and points feature-based methods, in order to obtain a more robustic and accurent approach.

In the following figure, we can notice a mixed approach that uses the GPS points motion features besides the infrastructure (ex. bus stops) features in order to determine useful labels for transportation mean analysis. (Ref. https://www.researchgate.net/publication/380529950_Identifying_Travel_Modes_Using_GPS_Trajectories_A_Short_Review)

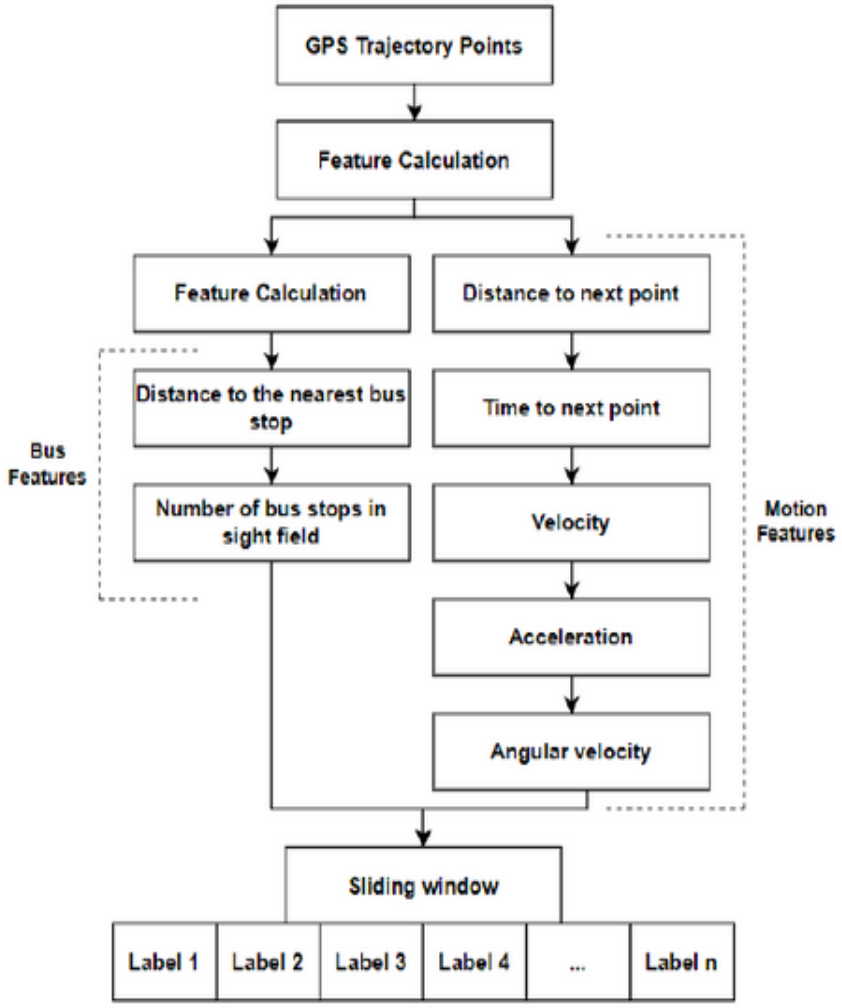

Fig. 1. Framework of feature frame construction (reproduced from [4].)

Figure 9.1: Feature frame extraction