Homework 2-Probe Data Analysis for Road Slope

Submitted By:
Animesh Patni (A20403240)
Abhijeet Ambekar(A20395747)
Vamsi K Kothapalli (A20395942)

Introduction: Probe Data

- Probe data is the byproduct of sensors and applications: hidden in logs, pings, and metadata lie tiny morsels of information like position and time. Individually, every record has little meaning, but the aggregate can be extremely valuable for augmenting other datasets.
- It (probe data) is generated by monitoring the position of individual vehicles like cars, buses, trains etc. over space and time using dedicated probe vehicles, mobile phones, sensors etc.
- The data thus obtained is a time series of locations. It is in the form of (Latitude, Longitude, TimeStamp).
- **Linked Data** is about using the Web to connect related data that wasn't previously linked, or using the Web to lower the barriers to linking data currently linked using other methods.
- The link data is information of road segments (links). It consist of data for the links that probe points can be map-matched to.
- It consist of many attributes like length, direction etc.
- It is useful in many ways:
  - Traffic speeds or travel time of road segments determination
  - Transportation planning
  - Map refinement/Creation
  - Intersection delay analysis

Introduction: Map-Matching:

- Map matching is the problem of how to match recorded geographic coordinates to a logical model of the real world, typically using some form of Geographic Information System. The most common approach is to take recorded, serial location points (e.g. from GPS) and relate them to edges in an existing street graph (network), usually in a sorted list representing the travel of a user or vehicle. Matching observations to a logical model in this way has applications in satellite navigation ,GPS tracking of freight and transportation engineering.
- Map-matching procedure integrates probe data with road link data to identify the correct link on which a vehicle is travelling and to determine the location of vehicle on a link.
- It is performed by obtaining probe data and thus mapping it to the map using the given information like latitude, longitude, timestamp, elevation etc.
- It helps in identifying inaccuracy of positioning, off the road locations, error of a GPS point etc. The error of GPS point is corrected by snapping the GPS point onto the road network.
- There are several algorithms present to perform map-matching which are:
  - Probabilistic theory
  - Kalman filter
  - Fuzzy logic etc.

Problem

- Input : Probe data and map
  - The raw probe points in Germany collected in 9 months
  - The link data for the links that probe points can be map-matched to

- Task :
  - Map match probe points along with all the required information -we have to map-match probe points to road links.
  - Derive road slope for each road link
  - Evaluate the derived road slope with the surveyed road slope in the link data file

APPROACH STAGE 1

- Map matching is the process of matching a set of real world coordinates from raw probe data to a digital map.
- It can be improved by providing the GPS directions of the probe, important for further digital analysis.
- Latitude and Longitude are given in both Probe Data and Link Data. This coordinated have to be compared and weight out or correct which coordinate would be right.
- In Map Match we use these Latitude and Longitude information of both these files and compare them with every other to find the one pair with minimum distance between them.
- This pair is resultant Map Matched Probe data points to corresponding Link.

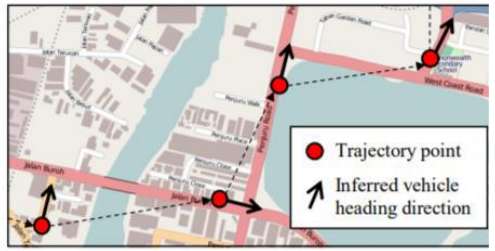


Fig. 2. At every stage, the vehicle heading direction is inferred based on the terminating direction of the last point and the historical path leading to the current point.

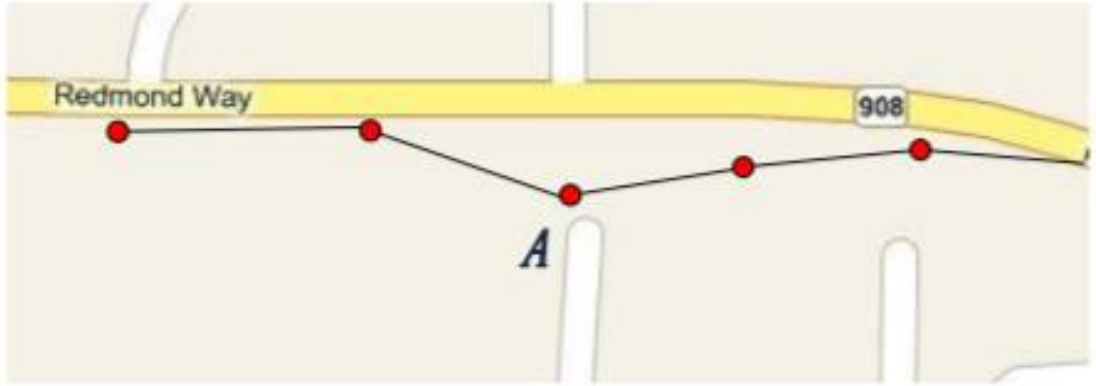


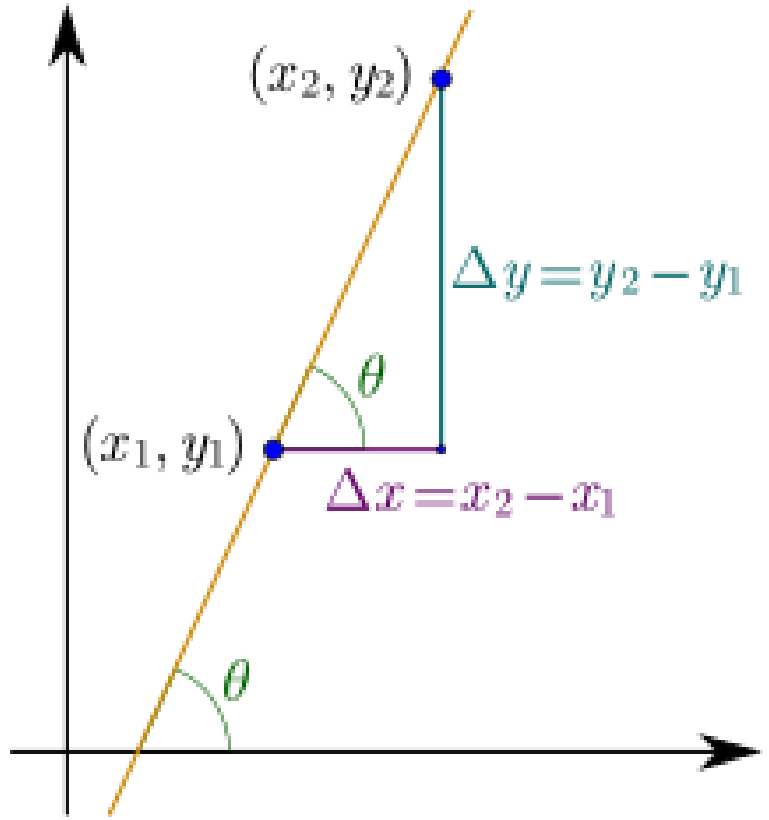
Figure 3. Map matching using topological information

APPROACH STAGE 2

Calculating Slope of Road

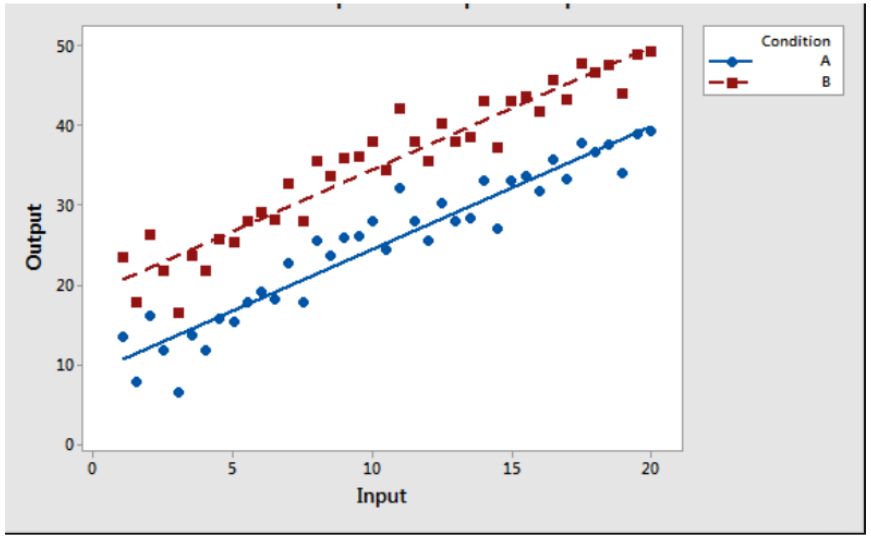
- To calculate the slope between two points here we use the matched points from previous step to calculate the slope of two consecutive points. The slope between two consecutive points (P1, P2) is = $RISE$ / RUN $i.e.$ $Y2-Y1$ / $X2-X1$
- where p1(x1,y1) p2(x2,y2) are two points. This is calculated for all the matched points.

- Here P1,P2 are the probe data points that have been map matched in a sequence.



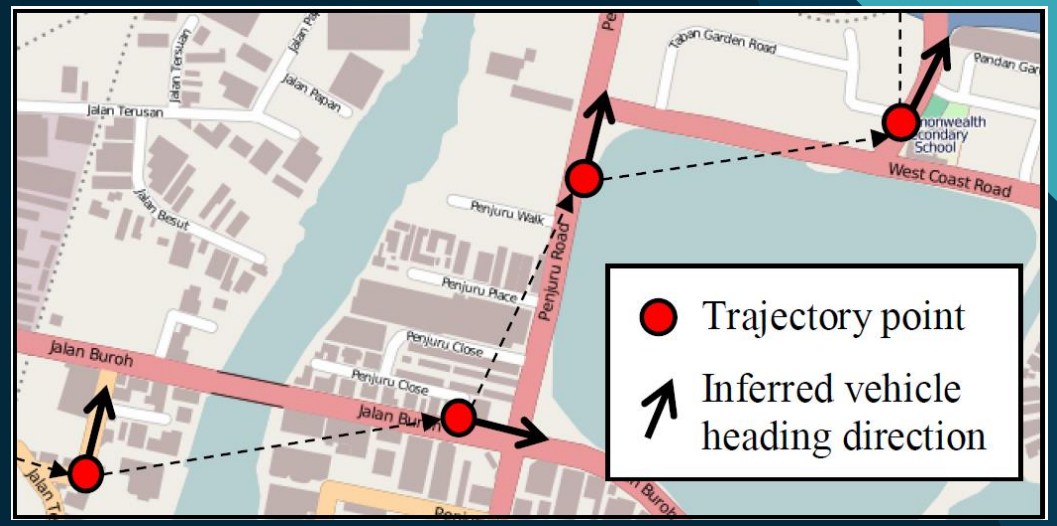
APPROACH STAGE 3

- **Comparing the road slope**
- Slope’s that results from all the probe data coordinated are aggregated.
- Aggregated slopes are then averaged, Slope is calculated for each Link data.
- This slope is then compared with the previous slopes that have be computed earlier to that point.



Step 1: Identifying probe points with respective link

- Each link consists of multiple link nodes(The red points shown on the map)
- We set a onset around each link node to detect probes around it to map it to the link.
- Every probe point we find out the closest link node(R1).
- This closest link(R1) will be mapped on to the road segment.



Step 2 : Identifying the map match probe point

- We find out the second closest link node(R2) by calculating perpendicular distances between all the possible closet points and choose the point(R2) with lowest perpendicular distance.
- We draw perpendicular line between the two closest link nodes(R1 and R2).
- Then, we draw a perpendicular from the R1 to R2. The point at which the perpendicular meets the line R1 and R2 is M, is the Map match probe point.

Step 3: Calculating Distance from Link and Distance from Reference

- We calculate the perpendicular distance from the map matched probe point on the link to the probe point, this is called “linkDistance” and “***linkDistance = PM***”
- We move from the first node in each link and check if the node is either ***L1*** or ***L2***.
  - If not, the pointer moves to the next node and sums up the distance between the previous node and itself to ***referenceDistance***
  - If so, the distance between the previous node and current node is summed up to the referenceDistance along with the distance between the link node and the point M.
- At the end of this step, we have the distance from each map matched probe point to the reference node, that is, referenceDistance. “***referenceDistance = AL1 + L1M***”

Step 4: Finding the direction in which the vehicle was travelling on the link.

- The direction of the probe is interpolated by the two successive points and the perpendicular line between them this will help us to decide the direction of the probe.
- We compare the time stamp of two successive probes to identify the direction of travel of the vehicle.

Step 5: Producing the output of map match probe points along with all essential information.

- All the essential columns for the required output are extracted from the **Partition6467ProbePoints.csv.** and **Partition6467LinkData.csv**.
- The values that were calculated such as “heading”, “distance from ref” and “direction” are also included in the final output files namely **MatchedPointsData.csv and SlopeEvaluation.csv**

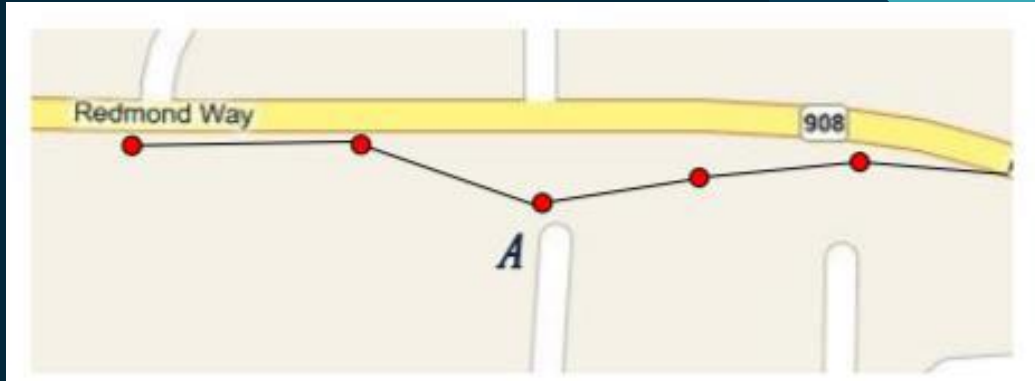
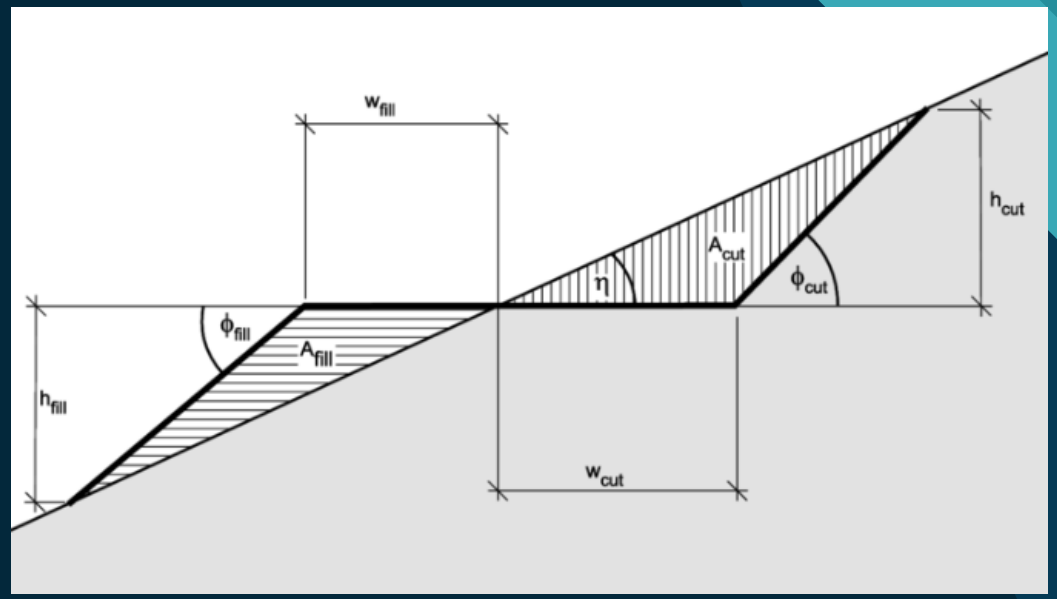


Figure 3. Map matching using topological information

Step 6: Calculating the road slope for each road link.

- Two successive map matched probe points **R1** and **R2,** are considered.
- Altitude at the probe is already known to us so we can calculate the distance between **R1** and **R2**.
- Now that we have the required values, altitude and the distance between the two probe points, we can calculate the slope between the points.



Step 7: Code Structure and Results

- MapMatchingScript.java is the main file. We have defined functions like readLinkData, readProbeData, calculateSlope etc. to perform the operations on the source input files.
- readLinkData takes a string as an input, containing the path to the source file: Partition6467LinkData.csv, performs all the calculations and stores the data into MatchedPointsData.csv using dumpMatchedPoints function.
  - Columns in the output csv are: SAMPLE ID, DATE TIME, SOURCE CODE, LATITUTE, LONGITUTE, ALTITUTE, SPEED, HEADING, LINK PVID, DIRECTION, DISTANCE FROM REF, DISTANCE FROM LINK, SLOPE
- readProbeData takes a string as an input, containing the path to the source file: Partition6467ProbePoints.csv, performs all the calculations and stores the output into SlopeEvaluation.csv using writeSlopeData function.
  - Columns in the output csv are: LINK PVID,CALCULATED MEAN SLOPE,PROVIDED MEAN SLOPE

Step 7 Continued: Screenshot of MatchedPointData.CSV

SAMPLE ID	DATE TIME	SOURCE C	LATITUTE	LONGITUT	ALTITUTE	SPEED	HEADING	LINK PVID	DIRECTION	DISTANCE	DISTANCE	SLOPE
3496	########	13	51.49687	9.386022	200	23	339	62007637	U	36.62946	0.066171	U
3496	########	13	51.49668	9.386157	200	10	129	62007637	T	12.94628	0.022547	0
3496	########	13	51.4967	9.386422	201	21	60	5.67E+08	U	79.3446	0.078774	U
3496	########	13	51.49675	9.38684	201	0	360	5.67E+08	T	43.97215	0.035345	0
3496	########	13	51.49686	9.387294	199	0	360	5.67E+08	U	43.97215	0.067524	-0.02742
3496	########	13	51.49693	9.387716	198	5	89	5.67E+08	F	66.00915	0.014136	-0.02742
3496	########	13	51.49696	9.387794	198	1	288	5.67E+08	F	72.0566	0.018485	0
3496	########	13	51.49695	9.387805	197	0	310	5.67E+08	F	72.74968	0.010143	-0.02742
3496	########	13	51.49695	9.387818	196	0	274	5.67E+08	F	73.5956	0.003552	-0.02742
3496	########	13	51.49694	9.38784	196	0	226	5.67E+08	F	75.05264	0.007127	0
3496	########	13	51.49694	9.387852	197	0	201	5.67E+08	F	75.80232	0.013423	0.027416
3496	########	13	51.49694	9.387855	197	0	182	5.67E+08	F	76.00882	0.01695	0
3496	########	13	51.49694	9.387857	197	0	232	5.67E+08	F	76.07503	0.018348	0
3496	########	13	51.49694	9.387859	197	0	202	5.67E+08	F	76.24583	0.02025	0
3496	########	13	51.49694	9.38786	197	0	199	5.67E+08	F	76.29457	0.020967	0
3496	########	13	51.49693	9.387863	197	0	179	5.67E+08	F	76.47546	0.022995	0
3496	########	13	51.49693	9.387865	197	0	184	5.67E+08	F	76.61108	0.024145	0
3496	########	13	51.49693	9.387867	197	0	199	5.67E+08	F	76.72774	0.026181	0
3496	########	13	51.49693	9.387869	197	0	178	5.67E+08	F	76.8161	0.027143	0
3496	########	13	51.49693	9.38787	197	0	183	5.67E+08	F	76.84612	0.028135	0
3496	########	13	51.49693	9.38787	197	0	194	5.67E+08	F	76.87295	0.028678	0
3496	########	13	51.49693	9.387871	197	0	176	5.67E+08	F	76.93867	0.029201	0
3496	########	13	51.49693	9.387872	197	0	172	5.67E+08	F	77.04597	0.029278	0
3496	########	13	51.49693	9.387873	197	0	177	5.67E+08	F	77.1144	0.029376	0
3496	########	13	51.49693	9.387874	197	0	195	5.67E+08	F	77.16264	0.029218	0
3496	########	13	51.49693	9.387875	197	0	185	5.67E+08	F	77.23204	0.02897	0
3496	########	13	51.49693	9.387875	197	0	214	5.67E+08	F	77.26502	0.028926	0
3496	########	13	51.49693	9.387876	197	0	190	5.67E+08	F	77.29037	0.02894	0
3496	########	13	51.49693	9.387876	197	0	179	5.67E+08	F	77.33098	0.028839	0
3496	########	13	51.49693	9.387876	197	0	183	5.67E+08	F	77.35633	0.028853	0
3496	########	13	51.49693	9.387877	197	0	207	5.67E+08	F	77.39177	0.028994	0
3496	########	13	51.49693	9.387877	197	0	191	5.67E+08	F	77.41885	0.028926	0
3496	########	13	51.49693	9.387878	197	0	176	5.67E+08	F	77.45847	0.029171	0
3496	########	13	51.49693	9.387878	197	0	176	5.67E+08	F	77.48973	0.029208	0
3496	########	13	51.49693	9.387879	197	0	184	5.67E+08	F	77.51164	0.029382	0
3496	########	13	51.49693	9.387879	197	0	185	5.67E+08	F	77.54708	0.029523	0

Step 7 Continued: Screenshot of SlopeElevation.CSV

LINK PVID	CALCULATED MEAN SLOPE	PROVIDED MEAN SLOPE
62007637	0.013707721	0
567329767	-1.87E-09	0
62007648	-0.027415393	0
78670326	0.01028079	0
51881767	-0.002741554	0.104333333
51881768	5.34E-09	-0.0165
51881825	0	-0.117
51881938	0	0.123333333
811768915	0	0.175333333
811768917	-0.002741556	0
778526449	0	0
51881935	0	0
51883233	0	1.491666667
67948960	-0.027414551	0.83725
67948524	0	0.424333333
51883237	0	0.978
51883238	0	0.10875
51883362	-0.027414991	1.15375
67948946	0	1.17875
67948945	0	0.09575
811593144	0.027415541	0
67948530	0	-0.6805
51883890	2.44E-08	-0.35825
51883887	0	0
51867694	0	0
554728228	0.00685355	0.942142857
554728229	0	0.2355
51867645	0	-0.102
762732456	0	-0.0185
51866678	-0.027415431	-0.021
51866677	0	-0.002
799517966	0.013707737	-0.055666667
799517967	0	-1.45E-18
51865408	0.027414392	-0.024
762732451	0	-0.012

Notes:

- The source code is present in different folder with Instructions to run the program in ReadMe file.
- The Resultant files containing evaluation results are present in folder.

References:

- Reducing Uncertainty of Low-Sampling-Rate Trajectories (Yu Zheng, MS Research)
- Online map-matching based on Hidden Markov model for real-time traffic sensing applications (C.Y. Goh, J. Dauwels, N. Mitrovic, M. T. Asif, A. Oran, P. Jaillet)
- Fast Viterbi Map Matching with Tunable Weight Functions (Yin Wang, HP Labs)
- Map-Matching for Low-Sampling-Rate GPS Trajectories (Yu Zheng, MS Research)