Exploring Weighted Kernel Density Estimation to Finding Road Accident Hotspots

Saahitya E
Computer Science Dept.
PES University
Bangalore, India
saahitya.e@gmail.com

Samarth M
Computer Science Dept.
PES University
Bangalore, India
samarthm241020@gmail.com

Satyam Shivam Sundaram
Computer Science Dept.
PES University
Bangalore, India
satyspy007@gmail.com

***Abstract*—Identification of Road Accident Hotspots is an important problem, with improvements in the identification of these hotspots directly correlating to better road safety management. We performed weigthed Kernel Density Estimation for the problem of finding road accident hotspots. We tryed out different parameters and different weighting criterions for generating hotspots from Kernel Density Estimation technique after weighting. We also observed the effects of using different weighting criterions in different regions.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

The newly adopted 2030 Agenda of Sustainable Development has set an ambitious target of halving the global number of deaths and injuries from road traffic crashes by 2020 [3]. Without sustained action, road traffic crashes are predicted to become the seventh leading cause of death by 2030 [3]. In India, the number of road crash increased by 31% from 2007 to 2017 and that road crashes have increased by 25.6% in the same period.

Road Accidents are thought to happen due to a combination of human error and mechanical failure. An important and underrated factor is geography and terrain (spatial factors) of a particular region or location. Even more so, these spatial factors compound human error or mechanical failure, for example, a drunk person driving on the road. Hence it is very important for road and highway agencies to consider these spatial factors while designing highways and roadways to be safer for citizens.

Cataloguing of accident hotspots is the process of spotting geospatial areas that have a high occurrence of road accidents. Cataloguing accident hotspots can be an effective and low-cost tool to identify high-risk accident-prone areas.

Proper identification of these accident hotspots or collision-prone areas can be used by city and highway or roadway authorities to take appropriate measures in these areas to reduce the accident rate after further case by case review and diagnosis. It is an interesting and important problem, because of the different approaches that have been used to solve it and the scope and use in real-world applications.

II. Summary of Literature Survey

Some of the relevant work in this field involves the use of kernel density estimation for detecting a region of interest i.e identifying the hotspots. One of the work presents a study aimed at comparing the outcome of two geostatistical-based approaches namely kernel density estimation (KDE) and kriging for recognizing crash hotspots in a road network [1]. Analysis of different time-of-day the accidents happened has also been performed [2]. Further, works also appropriated and designed the traffic accident density maps by the application of KDE [2].

III. Problem Statement

We seek to implement appropriately weighted Kernel Density Estimation Technique for Road Hot-spot Detection.

Road Accidents have a mild correlation to the location of the crash site. This means that an accident is more likely to happen in certain areas, than in some other areas. This is the intuition we want to utilize when we want to classify areas into hot-spots based on proximity to crash sites.

Road Accident hot-spot detection techniques have an inherent problem of considering all accident as similar, this means that a minor accident like a car on a slope bumping into another car's rear because a brake wasn't applied properly is the same as an accident where there was a human casualty involved. This approach tends to trivialize fatal accidents and give excessive importance to the already plentiful minor accidents.

This is obviously a big fallacy and a drawback as hot-spots detected in this context are not particularly useful or accurate in real-world applications. This is clearly not feasible for authorities that try to gain meaningful insight from these hot-spots, but who might not attain any accurate or useful information from it.

Alternatively, performing Kernel Density Estimation with only fatal accidents is not appropriate, this is because even minor accidents also show correlation to geospatial factors and should be considered while calculating hot-spots. Also, the number of fatal accidents is comparatively smaller and performing KDE could be considered misleading.

We seek to address this problem by using a modified Kernel Density Estimator that weighs the accidents appropriately by certain attributes or properties before doing a kernel density estimation(KDE) technique.

This data was collected from the kaggle website. The dataset was labelled ”UK Road Safety: Traffic Accidents and Vehicles”. Here we used the ”Accident_Information.csv” dataset particularly as it described road accidents with many attributes as geo-coordinates, weather conditions, road conditions, etc. The size of the dataset used was 600 Megabytes(MB).

Some Attributes of Dataset	Datatype
Accident_Index	integer
Accident_Severity	string
Date	date
Day	string
Latitude	numeric
Longitude	numeric
Local_Authority_.District.	string
Weather_conditions	string
Year	int

IV. Proposed System

A. Block Diagram

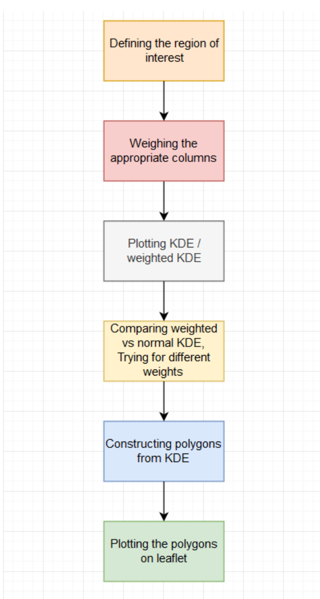

Fig. 1. Block Diagram of Proposed System

B. Explanation of Components of System

1) Choosing the Search Area: The First step is to define the search area we are interested in. The search area should be optimal, such that it can give us a good and meaningful visualization in limited time. The simplest way is to choose among different cities/districts. If the area of interest is smaller than a district, we can choose a center and a radius to define the search area and find all the accidents that have occurred in that circle.

2) Weighing: The second step is weighing. Some accidents could have been more severe than the other, hence weighing these accidents becomes an important step to find hotspots. First, we choose the attributes for weighing. Here, we have chosen the severity of the accident as the weighting parameter. Other options could be the number of casualties etc. There is no particular rule for assigning weights, it depends on the user’s view.

3) KDE / Weighted KDE: KDE is a smoothing problem. In this method, a kernel function is placed on every crash point resulting in a smooth, individual and continuous crash density surface. Now, a grid of cells is overlapped on the search area. For a given cell, density is estimated by summing the overlapping density surface resulted from each crash point. This procedure is repeated for all the cells in the grid, which results in hotspots. In the case of weighted KDE, the crash points with high severity are given higher scores. KDE was done using kde2d function from package MASS. This function takes three parameters, latitude, longitude and n. ’n’ defines the number of cells in the grid, the plot gets smoother by increasing the value of n. For weighted KDE, kde2d.weighted function was used. This function requires an additional parameter - weights (note: both kde2d and kde2d.weighted do not accept NA values. Hence, the rows with NA in latitude/longitude were excluded from the density estimation). Both the functions redefine the search area into an n x n square and the output is a matrix of size n x n where each value is the score of that cell.

4) Comparison: The plots of weighted vs normal KDE were compared with image function. Different weighing measures, for different values of n were explored.

5) Constructing polygons, final Visualization: To visualize the output of KDE, we have to create polygons of hotspots, to achieve this, first, contours were created around different levels of hotspots using contourLines function. These contours were then bound into spatial polygons using SpatialPolygons function of package sp. Finally, the polygons generated were plotted on leaflet maps. Leaflet provides free open street maps. The same procedure can be applied to google maps with qmap

function. Unlike leaflet, google maps is not entirely free and requires enabling API services from google cloud platform.

V. Experiments and Results

A. Method

We first tried working on the whole UK data by applying Kernel Density Estimation. But we were not able to perform it, as KDE has a high order of time complexity and is very computationally expensive. This is because KDE has to calculate distances for each crash point from each cell in the grid defined. So we then tried doing KDE in smaller regions.

We choose individual districts to perform Kernel Density Estimation. This was because the district contained a good number of points to perform KDE and we could set the resolution of the grid to a higher amount(as the region was smaller and the number of points is smaller in the district was smaller).

While choosing which districts we wanted to work on we calculated the ratio of the number of crash points with fatal or serious(separately) accidents versus the total number of crash points for each district. We then chose districts with skewness towards fatal or serious accidents in their crash by measuring for a high ratio.

Next, we had to select a parameter to weigh the kernel function while calculating distances upon, so looking at the dataset the most important thing that we observed that must be considered to weigh was the severity of the accident. Then we experimented how to weigh the parameter chosen. We experimented with linear weighting, quadratic weighting, bi-quadratic weighting, exponential weighting, etc.We used the normal function as the kernel function

Kernel Density Estimation returns to us a grid with kernel values calculated at each point in the grid. Then we are dividing these values into 2 classes with 50% of points in the grid in one class and the other 50% of points in another class. We are considering the class with higher values as the hotspots.

B. Case Study: Ryedale

We chose Ryedale as we noticed that the value of ratios of fatal accidents to the total number of accidents in this district is the highest in the dataset.Once we experimented with weighted kernel density estimation with different weighting criteria, we plotted Fig. 3 that was for did weighted for the parameter ”Severity” in the dataset with linear weights like 1, 2, 3 for factors ”Slight”, ”Serious”, ”Fatal” respectively. Also we similarly performed quadratic weighted KDE on the region of interest which is shown in Fig. 4.

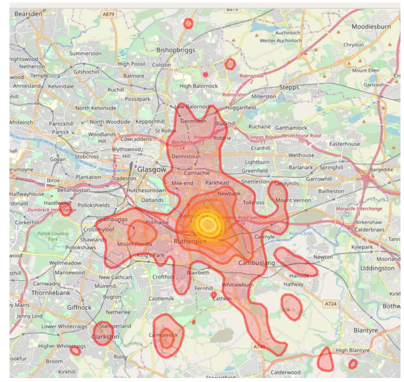

Fig. 2. Glasgow City Hotspots on map - KDE with exponential weighting i.e 1,10,100

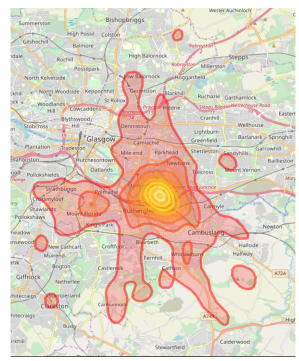

Fig. 3. Glasgow City Hotspots on map - Normal KDE

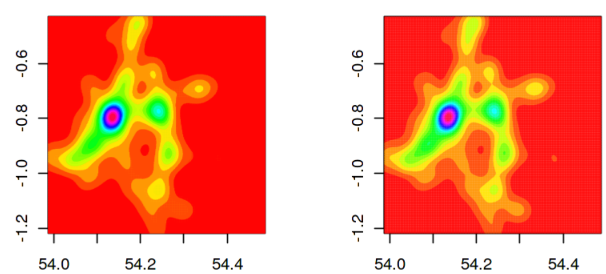

Fig. 4. Ryedale City Hotspots-(L)Normal KDE (R)KDE with linear weighting i.e 1,2,3

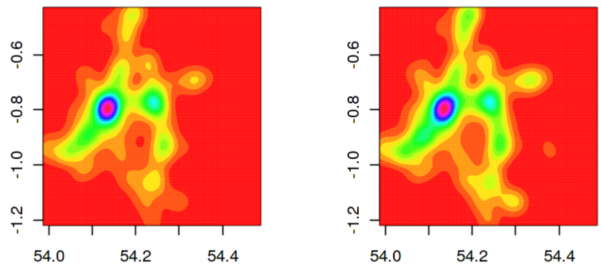

Fig. 5. Ryedal City Hotspots-(L)Normal KDE (R)KDE with quadratic weighting i.e 1,4,9

C. Case Study: Glasgow City

We chose Glasgow city because we noticed it featured among the top in the total number of accidents in any district. Also we noticed that the city is in Scotland and has a slate of weather conditions like snowing and raining. Also the city is mildly hilly. We experimented with linear, quadratic and exponential weighting and we have shown the heatmaps obtained in Fig .4, Fig .5 and Fig .6.
We also see that in Fig .1 and Fig .2, there are some hotspots in Fig .2 that are present but are not in Fig .1, On further examination by cross referencing with a scatter plot we were able to confirm that extra hotspots were detected becasuse of close proximity to fatal accidents crash points.

VI. Conclusions

We have found that doing weighting before performing Kernel Density Estimation has given us good results, and is especially effective in finding road accident hotspots. We have observed that there is no universal way to weight the appropriate parameter. Depending on the region, different weighting techniques seem to work. We also noted that the points in the grid that are towards the corners or edges have an unfair disadvantage in that they have comparatively less points that surround them and are close to them.

Acknowledgment

We would like to thank our Data Analytics professor Dr Gowri Srinivasa and the Teaching Assistants for giving us this opportunity to do this project and for their valuable input. We would also like to acknowledge the sources from which we

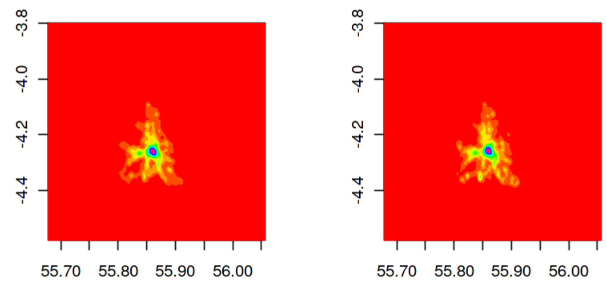

Fig. 6. Glasgow City Hotspots-(L)Normal KDE (R)KDE with biquadratic weighting i.e 1,16,81

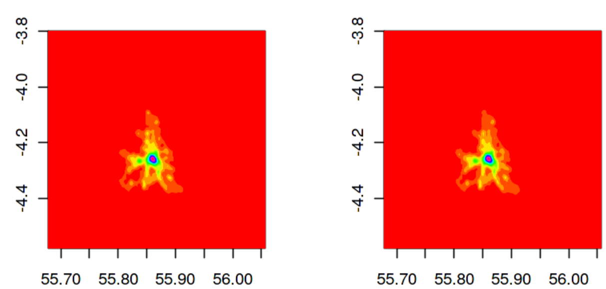

Fig. 7. Glasgow City Hotspots-(L)Normal KDE (R)KDE with quadratic weighting i.e 1,4,9

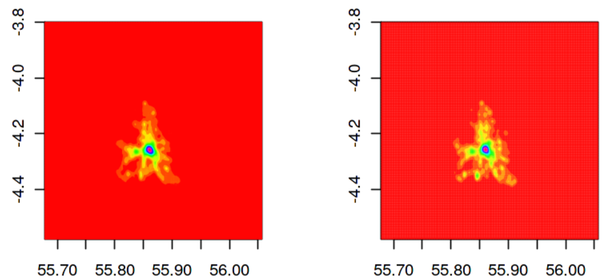

Fig. 8. Glasgow City Hotspots-(L)Normal KDE (R)KDE with exponential weighting i.e 1,10,100

obtained our data, the Kaggle user Thanasis for advancing the data-set and the website data.gov.uk for making this data publicly available.

References

[1] Identification of crash hotspots using kernel density estimation and kriging methods: a comparison - Lalita Thakali Tae J. Kwon Liping Fu

[2] Identification of hot-spots road locations of traffic accidents with pedestrain in urban areas - Svetalana Backalic, Bosko Matovic, Dragan Jovanovic

[3] https://unctad.org/en/PublicationsLibrary/dtltlb2017d4_en.pdf

VII. Contributions

Samarth MS was responsible for plotting the hotspots on the maps and choosing districts to do KDE for.
Saahitya was responsible for getting weighted KDE to work with the different parameters and the normal KDE
Satyam was reponsible for plotting timelapse and did explaratory data analysis.