A Take on Geo-informatics Systems with Pin-Codes as Atomic Unit

Siddharth Gupta, Abhinav Rana, M.Lakhsmi

Abstract: Rural areas in India may not necessarily have internet connection and what if they need to know about nearest hospitals, fire stations or if a message has to be broadcasted over a large section of this area. This paper tries to solve all of these problems by implementing a pin-code based approach to map these important locations. Now a days, more advanced technologies such as GPS are more useful in solving these problems in real-time but what this paper tries to do is solve them statically where technology can't take precedence over daily lives. It discusses a smart, fast algorithm with minimum memory overhead to solve these specific issues. Lack of data sets are a concern in this approach but to automate this we have used Image Processing to automatically detect boundaries of pin-codes on various sub-regions to form a universal Graph which can then be subjected to various algorithms like A*, DFS, BFS, Dijkstra which are local to each type of problem. This leaves scope for further research on developing models that can read images of maps more efficiently to create more accurate data sets which are accurate enough and void of longitudinal and latitudinal details.

Keywords: DFS, BFS, Dijkstra Which Are Local to Each Type of Problem.

I. INTRODUCTION

A lot of applications use latitude and longitude for the above-mentioned tasks but many applications require low accuracy and fast results thus low computing costs. Latitude and longitude approach may seem a logical choice when we have to work for services which require high accuracy but becomes futile when demographic data is to be studied. In Current systems, each location is a point on the surface of earth described by longitude, latitude and elevation. For long time haversine formula is being used but for finding distance between two points and then finding its nearest location is complex task [3] . So, a new take would be Geo-hash which has good complexity compared to haversine.Geo-hash divides the world into cubes and then those cubes are divided into sub-cubes according to the accuracy. To find the desired location one can traverse cubes recursively [4].

Siddharth Gupta*, Department of Computer Science and Engineering, SRM Institute of Science and Technology,Chennai (Tamil Nadu), India.Gsiddharth79@gmail.com

Abhinav Rana, Department of Computer Science and Engineering, SRM Institute of Science and Technology,Chennai (Tamil Nadu), India.Abhinav2106@outlook.com

Dr.M.Lakhsmi, Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai (Tamil Nadu), India.

This approach still contained some issues
- One of the issues is the edge case of water-bodies, so now let's say you are nearby some water body the geo-hash will look for hospitals in that water body which is again futile.
- Another limitation is that it still does not takes in count the population factor which is required in some applications

Our proposal gives an understanding of how population is distributed across land and maps population density to pin-codes for determining shortest routes for each task.

As the dataset is scarce, we needed image processing to read a map and establish a relationship of all pin-codes with each other in the form of graph or an adjacency list. After that we map population to these pin-codes as weight. On this acquired dataset we perform several algorithms to generate results which have less memory over-head and can be used in offline environment.

II. PROPOSED ALGORITHM

Retrieval Number: 100.1/ijsce.D34840310421
DOI:10.35940/ijsce.D3484.0310421
Journal Website: www.ijsce.org

13

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
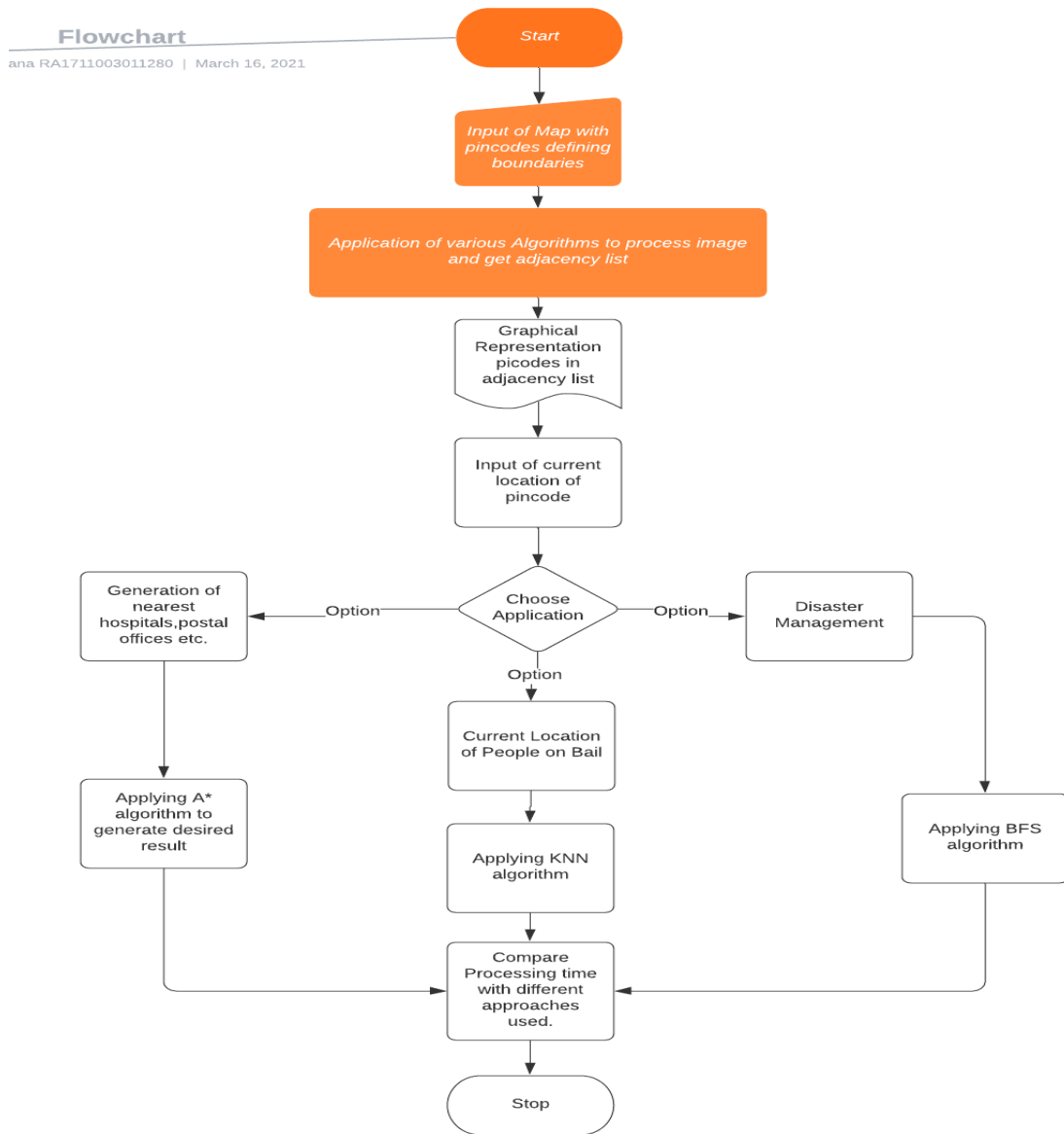*© Copyright: All rights reserved.*

FIG 1: The architecture diagram of all the modules.

The Whole concept can be break down into the following modules for better understanding:

- **Extraction of Data**: Using OCR engines and various pre-processing techniques to read map. Pre-processing techniques include Reading, Resizing, Removing noise, segmentation and morphology. For removing noise, we use Gaussian blur which is like viewing the images through translucent lens. In segmentation we differ the background from foreground and thus apply gaussian blur on it again to smooth the edges. After this step, markers are used to represent the regions of pin-codes as different objects. After this process we feed this polished image to OCR engine like OCR space to read the map and give us the links between pin-codes. All the code for above steps is written in python [6].
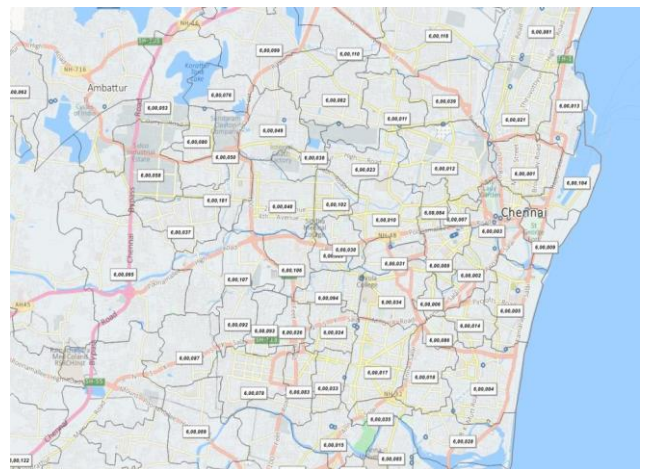

FIG 2: The image which was used for OCR to find adjacency list for graph.

Image recognition is applied on images like this and thus an adjacency list is created which will make the connected graph between the pin codes.

- **Creation of Meta-Data:** Populating existing data with web scraping for pin-code to population mapping. We Scraped websites to get dataset of Place name to population mapping and from our parent data set which contained all the pin codes and their respective place name we created a pin-code to placename mapping. After that through a python script we combined the two data set to create our final meta-data which contained information regarding the pin-code and their respective population. This meta data is crucial for our analysis of the area demographic. This meta data will serve as weights of our graph, the weight of the graph will be the average of population of the two pin code's population.

- **Applying Algorithms:** After obtaining this dataset, we can perform various approaches used in machine learning like A* algorithm which is based on heuristics to optimally find shortest path between two nodes. This can be used as an application for route detection [2]. We used BFS, DFS, Djikstra, Bellman-Ford, DLS techniques to perform various applications as intended in the flowchart.

- **Performance Analysis:** Analysing these algorithms on specific metrics. These metrics include execution-time, F1 score, mean absolute error and Mean squared error.One of the main performance metric will be performance in low bandwidth network. To get this metric we create our own unit which is execution time(in microseconds) per network speed(in kbps).Hence the model which has lower execution time per kbps will be a better model.

III. RESULTS

We were successfully able to create a console application in C++ by scraping data from two different sources. One being obtained from images of maps and other obtained from API provided by govt of India regarding the population density. Application parses the above data from default file handling mechanisms and create an adjacency list from these two CSV files. As explained in the above flow chart, we then create a custom algorithm to handle all test cases, unit cases and applications. For each application different algorithm with a specific primer was used. For instance, for generating nearest pin-codes in let's say 20 miles, we have different methods to do so but we found that A* was the best choice to do so. For broadcasting data to large set, we used a running total of population and distance as parameters and found that BFS was the fastest to do so. Additionally, we needed to weed out cycles from Graph to improve execution speed. We found that Union by rank techniques in Union Find Algorithm was fastest.

We conclude that this system is computationally fast but lacks severely in terms of Precision therefore rendering it useless in real world scenarios where Precision is key. It has some serious flaws as it can't detect rail-roads, water-bodies or map obstacles in front of it. But in places of rural areas where bandwidth is very low, this can be used effectively during disaster management by coupling this system with SMS service. We leave scope for further research in improvement of precision of this system.

IV. CONCLUSION

- One of the key opportunities that this model will bring us is to analyse the data faster, current system rely on highly complicated and sophisticated Geohash value which require high computing power to perform the operations . If not for so much sophisticated servers and high processing algorithms, current model wouldn't have been a reality , this model provides a low power lightweight solution which can be used in places where not much accuracy is needed.

- In developing countries high bandwidth is a luxury and so this model will come in handy in places where there is nohigh speed internet providing a low accuracy solution in case of any network outage also it will continue to work.

- Considering pin-code as an atomic measurement, it cannot be more precise than what is used by Google but is useful to find nearest neighbours or even generate a path in an efficient manner without the cost of huge memory header.

- A challenge would be to find an accurate data set of Indian pin-codes which will serve as the adjacency list for the graph as well as finding the meta data for the population of the graph.

- This model cannot be implemented globally because every country has its own way of denoting pin codes , Hence every country need to have its own different model.

REFERENCES

1. IpingSuprianaSuwardi, Dody Dharma, Dicky Prima Satya, DessiPuji Lestari School of Electrical Engineering & Informatics, ITB, Bandung, Indonesia iping, dody, dicky, dessipuji(2015): Index Based Spatial Data Model for Corporate
2. Prof. Nitin R.Chopde, Mr. Mangesh K. Nichat(2009): Landmark Based Shortest Path Detection by Using A* and Haversine Formula
3. C. C. Robusto(1957) : The Cosine-Haversine Formula
4. AnbaoWANG , Zhang Jun, Wenrong JIANG(2009): Useful Resources Integration Based on Google Maps
5. KarezAbdulwahhab Hamad, Mehmet Kaya.(2016) : A Detailed Analysis of Optical Character Recognition Technology.
6. Chirag Patel , Atul Patel , Dharmendra Patel (2012):Optical Character Recognition by Open source OCR Tool Tesseract: A Case Study.
7. B.B Chaudhary, U Pal (1998):A complete printed *Bangla* OCR system.

AUTHORS PROFILE

Siddharth Gupta, Siddharth gupta is currently a final year student at SRM institute of technology pursuing B.Tech in computer science. Currently he is working as a product engineering intern at Technovert Solutions, Hyderabad . This is his first official reasearch paper although he has done some research in distributed computing and algorithms.

Retrieval Number: 100.1/ijsce.D34840310421
DOI:10.35940/ijsce.D3484.0310421
Journal Website: www.ijsce.org

15

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication
© Copyright: All rights reserved.

A Take on Geo-informatics Systems with Pin-Codes as Atomic Unit

Abhinav Rana, Abhinav Rana is currently a final year student at SRM institute of science and technology pursuing a B.Tech in Computer Science. He is working as a platform engineer at Icertis, Pune. His primary academic interests include Algorithms, Distributed Computing and Software Engineering.

Dr.M .Lakshmiis currently a professor in Department of Computer Science and engineering. She has done her Doctorate from Sathyabama university in 2007 . Her primary areas of research include Wireless networks, AI and mining and deep learning.

16