

Optimization of Healthcare Network for improved service delivery

Aseer Ahmad Ansari*, Sabyasachi Chakraborty**, Satyabrata Aich**, Beom Su Kim**,
Hee-Cheol Kim**

*School of Computer Engineering, KIIT University, Bhubaneswar, India

**Department of Computer Engineering/Institute of Digital Anti-Aging Healthcare/u-HARC, Inje University, South Korea

a3ahmad.kiit@gmail.com, c.sabyasachi99@gmail.com, satyabrataaich@gmail.com,

beomsu91@naver.com, heeki@inje.ac.kr

Corresponding Author: heeki@inje.ac.kr

Abstract— The aim of this work is to optimize the healthcare network for improving service delivery by allocation and deallocation of healthcare resources as per need. The extrapolation of the need for facility and resources is the most challenging part of this work which requires knowledge of the population of various geographic regions/areas. The population is also looked at from the perspective of age-group, economy, the average fraction of population requiring healthcare and also keeping in mind any history of diseases or such prone areas. Speculating the thumb rule of the healthcare worker to patient ratio i.e. 1:2808 in order to make work smooth and patient care more robust for the staff members. The proposed efforts are to help the Washington State health ministry to amend, develop and improve the staff and resources in existing facilities

Keywords— population, migration, fertility, Distance Matrix, geocoding

I. INTRODUCTION

The commencement of work begins by acquiring the population of Washington State. Though there are multiple data sets available publicly the accuracy and geographical coverage for each data set with the benchmarking of integrity and quality needs to be ascertained before applying as a yardstick for the decision. This problem can be slightly relevant in counties within the state of Washington with changing economics and structural complexity in the area with the advancement in people's requirements of the public facility, especially healthcare. Thus, new methodologies for dynamic population data updating have to be devised using technology (imagery) or statistics (birth, death and migration) or amalgamation of both.

For sustainable development as a goal, the dynamic population data and changing needs of the end-users require proper mapping to drive a quantifiable parameter for inducing not only in the field of healthcare but anywhere / everywhere human is the focus to optimize demand and supply gap as a major achievement for each area-specific requirement. Changes in the lifestyle and economic needs lead to infrastructural changes in health support system requirements. The population

again has many dynamics - age, sex, economic level, etc. A proper step by step procedure would be - To first estimate the population at any instance on the basis of statistical methodology, we need to comprehend the population in terms of zip code, age-groups, economy, terrain, disease-prone areas and then, at last, apply the thumb rule of healthcare to patient ratio.

II. METHODOLOGY

A. Population Density: Publicly available data sources with the integrity of data up to 2010 are the US census bureau the current population of Washington state being 7.17 million and having a population density approximating to 113(p/mi²) and area of 66546mi² [1]. Also, from another source, the population of the state is 7,405,743 with a density of 103(p/mi²) spread over an area of 71,362 mi² [2]. The population as of the US census 2000 increased from 5,894,121 to 6,724,540 in the last US census 2010 with a percentage increase of 14.1%.

From the population data sources such as the US population data by zip code of 2000 and 2010 provided by the US Census Bureau and a list of zipcodes of the state of Washington [3] population density can be calculated simply by dividing the current population with total area. There are many important factors affecting the population of counties. These are fertility rate, mortality rate, immigration and emigration [4]. As of the **census conducted last in 2010**, population counts are taken but the factors controlling growth are getting more complex and unpredictable; so, the methods and parameters for calculating the growth need a high degree of approximation on the basis of various factors. The percent change in population mainly depends on the birth rate, death rate, immigration, emigration as well as unpredictable suicides. **In our work focusing on just the Washington state, even county to county migration flows will have a net effect over the population for each region.** Even though this happens at a rate higher than that can be recorded or updated, an understanding of the rate can be

formed from the recent most updated data of county to county migration flow^[5].

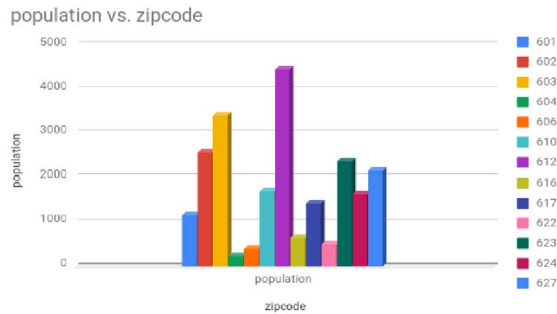


Figure 1. Variation in population by zip code

Figure 2 shows the variation in the population of a county by zip codes. This data substantiates the requirement of estimating population zip code wise to essentially arrive at the results accurate enough to redistribute healthcare resources.

The American Community Survey of the Census Bureau regularly gathers information previously contained only in the long form of the decennial census, such as ancestry, educational attainment, income, language proficiency, migration, disability, employment, and housing characteristics which are used by many public-sector, private-sector, and not-for-profit stakeholders to allocate funding, track shifting demographics, plan for emergencies and learn about local communities^[6]. Every time someone moves in or out of county resources overhead is created or removed. Besides this the difference in fertility of women in the united states recorded last as of 2016^[7] and mortality rate of the state added with the immigration and then removing the number of emigrations could give an estimable account of the current population. This can only act in optimization for estimating the population at any given instant.

Current_population = Base_year_population + (birth - death) + (immigration - emigration)

Current_population = Base_year_population + (fertility_rate * Base_year_population) - (mortality_rate * Base_year_population) + net_migration



Figure 2. Increase in migration over time in a US county

Figure 1 shows the change in net migration over a time period of a certain county. Rate over such accounted data can be calculated using exponentially weighted means.

Every census done periodically has a cut-off date that defines the timeline for future linear analysis within itself or with the outside world to maintain sanctity for its quality, accuracy and its integrity. Thus, to estimate the population the let's say six months after the cut-off date would require the base population or estimated population needs to be added with the factor of 1/2 achieved in a year. This rate of population enhancement is classically different and dependent on area-specific vital indicators (birth/death/fertility/mortality et) in counties of each state/state of Washington. Thus, in every county, the rate of enhancement needs to be calculated carefully for evaluation.

US census bureau has its cut-off date, 1 July. This means that the survey of the 2010 census has data intact up to 1 July 2010. Thus, to estimate the population at any instant, the growth rate will have to be multiplied by the amount of time from this cut-off date and added to the base population.

1. Related Work^[8]: Based on the planned goal and the needed material, there are two ways of assessing population: statistical modeling & areal interpolation. Areal interpolation methodology majorly involves transforming information from one part of spatial units to another for the zone transformation problem. Tools like interpolation or disaggregation are applied to input such as census population data to attain a proper accurate population surface. On the other hand, ascertaining the relationship between population and other variables is the statistical modeling approach to evaluate the total population for an area. Though the census data is not straightaway made use of as an input to the modeling instead, it makes use of socioeconomic variables such as correlation with urban areas, land use, dwelling units, image pixel characteristics for population estimation; it is only in model training purpose that the census population comes in picture. This approach majorly is devised to estimate the population of difficult areas but it also can be involved in the process of interpolating census population.

1) 2. Related Work^[9]: Remote sensing plays a crucial role as a self-reliant source of information majorly using optical satellite images. At various resolutions and time periods space-borne remote sensing arranges comprehensive geospatial report that wasn't obtained before., Remotely sensed data can generate advantageous input information for urban population assessment in amalgamation with geographic information system (GIS). Input requires inventory data and built-up area information on the socio-economic structure of the area. Although physical measures can directly be determined from its geometry visible, the socioeconomic parameters have to be estimated from an aggregation of both secondary and physical is.

3. Average Travel Time: The average aerial time between zip codes can be estimated using various open-source databases

and APIs. Likewise, **all combinations of zip codes are maintained by the national bureau of economic research**^[12]. Although the census bureau has the generalized representation of ZIP codes assigned to it as ZIP Code Tabulation Areas, ZCTAs are representative of geographical locations of populated areas. This might miss less populated areas with an inferior and depressed economy. Hence this work won't require the usage of the ZIP Code Tabulation Area (ZCTA) Distance Database. The primary and underlying estimate of aerial distance won't take into account the mobility mode and the terrain of the region. Thus, a more holistic and comprehensive approach to average out the travel time is by using maps services through the use of an API.

Travel Distance and duration among various origins and destinations can be computed using Google Distance Matrix services with a particular mode of travel^[12]. But since we require the use of the only average time of travel while taking into consideration the terrain and mode of travel and not the detailed route information including polylines and textual information, this API will do just fine without invoking the Distance services.

Google Geocoding API^[13] converts the address into geographic coordinates like latitude and longitude, which you can use to place markers on a map or position the map.

A plausible solution would be to use these two in conjugation. Since we are dealing with zip codes that cover an area usually using geocoding API would only pinpoint just one location with boundaries. These boundaries are provided in a northeast and southwest form of coordinates in the JSON output under the 'geometry' key.

```
import googlemaps
from datetime import datetime
gmaps = googlemaps.Client(key='API KEY')
# Geocoding an address
zipcode = #any healthcare facility zipcode
geocode_result = gmaps.geocode(zipcode)
```

JSON result of above request:

```
"geometry" : {
  "location" : {
    "lat" : 37.4224764,
    "lng" : -122.0842499
  },
  "location_type" : "ROOFTOP",
  "viewport" : {
    "northeast" : {
      "lat" : 37.4238253802915,
      "lng" : -122.0829009197085
    },
    "southwest" : {
      "lat" : 37.4211274197085,
      "lng" : -122.0855988802915
    }
  }
}
```

After acquiring latitude and longitude from the result and the boundaries. A set containing tuples of latitudes and longitudes can be formed. Since latitude provides a distance of 69 miles between each degree and so does longitude at maximum (since they are not equally distant). Each degree is further divided into 60 minutes which are again divided into 60 seconds. Simple calculations, one-minute equals 6068 ft (1.15 miles). Thus, a function to divide the boundary box into 1.15 miles separated coordinates can be written. Eventually, this can be understood as applying a grid over the concerned area and assuming each grid area to be covered by civilians in equivalent time from any point inside. During optimizations based on local area insights, we can further customize the grid area size or distance between coordinates.

```
boundaryNE_coordinates=(x1,y1)
boundarySW_coordinates=(x2,y2)
east_west_iter=(x2-x1)/one_minute
north_south_iter=(y2-y1)/one_minute
```

Thus, we would get a total of **east_west_iter * north_south_iter** coordinates. These will be different for different zip codes. For some, the number of in and destination coordinate computation would be highly intensive.

Thus, just to keep the problem less computationally intensive, for commencement we can have only the boundary coordinates and the pinpointed location.

```
def coordinates_in_boundary(northeast_lat, northeast_lng,
                           southwest_lat, southwest_lng):
    return [ (northeast_lat, northeast_lng),
             (southwest_lat, southwest_lng),
             (northeast_lat, southwest_lng),
             (southwest_lat, northeast_lng)]
```

With this function, we have 4 boundary coordinates and 1 marked coordinate. This can be done for every zip code for which the average time to travel is to be calculated from a source zip code. Now for all the facilities and their zip code, a dataset can be formed taking one as the source and other as the destination. For each of these combinations, the Distance Matrix API will be used. This API returns duration and distance for each origin, destination. Also, we can specify some optional parameters such as the mode of travel, the traffic_model(best_guess, pessimistic, optimistic).

```
"elements": [ {
  "status": "OK",
  "duration": {
    "value": 340110,
    "text": "3 jours 22 heures"
  },
  "distance": {
    "value": 1734542,
    "text": "1735 km"
  }
}]
```

After having the time calculated between the model's zip code. We can average the time for all the coordinates of source and destination and compute similarly for all zip codes.

But this computation can become intensive if for every source and destination zip code. Let's suppose we have 5 each destination and source coordinates. This would mean 25 combinations and for each let's say T time i.e. $25 * T$ time. This could require a heavy amount of time. Some ways of optimizing:

1. For countryside counties their zip code boundaries can sometimes pinpoint in the sea, we wouldn't need to compute the distance from there.
2. We can have a reference with aerial distance data between zip codes and can drop out the computation of those that exceed the aerial distance
3. At some places, the boundaries might point to non-populated areas (like mountains, hills, lakes). Their computation can be dropped.

III. OPTIMIZATION

For optimizing this work economy of areas, areas with disease history, age-group distribution can be taken into account.

A. Economy: The poor areas or the areas at the base of the pyramid might be more prone to facing health issues pertaining to their sanitation, medical care and also awareness of the preventive measures. Thus, diminished and subordinate economy calls for more requirements of healthcare. Hence the assumption that the population requiring healthcare at any instant would be half of the total population of that area, have to be either disregarded or outweighed.

B. Areas with disease history: The total population of the area should be considered for healthcare resources with a history of any contamination.

C. Age-group Distribution: This factor is critical to resource and service delivery. Taking into consideration the average life expectancy of Washington which is 78.1 years^[14], we can divide the age groups in the generation accordingly. Younger and older age groups are constantly prescribed check-ups. Thus, areas with a higher weight of such an age group would fully require healthcare services.

IV. CONCLUSIONS

With current population captured for counties would mean the net migration and fertility and mortality have been estimated. This number, when factored with the healthcare facilities located around, will help evaluate average travel time. Residing citizens when looked at from the perspective of age-group, economic level and need of healthcare will become a concrete commencement for allocating and deallocating of healthcare resources.

ACKNOWLEDGMENT

This research by the Ministry of Trade, Industry, and Energy (MOTIE), KOREA, through the Education Program for

Creative and Industrial Convergence (Grant Number N0000717)

REFERENCES

- [1] <http://worldpopulationreview.com/states/state-densities/>
- [2] "Idaho is Nation's Fastest-Growing State, Census Bureau Reports". *US Census Bureau* (Press release). 20 December 2017.
- [3] <https://www.reference.com/world-view/four-factors-can-affect-population-size-20646593bc3b4646>
- [4] <https://www.census.gov/data/tables/2016/demo/geographic-mobility/county-to-county-migration-2012-2016.html>
- [5] Eberstadt, Nicholas; Nunn, Ryan; Schanzenbach, Diane W.; Strain, Michael. *In Order That They Might Rest Their Arguments on Facts: The Vital Role of Government-Collected Data*
- [6] <https://www.census.gov/data/tables/2016/demo/fertility/women-fertility.html>
- [7] <https://www.researchgate.net/publication/245440370>
- [8] The application of optical satellite imagery and census data for urban population estimation: A case study for Ahmedabad, India Zur Erlangung des akademischen Grades einer DOKTORIN DER NATURWISSENSCHAFTEN von der Fakultät für Bauingenieur-, Geo- und Umweltwissenschaften des Karlsruher Instituts für Technologie (KIT)
- [9] <https://www.nber.org/data/zip-code-distance-database.html>
- [10] <https://developers.google.com/maps/documentation/javascript/distance-matrix>
- [11] <https://developers.google.com/maps/documentation/geocoding/intro#place-id>
- [12] <https://www.inc.com/bill-murphy-jr/scientists-just-ranked-all-50-states-by-life-expectancy-results-are-truly-stunning.html>
- [13] <https://www.kaggle.com/census/us-population-by-zip-code>
- [14] <https://www.kaggle.com/census/net-migration-flow-time-series-collection>
- [15] <https://www2.census.gov/programs-surveys/economic-census/2012/guidance/geonotes/us/2012metronotes-master.pdf>
- [16] <https://www.census.gov/geographies/reference-maps/2012/econ/census/united-states.html>
- [17] <https://www.census.gov/data/datasets/2017/demo/popest/nation-detail.html>



Aseer Ahmad Ansari, is currently pursuing his seventh semester in Computer Science & System Engineering from KIIT. He has published a few papers related to database models and data mining. He is the head instructor for the Data Science Group at his university. His research interests are Machine Learning, Deep Learning and Image Processing.



Sabyasachi Chakraborty is pursuing his Master's in computer engineering from Inje University. He has previously worked on multiple real-life projects related to computer vision, data sciences, biosignal analysis, and text mining. His research interest aligns with Computer Vision, Artificial Intelligence, BioSignal Processing, Algorithms, and Blockchain.



Satyabrata Aich is working as a research professor in the field of computer engineering. He has published many research papers in journals and conferences in the realms of Machine learning, Text mining, and Supply Chain Management. His research interests are natural language processing, Machine learning, supply chain management, Text mining, Bioinformatics, Artificial Intelligence, Blockchain.



Beom Su Kim is pursuing his Master's in digital anti-aging healthcare from Inje University. He has published many papers in the field of machine learning and health related informatics. His research interest aligns with biosignal analysis, machine learning, bioinformatics, artificial intelligence.



Hee-Cheol Kim received his BSc at the Department of Mathematics, MSc at the Department of Computer Science in SoGang University in Korea, and Ph.D. at Numerical Analysis and Computing Science, Stockholm University in Sweden in 2001. He is a Professor at the Department of Computer Engineering and Head of the Institute. Digital Anti-aging Healthcare, Inje University in Korea. His research interests include Machine learning, Text mining, Bio Informatics.