

ICCS240: Database Project

Population Estimation in Thailand

21 March 2020

Group Members:

Sirin Chankao (5980753)
Chakeera Wansoh (6080566)
Natthakan Euaumpon (6081213)

Overview:

This project use the different concepts of database management system and linear algebra. From linear algebra, we learned that eigenvector can be use to estimate the population with this idea in mind we decided to combine this with our DBMS knowledge. Therefore, the aim of this project is to estimate the future population of regions in Thailand by calculating from the previous years. Many of the project's informations (or evidences that the database exists) can be seen in the project's github. <https://github.com/chakeera/PopulationEstimation>

Theory:

For the sake of explanation, let assume that A is a matrix. This will work only if A is a symmetric matrix. We need to find all positive eigenvalues which are required for our numerical approach. We are going to implement this by combining Power Method with Rayleigh quotient iteration. This iteration has the following formula:

$$b_{i+1} = \frac{(A - \mu_i I)^{-1} b_i}{\|(A - \mu_i I)^{-1} b_i\|}$$

Where b_{i+1} is the next approximation of the eigenvector and I is the identity matrix. After that we can set the next approximation of an eigenvalue to the Rayleigh quotient:

$$\mu_{i+1} = \frac{b_{i+1}^* A b_{i+1}}{b_{i+1}^* b_{i+1}}$$

However, we need to compute all eigenvalues so we need to combine this technique with a deflation technique. Now we will get the largest possible eigenvalue. But we need to consider all of them. There are many deflation methods, but for this project, we are going to use Hotelling's deflation. This is one of the most popular techniques.

$$(A - \lambda_1 u_1 u_1^T)$$

Where u is elements in a set of N vectors, as eigenvectors are the set of N vectors of A. From the equation above, we will get the matrix that has the same eigenvectors and eigenvalues as matrix A. But the largest eigenvalue is replaced with 0. Thus we can combine this with power method and Rayleigh quotient to find the next biggest and so on until we find all the eigenvalues. Then we substitute the eigenvalues into a matrix to get our eigenvectors where it will be multiplied with nxn matrix that is the matrix of the offsprings and number of survivors and its result will

give us the estimated population value.

Database schema:

We have total three tables in our database with the following schema:
(These tables import data from dataset folder in github)

AgeStructure(Year INT, Age TEXT, Percentage NUMERIC)

RegionPopulation(Region TEXT, Year INT, Population NUMERIC)

TotalPopulation(Year INT, Population NUMERIC, Yearly_percent_change NUMERIC,
Yearly_change NUMERIC, Migrants INT, Median_Age NUMERIC,
Fertility_Rate NUMERIC, Density INT, Urban_percent NUMERIC,
Urban_Population NUMERIC, Country_Share NUMERIC, Global_Rank INT)

Backend:

Frontend:

For frontend, we use Java GUI to present our resulting data. There will be mainly three windows. The first window as shown in the first image below. This window will act as the menu window where user can select which option they wanted to use. The option includes ... and If the user select ..., the ... window will appear as shown in the second image below. If the user select, the ... window will appear as shown in the third image below.

Video:

link to video