The Doctors:
Global Disease Spreading Pattern and Prediction

Qinhan Liu HKUST

Ji Liu HKUST

Guanzhi Wang HKUST

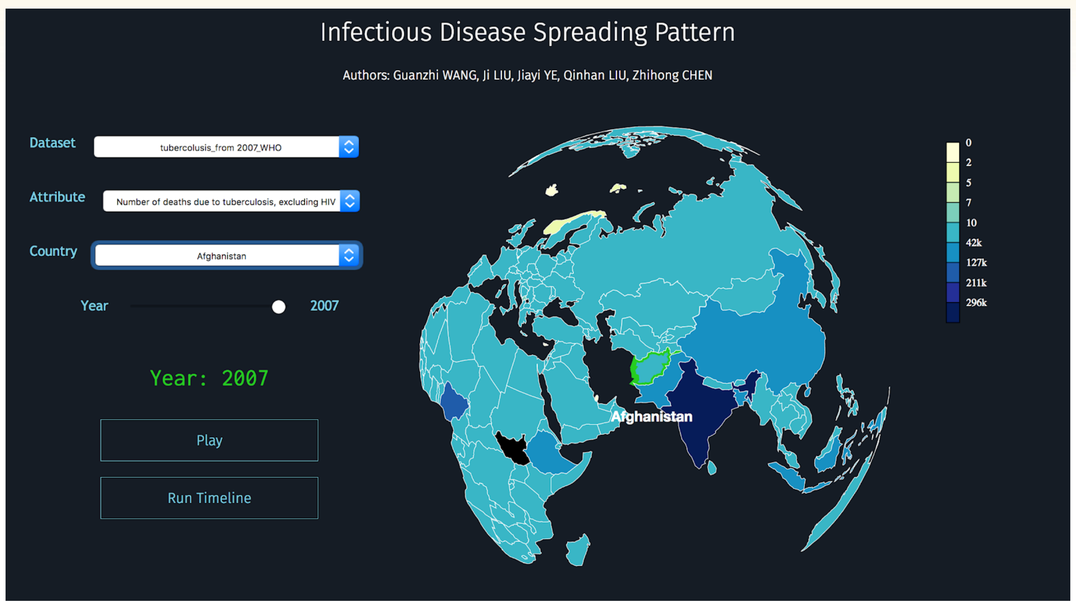
Jiayi Ye GT

Zhihong Chen CUHK

Summary

The Doctors aims to construct the **a 3D interactive map** of global infectious diseases **monitoring and prediction tools** in order to assist the public to be aware of or the government to improve the system of prevention, control and treatment on these diseases. A **heatmap** is formed from the geospatial time-serial data from **Google Flu data set** and **Kaggle Tuberculosis (TB) data set** to show the broadcasting patterns, and **KNN(10)** are implemented to predict the future trend based on local historical record.

Illustration



Method

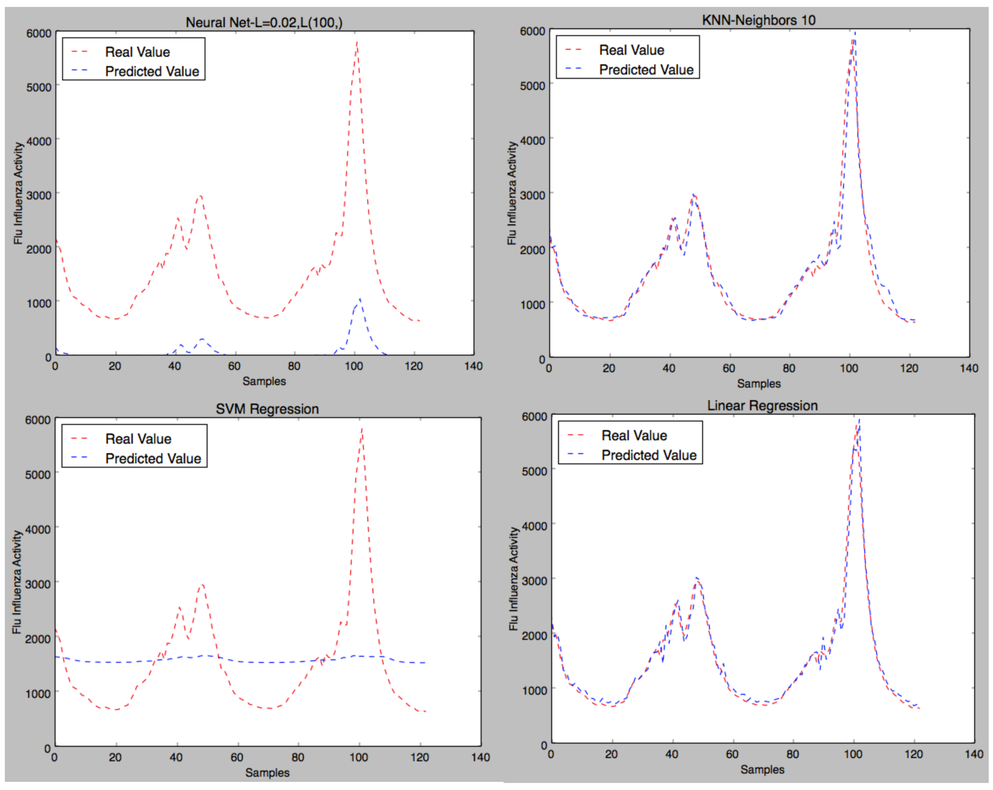
Data Cleaning: Two data sets, Google flu and Kaggle Tuberculosis are converted from .csv to .json format and then are regulated to fit the requirement of the map we implement.

Name	Spatial Scale	Time Scale	Size	Downloaded/ Scraped
Google Flu	Global	2007-2015	2M	Downloaded
Kaggle TB	Global	2007-2015	2M	Downloaded

Visualization: With **d3** being our tool, we applied the similar methodology from Homework 2 Q7 to this project. Beyond that, we have presented the data in **a three dimensional globe**, implemented the dragging and moving functionalities, enabled the drop-down menu for **country selection** as well as **attribute selection**, and added **dynamic timeline** for past data and prediction data. When the Country and Year attributes are not selected, the globe will display heatmap changes in time order.

Prediction: We tried 4 machine learning algorithms to learn the historical infection patterns to predict the future, they are Linear Regression, SVM Regression, Neural Networks, KNN(10). We split our datasets into training and test groups and perform weekly prediction.

Prediction Performance



Performance Evaluation

Cross-Validated Over All Countries	MAE	MSE	ESV	R2
Neural Network	391.89	666663.99	0.52	-2.00
K Nearest Neighbor	54.65	23938.63	0.86	0.85
Support Vector Machine	164.16	197631.38	0.31	0.16
Linear Regression	43.65	11493.58	0.90	0.90

Evaluation

Linear regression and KNN(10) outperform other two method and The Doctor can make good 1 week prediction. For now it is not realistic to make long term prediction with this setting. However, The Doctors has demonstrate how geo-health data visualization is done in the future.