Prediction and Research about future Environment Protection, Climate Change and Economy

Yedong Liu Shaalmali Raychaudhury
Illinois institute of Technology
3300 South Federal Street
Chicago, IL, 60616, USA
yliu275@hawk.iit.edu sraychau@hawk.iit.edu

Abstract—**This paper deals with the prediction and research about future environment protection, climate change and economy, How climate is changing day by day and how it is end up affecting human life, what are causing these changes and some statistical data.**

***Keywords* —- Data analysis, big data, economy, climate change, R, applied predictive modeling.**

I. Introduction

Let's start with what is climate? So Climate is defined as an area's long-term weather patterns. The easiest way to discuss climate is to consider the average temperature and precipitation for a certain period of time for a place. Other useful elements for getting knowledge about climate are like for example getting an idea about the type, precipitation timings, sunshine intensity, average of the wind speeds and directions, count of days above freezing, extreme weather, and local geography of the place concerned.

Climate is an important input to many human activities. As such it is to be expected that individuals will have a liking for a certain type of climate. These likings or preferences have been observed using a lot of different approaches including analyses of regional wage rates and prices of land in that particular region, the tendency to migrate, and analysis based on household consumption patterns.

1. Climate affects heating and cooling requirements. For example hot and humid regions need air conditioning and cool places need heating devices to keep warm.
2. Determines clothing.
3. Determines nutritional needs.
4. Limits recreational activities.

Tackling the problem of future climate change is one of the most challenging issues of this century and has major impacts on policies regarding development and management of the environment.

For determining, if climate change is good or bad, some indicators are needed. Measurements of the value of climate as a straight forward input for the well being of humans were mainly derived by using environmental valuation techniques such as the hedonic price approach or the household production function approach.

1.1 Hedonic Price Approach

The hedonic pricing approach is used to calculate economic values for ecosystem / environmental services that directly influence market prices. It is commonly used for variations in prices of housing that reflect the value of attributes of local environment. [1]

It can be used to calculate economic benefits or costs associated with:

1. environmental quality, also air pollution, water pollution, or noise
2. environmental amenities, such as views or proximity to sites for recreation.

The basic platform of the hedonic price approach is that marketed good pricing is related to its characteristics, or the services provided by it. For example, the price of a car implies the characteristics of that particular car—transportation, style, luxury, comfort, fuel economy, etc. Therefore, we can value the individual characteristics of a car by seeing how the expense people are willing to pay, as it changes when the characteristics or features provided change.

The hedonic pricing approach is often used for valuing environmental amenities which affect the price of residential properties.

1.2 Household Production Function

According to household production function approach, households combine goods marketed and nonmarket environmental goods to produce service flows which are of direct value to the household. This explains the reason, as an input to activities of household production, households might have their own preferences over the climate. [2]

Using techniques more often employed to account for differences in the composition of the demography of households household production function theory is used to estimate climate equivalence scales using expenditure data of household drawn from 51 Japanese cities over the period 2000-2009. The results indicate that warm temperatures result in a small but highly significant (statistically) reduction in the living cost. Combining these calculations with climate change cases associated with the IPCC A2, A1B, and B1 emissions cases other things being points in equal to a slight reduction in Japanese households' living cost.

1.3 Household Production Function

Since Jan. 1, United States of America set more than some 40k hot temperature records, but fewer than 6k cold temperature records, as per the National Oceanic and Atmospheric Administration. [3]

Global warming can be affecting formation of storm by decreasing the differences in temperature between the equator and the poles. That difference in temperature fuels the storms in mid-latitude affecting the Earth's most populated regions. Warm temperatures can be increasing the water vapor amount that enters the atmosphere. Result is a hotter environment with more humidity. At the equator, where environment are hot and humid already, this change is not expected to be too large. At the poles, however, the air is dry and cold, a small amount of extra heat and water vapor could raise temperatures to a great extent. "As a result, global warming might cause the temperature difference between the equator and the poles to decrease and with the decrease in difference, the number of storms should decrease too", as per George Tselioudis, who is research scientist at NASA Goddard Institute for Space Studies (GISS) and Columbia University.

But as the climate warming might decrease the total number of storms forming, it can also increase the number of storms more intense. With temperatures continuing to rise, more amount of water vapor can evaporate into the atmosphere, and water vapor ends up fueling storms. "If we are creating an atmosphere more loaded with humidity, any storm that does develop has greater potential to develop into an intense storm," says Tselioudis.

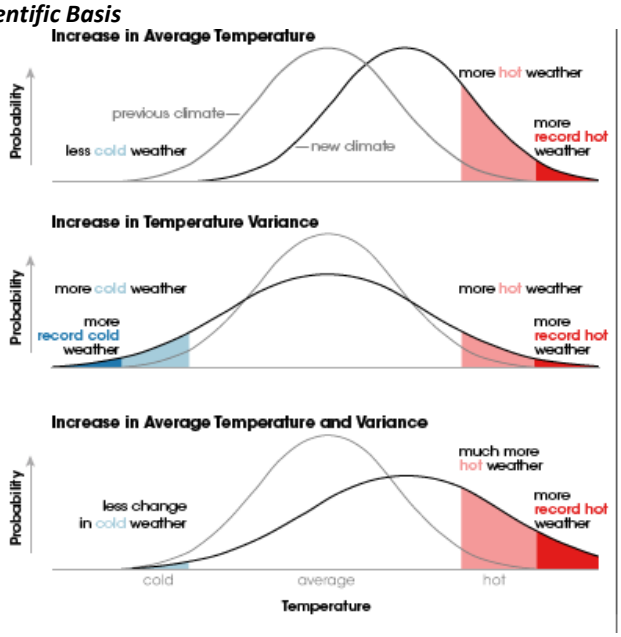
One way in which global warming can impact hurricanes is by making them all the more intense. More heat and water in the atmosphere and more warm sea surface temperatures can provide great fuel to increase the speed of winds of storms which are tropical. Warming that has already occurred since 1980, increased sea surface temperature by 0.3 degrees Celsius, which should increase the maximum potential hurricane wind speeds by 1 knot, as per hurricane intensity models. But, that small increase could not have been noticed or felt yet. "At present, hurricane intensity is calculated only to an accuracy of plus or minus 5 knots, so it is not possible to discern any change that might have occurred owing to warming that has already taken place," says Emanuel.

Coastal areas attract highest concentration of human economic and social activities at present, and these areas are the main victims of climate change effects. Consequently asset management is affected, both economic and social, representing a great challenge for the decision makers involved.

In the Caribbean and Bangladesh there is proof of both sensitivity declining to storms which are tropical and floods and increased resilience resulting from both transformation in economy and public actions for reduction in disaster.

The largest cluster of countries in high risk, vulnerable to these hazards of the climate, is in Sub-Saharan Africa. Risks originating from geophysical hazards need to be better noticed in urban areas which are highly exposed across the world as their potential costs are rising exponentially with development in economy.

Below is the figure adapted from ***Climate Change 2001: The Scientific Basis***



Now there is enough evidence for the ecological impacts of climate change at recent times, from polar terrestrial to marine tropical environments. Both flora and fauna responses span an array of ecosystems and organizational hierarchies, from the species to the community levels. Despite continuous uncertainty as to community and ecosystem issues under global change, review exposes a pattern which is coherent of ecological change across systems. Although it is an early stage in the noticed new trends of global warming, ecological responses to climate change at recent are already visible clearly.

1.4 Household Production Function

With the globalization going on, many countries began to industrialize, this process will benefit those countries but will also harm the environment.

While those countries are focusing in making money, they also have to spend a lot for protecting the environment.

The most important and obvious two are air pollution and water pollution.

Air pollution has been one of the most serious consequences of China's economic transformation and growth in the last three decades. Pollutant concentrations exceed all standards which are recommended by the World Health Organization (WHO) in virtually every important urban area. The large costs of air pollution are resulting in health impacts and loss of productivity in labor, running 6.5% of China's gross domestic product (GDP) every year between 2000 and 2010, and rising as China's population is becoming more productive and urbanized.

1.5 Health and Climate Change

There is this unanimous scientific conclusion that greenhouse gas emissions which are generated by human activity will end up changing Earth's climate. The recent (globally average) warming by around 0·5°Celsius is partly attributable to such emissions. Change in climate will affect human health in many way, mostly the bad way. [4]

Whether it's asthma, allergies or infectious disease, it is clear enough that global warming poses enough potential risks to human health.

Health officials expect the number of people who will contract infectious diseases to continue to rise with global temperatures. Infectious diseases that are mostly spread in mosquito borne, algal or fungal forms are more likely to breed and multiply because of the warming climates. Hot and humid climates provide the favorable habitat for these species and they are found mainly in regions that host such temperatures.

1.6 Health and Climate Change

The following are the three factors thought to be influencing climate change to human happiness:

1. Temp differences
2. Precipitation differences
3. GDP per capita

The measurement of per capita gross domestic product is given by the overall output of a country taking the gross domestic product (GDP) divided by the number of people in the country. The per capita GDP is very much useful when comparing one country to another as it shows the relative performance of the countries.

Human behavior to climate has been chalked out using three different models:

1. Neural Networks

   In cognitive science as machine learning, artificial neural networks (ANNs) are actually a family of models inspired by neural networks which are biological (the central nervous systems of animals, in particular the brain) which are used to estimate / approximate functions dependent on a large number of inputs and are unknown generally. Artificial neural networks are usually presented as collection of neurons which are interconnected which exchange messages with each other. The connections have weights in numeric that

can be tuned based on experience, making neural nets adaptive to inputs and learning capability.

2. Linear Regression

   Linear regression is a technique for modeling the relation between a scalar dependent variable denoted by x and one or more explanatory variables or not dependent variables denoted by y. The case of one variable which is explanatory is called simple linear regression. For more than one variable which is explanatory, the process is called multiple linear regression. [5]

   In linear regression, the relations are modelled using predictor functions which are linear whose unknown model parameters are calculated from the data. Such models are called linear models. Usually, the conditional mean of x given the value of Y and is assumed to be an affine function of Y; less often, the median or some other quantile of the conditional distribution of x given Y is expressed as a linear function of Y. Like all forms of regression analysis, linear regression stresses on the conditional probability distribution of x given Y, rather than on the joint probability distribution of x and Y, which is the domain of multivariate analysis.

   First type of regression analysis was Linear regression which was to be studied thoroughly, and to be used intensely in practical applications.

   As models which linearly depend on their not known parameters are more easily to fit than models which are non-linearly related to their parameters as the statistical properties of the estimators resulting are more easily to determine.

3. Single Trees

   Single trees help to partition the data into smaller groups that are all the more homogenous with respect to the response.

II. Hypotheses

Human beings have made many changes to the earth. Tremendous fossil fuel has being consumed every year by factories or transportations and turned into carbon dioxide in the environment. Trees are cut down, forest turned to agricultural land to feed increasing amount of people. Climate has been affected directly or indirectly by these behaviors.

The hypothesis is that these human behaviors have much influence on climate, which is specified as temperature in this research, and the climate can be an important factor of human happiness.

This research tries to build two predictive models, first is the impact of human behaviors on climate, the other is about how the climate influence human happiness.

III. Dataset

The dataset of human behaviors and climate change together with other demographical data are obtained from The World Bank, which is a website providing various data related to environment like fossil fuel consumption, carbon dioxide emission, agricultural land coverage etc. These data are public online, and can be download in csv format. Happiness indices are collected from the World Happiness Report 2013.

Because the impact of human behaviors are not obvious in the short time, we decided to collect data like temperature and precipitation change in 30 years, which is 1982 – 2012. And the happiness indices are determined in countries.

Agricultural land (% of land area) refers to the share of land area that is arable, under permanent crops, and under permanent pastures. Agricultural land (% of land area) is closely related to the climate change. The data is from 1980 to 2014, including more than 180 countries, therefore, the data resource is comprehensive. Each data has three effective numbers, and retains one decimal digit after decimal point. So, each data is accurate.

The data for Energy use (kg of oil equivalent per capita) is obtained from World Development Indicators of World Bank. Energy use refers to use of primary energy before transformation [6] to other end-use fuels, which is equal to indigenous production plus imports and stock changes, minus exports and fuels [7] supplied to ships and aircraft engaged in international transport. The data is available from 1960 to 2011.

The data for CO2 emissions in the unit of (kt) is from World Development Indicators of World Bank. It refers to carbon dioxide emissions stemming [8] from the burning of fossil fuels and the manufacture of cement. They include carbon dioxide produced during consumption of solid, liquid, and gas fuels and gas flaring [9]. The data is available from year 1960 to 2010. We took 30 years' difference of year 2010 and year 1981.

GDP per capita in current US$ is also from World Development Indicators of World Bank. The data is available from year 1960 to year 2013. GDP per capita is gross domestic product divided by midyear population. GDP is the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. [10] It is calculated without making deductions for depreciation of fabricated assets or for depletion and degradation of natural resources. [11] Data are in current U.S. dollars. [12]

The data for temperature and precipitation are obtained from World Bank Climate Data API. The Climate Data API provides access to historical average yearly values for temperature and precipitation. Both temperature and

precipitation data are annual country averages for 1901-2012.

We use 2012 and 1982 data and take the difference of the value of these two years to get the temperature change and precipitation change over 30 years period. There are three countries whose data are only available up till 2009 and thus lack 2012 data. They are Malta, Singapore and Democratic Republic of the Congo. Therefore we take these three counties out of all of our samples.

IV. Variables to be used in main analysis

4.1 Human Behavior Variables

Human behaviors data comprises of fossil fuel consumption, emission of carbon dioxide and change of agricultural land coverage. The difference are accumulated in 30 years from 1982-2012.

4.2 Climate Variables

Temperature and precipitation are two variables to represent the change of climate. Like the human behavior data, the change of temperature and precipitation are accumulated in 30 years.

4.3 Happiness Variable

The happiness of human is denoted by happiness index of 2010-2012 from World Happiness Report 2013. In this report, happiness index among 156 countries are represented in numbers.

4.4 GDP Variable

GDP per capita is used as a predictor of the outcome of happiness besides temperature and precipitation. In this analysis, GDP per capita of the year 2012 is used in consistent with the year of happiness index.

V. Research methods

Since data are obtained from different sources and some variable value are absent, data have to be preprocessed at first. After the final datasets have been determined, different predictive models will be trained and test on these data. Then the models will be compared in universal metrics such as Rsquared and RMSE, and for the two kinds of relationship the most suitable model would be determined respectively for our hypothesis.

R is a universal software for data analysis and model building, in this research models will be built, trained and tested in R.

VI. Results and Discussion

Different models have been used to analyze the data obtained before.

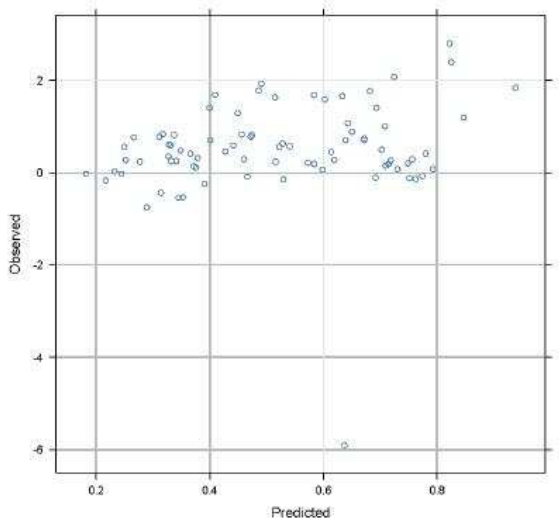
6.1 Linear Regression Analysis

Linear regression is one of the most straightforward predictive model. This model tries to find out the linear impact of human behavior to the climate change, and to the happiness.

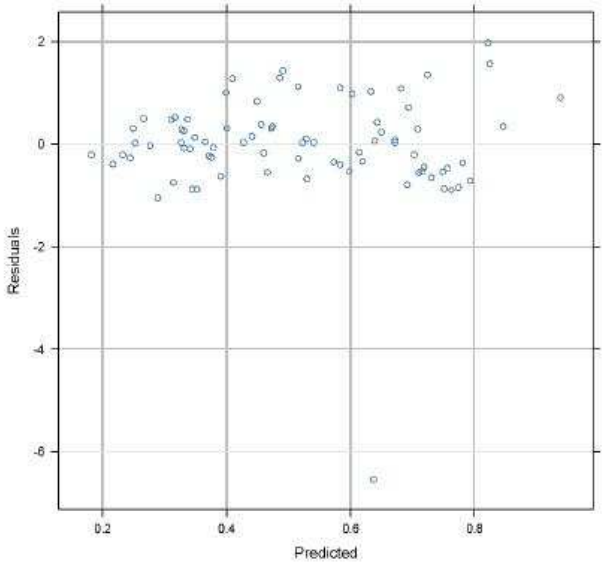
The data are split into training set and testing set first, the training data are used to build linear model, then analyze the performance of the model with the test data.

Linear regression analysis of human behavior to temperature

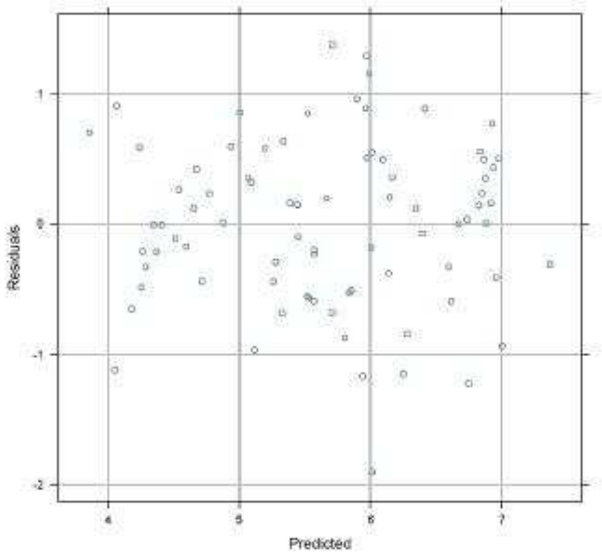
R-Squared: 0.199 RMSE: 0.860



The plot of predictive and observed outcomes



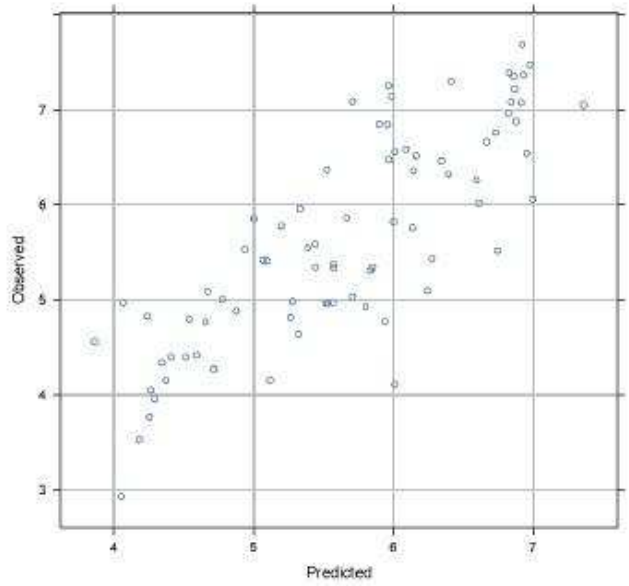
The plot of predictive value residuals



The plot of predictive value residuals

Linear regression analysis of climate change to human happiness

R-Squared: 0.668 RMSE: 0.659



The plot of predictive and observed outcomes

6.2 Single trees Analysis

Decision tree is a tree structure used in classification. Where each internal node of a property is on behalf of a test, the test results of each edge represents a leaf node which is on behalf of a class or the distribution of class. The top node is the root node. A decision tree shows that under what conditions would those values be. A decision tree is constructed using recursive top-down construction method. Multi-tree, for example, if a set of data there are several properties, the values of properties in a variety of this data set is subdivided into several subsets of the corresponding (branch), and then followed by recursive process of each subset. On the contrary, it is a leaf node. The results of the decision tree is a binary structure or tree, its input is a set of training data with the category tag. Internal node binary (non-leaf nodes) are generally expressed as a logical judgment, such as in the form of (a = b) the logic, which is a property of a, b is an attribute value of the property; the side of the tree is a logical analyzing the results of the branch. Multi-tree (ID3) internal nodes are property, and there are several property values. Leaf nodes of the tree are labeled categories.

Description of a given predictor variables X, a method for the conditional distribution of the variable Y, the use of spatial prediction binary tree recursively divided into several subsets, Y is distributed over the subset which is continuous and uniform, the tree leaf node corresponds to different area division, these divisions are made associated with each internal node with branching rule (Splitting rules) determined through the tree from the

root node to the leaf node, each predicted sample is assigned a leaf node, Y in the node the distribution is also determined.

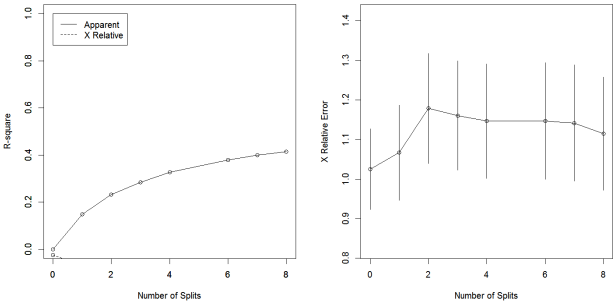
#Check the data(mydata)
> head(mydata)
> str(mydata)
'data.frame': 111 obs. of 6 variables:
> hap=rpart(Happiness/1~.,data=mydata)

#Regression tree model

> hap;printcp(hap);summary(hap,cp=0.1)
n=107 (4 observations deleted due to missingness)

Regression tree:
rpart(formula = Happiness/1 ~ ., data = mydata)

Variables actually used in tree construction:
[1] Argricultural.land.diff Energy.Usage Forest.land.diff
Precipatation.diff
[5] Temprature.diff
Variable importance Energy.Usage 31 Forest.land.diff 21
Argricultural.land.diff 21
Precipatation.diff 16 Temprature.diff 11
>plot(hap);text(hap)



#Illustration of different classification error

> par(mfrow=c(1,2));rsq.rpart(hap)

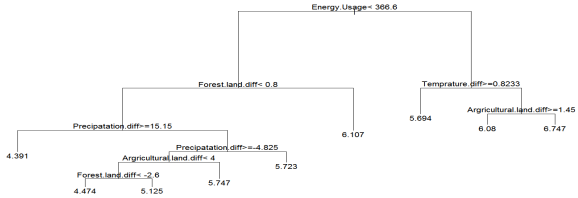
Regression tree:

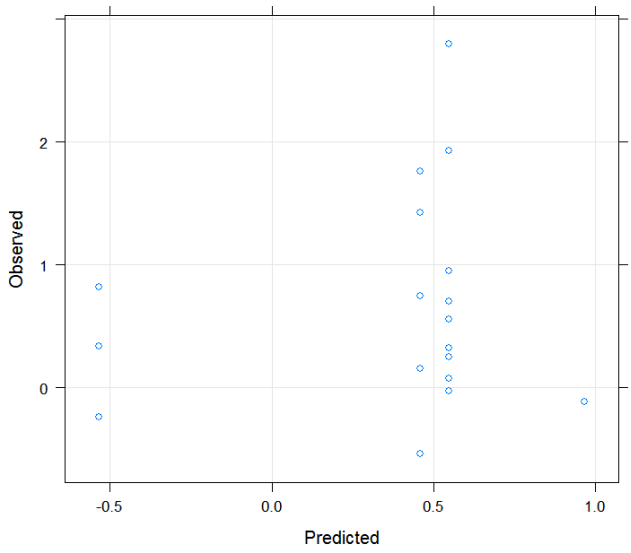
rpart(formula = Happiness/1 ~ ., data = mydata)

Variables actually used in tree construction:

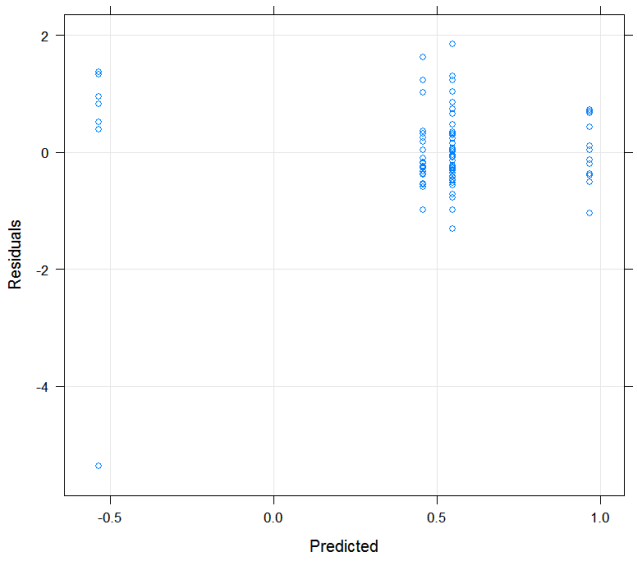
[1] Argricultural.land.diff Energy.Usage Forest.land.diff Precipatation.diff

[5] Temprature.diff



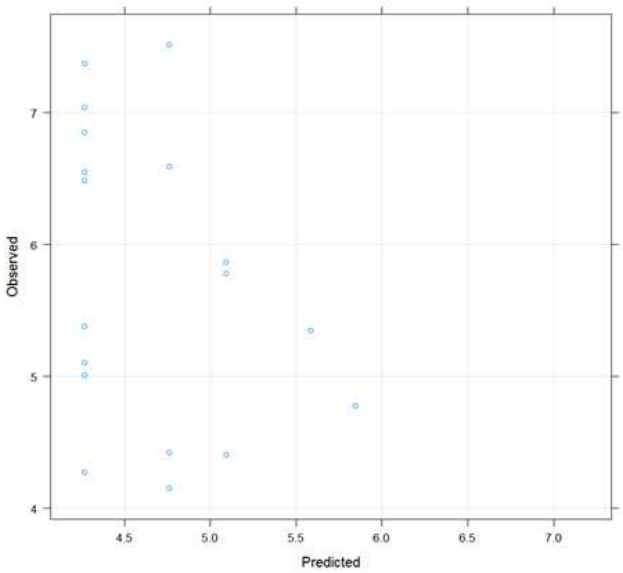


Human behavior and Climate Change Observed outcome

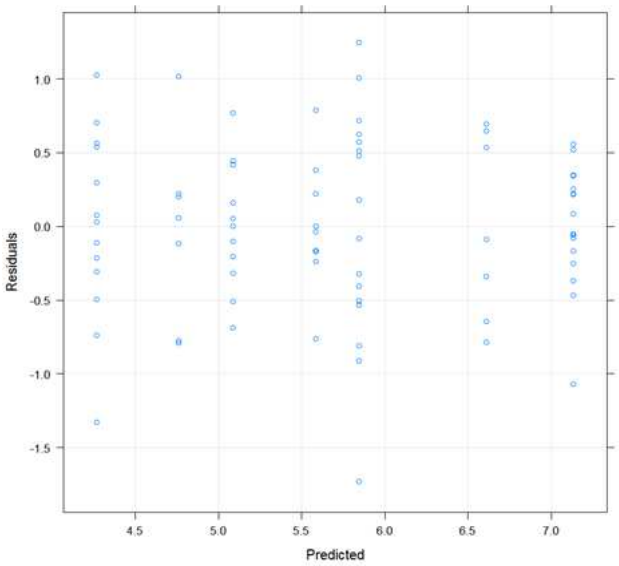


Human behavior and climate change Predict outcome

RMSE=0.8904103; R-squared=0.1948958



Climate change and Happiness Observe outcome



Climate Change and Happiness predict outcome

RMSE=0.7687188; Rsquared=0.6035076

6.3 Non-Linear Analysis

Nueral Networks

Neural networks are inspired by theories about how the brain works. The outcome is modeled by an intermediary set of unobserved variables. [13] Each hidden unit is a linear combination of some or all of the predictor variables. The linear combination is actually evolved from a nonlinear function, such as the logistic function.

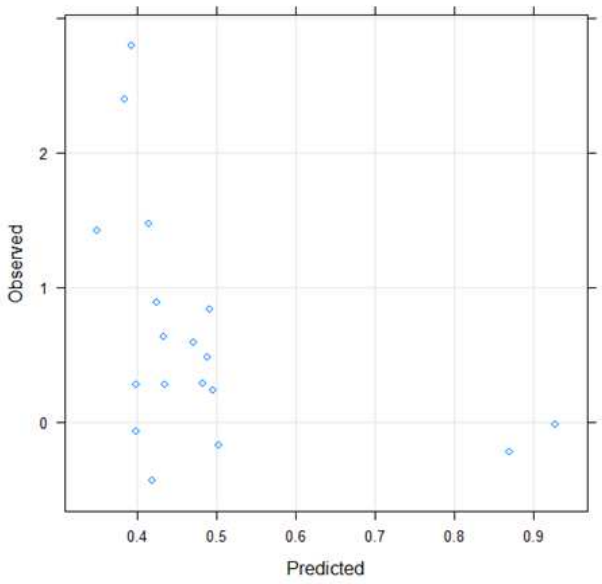
A approach to moderating over-fitting in neural networks is to use weight decay. It is a penalization method to regularize the model.

The data was divided into training set and testing set. The training data was used to build the model and the testing data was used to examine the performance of the built model. Three different weight decay values were evaluated (lambda=0.00, 0.01, 0.10) together with one layer ranging between 1 and 5 units. Different parameter of lambda and the number of hidden units were then used to form models and these models were compared based on their RMSE. We chose the model with the least RMSE.

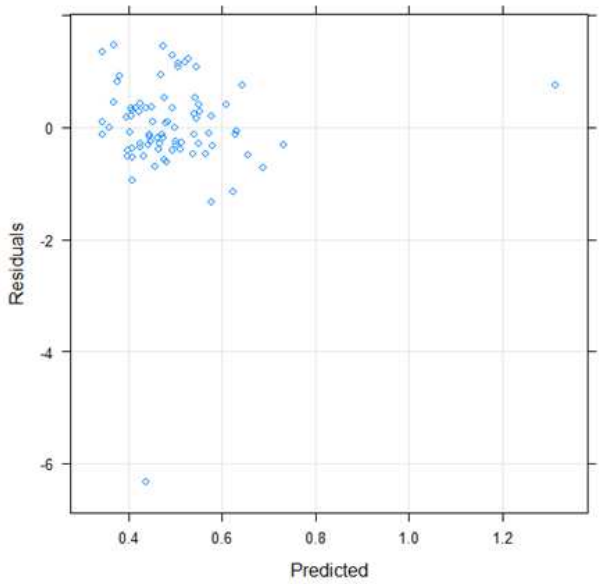
Neural Networks analysis of human behavior to temperature

R-Squared: 0.188 RMSE: 0.947

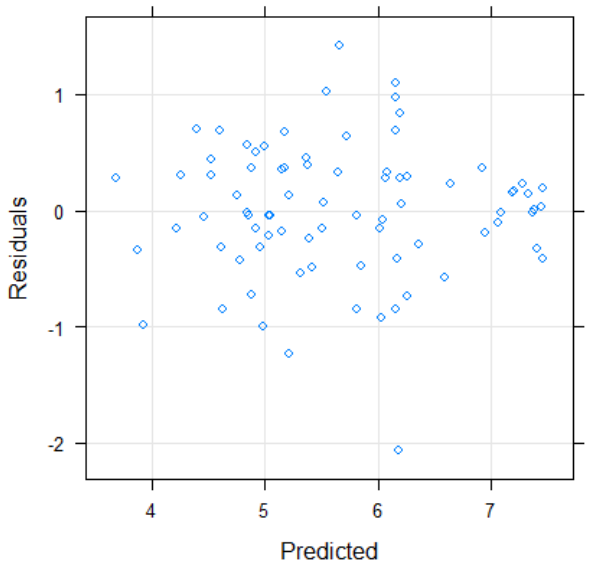
size=1, deday=0.1



The plot of predictive and observed outcomes



The plot of predictive value and residuals

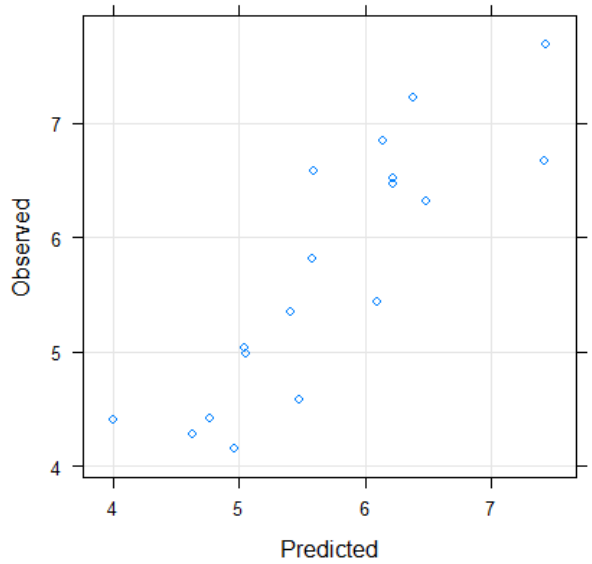


The plot of predictive value and residuals

Neural Networks analysis of climate change to happiness

R-Squared: 0.747 RMSE: 0.547

size=3, decay=0.01



The plot of predictive and observed outcomes

VII. CONCLUSION

After a comparison across these four models, it can be seen from their R-Squared and RMSE that the temperature change of a country has a weak correlation with the agriculture land, energy use and CO2 emissions of that country. On the other hand, happiness of a country is strongly correlated to the GDP per capita of that country. Surprisingly, happiness is weakly positively correlated with precipitation, indicating that the country with more precipitation makes its people happier. However, the result shows that happiness has little correlation with temperature.

REFERENCES

[1]http://www.ecosystemvaluation.org/hedonic_pricing.htm
[2]https://ideas.repec.org/p/kie/kieliw/1693.html
[3]http://earthobservatory.nasa.gov/Features/RisingCost/rising_cost5.php
[4]http://environment.nationalgeographic.com/environment/global-warming/pollution-overview/
[5]https://en.wikipedia.org/wiki/Linear_regression
[6]https://www.destatis.de/EN/FactsFigures/CountriesRegions/InternationalStatistics/Topic/DefinitionsGlossary/PrimaryEnergyUsePerCapita_c.html?view=getColorboxEntry
[7]http://data.worldbank.org/indicator/EG.IMP.CONS.ZS
[8]http://hdr.undp.org/en/content/carbon-dioxide-

emissions-capita-tonnes
[9]http://data.worldbank.org/indicator/EN.ATM.CO2E.PC
[10]http://data.worldbank.org/indicator/NY.GDP.PCAP.CD/countries/HT-xj?display=graph
[11]http://data.worldbank.org/indicator/NY.GDP.PCAP.CD/countries/HT-xj?display=graph
[12]http://data.worldbank.org/indicator/NY.GDP.PCAP.CD/countries/HT-xj?display=graph
[13]https://books.google.com/books?id=xYRDAAAAQBAJ&pg=PA141&lpg=PA141&dq=%22of+unobserved+variables+(called+hidden+variables+or+hidden%22&source=bl&ots=qQseENkBZ5&sig=Lr-obMV16DpyEPImSz384UyixpY&hl=en&sa=X&ved=0ahUKEwjjkf2F48PMAhUEFx4KHZ2XCUIQ6AEIHTAA#v=onepage&q&f=false