Statistical Overview and Prediction of Election Data

Students: NG Ching Tao (CityU), Jorge Garcia (NMSU), Frank Betancourt (UTK)
Mentor: Dr. Kwai WONG (UTK)

Introduction

The objectives of the project:

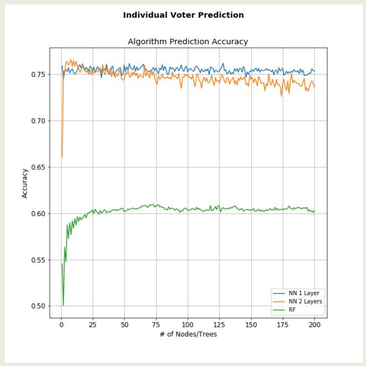
- Find out relationship between factors affecting voting decision and voting result
- Use machine learning algorithms to experiment with feasibility of predicting electoral results
- Visualize election data result to public

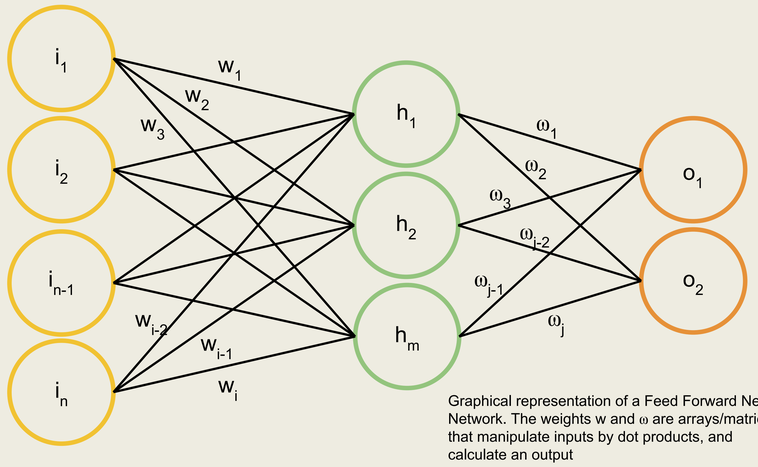
The datasets are from:

- Public Opinion Program, the University of Hong Kong (HKUPOP)
- ANES Time Series Study (ANES)
- American Community Survey (US Census Bureau)

Approach

- US & HK similar 2-party system
- Recollection of election public opinion data
- Find out the Pearson correlation between the factors and voting result
- Parameterize nominal data by one-hot-encoding parameters
- Predict probable voter outcome with a Feed Forward Neural Network (2 Layers, 10 Nodes) using the Keras library
- Count total votes for each party with weighted population
- Repeat predictions 100 times to obtain a sample
- Approximate sample with a cont. probability distribution
- Obtain appropriate moments of the distribution to describe the measurements (average and variance)

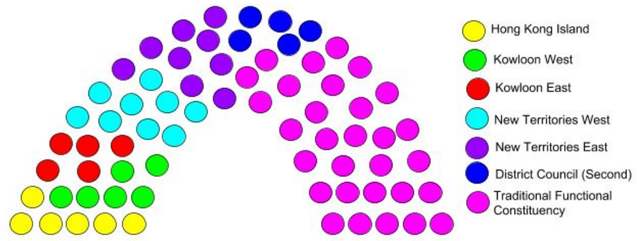




Graphical representation of a Feed Forward Neural Network. The weights w and ω are arrays/matrices that manipulate inputs by dot products, and calculate an output

Hong Kong data

The Hong Kong Legislative Council Election has two parts, which are the Geographical constituency and Functional constituency.



In Geographical constituency, the dominant factors are:

- Political inclination
- Emphasis on relationship with Central government
- Voting decision
- Education level

In District Council (Second) Functional constituency, the dominant factors are: Preferences of candidates, Education level, Voting decision

- In 2008, it was joining July first demonstration in Hong Kong Island

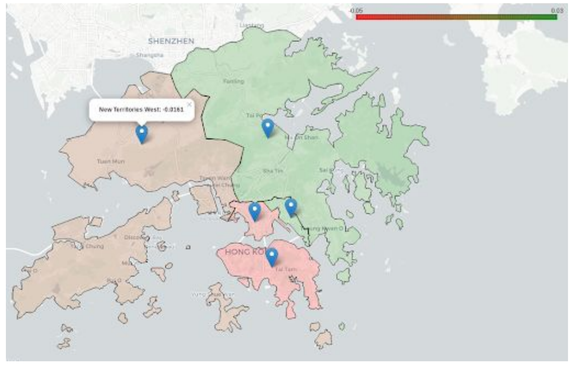
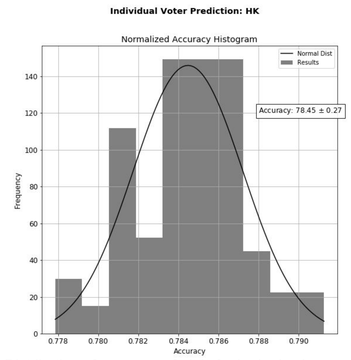


Figure: Chloropleth map showing the difference in correlation between factors and voting result for the 5 geographical constituencies in Hong Kong



Distribution of accuracy of trials for both elections are best approximated with a normal distribution

Predicting Legislative Council elections:

- Use opinion surveys from 2008 and 2012 to train a prediction model.
- Test prediction accuracy with 2016 opinion survey.
- Count predicted votes and compare to actual results from 2016.

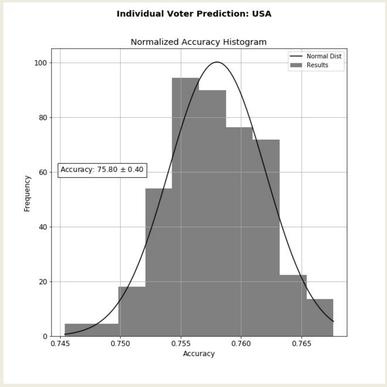
Most predictions are well within a σ of actual results, but uncertainties can declare some predictions inconclusive.

	2016 Legislative Council Election			
	Pro-Government		Pro-Choice	
District	Prediction	Actual	Prediction	Actual
Hong Kong Island	$61.14 \pm 5.59\%$	48.97%	$38.86 \pm 5.59\%$	51.03%
Kowloon W	$38.06 \pm 11.39\%$	36.91%	$61.94 \pm 11.39\%$	63.09%
Kowloon E	$49.69 \pm 8.88\%$	49.14%	$50.31 \pm 8.88\%$	50.86%
New Territories W	$52.76 \pm 7.93\%$	44.27%	$47.24 \pm 7.93\%$	55.73%
New Territories E	$44.56 \pm 9.04\%$	40.19%	$55.44 \pm 9.04\%$	59.81%

U.S. Data

Can election outcomes be predicted using state census data?

The ANES Time Series studies provide rich information on individuals opinion of public and political matters (eg. elections). The Clinton era was chosen as the beginning of modern american politics to serve as the basis of training.



Predicting US House of Representatives elections:

- Identify parameters also existent in US Census Bureau surveys (ACS)
- Homogenize data to have same format
- Train NN model using ANES surveys from 1992 - 2014
- Test prediction accuracy with 2016 survey data
- Predict 2016 election outcome in state by evaluating ACS data using trained model. Compare to actual results.

Demographics can only go so far. Variance in estimates can render some predictions inconclusive.

	2016 House of Representatives Election			
	Democrat		Republican	
State	Prediction	Actual	Prediction	Actual
CA	$76.43 \pm 6.83\%$	62.31%	$23.57 \pm 6.83\%$	36.89%
TX	$37.18 \pm 5.75\%$	37.1%	$62.82 \pm 5.75\%$	57.2%
AL	$41.41 \pm 4.36\%$	32.91%	$58.59 \pm 4.36\%$	64.67%
MN	$54.51 \pm 8.45\%$	50.23%	$45.49 \pm 8.45\%$	46.73%
FL	$33.97 \pm 4.86\%$	45.21%	$66.03 \pm 4.86\%$	54.71%

Future Work

- Maximize voter prediction accuracy by identifying most relevant factors impacting vote outcome
- Use of other machine learning algorithms or neural network architectures may provide better results
- Develop an interactive visualization of election results for the US and HK

Acknowledgements

This project was sponsored by the National Science Foundation through Research Experience for Undergraduates (REU) award, with additional support from the Joint Institute of Computational Sciences at University of Tennessee Knoxville. This project used allocations from the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation.
In addition, the computing work was also performed on technical workstations donated by the BP High Performance Computing Team.