Statistical Overview and Visualization of Hong Kong Legislative Council Election

Mentor: Dr. Kwai Wong

Project by:
NG Ching Tao
Jorge Garcia
Frank Betancourt
Hong Kong Legislative Council Election

- Part 2 Other Functional Constituency is not included
- Two parties of candidate:
  - Pro- government camp
  - Pro- choice camp
R programming

E.g.

cor.test(RP2012_sorted$q7, RP2012_sorted$R3)

    Pearson's product-moment correlation

    sample estimates:

        cor

    0.06245718
Election Result

2008

Democratic Alliance for the Betterment of Hong Kong
Hong Kong Federation of Trade Union
Civic Party
Liberal Party
Labour Party
Pro-government individuals and others

2012

Democratic Alliance for the Betterment of Hong Kong
Pro-government individuals
Civic Party
Liberal Party
Economic Synergy
People Power
Pro-choice individuals

2016

Democratic Alliance for the Betterment of Hong Kong
Pro-government individuals
Civic Party
Liberal Party
Economic Synergy
People Power
Pro-choice individuals

Number of seats

Democratic Party
Hong Kong Federation of Trade Union
Liberal Party
Economic Synergy
People Power
Pro-choice individuals

New People’s Party
Hong Kong Federation of Trade Union
Hong Kong Professional General Union
League of Social Democrats

New Forum

Democratic Party
Pro-choice Independent

## Data Correlation

In election period, the dominant factors are

<table>
<thead>
<tr>
<th>Year</th>
<th>Hong Kong Island</th>
<th>Kowloon West</th>
<th>Kowloon East</th>
<th>New Territories West</th>
<th>New Territories East</th>
<th>District Council (Second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Political Inclination</td>
<td>Political Inclination</td>
<td>Political Inclination</td>
<td>Political Inclination</td>
<td>Political Inclination</td>
<td>/</td>
</tr>
<tr>
<td>2012</td>
<td>Emphasis on relationship with Central government raised by candidate</td>
<td>Follow strategic plan raised by candidate</td>
<td>Emphasis on relationship with Central government raised by candidate</td>
<td>Voting decision</td>
<td>Education level</td>
<td>Political Inclination</td>
</tr>
<tr>
<td>2016</td>
<td>Voting decision</td>
<td>Voting decision</td>
<td>Education level</td>
<td>Voting decision</td>
<td>Voting decision</td>
<td>Voting decision</td>
</tr>
</tbody>
</table>
Data Correlation

On election day, the dominant factors are

<table>
<thead>
<tr>
<th>Year</th>
<th>Hong Kong Island</th>
<th>Kowloon West</th>
<th>Kowloon East</th>
<th>New Territories West</th>
<th>New Territories East</th>
<th>District Council (Second)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2008</td>
<td>Join 1st July Demonstration</td>
<td>Occupation</td>
<td>Why candidate attracts</td>
<td>Duration of being voter</td>
<td>Education level</td>
<td>/</td>
</tr>
<tr>
<td>2012</td>
<td>Channels of knowing candidates</td>
<td>Why candidate attracts</td>
<td>Voting decision</td>
<td>Reason of voting</td>
<td>Voting decision</td>
<td>Age</td>
</tr>
</tbody>
</table>
Manipulating the Data

- Three surveys:
  - 2008: 15,323 citizens
  - 2012: 16,253 citizens
  - 2016: 10,601 citizens

- Survey Question → Parameter
  - Each possible answer represented by a number (1, 2, 3..)
  - Same questions throughout years, different answer layout (Ugh!)

- Homogenize Data: *(Very boring and tedious)*
  - Result: 42,177 citizens total

Can now play with different models...
What is One Hot Encoding?

Cardinality:

- **Education:**
  - 0 - No Information
  - 1 - Elementary School
  - 2 - High School
  - 3 - Higher Education

One Hot Encoded:

- **Education**
  - [0 0 0 0]
  - Index corresponds to answer!

Eg: [0 0 1 0]

High School education
Evaluation of Algorithms

General Election, 2007 2012 Training

Not One Hot Encoded

Sample Size: 20705
- Training: 80%
- Testing: 20%
# of Parameters: 7
NN Layers: 1

One Hot Encoded

Training Sample Size: 20705
- Training: 80%
- Testing: 20%
# of Parameters: 39
NN Layers: 1
Legislative Council Election 2016

Total seats: 35

Total number of votes: 2,064,077

Pro-Choice: Pan-Democrats & Localists

Prediction Results 2016

Neural Network: 1 Layer, 15 Nodes
- Pro-Government: 48.14% of votes
- Pro-Choice: 51.56% of votes

Random Forest: 30 Trees
- Pro-Government: 37.80% of votes
- Pro-Choice: 49.62% of votes

Actual Results 2016

- Pro-Government: 871,016 (42.20%) votes
- Pro-Choice: 1,193,061 (57.80%) votes
Future Work

Improve prediction accuracy:
- Additional NN layers (DNN)
- Use of other machine learning algorithms and NN architectures
- Approach as a regression problem? (Predict probability of side winning)

Attempt to use model in other elections:
- Plenty of U.S. election data
Visualization Goals

Development of additional software to interface with Dark Matter as a plugin

Create 3D visualizations, and be able to interact with how the visualizations are displayed with graphical or scripted controls in Dark Matter
Data Visualization

-Still no access to DarkMatter

-Focusing on VR/Google headsets for prototyping of visualization

Figure 1: Projection of Hong Kong Map on to a sphere

Figure 2: View inside of sphere