CS-E4870

Research Project in Machine Learning and Data Science

Ananth Mahadevan

Department of Computer Science, Aalto University
ananth.mahadevan@aalto.fi

Abstract

1 Introduction

- how wikipedia is a large encyclopedia
- maintained by a small group of Administrators
- They undergo an election like process of RfAs
- How this is an important online social election framework
- How it has been studied in previous works
- What we aim to do by using a social network and theories of democracy

Wikipedia is the largest online encyclopedia containing over 5 million pages of content. It is one the most popular websites on the Internet. Wikipedia has a diverse collection of articles from many different topics and is constantly being updated. Although Wikipedia started out as an open platform where anyone could create and articles, this lead to many factual errors and biased articles. Wikipedia started to incorporate elements hierarchy gradually over time. In the English version of Wikipedia all editors to have a registered account and pages that are controversial and of a sensitive nature are protected by administrators.

Although there are over 38 million registered editors[1] only around 130 thousand are regular contributors to articles and the discussion forums.

2 Literature review

- election prediction using candidate stats
- election analysis using voter and candidate info
- prediction using communication and how close
- Signed edge prediction and difficulties

3 Dataset

- explain RfA data collection
  - existing SNAP data and limitations
  - XML parsing
  - regex and string matching
  - date parsing
- Social interactions
  - User contributions
  - wealth and diversity of info
  - creating underlying network

4 Viscous Democracy

Brief explanation of viscous democracy

5 Proposed Model

Use viscous democracy models using heuristic delegation functions on social network to predict elections separately

6 Implementation

directed graph concepts and delegation function considerations. Agony and hierarchy. local and global top k delegates.

[1] From here all information is for the English version of Wikipedia

7 Results

The quality of predictions using local or global important editors.

8 Conclusions

How we can instead try and model individual voter behaviour. Find a more robust ML framework to learn an optimal delegation function.

References