Abstract

Abstract

Contents

1 Introduction

In Republic of Lithuania, public elects their representatives to a parliament in which new legislation is considered. By this election each citizen delegates specifics of legislation process to their representatives so they don't have to actively participate in the process.

However, a problem arises once citizen wants to validate what his representative has been doing. Single term of office NEEDS-CITATION involves thousands of complicated laws and votes. To analyze everything becomes almost an impossible task for a single citizen who is out of the loop.

To make it easier there are journalists, politologists and other personas who review new legislation, current issues. Delegates themselves also do press conferences, debates where they state their intentions, comment on their actions. However, this requires citizens to trust that journalists and delegates only state truth, don't omit important information and don't have other hidden agendas. Study done about intrinsic honesty showed that the more society is corrupt - the more people lie in a simple dice game. This applies to politicians too and citizens trust in parliament is relatively low. Therefore, anything that can be done to better observe representatives is useful.

2 Analysis

2.1 Literature analysis

There is a decent amount of previous work analyzing voting on roll call data. Research seems to be focused on specific elections that happened in the past.

In *Spatial Models of Parliamentary Voting* author discusses how voter's positions on specific issues can be captured by his position on one or two dimensions such as liberalism or conservatism. This constraint means there are two spaces - one with few dimensions - basic or ideological. The other - high dimensional space which represents remaining issues. This breakthrough might suggest Multidimensional scaling as a good performance method for visualization and analysis as majority of data is encoded in few dimensions.

There is also research done specifically on Parliament of the Republic of Lithuania (LRS) data. One such is *On Structural Analysis of Parliamentarian Voting Data* [1]. In this paper authors discuss about data reduction to dissimilarity matrix, vote encoding, Multidimensional scaling (MDS) and its performance on specific dimensions. Authors focus on specific elections and term of office which is different from our goal.

2.2 Lithuania's parliament open data analysis

2.3 Problem analysis

Problem can be divided into two pieces:

- Research part, including MDS and k means classification
- Software development

If we consider what is the output of parliaments during their term of office in parliament - it would be their voting outcomes. Let's say we have a set of all votes made by parliament members $V = \{v_1, v_2, ...v_n\}$ where each vote v_i has a tuple of parameters $P = \{timestamp, term of office, parliament member id, voting outcome, ...\}$. Our goal is visualize set V in a way that similar voting patterns of different members are visible.

With MDS XX

With k means XX

2.4 Method analysis

2.4.1 MDS

2.4.2 Unsupervised learning: *k*-means clusterization

Due missing labels of this dataset we are forced to look into unsupervised machine learning methods. *k-means* clusterization [**?**] is one such method and it fits well with our features. This is due TFIDF which transforms text into vectors be used with *k-means*. For distance between comments we can use cosine similarity [**?**] due its nature of performing better when comparing texts. To compare we can try with Euclidian and Manhattan distances too.

2.5 Visualizations

3 Software design

3.1 Components of system

3.1.1 Data flow diagram

Data flow diagram

3.2 Tools

3.2.1 Programming language: *S*cala

3.2.2 Database: *M*ySQL

3.3 Downloader

3.4 Coordinator

3.5 API server

3.6 Repository

3.7 User interface (Frontend)

4 Description of experimental research

4.1 Data

4.1.1 Data semantics

4.1.2 Data statistics

4.2 Experiments

4.2.1 MDS

4.2.2 Unsupervised learning: k-means clusterization

4.3 Results

References

[1] T. Krilavicius and A. Zilinskas, “On structural analysis of parliamentarian voting data,” *Informatica, Lith. Acad. Sci.*, vol. 19, no. 3, pp. 377–390, 2008.

A Appendix