

Probabilistic Measures of Regime Type

Shahryar Minhas

sfm12@duke.edu

Duke University

August 6, 2014

Goal

Create country/year probabilistic measures of archetypical regime types, specifically:

- Democracy
- Militaristic
- Monarchical
- One-Party

Method: Supervised Approaches

- Bernoulli Naive Bayes
- Support Vector Machines (SVM)

Textual Data

- Scraped all country level reports for 1999-2013 from:
 - State Department Human Right Reports
 - Freedom House Freedom in the World Reports
- Cleaning textual data:
 - Removed numbers, punctuation, and stopwords
 - Tokenized results
 - Lemmatized tokens to group together inflected forms

Datasets used for Labels

- Marshall et al. 2014 (Polity)
- Freedom House 2014 (FH)
- Geddes et al. 2014 (GWF)
- Hadenius et al. 2012 (ADR)

Label Rules

Democracy

If Polity equals 10 and FH equals Free then we set Democracy for country-year equal to one otherwise zero

Military, Monarchy, & Party

If GWF and ADR dataset agree on coding for country-year then variable equals one otherwise zero

Cross-Validation Strategy

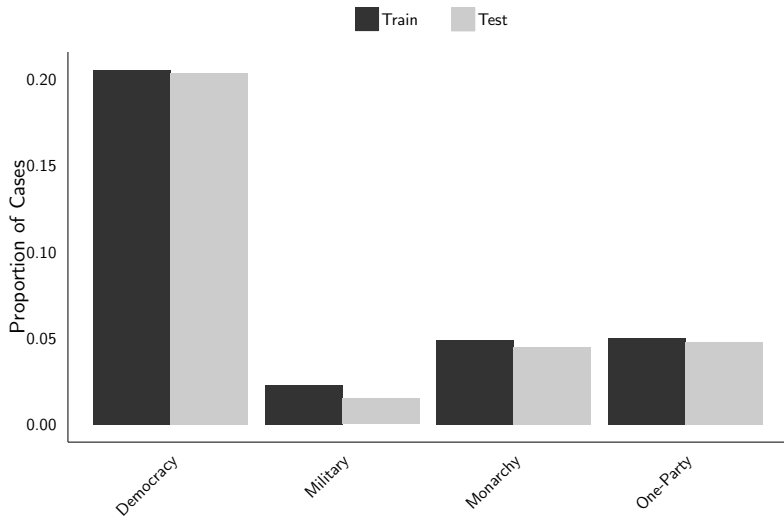
Democracy

- Training dataset: 1,557 cases from 1999-2008
- Test dataset: 707 cases from 2009-2013

Military, Monarchy, & Party

- Training dataset: 1,138 cases from 1999-2006
- Test dataset: 583 cases from 2007-2010
- Shorter timeline because GWF and ADR datasets end at 2010

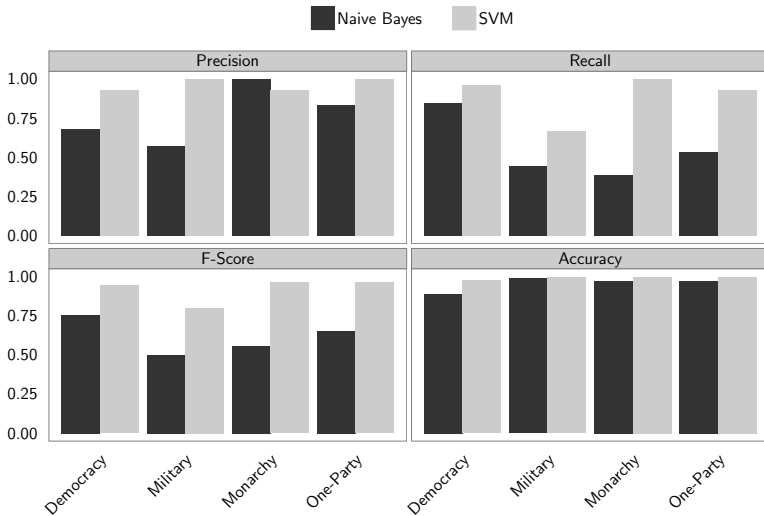
Label Distributions



Features for Supervised Models

- Vectorized tokens from combined State and FH reports for a given country-year
- Experimented with additional features but supervised models with just vectorized tokens performed the best (in terms of precision, recall and accuracy), see below for other features attempted:
 - Word count of combined tokens
 - Posterior topic predictions from an unsupervised LDA

Out-of-Sample Performance



Performance by Class for Democracy

Naive Bayes

	Precision	Recall	F-score	Support
Not Democracy	0.96	0.90	0.93	563
Democracy	0.68	0.85	0.76	144

SVM

	Precision	Recall	F-score	Support
Not Democracy	0.99	0.98	0.98	563
Democracy	0.93	0.96	0.94	144

Performance by Class for Military

Naive Bayes

	Precision	Recall	F-score	Support
Not Military	0.99	0.99	0.99	574
Military	0.57	0.44	0.50	9

SVM

	Precision	Recall	F-score	Support
Not Military	0.99	1.00	1.00	574
Military	1.00	0.67	0.80	9

Performance by Class for Monarchy

Naive Bayes

	Precision	Recall	F-score	Support
Not Monarchy	0.97	1.00	0.99	557
Monarchy	1.00	0.38	0.56	26

SVM

	Precision	Recall	F-score	Support
Not Monarchy	1.00	1.00	1.00	557
Monarchy	0.93	1.00	0.96	26

Performance by Class for One-Party

Naive Bayes

	Precision	Recall	F-score	Support
Not One-Party	0.98	0.99	0.99	555
One-Party	0.83	0.54	0.65	28

SVM

	Precision	Recall	F-score	Support
Not One-Party	1.00	1.00	1.00	555
One-Party	1.00	0.93	0.96	28

Next Steps

- Incorporate additional textual data from Human Rights Watch
- Compare performance of SVM with supervised LDAs