Democratic Regression

Timothy Baldwin

(joint work with Shivashankar Subramanian and Trevor Cohn)

THE UNIVERSITY OF MELBOURNE

ACL2018

15 – 20 JULY / MELBOURNE AUSTRALIA

Talk Outline

Computational Political Science

- Burgeoning area of political science/computation social science
- Political Science has always been text/data-rich, so unsurprising that political scientists are increasingly looking to NLP/ML methods
- Example applications:
    - election outcome forecasting [Tumasjan et al., 2010, Chung and Mustafaraj, 2011]
    - analysing political debates [Malouf and Mullen, 2008, Prabhakaran et al., 2014, Habernal and Gurevych, 2015]
    - congressional vote prediction [Thomas et al., 2006, Burfoot et al., 2011]
    - political agenda analysis over press releases [Grimmer, 2010]
- Loads of annotated text datasets in political science; natural target for (high-accuracy) NLP

Text Regression

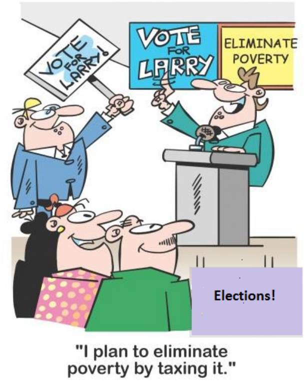
- Task = predict real-valued target variable from text, e.g.:
  - box office revenue from movie reviews [Joshi et al., 2010, Bitvai and Cohn, 2015b]
  - risk assessment from financial report [Kogan et al., 2009]
  - loan amount from micro-finance loan application [Bitvai and Cohn, 2015a]
- (Surprisingly?) understudied problem in NLP

Talk Outline

- Computational political science + text regression = "demographic regression" ... in form of two regression problems from the political science domain:
  1. predict the political leaning of a party on the basis of its manifesto [Subramanian et al., 2018]
  2. predict the popularity/outcome of an online petition [Subramanian et al., to appear]
- In both cases, various data quirks make for interesting modelling challenges

Talk Outline

Manifestos



- Critical artefact for understanding party positions in a representative democracy
- Comparative Manifesto Project (CMP) = large-scale collection of digitised manifestos [Merz et al., 2016]
  - long documents, across a wide array of languages
- Two major tasks:
  1. analysing party positions on various policy issues (fine-grained)
  2. for quantifying a party's position on the left–right spectrum (coarse-grained)

Source(s): `https://www.cartoonstock.com/cartoonview.asp?catref=rbin20`

Green Party UK, 2015

Invest in an £85 billion public programme of renewable electricity generation, flood defences and building insulation. Support local sustainable agriculture, with respect for animals and wild places. Cut emissions by providing cheaper public transport and encouraging cycling and walking. End privatisation in the National Health Service, provide proper funding and free social care for the elderly. Double Child Benefit.

Green Party UK, 2015

Invest in an £85 billion public programme of renewable electricity generation, flood defences and building insulation. Support local sustainable agriculture, with respect for animals and wild places. Cut emissions by providing cheaper public transport and encouraging cycling and walking. End privatisation in the National Health Service, provide proper funding and free social care for the elderly. Double Child Benefit.

Policy issue ($\times 56$, under 7 major policy categories) + Positions ( )

Green Party UK, 2015

Invest in an £85 billion public programme of **renewable electricity generation**, flood defences and building insulation. \ *Support* local **sustainable agriculture**, with respect for animals and wild places. \ *Cut* **emissions** by providing cheaper public transport and encouraging cycling and walking. \ End privatisation in the **NHS**, *provide proper funding* and free social care for the elderly. \ *Double* **Child Benefit**.

Environmental Protection Welfare State

Green Party UK, 2015

Invest in an £85 billion public programme of **renewable electricity generation**, flood defences and building insulation. \ *Support* local **sustainable agriculture**, with respect for animals and wild places. \ *Cut* **emissions** by providing cheaper public transport and encouraging cycling and walking. \ End privatisation in the **NHS**, *provide proper funding* and free social care for the elderly. \ *Double* **Child Benefit**.

Environmental Protection 👍 Welfare State 👍

Green Party UK, 2015

Invest in an £85 billion public programme of **renewable electricity generation**, flood defences and building insulation. \ *Support* local **sustainable agriculture**, with respect for animals and wild places. \ *Cut* **emissions** by providing cheaper public transport and encouraging cycling and walking. \ End privatisation in the **NHS**, *provide proper funding* and free social care for the elderly. \ *Double* **Child Benefit**.

Environmental Protection 👍 Welfare State 👍

- "smoothness" within document, in terms of contiguous regions of different classes, and transitions between classes

Manifesto Positioning

RILE [Volkens et al., 2013]

- Political scientists map fine-grained positions to left/right/neutral
- RILE is the difference between the count of **right-** and **left-leaning** sentences in a manifesto ($R - L$).

- For example:
  - Military = left
  - Military = right
  - Environmental protection = neutral.
- Used for **document-level** supervision only, as silver-standard heuristic

Manifesto Positioning

CHES

- A more rigorous means of assessing the political learning of a party is expert surveys (e.g. the Chapel Hill Expert Survey = "CHES") — context + time dependent, but asynchronous with elections (run every 4 years)

- Used for **document-level** evaluation only, as gold-standard

Data Setting: It's Complicated

- For all manifestos, we have a RILE score ... but only used for supervision/hyperparameter tuning
- For *some* training manifestos, we have sentence-level labels (despite the fact that, yes, RILE presupposes sentence-level annotation ...)
- Sentence-level labels in form of 7 coarse-grained and 56 fine-grained classes; RILE score directly derived from (subset of) sentence labels
- For recent elections (post 1999), we have CHES scores
- Manifestos in many different languages
- Structure of manifestos varies considerably

Contributions

1. Hierarchical joint-model
   - joint-structured objective capturing task dependencies
   - leverages additional document-level annotations for sentence-level task
2. Use context and temporal dependencies to smooth estimates
   - improved document-level predictions using Probabilistic Soft Logic, especially wrt CHES

Proposed Approach

Two-Stage Approach

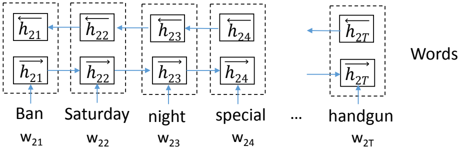
1. Hierarchical bi-LSTM to model sentence-level classification and document-level regression tasks
   - objective captures task dependency and consistency
   - leverage additional document-level annotations for sentence-level task
2. Manifesto position recalibration: Probabilistic Soft Logic (PSL) to incorporate context and temporal dependencies

Proposed Approach

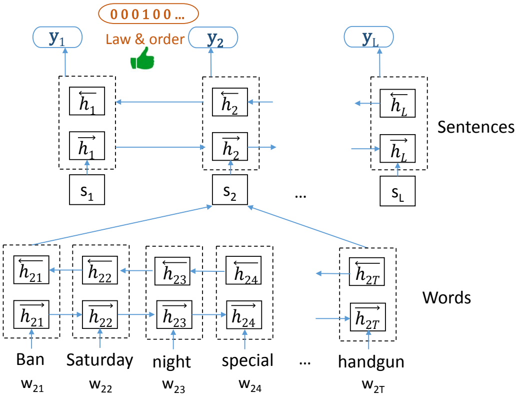
Two-Stage Approach

1. Hierarchical bi-LSTM to model sentence-level classification and document-level regression tasks
   - objective captures task dependency and consistency
   - leverage additional document-level annotations for sentence-level task
2. Manifesto position recalibration: Probabilistic Soft Logic (PSL) to incorporate context and temporal dependencies

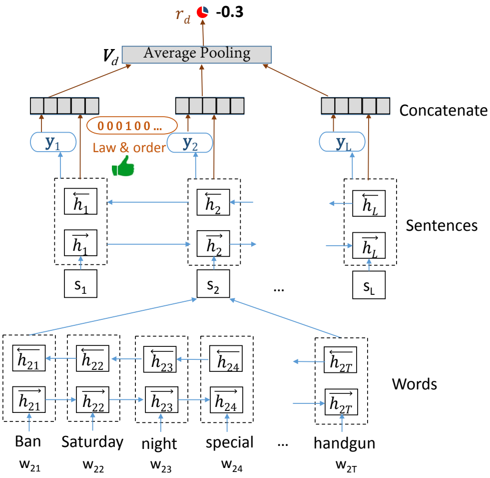
Hierarchical bi-LSTM



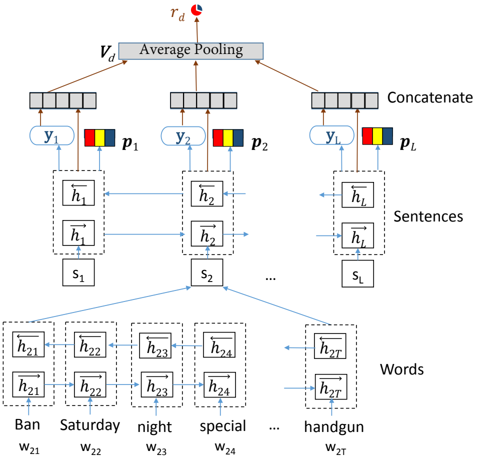
Hierarchical bi-LSTM



Hierarchical bi-LSTM



Hierarchical bi-LSTM (Dependency + Consistency)



Recap: Objective Functions

1. Single-task loss, based on sentence-level classification [**Joint**$_{sent}$]
2. Single-task loss, based on document-level regression [**Joint**$_{doc}$]
3. Multi-task loss, which captures the dependency between sentence-level classification and document-level regression [**Joint**]
4. Multi-task loss, which captures the dependency between sentence-level classification and document-level regression, with additional term for deviation between sentence- and document-level positions [**Joint**$_{struc}$]

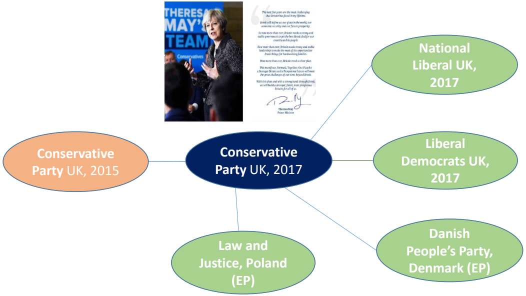
$$\mathcal{L}_{struc} = \frac{1}{|D|} \sum_{d=1}^{|D|} \left(\frac{1}{L_d} \sum_{i \in d} (p_{i_{right}} - p_{i_{left}}) - r_d \right)^2$$

Proposed Approach

Two-Stage Approach

1. Hierarchical bi-LSTM to model sentence-level classification and document-level regression tasks
   - objective captures task dependency and consistency
   - leverage additional document-level annotations for sentence-level task
2. Manifesto position recalibration: Probabilistic Soft Logic (PSL) to incorporate context and temporal dependencies

Probabilistic Soft Logic



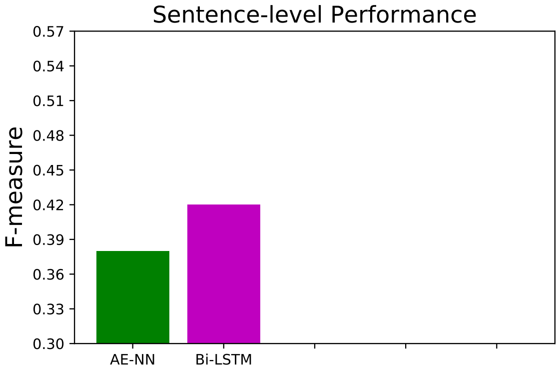
- Temporal dependency [temp]
- Coalition (regional vs. European) [coal]
- Manifesto similarity [**sim**]
- Location-weighted Right vs. Left [**sloc**] — local feature

Experimental Setting

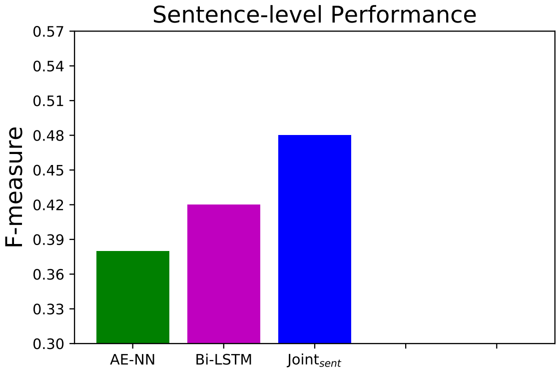
Dataset

- 1004 manifestos from 12 European countries, written in 10 different languages
  - Denmark, Netherlands, Ireland, United Kingdom, Finland, France, Austria, Germany, Italy, Portugal, Spain, Sweden
- 272 annotated with sentence-level labels
  - 320k (out of 650k) annotated sentences

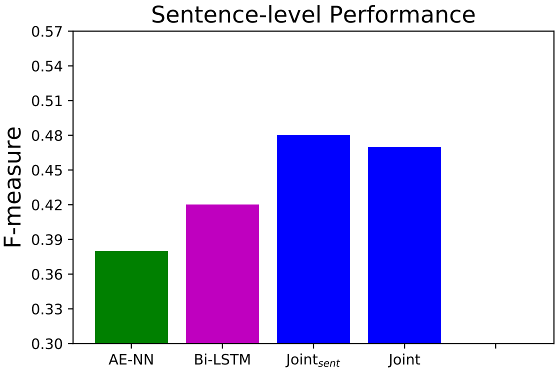
Sentence-level Results



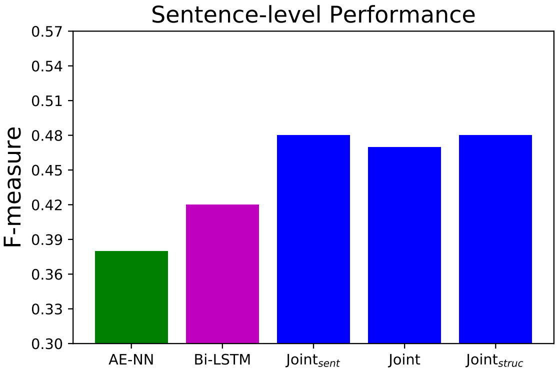
Sentence-level Results



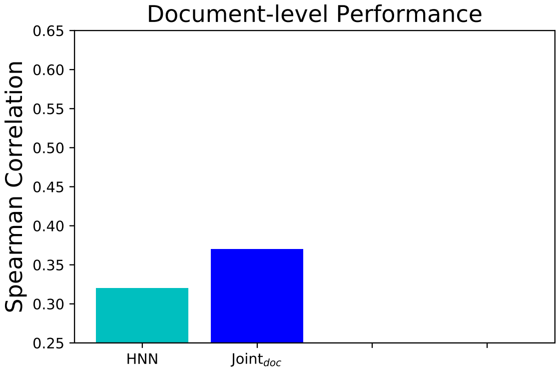
Sentence-level Results



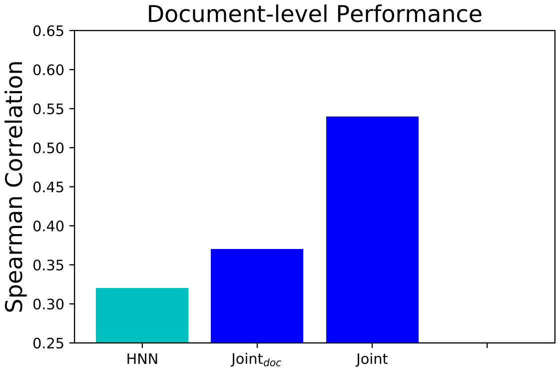
Sentence-level Results



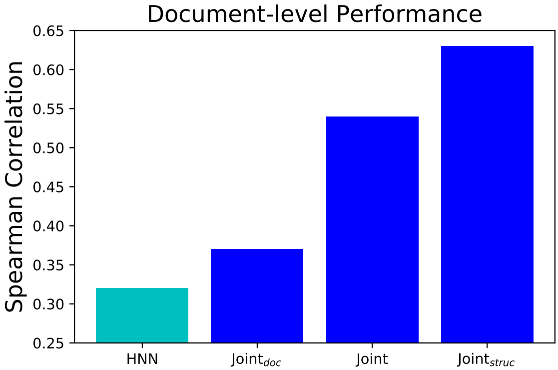
Sentence-level Results (RILE)



Sentence-level Results (RILE)



Sentence-level Results (RILE)



Sentence-level Performance — Reduced Training Data

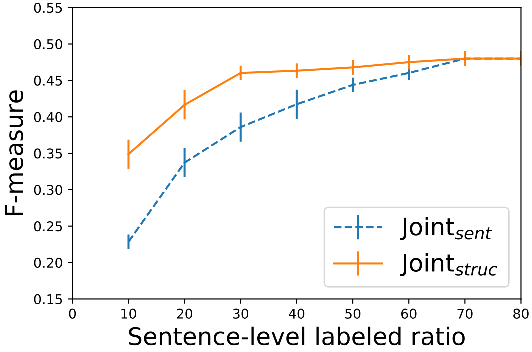
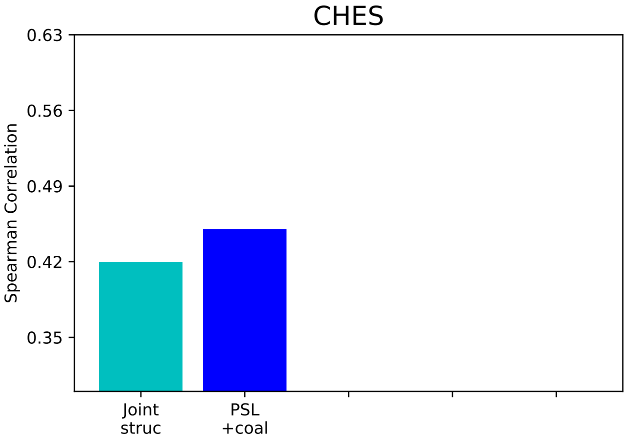
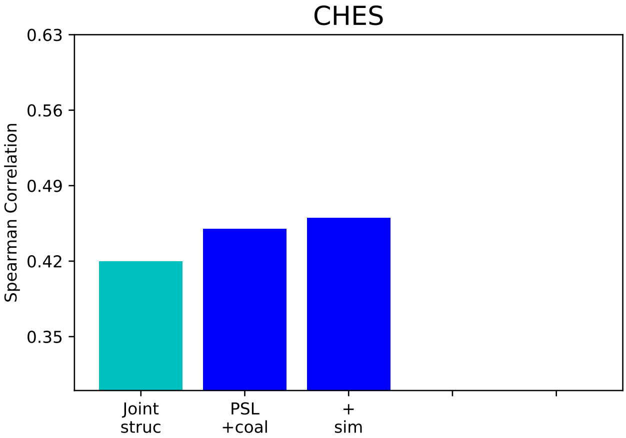


Figure: Varying the ratio of documents with sentence-level annotations

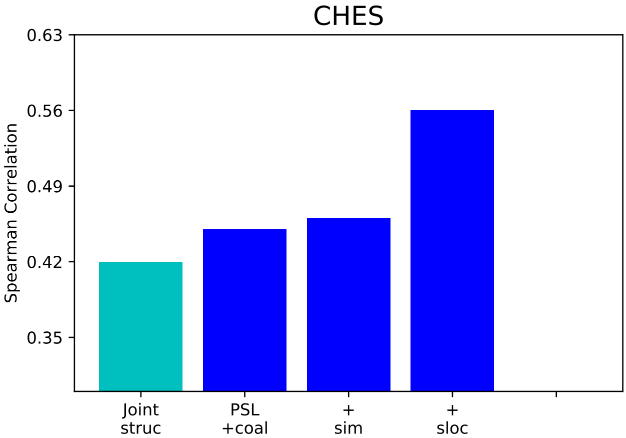
Document-level Results (CHES) — PSL



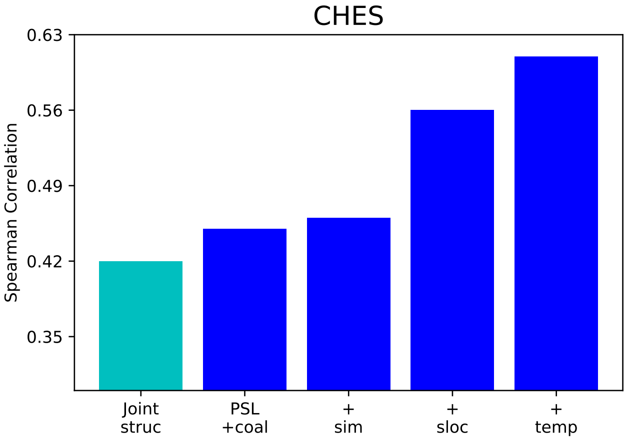
Document-level Results (CHES) — PSL



Document-level Results (CHES) — PSL



Document-level Results (CHES) — PSL

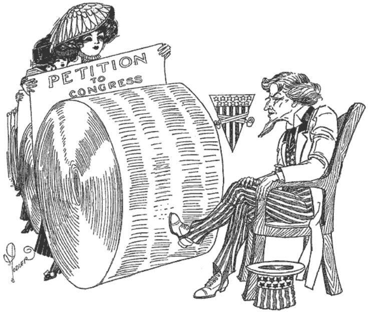


Conclusions

- Task = predict political leaning of party on the basis of their political manifesto
- Modeling the structure in text; task dependency and consistency boost performance
- Incorporating domain knowledge with Probabilistic Soft Logic provides significant gains
- Applicable to other multi-grain tasks such as stance classification
- PSL has applications for other computational social science tasks to encode domain information

Talk Outline

Petitions



- Instrument for "advocacy democracy", to request change/action of government or public entity
- Focus on government-run petition sites, in the UK and USA, with critical property that once signatures have crossed a certain threshold, a government response is guaranteed
- Two tasks:
  1. predict raw popularity of a given petition from its textual content
  2. predict whether a petition will be popular enough to elicit response

Source(s): `https://www.pinterest.com/pin/265853184226522088/`

Background: UK Parliament Petitions ("UK Petitions")

- Run by UK government, and open to all British citizens and UK residents
- Principal content: title, body, "more details"
- Guaranteed government response if petition gets more than 10k signatures
- Debate in House of Commons if petition gets more than 100k signatures
- Notable petitions:
  - require referendums to have supermajority
  - ban Trump from entering the UK ($\times 2$)

Petitions

UK Government and Parliament

148
petitions got a response from the Government

24
petitions were debated in the House of Commons

Search petitions

Popular petitions

Reject calls to add Staffordshire Bull Terriers to the Dangerous Dogs Act

961 signatures in the last hour

Prevent avoidable deaths by making autism/learning disability training mandatory

176 signatures in the last hour

anybody caught carrying a knife/gun serve a mandatory life with 10 year minimum

121 signatures in the last hour

View all open petitions

Background: We the People ("US Petitions")

- Run by White House
- Principal content: "type", title, description, tags
- Guaranteed response if petition gets more than 100k signatures within 30 days
- Deleted/unsearchable if $<$ 150 signatures
- Notable petitions:
  - create a Death Star
  - gun control (after Newtown mass-shooting)
- Semi-decommissioned under current administration

Petition the White House on the Issues that Matter to You

Create a Petition

How Petitions Work

1

Create a Petition

Call on the White House to take action on the issue that matters to you.

2

Gather Signatures

Share your petition with others, build a community for the change you want to make.

3

100,000 Signatures in 30 Days

Get an official update from the White House within 60 days.

MORE ON HOW IT WORKS

Basic Methodology

- Use CNN [Kim, 2014], with max-pooling layer tanh, and final MLP with exponential linear unit activation
- Inputs = GloVe pre-trained word embeddings [Pennington et al., 2014], with updating
- Training: minimise MSE over log-transformed vote counts

Auxiliary Loss

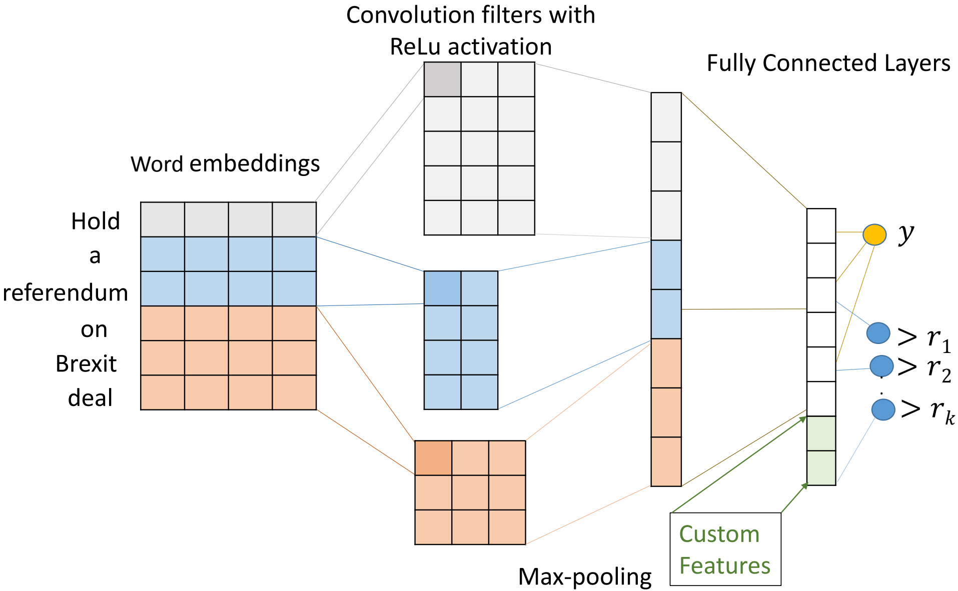
- Optionally include auxiliary ordinal regression loss, capturing "step function" actions associated with petitions, to sensitise model to critical numeric break points
  - UK Petitions: $O = \{10, 100, 1000, \mathbf{10000}, \mathbf{100000}\}$
  - US Petitions: $O = \{1000, 10000, \mathbf{100000}\}$
- Transform ordinal regression problem into series of binary classification problems, one for each threshold $\in O$, each with a sigmoid activation

Hand-engineered Features

- Additionally experiment with side information, via hand-engineered features commonly seen in computational social science research to: (a) determine what sorts of features the model is able to (latently) learn; and (b) assess complementarity with learned representations
- Feed in to model via single tanh layer, and concatenate with document representation from CNN

Summary of Features

- Additional Information (ADD): does the petition have additional details?
- Ratio of indefinite (IND) and definite (DEF) articles
- Ratio of 1P-sing pronouns (FSP), 1P-plural pronouns (FPP), 2P pronouns (SPP), 3P-sing pronouns (TSP), and 3P-plural pronouns (TPP)
- Ratio of subjective words (SUB) and difference between count of +ve and -ve words (POL)
- Ratio of biased words (BIAS) from bias lexicon [Recasens et al., 2013]
- POS class features
- Number of named entities
- Freshness (FRE): average weighted cos similarity with previous petitions
- Policy category popularity score (CSC): commonality of the petition's policy issue
- Political bias and polarity: relative leaning/polarity based on RILE scores

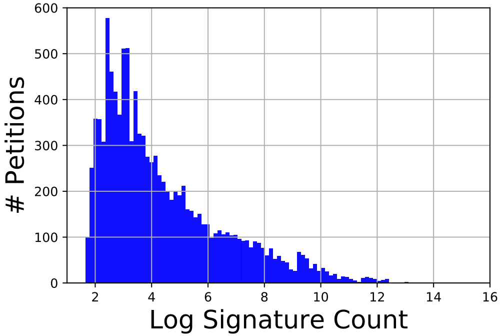


Convolution filters with ReLu activation
Fully Connected Layers
Word embeddings
Hold
a
referendum
on
Brexit
deal
y
$> r_1$
$> r_2$
$> r_k$
Custom Features
Max-pooling

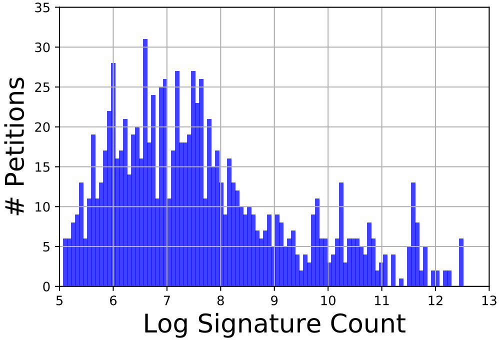
Datasets

- UK Petitions:
  - 10950 petitions, with 31m+ signatures (2011–2017)
- US Petitions
  - 1023 petitions, with 12m+ signatures (2011–2017)
- In each case, split the data chronologically into train/dev/test splits based on 80/10/10 breakdown

Dataset Distribution



UK Petitions



US Petitions

Results

Approach	UK Petitions		US Petitions	
	MAE	MAPE	MAE	MAPE
SVR_{BoW}	1.53	45.35	1.39	20.37
SVR_{feat}	1.54	46.96	1.40	20.48
$\text{SVR}_{\text{BoW+feat}}$	1.52	44.71	1.39	20.38
$\text{CNN}_{\text{regress}}$	1.44	36.72	1.24	14.98
$\text{CNN}_{\text{regress+ord}}$	1.42	33.86	1.22	14.68
$\text{CNN}_{\text{regress+ord+feat}}$	1.41	32.92	1.20	14.47
$\text{CNN}_{\text{regress+feat}}$	1.43	35.84	1.23	14.75
$\text{CNN}_{\text{regress+ord+feat}}$ + Additional hidden layer	**1.40**	**31.68**	**1.16**	**14.38**

$$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i}$$

Feature Analysis

		UK Petitions		US Petitions	
Feature	Description	p	p_{hidden}	p	p_{hidden}
ADD	Additional details	***			
IND	Indefinite articles	*		*	*
DEF	Definite articles	***	*		
SUBJ	Subjective words	*		*	***
POL	Polarity		*		
BIAS	Biased words	*		*	
NNC	Nouns		***		**
VBC	Verbs		**	*	***
ADC	Adjectives	***	***		
NEC	Named entities	***	***		*
FRE	Freshness	***	*	**	*
CSC	Policy category popularity	***			
PBIAS	Political bias			**	
L–R	Left–right scale	*		**	

Future Work

- Petitions aren't authored/signed in a vacuum (e.g. Newtown mass-shooting, Trump referendum), and blindingly obvious that we need world context ... but what/how?
- More modelling of time series of signatures to model site popularity/scaling factor, seasonality, ...

Summary

- Task of predicting the popularity (in signatures) of an online petition via text regression
- Data from official UK and US petition sites
- Text regression model, with optional auxiliary ordinal regression loss and side information in the form of hand-engineered features ... both of which led to minor improvements
- Code and data available at:

  `http://github.com/shivashankarrs/Petitions`

Acknowledgements

- Joint work with Shivashankar Subramanian and Trevor Cohn
- This work was supported by the Australian Research Council


Springfield Shopper
DAILY NEWS
50¢
SIMPSON WINS IN LANDSLIDE
Says "Crazy Promises" The Key

Thank You!

References

Zsolt Bitvai and Trevor Cohn. Predicting peer-to-peer loan rates using Bayesian non-linear regression. In *Proceedings of the 29th Annual Conference on Artificial Intelligence (AAAI-15)*, pages 2203–2209, Austin, USA, 2015a.

Zsolt Bitvai and Trevor Cohn. Non-linear text regression with a deep convolutional neural network. In *Proceedings of the Joint conference of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing of the Asian Federation of Natural Language Processing (ACL-IJCNLP 2015)*, pages 180–185, Beijing, China, 2015b.

Clinton Burfoot, Steven Bird, and Timothy Baldwin. Collective classification of congressional floor-debate transcripts. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011)*, pages 1506–1515, Portland, USA, 2011.

Jessica Elan Chung and Eni Mustafaraj. Can collective sentiment expressed on Twitter predict political elections? In *Proceedings of the 25th Annual Conference on Artificial Intelligence (AAAI-11)*, pages 1770–1771, San Francisco, USA, 2011.

Justin Grimmer. A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in Senate press releases. *Political Analysis*, 18(1):1–35, 2010.

Ivan Habernal and Iryna Gurevych. Exploiting debate portals for semi-supervised argumentation mining in user-generated web discourse. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015)*, pages 2127–2137, Lisbon, Portugal, 2015.

References

Mahesh Joshi, Dipanjan Das, Kevin Gimpel, and Noah A. Smith. Movie reviews and revenues: An experiment in text regression. In *Proceedings of Human Language Technologies: The 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL HLT 2010)*, pages 293–296, Los Angeles, USA, 2010.

Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1746–1751, Doha, Qatar, 2014.

Shimon Kogan, Dimitry Levin, Bryan R Routledge, Jacob S. Sagi, and Noah A. Smith. Predicting risk from financial reports with regression. In *Proceedings of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies 2009 (NAACL HLT 2009)*, pages 272–280, Boulder, USA, 2009.

Robert Malouf and Tony Mullen. Taking sides: Graph-based user classification for informal online political discourse. *Internet Research*, 18(2):177–190, 2008.

Nicolas Merz, Sven Regel, and Jirka Lewandowski. The manifesto corpus: A new resource for research on political parties and quantitative text analysis. *Research & Politics*, 3(2), 2016.

Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1532–1543, Doha, Qatar, 2014.

References

Vinodkumar Prabhakaran, Ashima Arora, and Owen Rambow. Staying on topic: An indicator of power in political debates. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*, pages 1481–1486, Doha, Qatar, 2014.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL 2013)*, pages 1650–1659, 2013.

Shivashankar Subramanian, Trevor Cohn, and Timothy Baldwin. Hierarchical structured model for fine-to-coarse manifesto text analysis. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics — Human Language Technologies (NAACL HLT 2018)*, pages 1964–1974, New Orleans, USA, 2018.

Shivashankar Subramanian, Timothy Baldwin, and Trevor Cohn. Content-based popularity prediction of online petitions using a deep regression model. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL 2018)*, Melbourne, Australia, to appear.

Matt Thomas, Bo Pang, and Lillian Lee. Get out the vote: Determining support or opposition from congressional floor-debate transcripts. In *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP 2006)*, pages 327–335, Sydney, Australia, 2006.

References

Andranik Tumasjan, Timm Oliver Sprenger, Philipp G. Sandner, and Isabell M. Welpe. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM 2010)*, pages 178–185, Washington D.C., USA, 2010.

Andrea Volkens, Judith Bara, Budge Ian, and Simon Franzmann. Understanding and validating the left-right scale (RILE). In *Mapping Policy Preferences From Texts: Statistical Solutions for Manifesto Analysts*, chapter 6. Oxford University Press, 2013.