

Popularity Prediction of Online Petitions using a Multimodal Deep Regression Model

Kotaro Kitayama[♠]

Shivashankar Subramanian[♡]

Timothy Baldwin[♡]

♠ Tohoku University

♡ The University of Melbourne

kitayama@ecei.tohoku.ac.jp

{shivashankar.subramanian, tbaldwin}@unimelb.edu.au

Abstract

Online petitions offer a mechanism for people to initiate a request for change and gather support from others to demonstrate support for the cause. In this work, we model the task of petition popularity using both text and image representations across four different languages. We evaluate our proposed approach using a dataset of 75k petitions from Avaaz.org, and find strong complementarity between text and images.

1 Introduction

A petition is a formal request for change or an action to any authority, co-signed by a group of supporters (Ergazakis et al., 2012). The targets of petitions are usually government agencies and business organizations.

In this work we study petitions from Avaaz.org, a popular petition platform available across six continents, with support for seventeen different languages. Avaaz provides a platform for petitions and funding campaigns. An example petition is given in Figure 1, opposing the Anti-Counterfeiting Trade Agreement (ACTA) and supporting a free and open internet. The popularity of a petition, in terms of the number of signatures it attracts, is critical to its success, and predicting popularity can help petition organizers to enhance their engagement strategy by optimising the petition content. In particular in this work, we target the task of predicting petition popularity at the time of submission (independent of any social media or direct signature signal).

While existing work on petitions has focused on their text content (Elnoshokaty et al., 2016; Subramanian et al., 2018), images are also a key ingredient.¹ Additionally, despite petitions being popular in many different languages, there has been

¹https://secure.avaaz.org/en/community_petitions/how_to_create_petition/

no work on multilingual modeling. From a social science viewpoint, multilingual analysis can contribute to an understanding of issues present in different languages (or regions).

Previous research has shown that, other than petition content, metadata is also effective in modelling its popularity. Elnoshokaty et al. (2016) showed that the category of a petition has an influence on its popularity and success, e.g., *human trafficking* related petitions get more signatures than *health* related ones. Vidgen and Yasseri (2019) investigated the interaction between topics and geographic features, and showed that some issues receive broad geographic support (e.g., *law & order*, *work & play*) but others are far more local (e.g., *animals & nature*, *driving*). Since we aim to model petition popularity across multiple countries, we study the utility of the author’s country information.

Our contributions in this work are as follows: (1) we propose a multimodal regression approach for petition popularity prediction task using text and image features; (2) we experiment in both mono- and multi-lingual data settings, to evaluate the impact of training data from other languages; and (3) we develop a novel multimodal, multilingual dataset for the task.²

2 Related Work

The majority of work on modeling petition popularity has focused on predicting popularity growth over time based on an initial popularity trajectory (Hale et al., 2013; Yasseri et al., 2017; Proskurnia et al., 2017), e.g. given the number of signatures a petition gets in the first x hours, predicting the total number of signatures at the end of its lifetime. Since the popularity of a petition also depends on its author’s campaign strategies, Asher et al. (2017)

²The dataset and all code associated with this paper will be available at: <https://github.com/kkitayama/petitions-with-image>.

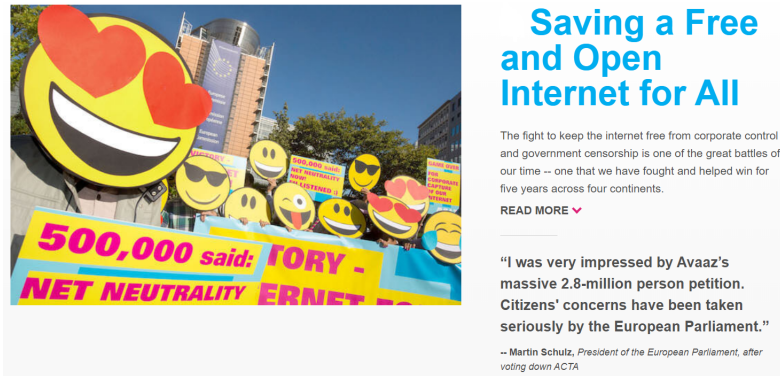


Figure 1: An example petition from Avaaz, which led to the European Parliament abandoning ACTA.

and Proskurnia et al. (2017) examined the impact of sharing petitions on Twitter, as a time series regression task. However, none of this work analyzed the petition’s content, which is a primary focus in this work, in addition to making the prediction at the time of *submission* rather than based on early social indicators.

Elnoshokaty et al. (2016) analyzed Change.org petitions by performing correlation analysis of popularity against the petition’s category, target goal set, and the distribution of words in General Inquirer categories (Stone et al., 1962). Subramanian et al. (2018) is the closest work to this paper, which targets UK and US government petitions, and poses popularity prediction as a text regression task using a convolutional neural network model.

Though petition platforms such as Avaaz.org and Change.org are popular in different countries and languages, almost all the existing work has focused on the monolingual setting (almost exclusively on English). With the increasing use of petitions across the globe, it is necessary to model petitions across different languages (Aragón et al., 2018). Lastly, in addition to textual content, the choice of images and other multimodal information has been shown to have impact on the popularity of social media posts (Meghawat et al., 2018; Wang et al., 2018; Chen et al., 2019), but not utilized in the context of online petitions.

3 Dataset

We use petitions from the Avaaz.org dataset of Aragón et al. (2018), which consists of over 360k petitions in more than fifty languages. Each petition is made up of its textual content, author details, country, count of shares on social media, language, and other metadata. For our work, we use the top four languages based on the raw count of petitions:

	Train	Dev	Test
English	14,262	1,800	1,800
Portuguese	21,888	2,700	2,700
French	13,647	1,700	1,700
Spanish	10,118	1,300	1,300

Table 1: Data split across languages

English, French, Portuguese, and Spanish. We extended the filtered dataset by crawling the image content for the petitions in those four languages from Avaaz.org. We removed petitions with empty content or where there is not a single majority language for all sentences (based on langid.py: Lui and Baldwin (2012)). The resulting dataset has a total of around 75k petitions, nearly 45% of the which have default images.³ The distribution across languages is given in Table 1.

4 Methodology

4.1 Model

We evaluate three classes of model: text-only, image-only, and combined text, image and meta-data. In each case, we regress over the petition signatures, and use fully-connected hidden layers with a ReLU activation function before the final output layer. Note that we log-transform the signature count, consistent with previous work (Elnoshokaty et al., 2016; Proskurnia et al., 2017; Subramanian et al., 2018).

4.1.1 Text-only model

We employ two different text-only model architectures: (1) a CNN regression model (Bitvai and

³https://avaazdo.s3.amazonaws.com/do_generic*

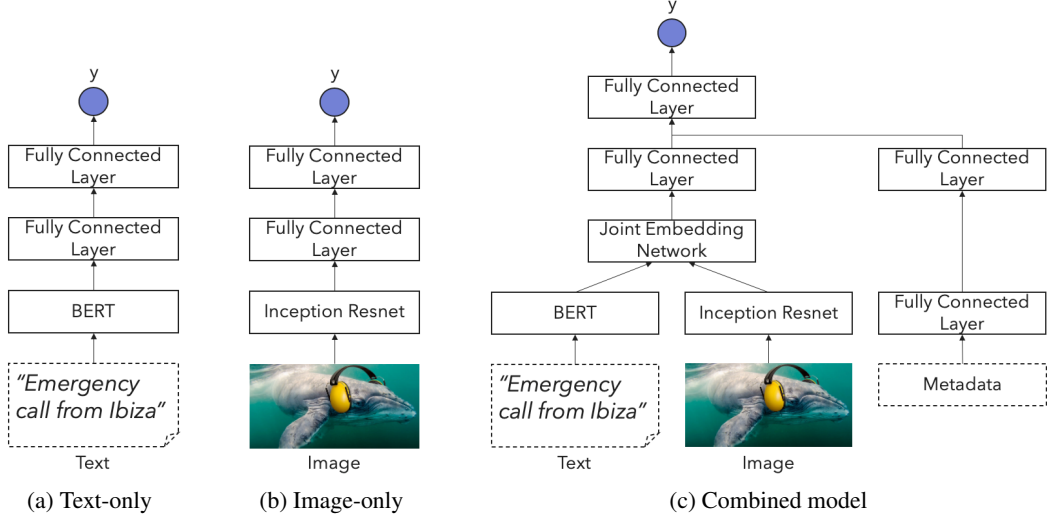


Figure 2: Overview of the models, where y denotes the signature count.

Cohn, 2015) based on the method of Subramanian et al. (2018); and (2) a BERT (Devlin et al., 2019) regression model, where the [CLS] encoding of the final layer is used as the text representation.

For our monolingual experiments over English petitions, for the CNN model we use GloVe (Pennington et al., 2014) word embeddings, and for BERT we use the pre-trained BERT-base English model (Devlin et al., 2019). For the multilingual experiments, we use the pre-trained multilingual BERT model (“mBERT”: Devlin et al. (2019)). An overview of the text model architecture is presented in Figure 2a.

4.1.2 Image-only model

For the image-only model, we use Inception-ResNet v2 (Szegedy et al., 2017) pre-trained on ImageNet, and extract the image representation from the penultimate layer. An overview of the image-only model is presented in Figure 2b.

4.1.3 Combined model

In the combined model, we use text, image, and metadata features, as detailed in Figure 2c, adopting the approach of Wang et al. (2018). We use text features extracted from the text-only model based on CNN or BERT, and image features from the image-only model. In both cases, we freeze all model parameters. Following that, text and image features are jointly embedded using a fully connected layer, a ReLU activation layer, a second fully connected layer, a batch normalization layer, and an L2 normalization layer (referred to as Joint Embedding Network in Figure 2c). Finally, the joint embedding and metadata embeddings are

combined together using a fully connected layer.

As metadata, we use author’s country information from the original dataset, encoded as a one-hot vector. Although the original data includes other metadata such as social media likes, we do not use it as it would not be available at the time of authoring.

4.2 Loss

We evaluate two types of regression loss functions. First we employ mean squared error in log-space (“MSLE”) as used by Subramanian et al. (2018), and calculated as:

$$\frac{1}{N} \sum_{i=1}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2$$

where y_i is the actual signature count and \hat{y}_i is the predicted signature count. Second, we use mean absolute percentage error, again in log-scale (“MAPE”), jointly with MSLE. MAPE helps to capture the deviation between actual and predicted values, relative to the actual ones. An intuitive reason to use MAPE is to directly capture the expected model behavior consistent with the evaluation metric (see Section 5). The joint loss is computed as follows:

$$\frac{1}{N} \sum_{i=1}^N (\log(y_i + 1) - \log(\hat{y}_i + 1))^2 + k \times 100 \times \frac{1}{N} \sum_{i=1}^N \left| \frac{\log(y_i + 1) - \log(\hat{y}_i + 1)}{\log(y_i + 1)} \right|$$

where y_i is the actual signature count and \hat{y}_i is the predicted signature count; k is a hyper-parameter,

	MSLE		MSLE + MAPE	
	MAPE↓	ρ ↑	MAPE↓	ρ ↑
IMG	50.6	0.235	37.1	0.200
TXT _{CNN}	40.9	0.363	36.0	0.354
TXT _{CNN} +IMG	39.7	0.392	35.6	0.375
TXT _{CNN} +IMG +CN	39.2	0.381	35.9	0.360
TXT _{BERT}	42.4	0.385	35.1	0.375
TXT _{BERT} +IMG	40.7	0.405	35.4	0.388
TXT _{BERT} +IMG +CN	41.1	0.397	34.9	0.403

Table 2: Monolingual English results; best results in bold. “TXT” = text, “IMG” = image, “CN” = country.

	MSLE		MSLE + MAPE	
	MAPE↓	ρ ↑	MAPE↓	ρ ↑
IMG	49.2	0.234	36.7	0.212
TXT _{mBERT}	46.8	0.277	35.8	0.239
TXT _{mBERT} +IMG	43.6	0.306	36.3	0.298
TXT _{mBERT} +IMG +CN	44.0	0.305	36.8	0.300

Table 3: Multilingual results (all languages); best results in bold. “Txt” = text, “IMG” = image, “CN” = country.

tuned on the development data.

5 Experiments

5.1 Settings

We use the predefined training/validation/test random splits (Table 1) for the experiments. We use mean absolute percentage error (MAPE) and Spearman’s rank correlation as evaluation metrics, as commonly used by other popularity prediction tasks (Subramanian et al., 2018; Wang et al., 2018). For MAPE, lower is better, and a perfect system will score 0%; and for Spearman’s rank correlation (“ ρ ”) higher is better, and a perfect system will score 1.0.

For the CNN model, we set the number of filters to 100 and the dimensionality of the fully connected layer to 100, and for BERT monolingual and multilingual models, we set the dimensionality of the hidden layers to 300 and 500, respectively. For the image-only model (IMG), we set the dimensionality of the fully connected layer to 500 and 200 in the monolingual and multilingual models, respectively. All hyper-parameters were tuned on the development data.

5.2 Results

First, we evaluate the models using English data, and present the results in Table 2. For text-based

	MSLE		MSLE + MAPE	
	MAPE↓	ρ ↑	MAPE↓	ρ ↑
English	48.7	0.246	35.6	0.276
Portuguese	47.2	0.280	35.6	0.276
French	43.4	0.331	35.9	0.258
Spanish	47.6	0.199	35.2	0.170

Table 4: Multilingual results (per language), based on TXT_{mBERT} +IMG

modelling (“TXT”), we use CNN and BERT. For image-based modeling (“IMG”), we use ResNet encodings. From the results, it is evident that TXT is more discriminative than IMG, which provides the best standalone performance. But the combined text and image model performs better than the text-only model. In terms of the two loss functions, Spearman’s ρ is largely the same as with simple MSLE, but the combined MSLE + MAPE loss predictably leads to substantial improvements in MAPE, especially for the image-only model. The inclusion of country data (“+CN”) leads to marginal improvements.

In the multilingual setting, we use multilingual BERT (“mBERT”) to represent text for all the languages. mBERT trains a single BERT model for over 100 languages with a large shared vocabulary (Devlin et al., 2019). Employing monolingual BERT trained on each language, as well as using cross-lingual language models (Conneau and Lample, 2019) are valid alternate approaches, which are left for future work. Here, we observe a similar overall trend where text-only models are slightly better than image-only models, and despite the almost fourfold increase in training data, the absolute results are worse than the monolingual results (Table 2) in the case of English in Table 4. In terms of loss, the performance for French with MSLE is superior to other languages. Lastly, Spearman’s ρ for Spanish is quite a bit lower than the other languages, a result which requires further analysis.

6 Conclusions

We proposed an multimodal, multilingual approach to the task of petition popularity prediction, and found that while the text-only model was superior to the image-only model, the combination of multimodal features performed the best. Exploring further choices of metadata, and alternate ways to model multilingual text is left to future work.

References

- Pablo Aragón, Diego Sáez-Trumper, Miriam Redi, Scott A. Hale, Vicenç Gómez, and Andreas Kaltenbrunner. 2018. Online petitioning through data exploration and what we found there: A dataset of petitions from avaaz.org. In *ICWSM-18 - 12th International AAAI Conference on Web and Social Media*.
- Molly Asher, Cristina Leston Bandeira, and Viktoria Spaiser. 2017. Assessing the effectiveness of e-petitioning through Twitter conversations. *Political Studies Association (UK) Annual Conference*.
- Zsolt Bitvai and Trevor Cohn. 2015. Non-linear text regression with a deep convolutional neural network. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 180–185.
- Junhong Chen, Dayong Liang, Zhanmo Zhu, Xiaojing Zhou, Zihan Ye, and Xiuyun Mo. 2019. Social media popularity prediction based on visual-textual features with xgboost. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2692–2696.
- Alexis Conneau and Guillaume Lample. 2019. Cross-lingual language model pretraining. In *Advances in Neural Information Processing Systems*, pages 7059–7069.
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186.
- Ahmed Said Elnoshokaty, Shuyuan Deng, and Dong-Heon Kwak. 2016. Success factors of online petitions: Evidence from change.org. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 1979–1985.
- Kostas Ergazakis, Dimitrios Askounis, Panagiotis Kokkinakos, and Anastasios Tsitsanis. 2012. An integrated methodology for the evaluation of e-petitions. In *Empowering Open and Collaborative Governance*, pages 39–59. Springer.
- Scott A Hale, Helen Margetts, and Taha Yasseri. 2013. Petition growth and success rates on the UK No. 10 Downing Street website. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 132–138.
- Marco Lui and Timothy Baldwin. 2012. `langid.py`: An off-the-shelf language identification tool. In *Proceedings of the ACL 2012 System Demonstrations*, pages 25–30, Jeju Island, Korea. Association for Computational Linguistics.
- Mayank Meghawat, Satyendra Yadav, Debanjan Mahata, Yifang Yin, Rajiv Ratn Shah, and Roger Zimmermann. 2018. A multimodal approach to predict social media popularity. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)*, pages 190–195.
- Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1532–1543.
- Julia Proskurnia, Przemyslaw A. Grabowicz, Ryota Kobayashi, Carlos Castillo, Philippe Cudré-Mauroux, and Karl Aberer. 2017. Predicting the success of online petitions leveraging multidimensional time-series. In *Proceedings of the 26th International Conference on World Wide Web*, pages 755–764.
- Philip J. Stone, Robert F. Bales, J. Zvi Namenwirth, and Daniel M. Ogilvie. 1962. The General Inquirer: A computer system for content analysis and retrieval based on the sentence as a unit of information. *Systems Research and Behavioral Science*, 7(4):484–498.
- Shivashankar Subramanian, Timothy Baldwin, and Trevor Cohn. 2018. 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 (Volume 2: Short Papers)*, pages 182–188.
- Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. 2017. Inception-v4, Inception-ResNet and the impact of residual connections on learning. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 4278–4284.
- Bertie Vidgen and Taha Yasseri. 2019. What, when and where of petitions submitted to the UK government during a time of chaos. *arXiv preprint arXiv:1907.01536*.
- Ke Wang, Mohit Bansal, and Jan-Michael Frahm. 2018. Retweet wars: Tweet popularity prediction via dynamic multimodal regression. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1842–1851.
- Taha Yasseri, Scott A Hale, and Helen Z Margetts. 2017. Rapid rise and decay in petition signing. *EPJ Data Science*, 6(1):1–13.