

A Dataset for Troll Classification of Tamil Memes

A World
Leading SFI
Research
Centre



Insight

SFI RESEARCH CENTRE FOR DATA ANALYTICS

Shardul Suryawanshi, Bharathi Raja Chakravarthi,
Pranav Varma, Mihael Arcan, John P. McCrae, Paul
Buitelaar
Data Science Institute,
National University of Ireland Galway
(shardul.suryawanshi@insight-centre.org)

HOST INSTITUTIONS



PARTNER INSTITUTIONS



FUNDED BY:



Outline

- Troll Meme
- Challenges
- Dataset Annotation
- Experimental Setup
- Methodology
- Result
- Conclusion and Future work

Troll Meme

- Troll meme contains
 - offensive text and non-offensive images
 - offensive images with non-offensive text
 - sarcastically offensive text with non-offensive images
- It provokes, distracts, and has a digressive or off-topic content
- and intends to demean or offend an individual or a group.

ஒரு Packet காற்று வாங்கினால்



facebook.com/Trollpettai

ஐந்து Chips முற்றிலும் இலவசம்

Translation: "If you buy one packet of air, then 5 chips free"

Challenges: Context

- Same image but different text



Translation: "can not understand what you are saying"



Translation: "I am confused"

Challenges: Data imbalances and Low Resource

- After collection, number of troll memes were more than not-troll memes
- Hence, added images from Flickr [1] dataset in not-troll category
- Due to lesser data, we used ImageNet weights for fine tuning



Example from Flickr dataset

[1] <https://www.kaggle.com/hsankesara/flickr-image-dataset>

Challenges: Emotional Toll on Annotators

- Voluntary annotators were onboarded
- To reduce the burden of annotation, annotators were allowed to leave at their will

Dataset Annotation

- Amongst several volunteers, only native Tamil speakers were selected
- Substantial agreement between annotators (Cohen's kappa = 0.62)
- Data Statistics
 - Total memes: 2,969
 - # troll: 1,951
 - # not-troll: 1,018

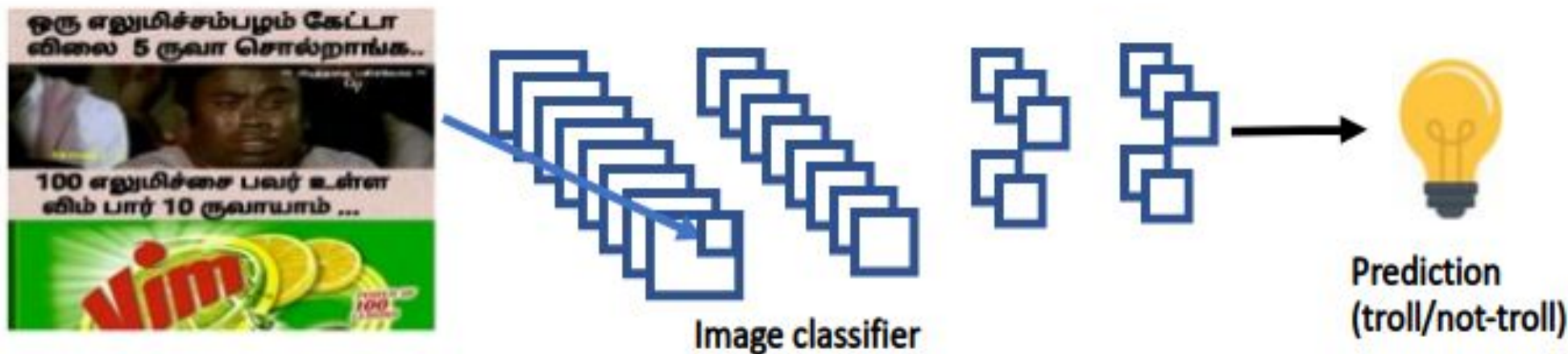
Experimental Setup

- ResNet and MobileNet classifier trained on
 - Imbalanced dataset
 - *TamilMemes*
 - *TamilMemes + ImageNet**
 - *TamilMemes + ImageNet* + Flickr30k*
 - Balanced dataset
 - *TamilMemes + ImageNet* + Flickr1k*

(*pre-trained on ImageNet weights)

Methodology

- Benchmark results using convolutional neural network (CNN) for image classification.



Result: ResNet [2]

variation	TamilMemes			TamilMemes + ImageNet		
	Precision	Recall	F1-score	Precision	Recall	F1-Score
troll	0.37	0.33	0.35	0.36	0.35	0.35
not-troll	0.68	0.71	0.70	0.68	0.69	0.68
macro-avg	0.52	0.52	0.52	0.52	0.52	0.52
weighted-avg	0.58	0.58	0.58	0.57	0.57	0.57
	TamilMemes + ImageNet + Flickr1k			TamilMemes + ImageNet + Flickr30k		
troll	0.30	0.34	0.32	0.36	0.35	0.35
not-troll	0.64	0.59	0.62	0.68	0.69	0.68
macro-avg	0.47	0.47	0.47	0.52	0.52	0.52
weighted-avg	0.53	0.51	0.52	0.52	0.52	0.52

[2] He, Kaiming, et al. "Deep residual learning for image recognition." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Result: MobileNet [3]

variation	TamilMemes			TamilMemes + ImageNet		
	Precision	Recall	F1-score	Precision	Recall	F1-Score
troll	0.28	0.27	0.28	0.34	0.43	0.38
not-troll	0.64	0.66	0.65	0.67	0.58	0.62
macro-avg	0.46	0.46	0.46	0.50	0.51	0.50
weighted-avg	0.52	0.53	0.52	0.56	0.53	0.54
	TamilMemes + ImageNet + Flickr1k			TamilMemes + ImageNet + Flickr30k		
troll	0.33	0.55	0.41	0.31	0.34	0.33
not-troll	0.66	0.45	0.53	0.65	0.62	0.64
macro-avg	0.50	0.50	0.47	0.48	0.48	0.48
weighted-avg	0.55	0.48	0.49	0.54	0.53	0.53

[3] Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).

Overall Results

- **Macro averaged F1-score** with or without data imbalance ranged from **0.47 to 0.58**
- Overall the precision for troll class identification lies in the range of 0.28 and 0.37
- ResNet is not hampered by imbalanced settings
- MobileNet shows poor performance in imbalanced settings

Conclusion and Future Work

- Image classifier does not give significant result
- Text embedded on meme gives it meaning
- This text is code-mixed with English
- It is challenging to train classifier just on the basis of image
- Rather same meme could be used in different context
- We plan to use OCR technique to capture textual data and treat this problem in multimodal way

Thank you !!

Questions?