EMNLP 2024

ArMeme: Propagandistic Content in Arabic Memes

Firoj Alam, Abul Hasnat, Fatema Ahmad, Md Arid Hasan, Maram Hasanain

{fialam,fakter,mhasanain}@hbku.edu.qa,
hasnat@blackbird.ai, arid.hasan@unb.ca

November 12, 2024

Motivation

- Memes have become a significant medium for cultural and political expression.
- They are a source of misleading information for audiences on social media.
- Research on medium- to low-resource languages is relatively limited.
- This research mainly focuses on Arabic memes and identifying propagandistic content across different modalities.

Motivation

Translation:

DR. (orange): do you think your wife is controlling you?

wife (white) : No, i don't think so

a) Propagandistic.

Translation:

It scares them that you are retarded.

You mean being different?

No you are retarded, and it is scaring us all.

b) Not-propagandistic.

Motivation

Translation:
when I enter the bathroom
everyone:
Let me in
LET ME IIIIIIIIIIIIIIIIIIIIN!

لما ادخل الحمام
كل الناس:

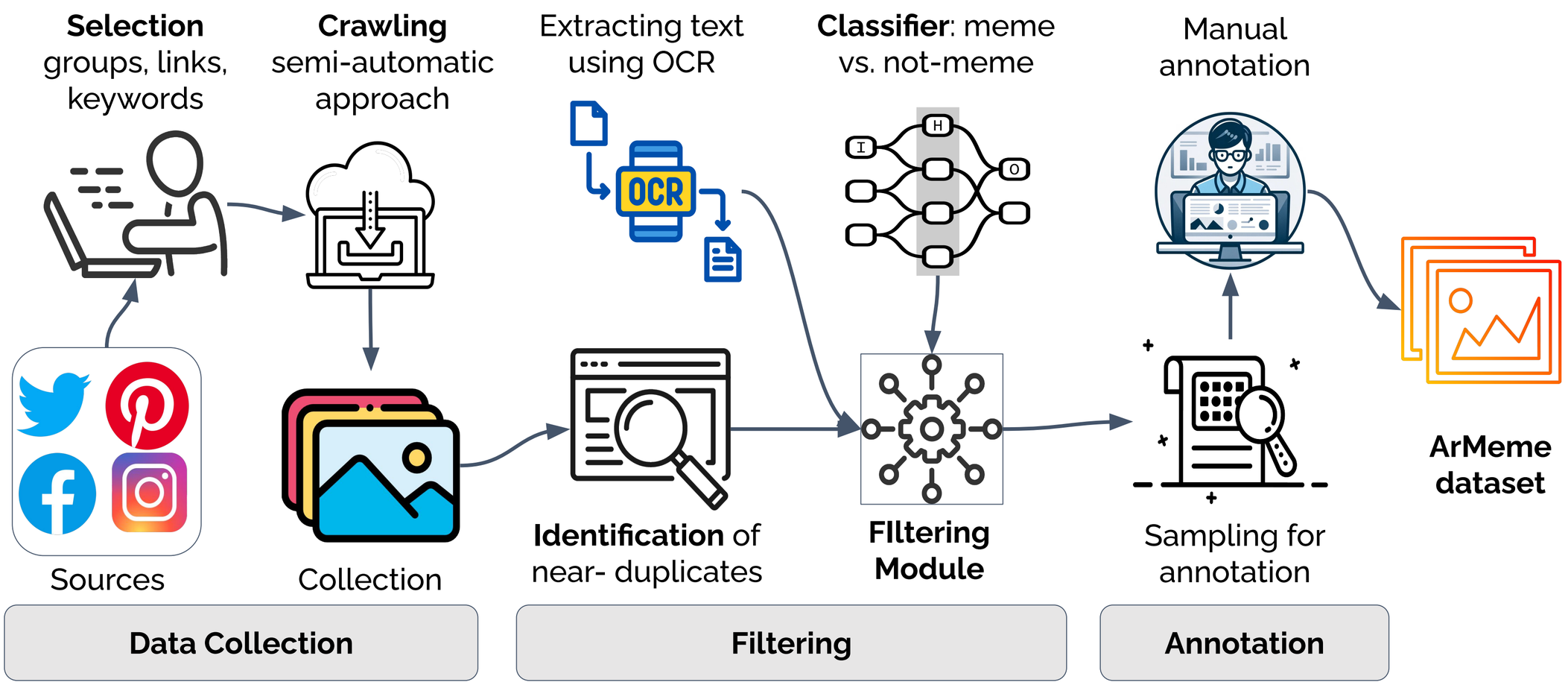
c) Other

d) Not-meme

Contributions

- The ***first Arabic meme dataset*** with manual annotations defining four categories.
- A detailed description of the data collection procedure, which can assist the community in future data collection efforts.
- An ***annotation guideline*** that will serve as a foundation for future research.
- Experiments:
  - **Text modality:** training classical models and fine-tuning monolingual vs. multilingual transformer models.
  - **Image modality:** fine-tuning CNN models with different architectures.
  - **Multimodality:** training an early fusion-based model.
- Evaluating different LLMs in a zero-shot setup for all modalities.

ArMeme Dataset Development



Dataset Statistics

Source	# of Group	# of Images
Facebook	19	5,453
Instagram	22	107,307
Pinterest	-	11,369
Twitter	-	5,369
Total		**129,498**

Source	Not propaganda	Propaganda	Not-meme	Other	Total
Facebook	464	332	58	144	998
Instagram	2,052	637	46	60	2,795
Pinterest	1,245	414	147	78	1,884
Twitter	3	5	38	2	48
Total	**3,764**	**1,388**	**289**	**284**	**5,725**

Dataset Statistics

Class label	Train	Dev	Test	Total
Not propaganda	2,634	384	746	3,764
Propaganda	972	141	275	1,388
Not-meme	199	30	57	286
Other	202	29	56	287
Total	**4,007**	**584**	**1,134**	**5,725**

Experiments

- **How effective are SLMs and LLMs at capturing propagandistic content across different modalities?**
- **Are multimodal models more effective?**

Unimodal – Text:

- **Text based SLMs**
  - Monolingual: AraBERT, Qarib
  - Multilingual: mBERT, XLM-r
- **Text based LLMs:** GPT-4v, GPT-4o

Unimodal – Image:

- **Embedding based:** Extracted embedding + SVM
- **CNN based models:** MobileNet, ResNet18, ResNet50, Vgg16, EfficientNet
- **LLMs:** GPT-4v, GPT-4o

Multimodal:

- **Fusion:** Embeddings from different modalities followed by fusing at the embedding level with SVM as a classifier
- **LLMs:** GPT-4v, GPT-4o, Gemini

Ablation Study

- **Binary classification** using *ArAIEval 2024 dataset*.
- **LLMs:** GPT-4v, GPT-4o, Gemini

Results

Unimodal -- Text

Model	Acc	W-P	W-R	W-F1	M-F1
Ngram	0.669	0.624	0.669	0.582	0.280
AraBERTV2	0.688	0.670	0.688	0.666	0.511
mBERT	0.707	0.688	0.707	0.675	0.487
Qarib	0.697	0.688	0.697	**0.690**	0.551
XLM-r-base	0.699	0.676	0.699	0.678	0.489
XLM-r	0.698	0.653	0.698	0.656	0.418
GPT-4v	0.664	0.620	0.664	0.624	0.384
GPT-4o	0.573	0.611	0.573	0.579	0.350

- **Text based model: Qarib model** outperforms all other text based models (0.69 weighted F1).

Results

Unimodal -- Image

Model	Acc	W-P	W-R	W-F1	M-F1
ConvNeXt + SVM	0.655	0.608	0.655	0.614	0.405
densenet	0.667	0.586	0.667	0.588	0.329
mobilenet_v2	0.660	0.618	0.660	0.620	0.426
squeezenet	0.667	0.599	0.667	0.595	0.325
resnet18	0.656	0.597	0.656	0.593	0.358
resnet50	0.660	0.638	0.660	**0.637**	0.434
resnet101	0.677	0.604	0.677	0.612	0.359
vgg16	0.656	0.597	0.656	0.593	0.358
efficientnet_b1	0.658	0.558	0.658	0.572	0.298
efficientnet_b7	0.660	0.597	0.660	0.595	0.352
GPT-4v	0.565	0.551	0.565	0.545	0.223
GPT-4o	0.693	0.627	0.693	**0.634**	0.305

- **Image based model:** ResNet50 is close to GPT-4o, however, it is better with M-F1.

Results

Multimodal

Model	Acc	W-P	W-R	W-F1	M-F1
ConvNeXt + AraBERT + SVM	0.683	0.655	0.683	**0.659**	0.513
Gemini	0.519	0.551	0.519	0.521	0.276
GPT-4-v	0.681	0.461	0.330	0.619	0.340
GPT-4o	0.653	0.443	0.354	0.639	0.363

- **Multimodal model:** Embedding based approach is outperforming all other models.

Summary and Future Work

Summary

- ***ArMeme,*** Arabic propaganda detection dataset
  - Developed a detail annotation guidelines in English and in Arabic
- Experiments focusing on different modalities (text, image and multimodal)

Future work

- Extend to fine-grained techniques
- Offer span level detection tasks

Dataset Availability

- Dataset is released under CC-BY-NC-SA through https://huggingface.co/datasets/QCRI/ArMeme

Acknowledgments

This publication was made possible by NPRP grant 14C-0916-210015. Part of this work was also funded by Qatar Foundation's IDKT Fund TDF 03-1209-210013.

Thank you!