Beyond Words: Analyzing Social Media with Text and Images

Danae Sanchez Villegas
Department of Computer Science, University of Sheffield

October 2023

Analyzing Social Media with Text and Images

Multimodal posts offer a creative and engaging means of communication for users.

Applications in natural language processing

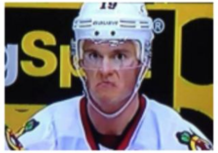
- Sentiment analysis
- Rumor detection and fact checking
- Sarcasm Detection

This is relevant again

Analyzing Social Media with Text and Images

Modeling text-image pairs from social media posts presents particular challenges.

- While image captions have a clear visual-language connection, image-text relationships in social media posts may no be apparent

Image	Text (Post)	Image-Text Relation in Post	Image Caption
	When @USER gets more followers than you in 12 hours	The image complements the text to provide meaning of the post	A close up of a hockey player wearing a helmet
	My baby approves	The image does not add to the meaning of the post and the text does not provide a description of the image	A gray and white chicken standing in the dirt

Analyzing Social Media with Text and Images

Crucial to advancing natural language understanding:

- Enhances the understanding of the user's intentions, emotions, and opinions.
- Disambiguating the intended meaning
- Visual context can help handling noisy text (e.g., abbreviations and typos)

Analyzing Social Media with Text and Images

Introducing challenging tasks as well as methods to gain a better understanding of multimodal content in the context of social media.

Point-of-interest Type Prediction

Next stop: NYC ✈️

Online Political Advertising

WE CAN'T LET JOE BIDEN WIN! VOTE EARLY

Influencer Content Analysis

Cherry tree hill is hands down the best view in #Barbados. #VisitBarbados

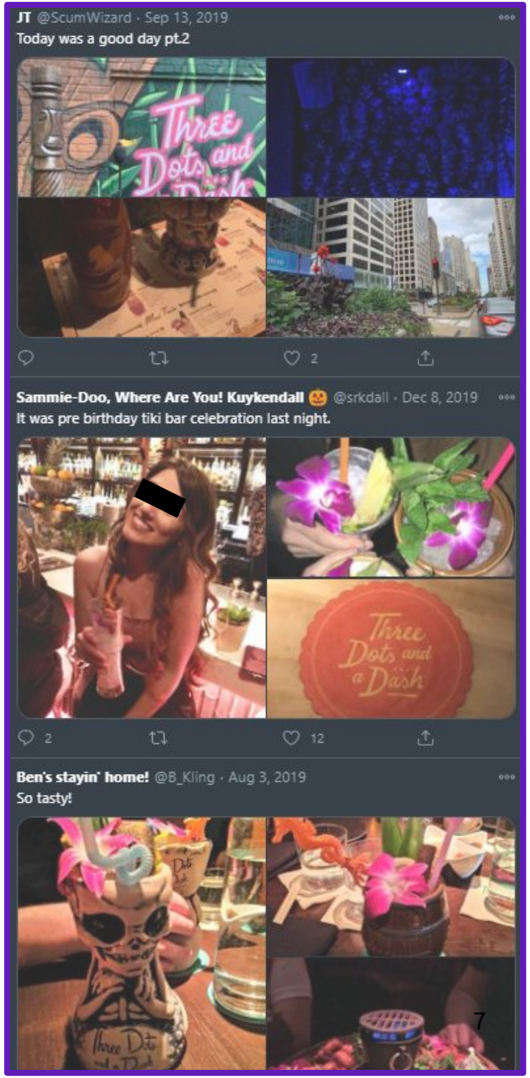
Point-of-interest Type Prediction

Sánchez Villegas, Danae, et al., "*Point-of-Interest Type Inference from Social Media Text*", **AACL 2020**

Sánchez Villegas, Danae and N. Aletras, "*Point-of-Interest Type Prediction using Text and Images*", **EMNLP 2021**

Point-of-Interest (POI)

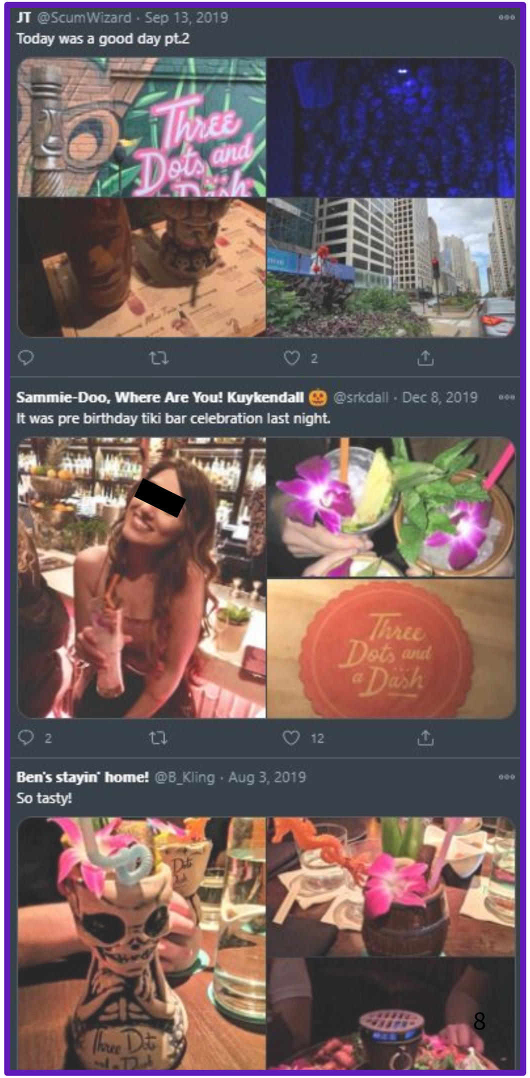
A place or point-of-interest is a physical space infused with human meaning and experiences that facilitate communication.



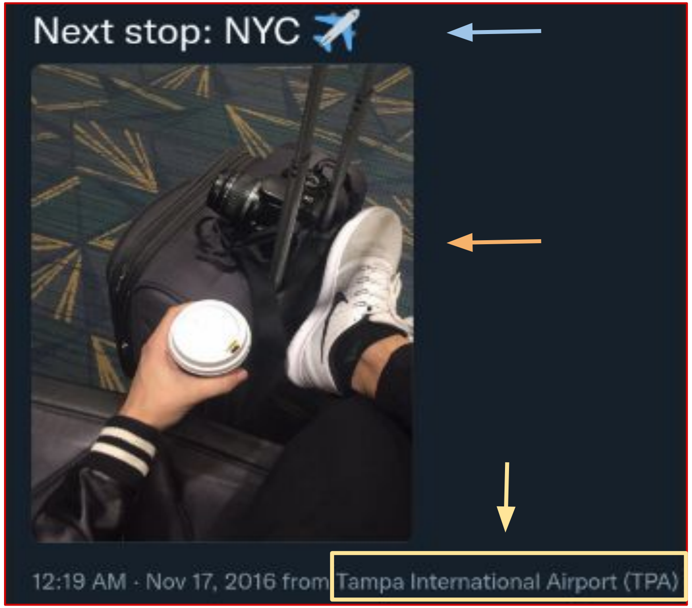
Point-of-Interest (POI)

A place or point-of-interest is typically described as a physical space infused with human meaning and experiences that facilitate communication.

Social networks allow users to post from different POIs



Points-of-Interest (POI) in Social Media



The multimodal content of social media posts such as:

- text and emojis
- images

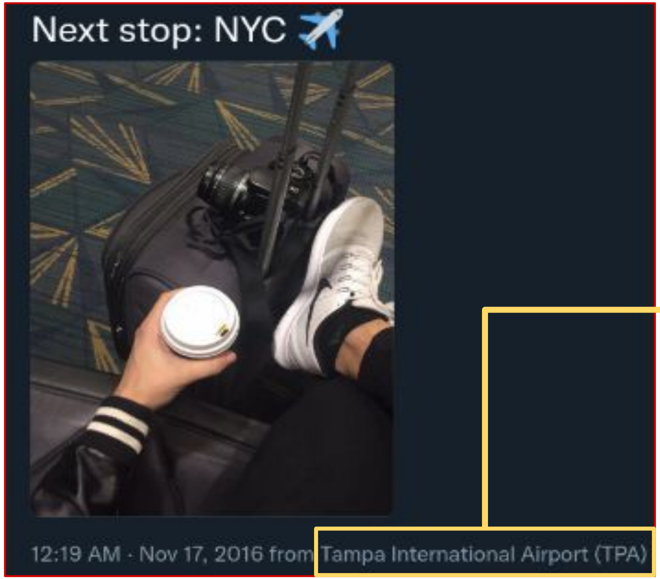
generated by users from specific

places (points-of-interest)

contribute to shaping a place’s identity

Points-of-Interest Type Prediction

Multi-class classification task performed at the social media post level.



- Arts & Entertainment
- College & University
- Great Outdoors
- Nightlife Spot
- Professional
- Shop & Service
- Travel & Transport

Applications

- ➢ POI Visualization
- ➢ POI Recommendation
- ➢ Social and cultural geography

Distinct from geo-location prediction:

- ➢ Predict type of place (POI)
- ➢ Rather than / irrespective of the exact location / coordinates

Text and Labels

- Text and Labels
- 196, 235 tweets written in English

  Each tweet is labeled with one out of the eight POI broad type categories:
  - 8 primary top-level POI categories in ‘Places by Foursquare’

Travel & Transport	Arts & Entertainment	Great Outdoors	Shop & Service
Professional	College & University	Nightlife Spot	Food

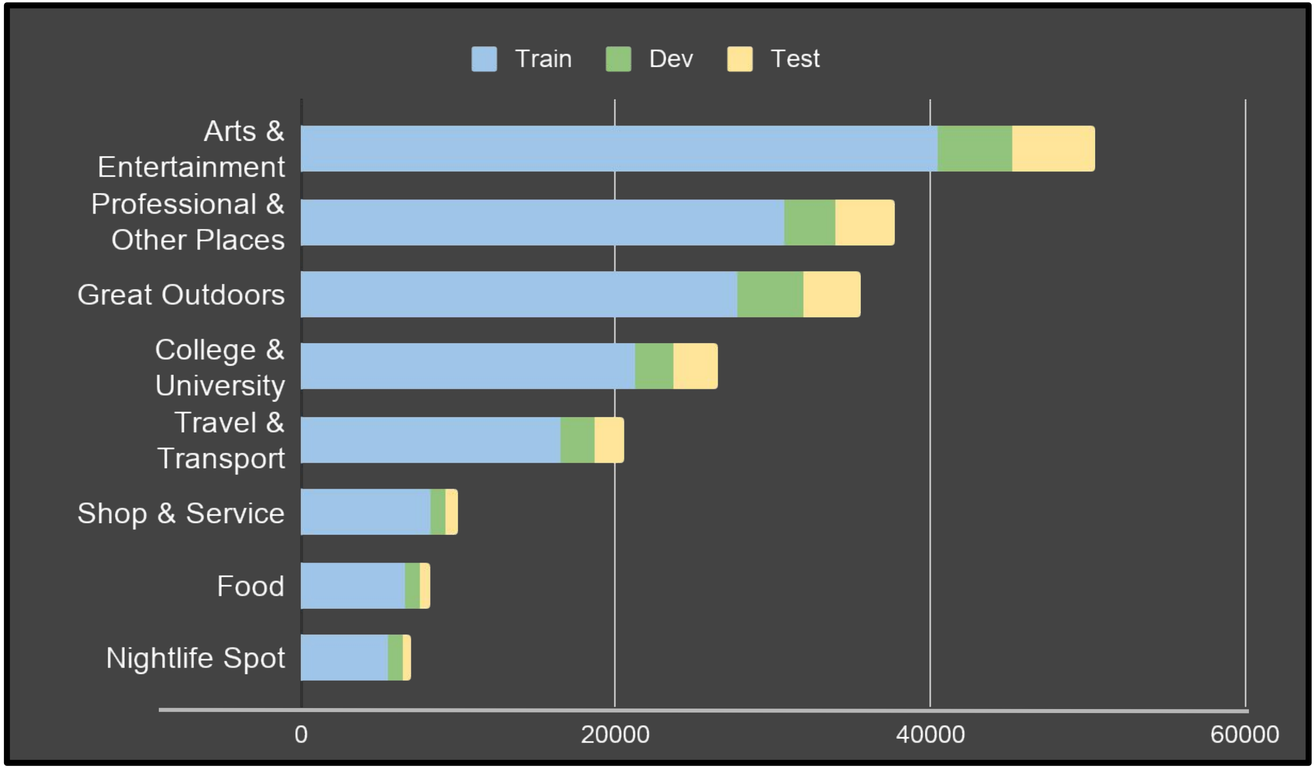
Image Collection

- Collect the images that accompany each text post in the data set
- 91, 095 text-image pairs
- Most common objects in image content of tweets

Food

Outdoors

Data



Train
Dev
Test
Arts & Entertainment
Professional & Other Places
Great Outdoors
College & University
Travel & Transport
Shop & Service
Food
Nightlife Spot
0
20000
40000
60000

Multimodal POI Type Prediction

MM-Gated-XAtt

Combine text (BERT) and image representations (Xception)

Multimodal POI Type Prediction

MM-Gated-XAtt

Combine text (BERT) and image representations (Xception)

1) Weighting strategy to assign more importance to the most relevant modality and suppress irrelevant information
    a) **MM-Gate**: gated multimodal fusion (Arevalo et al., 2020) to control the contribution of text and image to the POI type prediction.

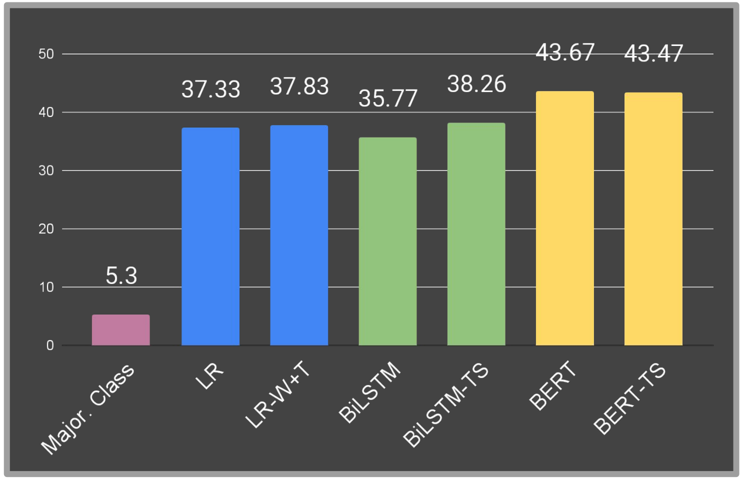
Multimodal POI Type Prediction

MM-Gated-XAtt

Combine text (BERT) and image representations (Xception)

1) Weighting strategy to assign more importance to the most relevant modality and suppress irrelevant information
   a) **MM-Gate**: gated multimodal fusion (Arevalo et al., 2020) to control the contribution of text and image to the POI type prediction
2) Capture interactions between text and image
   a) **MM-XAtt**: cross-attention mechanism (Tsai et al., 2019; Tan and Bansal, 2019) to combine text and image information

Models - Baselines



Unimodal Models

Text } BERT (best text-only model) in Sánchez Villegas et al., 2020

Image } ResNet, EfficientNet, Xception

- LR: Logistic Regression
- TS/T: Temporal Features

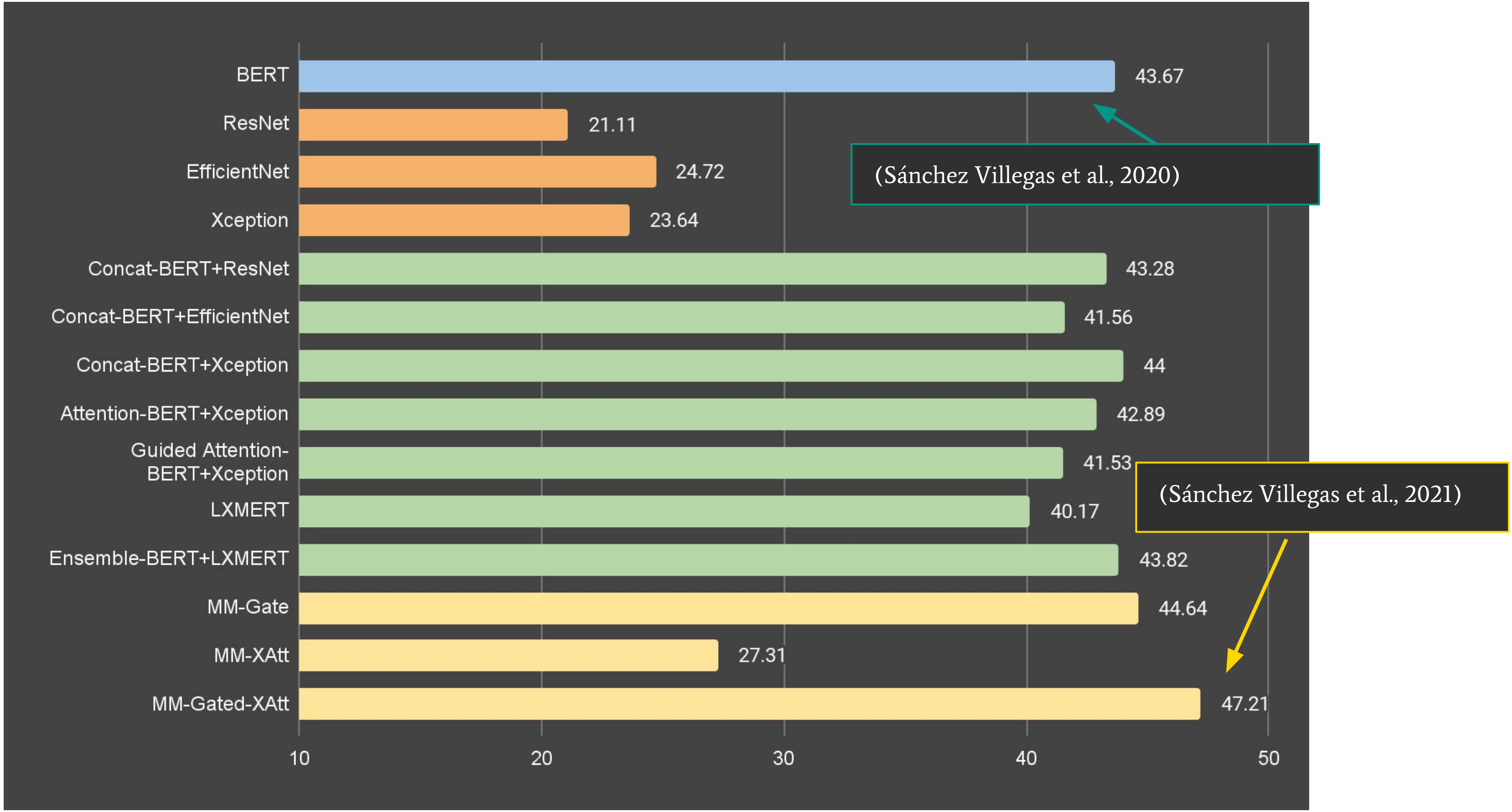
Models - Baselines

Multimodal

Models

- Concat-BERT+ResNet
- Concat-BERT+EfficientNet
- Concat-BERT+Xception
- Attention-BERT+Xception
- Guided Attention-BERT+Xception
- LXMERT
- Ensemble-BERT+LXMERT

F1


BERT
43.67
ResNet
21.11
EfficientNet
24.72
Xception
23.64
Concat-BERT+ResNet
43.28
Concat-BERT+EfficientNet
41.56
Concat-BERT+Xception
44
Attention-BERT+Xception
42.89
Guided Attention-BERT+Xception
41.53
LXMERT
40.17
Ensemble-BERT+LXMERT
43.82
MM-Gate
44.64
MM-XAtt
27.31
MM-Gated-XAtt
47.21
10
20
30
40
50
(Sánchez Villegas et al., 2020)
(Sánchez Villegas et al., 2021)

Error analysis

Most errors occur identifying for POI categories where people might perform similar activities in each of them

Error Analysis

Food and **Shop & Service**

celebrate the fruits of #fermentation's labor at #bostonfermentationfestival! next sun 10-4

True: Shop & Service

MM-Gated-XAtt (Ours): Food

Error Analysis

Food and **Nightlife Spot**

True: Nightlife Spot

MM-Gated-XAtt (Ours): Food

Summary 🙂

- We presented the first study on point-of-interest type prediction from social media content
- Released a data set with tweets mapped to their POI category → https://archive.org/details/poi-data
- Trained predictive models to infer the POI category
- Visual information is beneficial for POI type prediction
- Model performance may improve if more contextual information about the places is available
  - e.g. finer subcategories of a type of place
  - how POI types are related to one another

DSV is supported by the Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by the UK Research and Innovation grant EP/S023062/1.

Online Political Advertising Analysis

Sánchez Villegas, Danae, et al.“*Analyzing Online Political Advertisements*”, **Findings of ACL 2021**

Motivation

- Online advertising is an integral part of modern digital election campaigning
- The 2020 U.S. election campaign spending hit a record $10.8 billion[1]

Source: https://twitter.com/OpenSecretsDC/status/1321589058993332224

[1]https://www.cnbc.com/2020/10/01/election-2020-campaign-spending-set-to-hit-record-11-billion.html

Motivation

Third-party advertising had an increased presence in 2018 and 2020 US elections

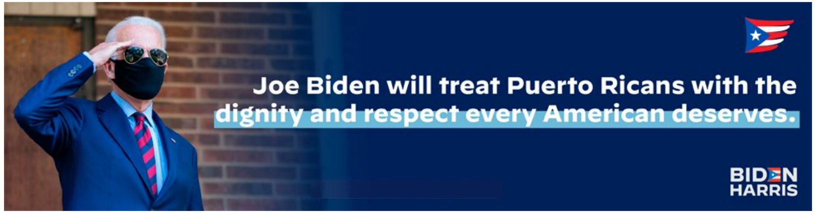
Almost half of the third-party sponsored ads were funded by dark-money sources

Freedom Club is the premier non-profit organization making a difference in Minnesota. Not only do our members talk about the problems facing our state and nation, but we also put our money where our mouth is and lead the way.

Source: https://www.freedomclub.mn/

Motivation

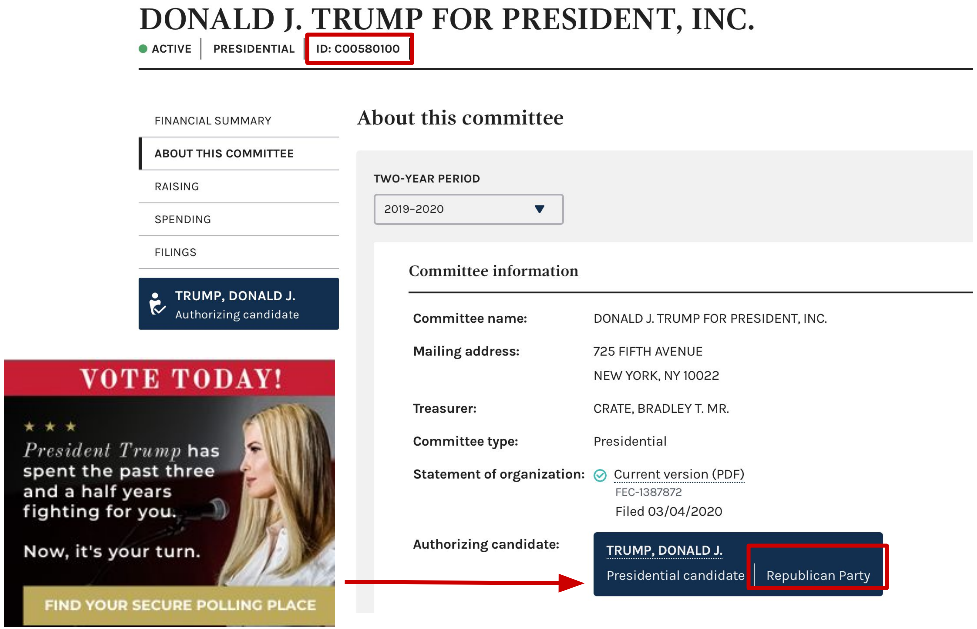
- Serious implications about transparency and accountability
  - How voters were targeted?
  - By whom?



Task 1

Political Ideology Prediction

- Label an ad according to the dominant political ideology of the party that sponsored the ad either as: Conservative or Liberal



Task 2

Ad sponsor type Prediction

- Classify an ad according to the type of the organization that sponsored the ad as: Political Party or Third-Party
    - Political Party: official political committees
    - Third-Party sponsors: not-for-profit organizations and businesses

→ DONALD J. TRUMP FOR PRESIDENT INC.
FEC ID: C00580100 → Political Party

official political committees

→ FREEDOM CLUB
EIN ID: 80-0684337 → Third-Party

not-for-profit organizations/ businesses

Collecting Ads

Political Advertising on Google US (2018-2020)

Ads

Text:
FIGHTING FOR WORKING FAMILIES, FOR GOOD JOBS, AND FAIR PAY.

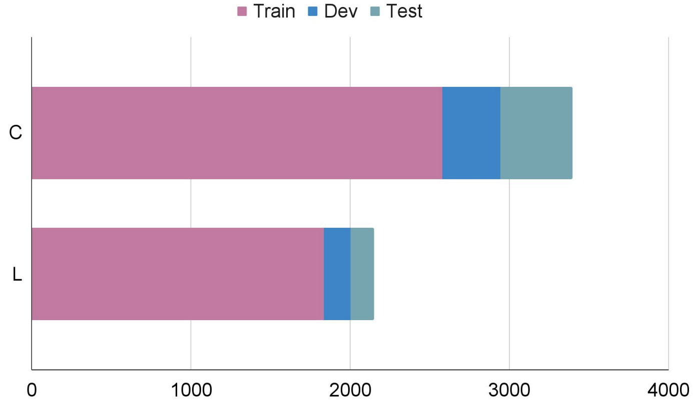
Densecaps:
the man is wearing glasses,, the background is blue

Eliminate duplicates
Filter English only

Data Splits

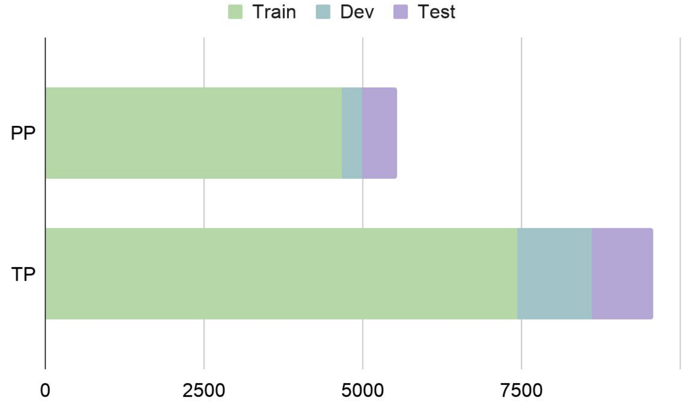
Conservative/Liberal

- Train 79.51%
- Dev 9.63%
- Test 10.86%



Political Party (PP)/Third-Party (TP)

- Train 79.98%
- Dev 10.00%
- Test 10.02%



Models

Text-only

- BERT_{D}
- BERT_{IT}
- BERT_{IT+D}

Image-only

- EfficientNet

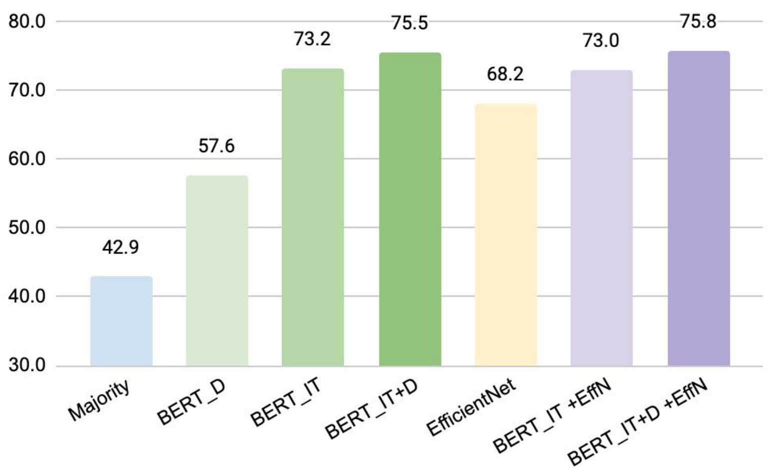
Text & Image

- BERT_{IT}+EffN
- BERT_{IT+D}+EffN

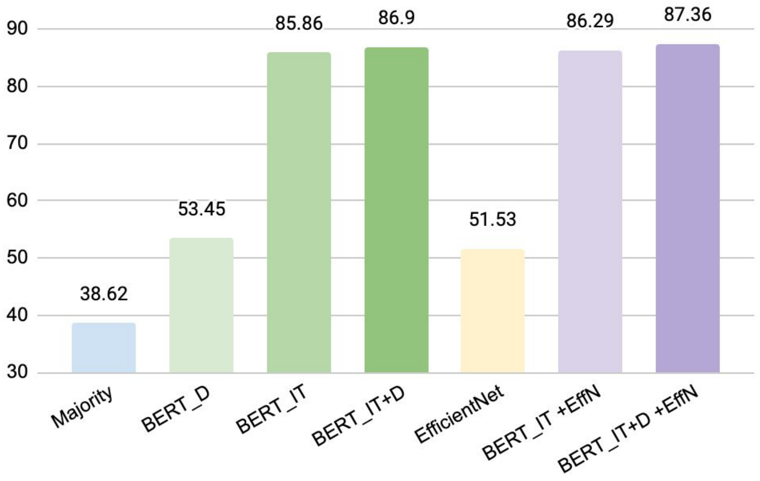
★ IT: Image Text
★ D: Densecaps

Data Splits

Conservative/Liberal



Political Party (PP)/Third-Party (TP)



F1

Error Analysis
Conservative/Liberal

Conservative

Pred ($BERT_{IT+D}$): Liberal

Densecaps:

'the sign is blue',
'a blue and white stripe shirt' ,
'a man wearing a hat',
*'**a man is holding a horse**',*

Pred ($BERT_{IT}$): Conservative

Error Analysis
Political Party/Third Party

Political Party

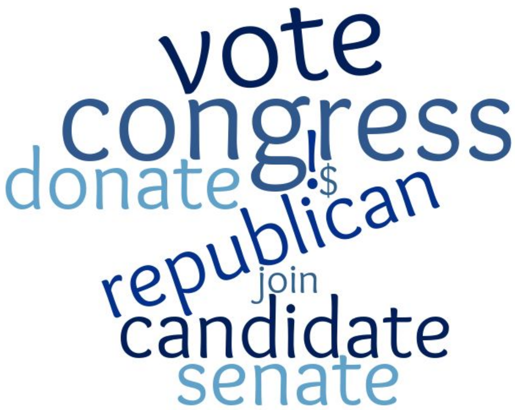
Pred ($BERT_{IT+D}$+EffN): Third-Party

Confronting

Negative style

Negative campaigning

Linguistic Analysis
Political Party/Third Party



Political Party

Third-Party

Linguistic Analysis
Political Party/Third Party

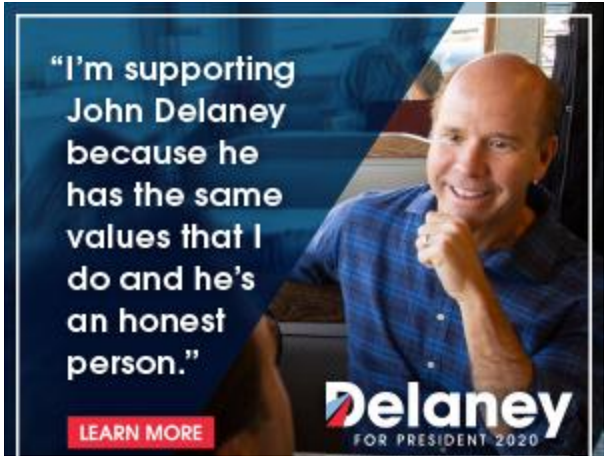
Third-Party

Linguistic Analysis
Conservative/Liberal

Conservative

Liberal

Linguistic Analysis
Conservative/Liberal



Liberal

Summary

- We presented the first study on Political Ideology and Ad Sponsor Type Prediction
- Built a dataset with ads mapped to their category → https://archive.org/details/pol_ads
    - Political Ideology
    - Ad sponsor Type
- Trained predictive models using
    - Text
    - Image descriptions
    - Image
- Analysis of the Ad content

DSV is supported by the Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by the UK Research and Innovation grant EP/S023062/1. NA is supported by a Leverhulme Trust Research Project Grant. This work complies with the Terms of Service of the Google Political Ads Dataset

Influencer Content Analysis

Sánchez Villegas, Danae, et al., "*A multimodal analysis of influencer content on twitter*", in **AACL 2023** (accepted)

Social Media Influencers

Social media influencers are content creators who have established credibility in a specific domain (e.g., fitness, technology), are followed by a large number of accounts and can impact the buying decisions of their followers.

Influencer Marketing

- Influencer marketing is more effective than traditional paid advertising.
- Online creators can help brands reach new, engaged audiences through endorsements and product placements, leveraging the trust these influencers have built with their followers.

Influencer Marketing

Influencer marketing is dominated by native advertising

- there is no obvious distinction between commercial and non-commercial content

Chunky knits and dainty jewels. 🍂🐺🍂This is my favorite vintage sweater 🍂🐺🍂#lovechupi

Detecting commercial content

Automatically identifying commercial content by influencers is important

- Transparency: it helps ensure transparency in advertising and marketing.
- Consumer Protection: it protects consumers from deceptive advertising.
- Regulatory Compliance: some countries have laws and regulations governing advertising and disclosure requirements for influencers and brands.
- Analysis of commercial language characteristics on a large scale.

Detecting commercial content

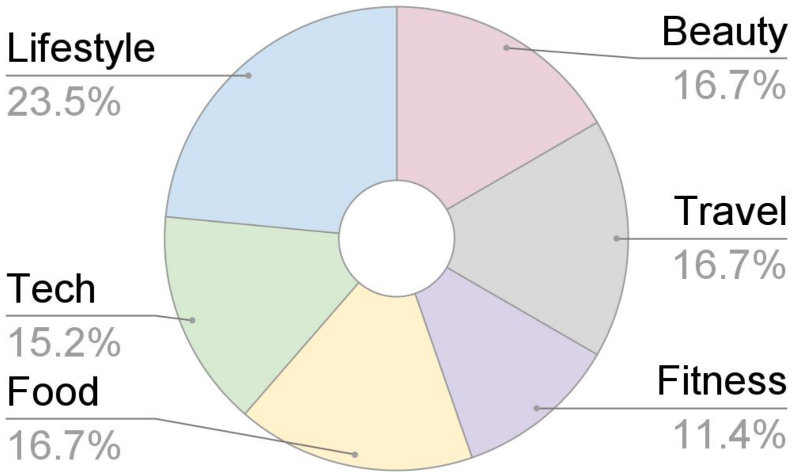
Automatic detection of influencers commercial content is difficult.

- Disclosure guidelines (including keywords such as #ad, #sponsored) are not always followed
- Brand cues may appear in different modalities such as text and images

Multimodal Influencer Content Dataset (MICD)

A large publicly available dataset of 14, 384 text-image pairs and 1, 614 text-only influencer tweets written in English.

- 132 Influencer Accounts
- 6 domains
- Jan 2015- Aug 2021



Multimodal Influencer Content Dataset (MICD)

Tweets are mapped into commercial and non-commercial categories

- Keyword-based Weak Labeling (train & dev sets)
- Human Data Annotation (test sets)

Multimodal Influencer Content Dataset (MICD)

Keyword-based Weak Labeling

Extend the keyword lists (verified by members of a national consumer authority)

- Disclosure terms: #ad, #sponsored
- Terms relevant to different business models:
  - Gifting: #gift
  - Endorsements: #ambassador
  - Affiliate marketing: #aff
- All of the keywords used for data labeling are removed for the experiments

Multimodal Influencer Content Dataset (MICD)

Human Data Annotation (test sets)

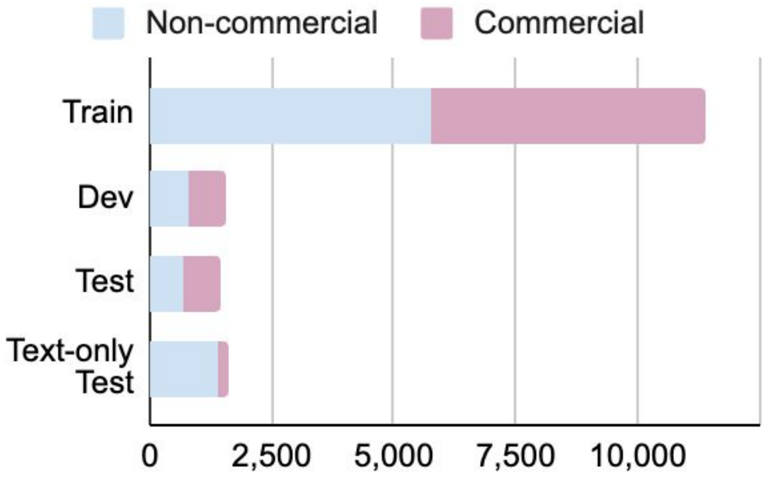
- Four annotators with a substantial legal background and knowledge of advertising regulation
- The inter-annotator agreement between two annotations across all tweets is 0.78 Cohen's-Kappa – substantial agreement —

Multimodal Influencer Content Dataset (MICD)

Data Splits

Account-level splits

Split	Total
Train	11,377 (79.1%)
Dev	1,572 (10.9%)
Test	1,435 (10%)
Text-only Test	1,614
All	15,998



Multimodal Influencer Content Dataset (MICD)

Dataset	Publicly Available	Posts w/o brand mentions	Human Annotation	Keyword Matching	No. of Commercial Keywords	Platform	Modality	Time Range	Domains
Han et al. (2021)	✗	✗	✗	✗	0	Twitter	Text	not specified	fashion
Zarei et al. (2020)	✗	✓	✗	✓	7	Instagram	Text	Jul 2019 - Aug 2019	not specified
Yang et al. (2019)	✗	✗	✗	✓	3	Instagram	Text & Image	not specified	not specified
Kim et al. (2021b)	✓	✓	✗	✓	3	Instagram	Text & Image	not specified	not specified
Kim et al. (2020)	✓	✗	✗	✓	1	Instagram	Text & Image	Oct 2018 - Jan 2019	beauty, family, food, fashion, pet, fitness, interior, travel,
MICD (Ours)	✓	✓	✓	✓	26	Twitter	Text & Image	Jan 2015 - Aug 2021	beauty, travel, food fitness, technology, lifestyle

Comparison of existing datasets for influencer content analysis

Influencer Content Classification Models

Prompting

- Flan-T5 (zero-shot, few-shot)
- GPT-3 (zero-shot, few-shot)

Text-only

- BiLSTM-Att
- BERT
- BERTweet

Image-only

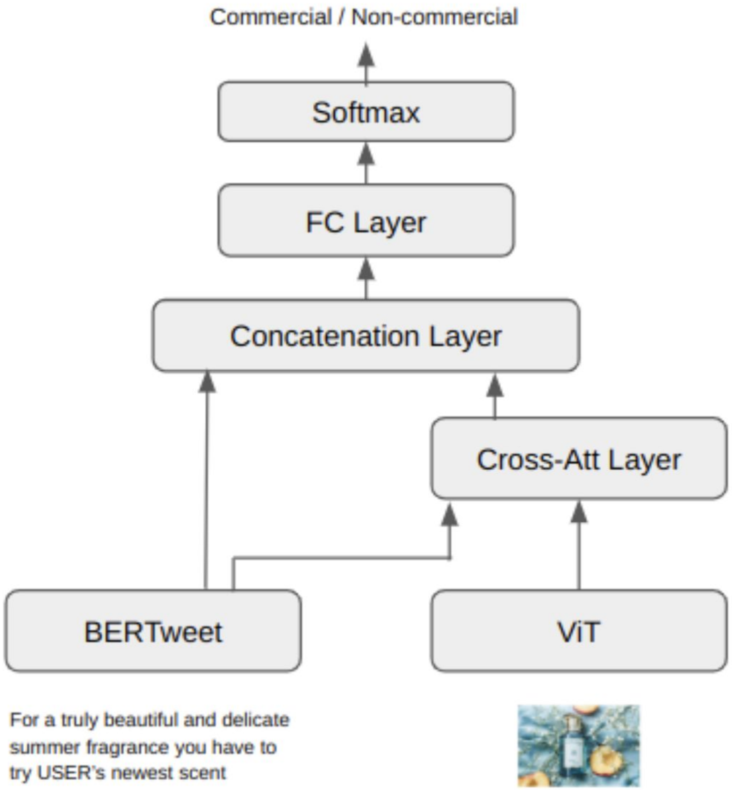
- ResNet
- ViT

Text & Image

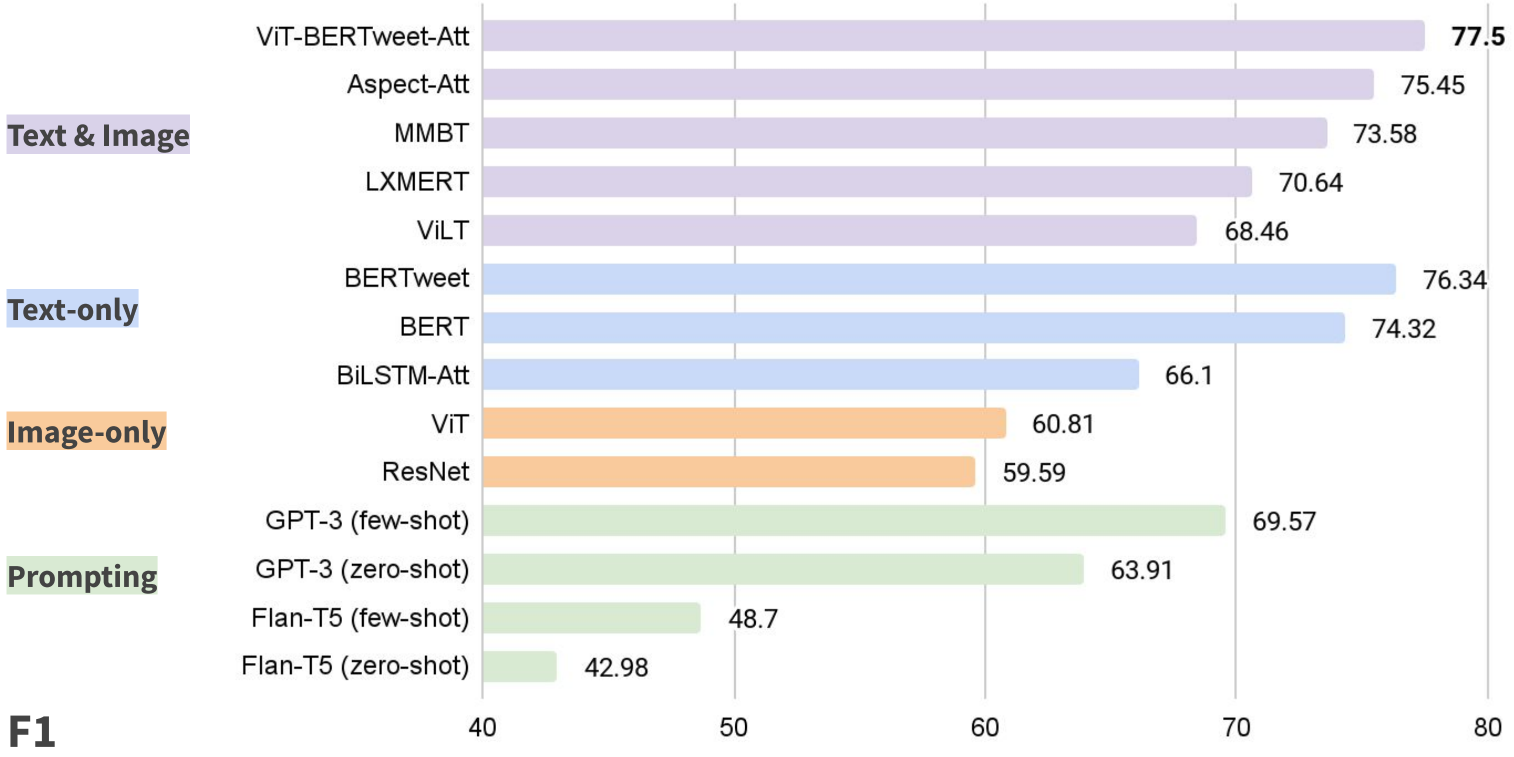
- ViLT
- LXMERT
- MMBT
- Aspect-Att
- ViT-BERTeet-Att (Ours)

ViT-BERTweet-Att

Combine unimodal pretrained representations via cross-attention fusion strategy so that text features can guide the model to pay attention to the relevant image regions.

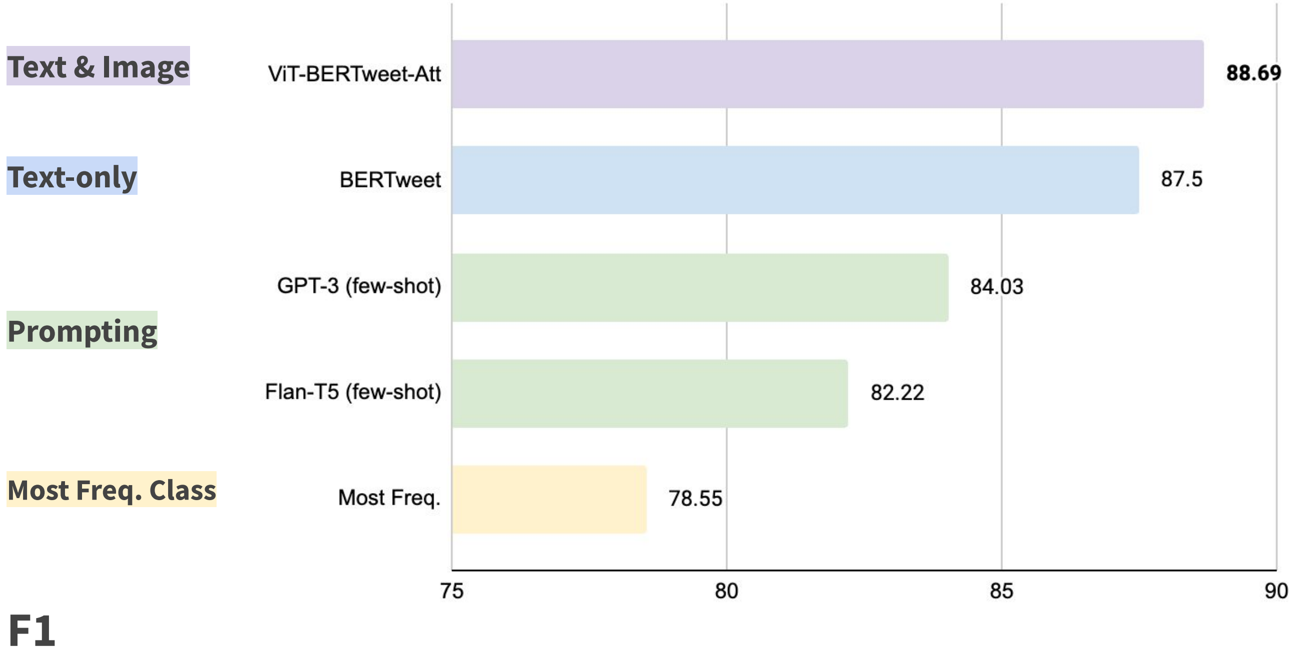


Identifying Commercial Influencer Content



Identifying Commercial Influencer Content

Text-only Test Set



Analysis

- Multimodal modeling captures context beyond keyword-matching.

`Just seen that Pepsi ad...awkward.`

ViT-BERTweet-Att: NC

- Multimodal modeling aids in the discovery of undisclosed commercial posts

chunky knits and dainty jewels. This is my favorite vintage sweater #lovechupi

Actual: C
BERTweet: NC
ViT-BERTweet-Att: C

Analysis

Challenging cases for text and multimodal models:

- Posts that describe their “personal" experiences, particularly while traveling
- Posts include “natural photos” rather than product promotions

Cherry tree hill is hands down the best view in #Barbados. #VisitBarbados

Actual: C
BERTweet: NC
ViT-BERTweet-Att: NC

Summary

- Introduced a novel dataset of multimodal influencer content consisting of tweets labeled as commercial or non-commercial.
- First dataset to include high quality annotated posts by experts in advertising regulation.
- Experiments including vision, language and multimodal approaches for identifying commercial content
- Multimodal modeling is useful for identifying commercial posts
  - Reducing the amount of false positives
  - Capturing relevant context that aids in the discovery of undisclosed commercial posts.
- Dataset: https://github.com/danaesavi/micd-influencer-content-twitter

DSV and NA are supported by the Leverhulme Trust under Grant Number: RPG#2020#148. NA is also supported by ESRC (ES/T012714/1). DSV is also supported by the Centre for Doctoral Training in Speech and Language Technologies (SLT) and their Applications funded by the UK Research and Innovation grant EP/S023062/1. CG is supported by the ERC Starting Grant research project HUMANads (ERC-2021-StG No 101041824) and the Spinoza grant of the Dutch Research Council (NWO), awarded in 2021 to José van Dijck, Professor of Media and Digital Society at Utrecht University

Future Work

Future Work

- Point-of-interest Type Prediction
    - More specific subcategories, incorporating user and network information

Future Work

- Point-of-interest Type Prediction
    - More specific subcategories, incorporating user and network information
- Online Political Advertising Analysis
    - Include other modalities such as speech and video
    - Extend the work to different regions, languages and platforms

Future Work

- Point-of-interest Type Prediction
  - More specific subcategories, incorporating user and network information
- Online Political Advertising Analysis
  - Include other modalities such as speech and video
  - Extend the work to different regions, languages and platforms
- Influencer Content Analysis
  - Modeling influencer content in multilingual settings across platforms
  - Political advertising by influencers

THANKS