Fake News Detection Project 2

Marcel Affi, Mateusz Wójcik,
Arkadiusz Zalas

Agenda

- Project Contributions
- Preprocessing Techniques
- LIAR-PLUS dataset
- CT-FAN dataset
- Word2vec with LSTM model
- BERT Experiments
- SentiGAN
- CatGAN
- Improvements & Further Research

Project Contributions

We continued working on fake news classification task with the following extensions.

- multi-class classification with the degree of certainty that a piece of news is fake (for example, 0-5 scale)
- datasets with extended input, for example justification
- providing analysis how the choice of inputs influence the models' performance
- application of generative models: SentiGAN and CatGAN

Preprocessing Techniques

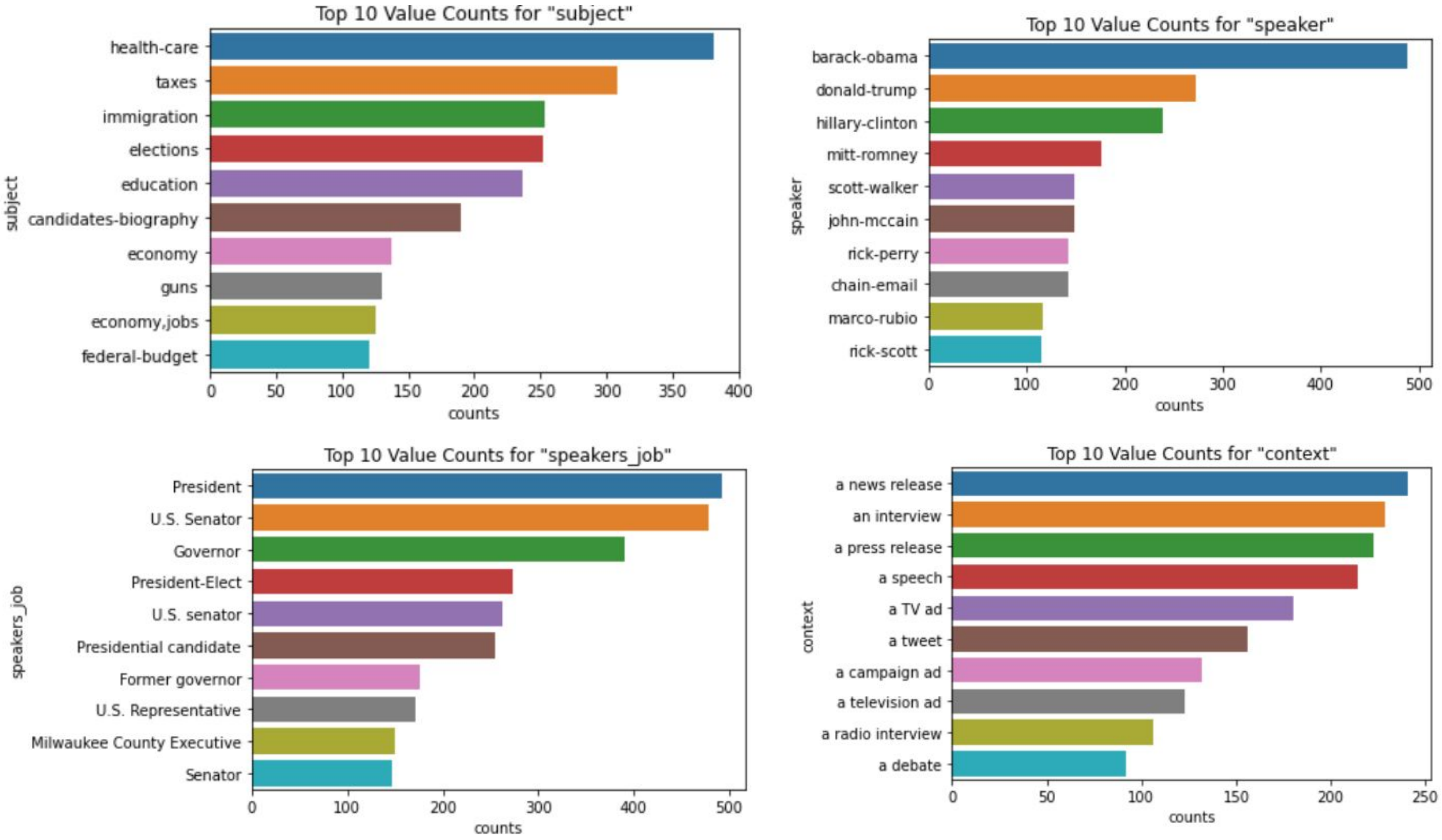
Sets of functions that we applied to prepare data for models

- Using regular expressions to remove:
  - punctuations
  - special symbols like $, #, &
  - digits <- we only replaced digits with words
  - URLs
  - wide spaces
  - single characters
- removing stop words
- lemmatization from different sources (NLTK, SpaCy)
- tokenization suited for a particular architecture
- other techniques required for the models, e.g. padding

LIAR-PLUS Dataset

The extension of the dataset manually labeled by PolitiFact consisting of short statements, published in **2018**.

- multiple labels: false, pants-fire, barely-true, half-true, mostly-true, true
- around 12800 of observations
- well-balanced set, ranging from ~2000 to ~2600 instances per class
- attributes such as: label, statement, subject, speaker, speaker's job title, party affiliation...
- evidence sentences that have been automatically extracted from full-text verdict reports written by journalists in Politifact

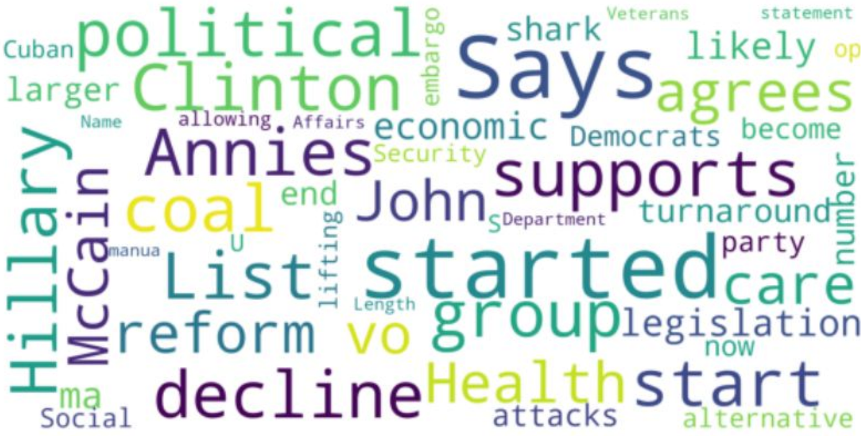

Top 10 Value Counts for "subject"
subject
health-care
taxes
immigration
elections
education
candidates-biography
economy
guns
economy,jobs
federal-budget
0
50
100
150
200
250
300
350
400
counts
Top 10 Value Counts for "speaker"
speaker
barack-obama
donald-trump
hillary-clinton
mitt-romney
scott-walker
john-mccain
rick-perry
chain-email
marco-rubio
rick-scott
0
100
200
300
400
500
counts
Top 10 Value Counts for "speakers_job"
speakers_job
President
U.S. Senator
Governor
President-Elect
U.S. senator
Presidential candidate
Former governor
U.S. Representative
Milwaukee County Executive
Senator
0
100
200
300
400
500
counts
Top 10 Value Counts for "context"
context
a news release
an interview
a press release
a speech
a TV ad
a tweet
a campaign ad
a television ad
a radio interview
a debate
0
50
100
150
200
250
counts

Example Justification

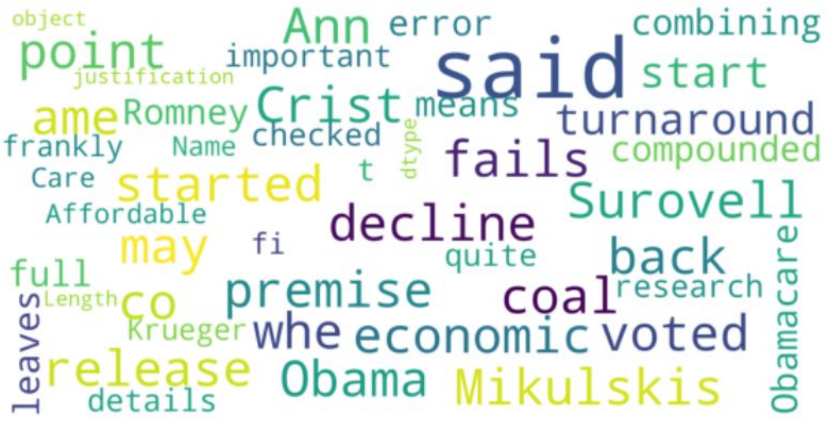
Statement: Oregon is the only state out of the 50 states in the USA that continues to pay 100% of the medical benefits for its employees and their families.

Label: false

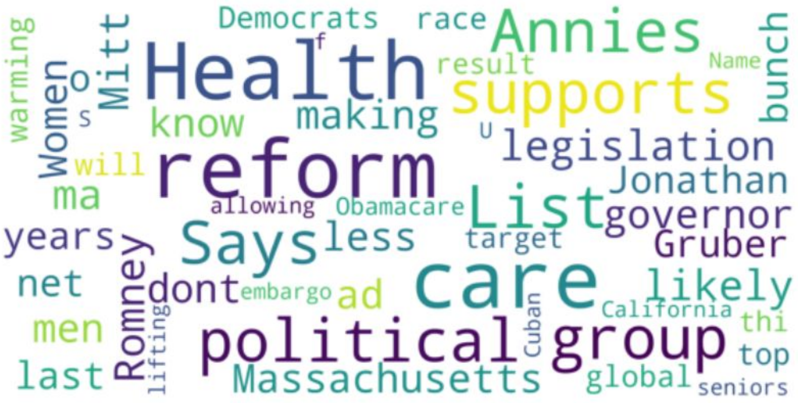
Justification: Oregon is one of two states that covers a full range of benefits -- health, dental and vision -- but it’s one of at least four that covers the premiums for its lowest-cost health plan. Richardsons larger point, that Oregon is in a shrinking group of states that do so is certainly a strong and valid one. However, he undercuts his argument by resorting to hyperbole.



"statement" word cloud



"justification" word cloud



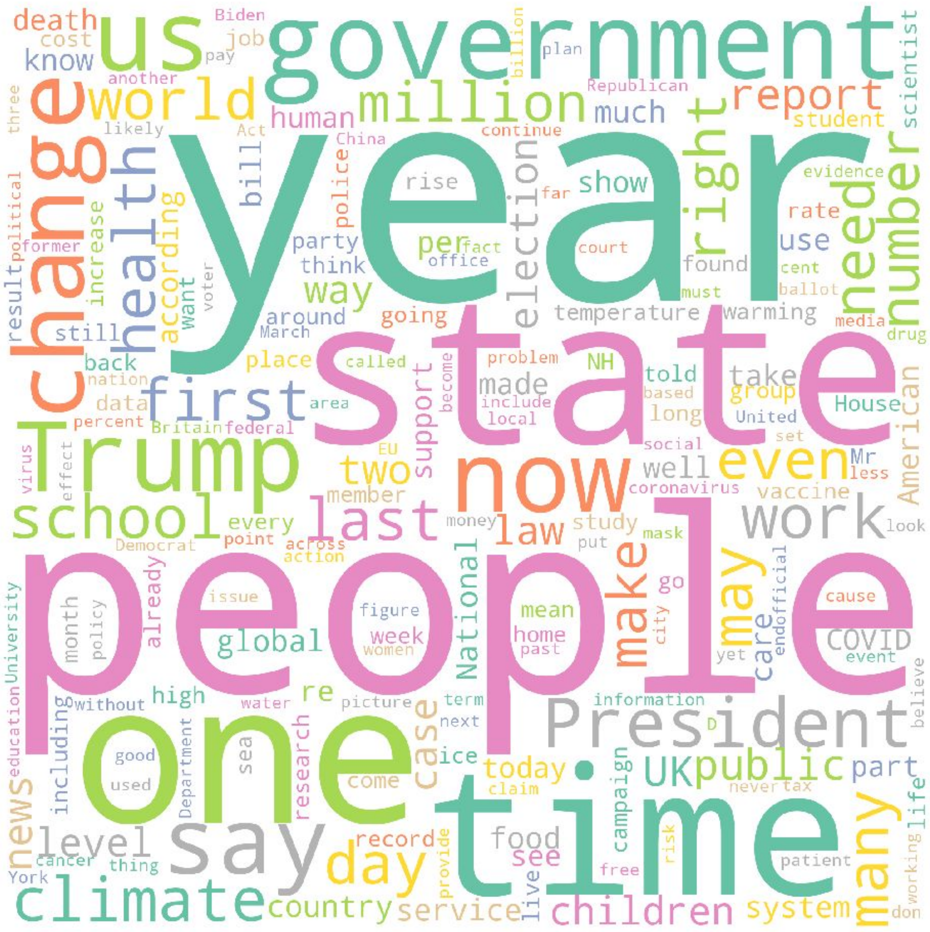
"false" word cloud

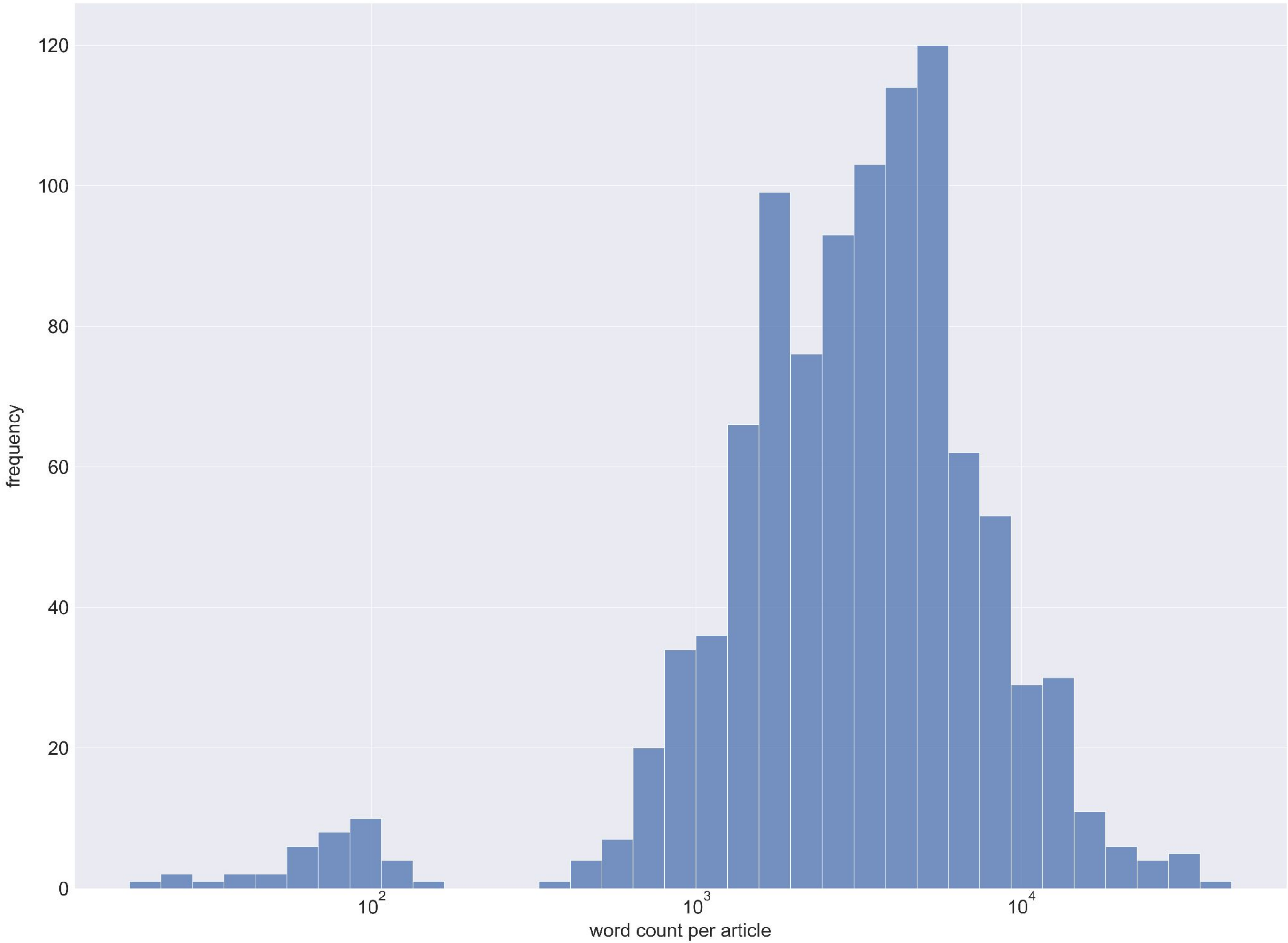
Example word clouds for the statement variable, justification variable and statements in false class

CT-FAN Dataset

The dataset consisting of news articles data, that was published in 2022. The dataset was collected from 2010 to 2022, including various topics related to elections, COVID-19 etc.

- in total **1264** articles
- attributes: id, title, text, rating (label)
- multiple labels: False (46%), Partially False (28%), True (17%) and Other (9%)
- news articles also in German language


death
cost
know
us
Biden
job
pay
government
billion
plan
scientist
another
Republican
three
world
million
report
much
change
human
likely
Act
right
student
China
continue
health
year
bill
police
evidence
political
rise
show
far
rate
need
number
former
according
increase
party
per
fact
court
use
office
think
found
cent
result
want
voter
way
election
must
ballot
around
going
temperature
warming
media
still
March
drug
state
back
place
called
problem
NH
nation
told
take
become
made
data
first
area
include
based
group
American
House
percent
Britain
federal
local
long
United
Trump
EU
social
set
virus
effect
two
support
now
well
even
Mr
less
member
coronavirus
vaccine
school
last
work
every
money
study
look
Democrat
point
across
law
put
mask
action
make
may
people
official
already
issue
figure
mean
cause
University
month
policy
go
week
home
city
global
National
end
COVID
care
women
past
yet
event
believe
high
re
without
water
picture
term
information
education
including
next
President
D
Department
case
good
one
sea
research
ice
today
UK
public
part
used
come
campaign
claim
never
tax
life
news
say
level
cancer
thing
record
provide
food
day
time
see
risk
many
patient
York
free
working
climate
country
service
live
children
system
don


120
100
80
60
40
20
0
frequency
10^2
10^3
10^4
word count per article

Word2vec embeddings with LSTM model

- applied all preprocessing techniques mentioned earlier
- clipped and padded length of articles to 3000
- 100-dim vectors
- NN model:
    - word2vec based embedding layer
    - LSTM layer with 32 units
    - dropout layers
    - Adam optimizer

Word2vec embeddings with LSTM model

Obtained results for 10 epochs of training:

Dataset	Accuracy	Balanced accuracy
LIAR-PLUS (6 classes): S	0.19079	0.17397
LIAR-PLUS (6 classes): S+J	0.19205	0.18438
CT-FAN (4 classes)	0.45810	0.31610

BERT - Bidirectional Encoder Representations from Transformers

- We used pretrained *BertForSequenceClassification* with *BertTokenizer* implemented in Transformers library in Python
- The architecture consisted of 124 layers, 1024 hidden layers, 16 heads, about 450 MB
- pretrained on uncased vocabulary
- the model was trained on Nvidia RTX 3060

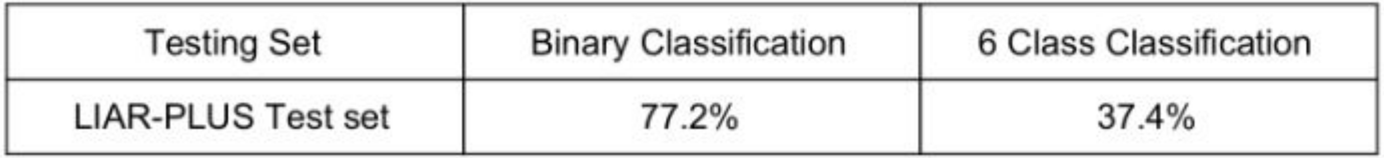
Hyperparameters and other features:

- Maximum length of tokens - 512 (a popular length for pretrained transformers)
- Fixed seed for both dataloaders and the model
- **training-validation-test split proposed by the researchers**
- AdamW optimizer with **0.0001** learning rate with **StepLR scheduler**
- **10–20 epochs of training**

BERT - Training on LIAR Dataset

As a point of reference, we also trained our BERT model on LIAR-PLUS dataset with using all preprocessing techniques. After 20 epochs, we obtained the following results:

- Accuracy: 0.61510 -> **0.62344**
- F1: 0.62789 -> **0.61682**



Testing Set	Binary Classification	6 Class Classification
LIAR-PLUS Test set	77.2%	37.4%

https://github.com/manideep2510/siamese-BERT-fake-news-detection-LIAR
https://aclanthology.org/W18-5513/

Multiclass Classification on LIAR-PLUS Dataset

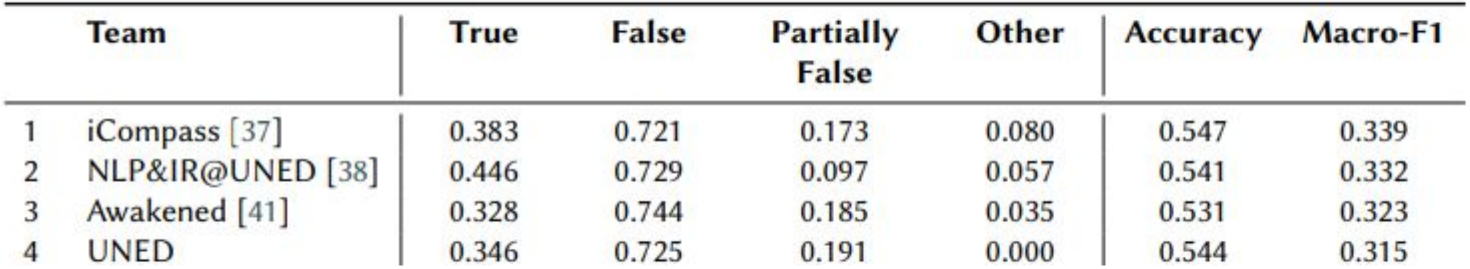
Scenario	Balanced Accuracy	Weighted F1-Score	MAE	MSE
80-20 training on Valid: S+J	0.78586	0.79560	0.43828	1.29297
80-20 training on Test: S+J	0.84140	0.83385	0.36709	1.11076

Statement versus Statement & Justification

Scenario	Balanced Accuracy	Weighted F1-Score	MAE	MSE
Train: S	0.93465	0.93677	0.11748	0.28736
Train: S+J	0.91823	0.92099	0.14016	0.34311
Validation: S	0.23441	0.24383	1.68047	4.84922
Validation: S+J	**0.24940**	**0.25722**	1.61875	4.72031
Test: S	**0.23825**	**0.23796**	1.71851	5.05680
Test: S+J	0.20855	0.21413	1.79763	5.44256

Multiclass Classification on CT-FAN Dataset

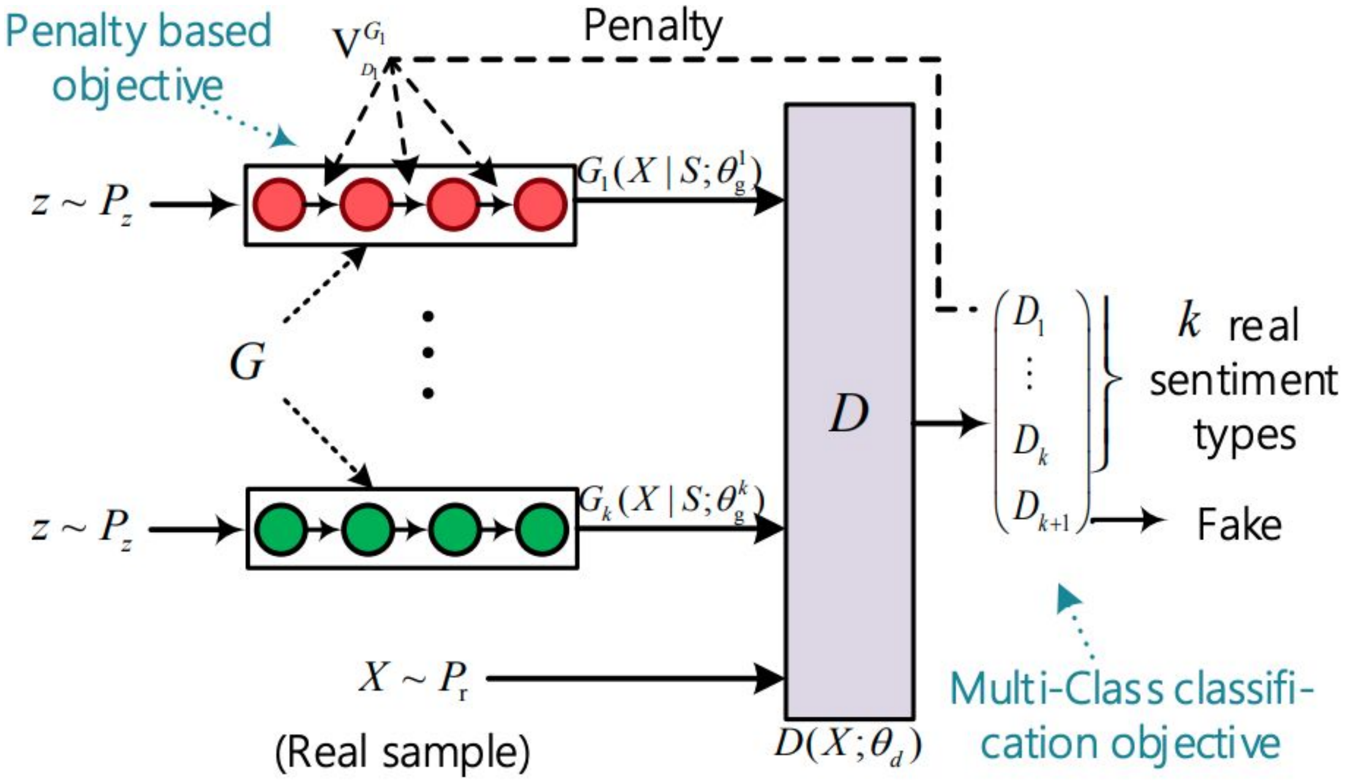
Scenario	Balanced Accuracy	Weighted F1-Score	MAE	MSE
Train Dataset	0.83937	0.88907	0.15348	0.26899
Test Dataset	**0.40765**	**0.53178**	0.83333	1.64103



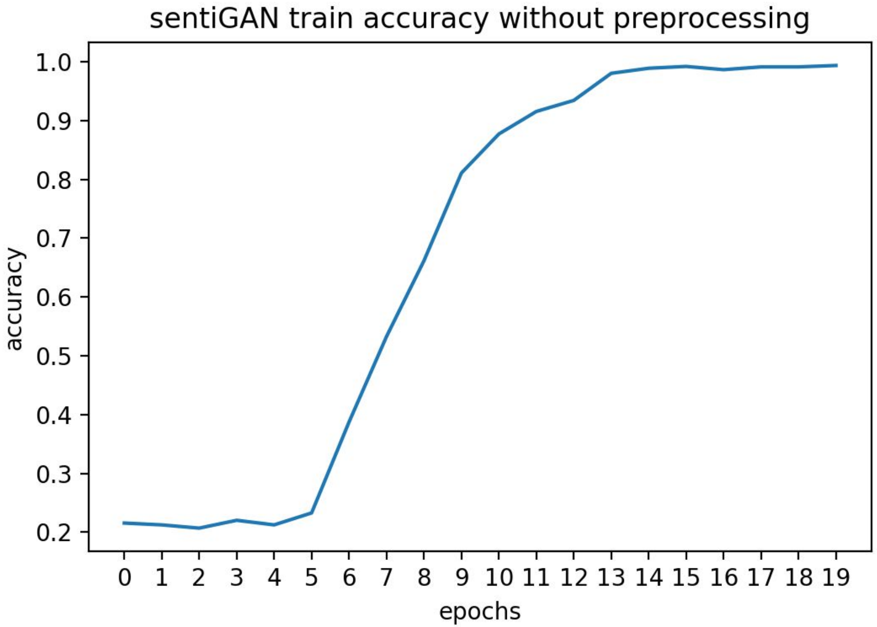
	Team	True	False	Partially False	Other	Accuracy	Macro-F1
1	iCompass [37]	0.383	0.721	0.173	0.080	0.547	0.339
2	NLP&IR@UNED [38]	0.446	0.729	0.097	0.057	0.541	0.332
3	Awakened [41]	0.328	0.744	0.185	0.035	0.531	0.323
4	UNED	0.346	0.725	0.191	0.000	0.544	0.315

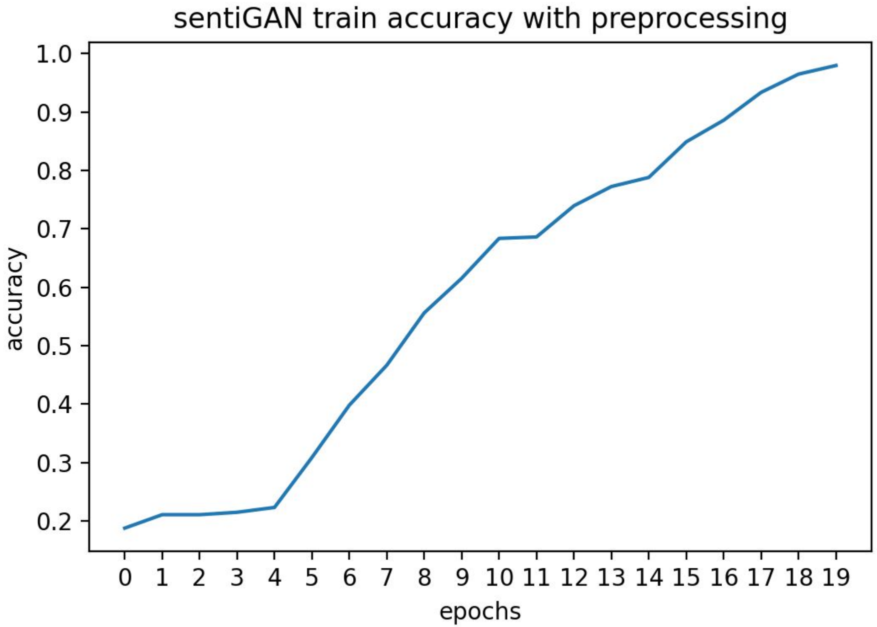
Overview of the CLEF-2022 CheckThat! Lab: Task 3 on Fake News Detection

SentiGAN

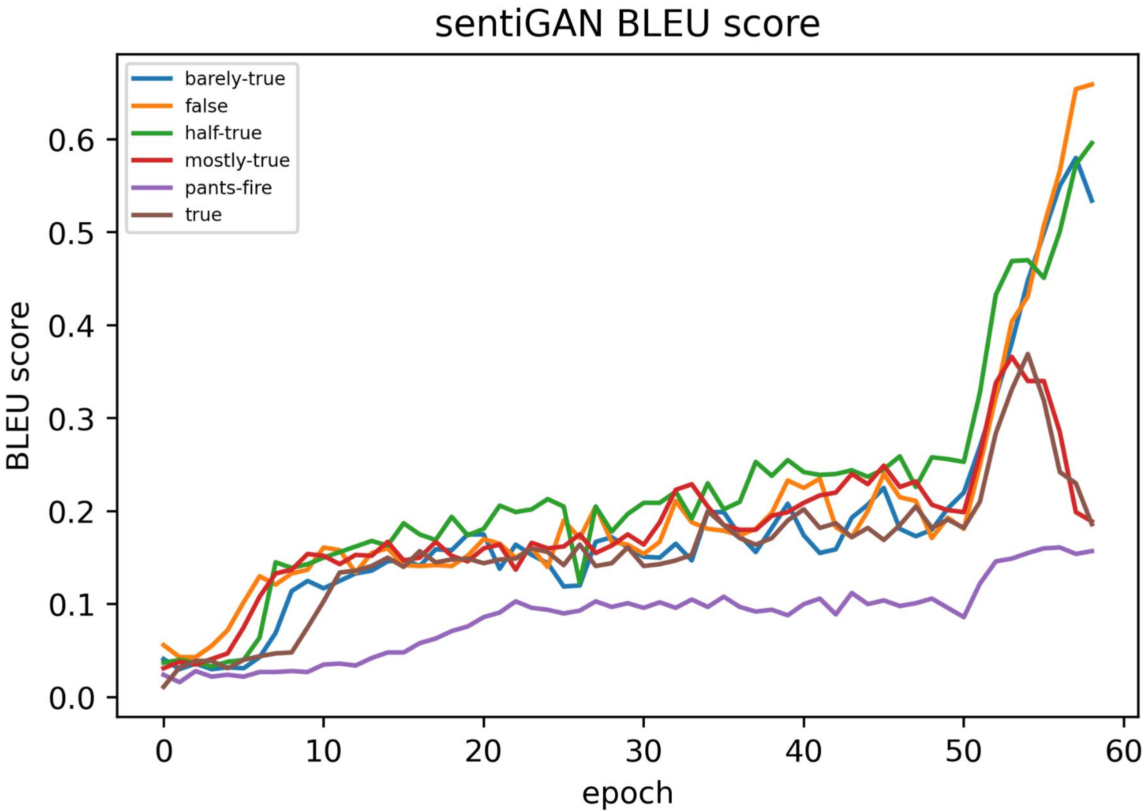

Penalty based objective
$V_{D_1}^{G_1}$
Penalty
$z \sim P_z$
$G_1(X \mid S; \theta_g^1)$
G
$z \sim P_z$
$G_k(X \mid S; \theta_g^k)$
$X \sim P_r$
(Real sample)
D
$D(X; \theta_d)$
D_1
D_k
D_{k+1}
k real sentiment types
Fake
Multi-Class classification objective

SentiGAN - Train Metrics (LIAR)





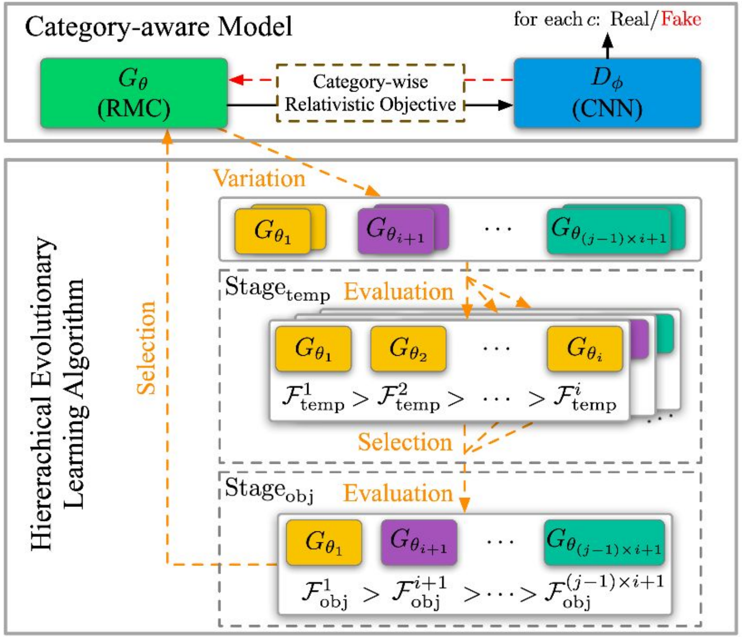
SentiGAN - BLEU Score (LIAR)



SentiGAN - Results (LIAR)

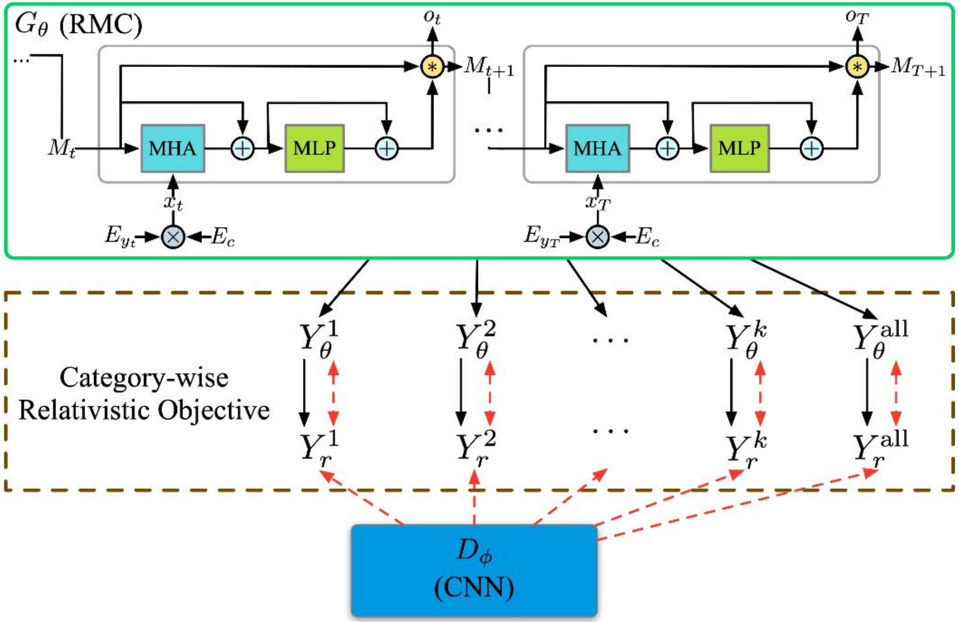
Label	barely-true	false	half-true	mostly-true	pants-fire	true	**AVG**
Accuracy (no preprocessing)	79.28	12.55	2.07	0.31	24.32	15.72	**22.37**
Accuracy (with preprocessing)	0	11.25	75.26	0	0	0	**14.4**

CatGAN



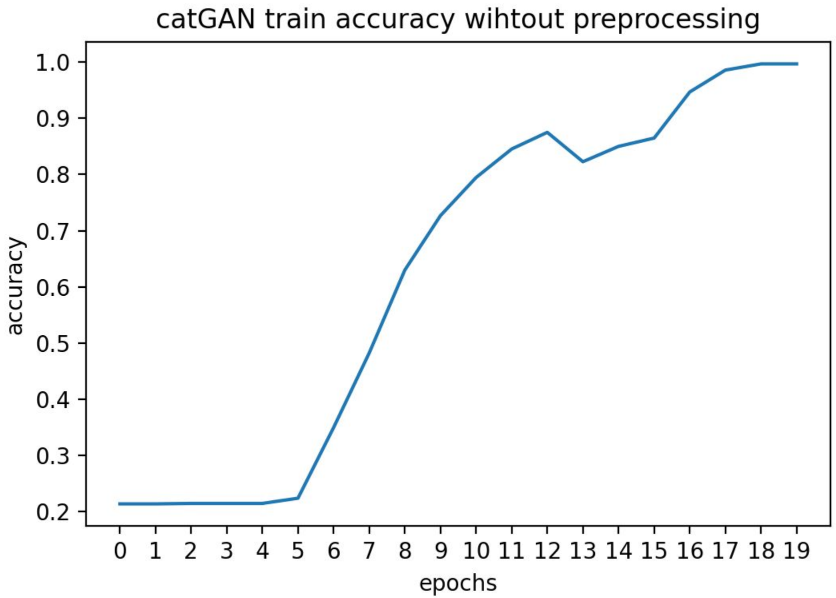
Category-aware Model
for each c: Real/Fake
G_θ (RMC)
Category-wise Relativistic Objective
D_ϕ (CNN)
Variation
G_{θ_1}
$G_{\theta_{i+1}}$
$\cdots$
$G_{\theta_{(j-1)\times i+1}}$
Stage$_{\text{temp}}$
Evaluation
G_{θ_1}
G_{θ_2}
$\cdots$
G_{θ_i}
$\mathcal{F}^1_{\text{temp}} > \mathcal{F}^2_{\text{temp}} > \cdots > \mathcal{F}^i_{\text{temp}}$
Selection
Stage$_{\text{obj}}$
Evaluation
G_{θ_1}
$G_{\theta_{i+1}}$
$\cdots$
$G_{\theta_{(j-1)\times i+1}}$
$\mathcal{F}^1_{\text{obj}} > \mathcal{F}^{i+1}_{\text{obj}} > \cdots > \mathcal{F}^{(j-1)\times i+1}_{\text{obj}}$
Hiererachical Evolutionary Learning Algorithm
Selection

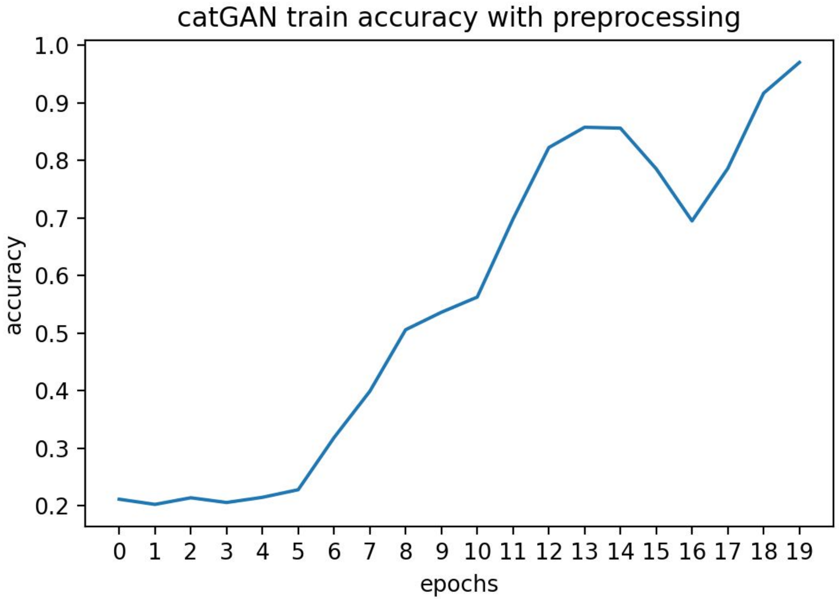
(a) CatGAN



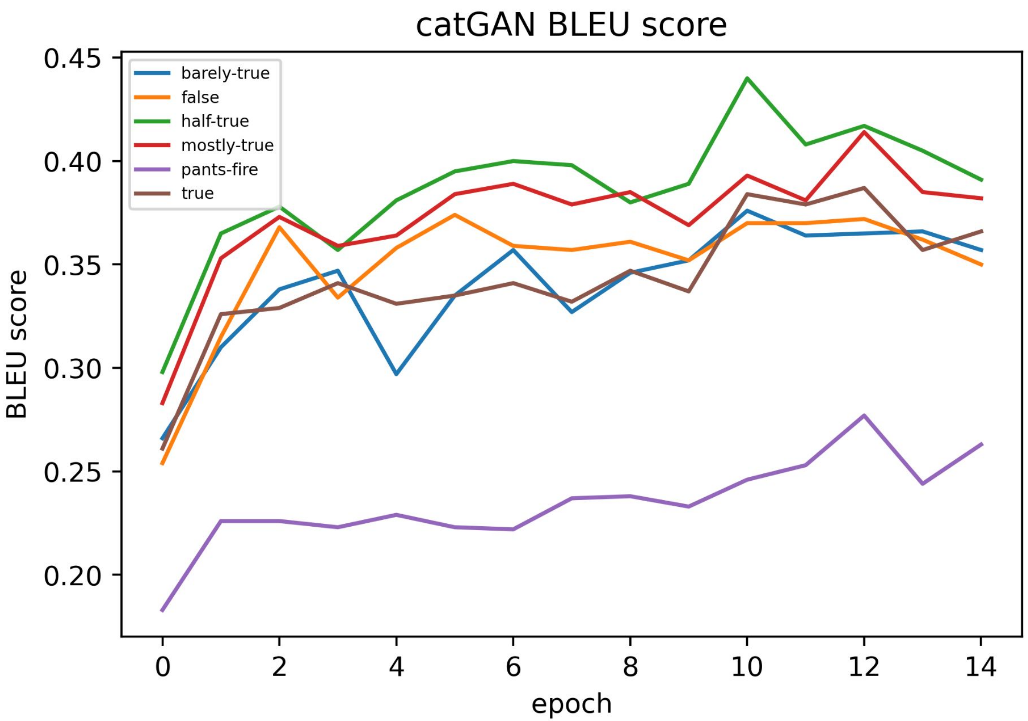
(b) Category-aware Model

CatGAN - Train Metrics (LIAR)





CatGAN - BLEU Score (LIAR)



CatGAN - Results (LIAR)

Label	barely-true	false	half-true	mostly-true	pants-fire	true	**AVG**
Accuracy (no preprocessing)	8.89	21.12	35.96	18.11	0	19.11	**17.19**
Accuracy (with preprocessing)	0	2.86	96.63	0	0	0	**16.58**

Potential Improvements & Further Research

- Make the evaluation and testing **consistent** for all of the architectures used
- Applying **ensemble learning** with models of different complexity and word embeddings
- Using **multiple architectures for each of the inputs** (for example metadata, statement, justification in LIAR-PLUS)
- Training on more **powerful hardware** with **more epochs** and **training optimisation** (early stopping, different schedulers, and hyperparameters)

Thank you for your attention!

Shall you have any questions,
please do not hesitate to ask.