Transformer-based Multitask Learning German Sexism Detector

GERMEVAL2024 GERMS-DETECT (CLOSED TRACKS 1 & 2)

LMR

FAKHRI MOMENI, M. TAIMOOR KHAN

GESIS – LEIBNIZ INSTITUTE FOR THE SOCIAL SCIENCES, GERMANY

Outline

Introduction

Challenge task

Proposed approach

Results

Conclusion

Limitations and future work

Introduction

Sexism/Misogyny

- Sexism is prejudice, stereotyping or discrimination (typically against women) based on their sex or offensive use of language against women.
- Misogyny is dislike, contempt for, or prejudice against women

It can be in the form of explicit jokes or suggested in a subtle way with implied context

Sexism detection: Analyzed as a prediction problem

Task (closed task1 and task2)

This shared task is about the detection of sexism/misogyny in comments posted in (mostly) German language to the comment section of an Austrian online newspaper.

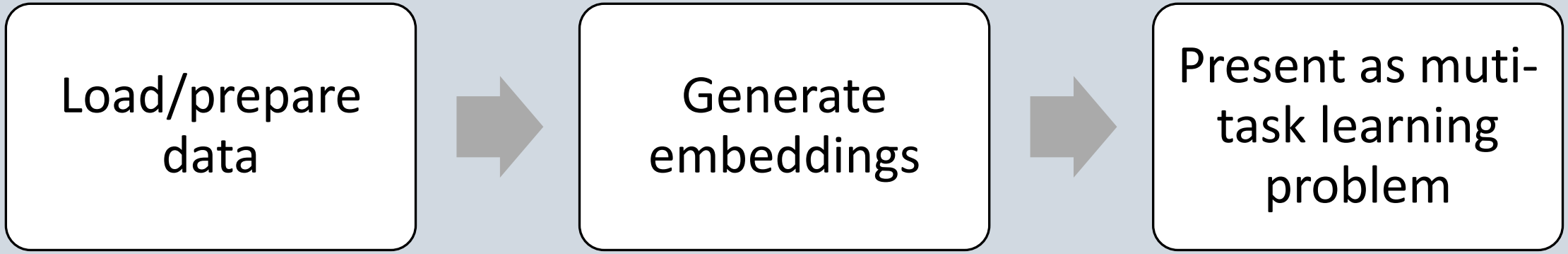
Challenges:

- Annotations instead of labels (may very and have a tie too)
- Multiple prediction tasks
  - Binary class classification
  - Multi class classification
  - Regression
- Cannot use pretrained models

Evaluation metric:

- Macro F1 for classification (closed task 1)
- Jensen-Shannon distance for binary and multi score distribution (closed task 2)

Proposed Approach



Intuition for Multi-task learning

- Same input data for all tasks
- Similar tasks (in nature)
- Can benefit from sharing jointly learnt features through common layers

Preparing data

Preprocessing

- Noise filtered (numbers and special characters)
- Training samples < 2 words filtered

Aggregating annotations [0-kein, 1-Gering, 2-Vorhanden, 3-Stark, 4-Extrem]

- Classification
  - Binary majority (0 vs 1,2, 3, 4) annotations
  - Binary one (at least one 0 annotation)
  - Binary all (no 0 annotation)
  - Multi majority (annotation in majority)
  - Disagree binary (0 vs 1,2,3,4) annotation
- Distribution prediction
  - Binary probability distribution (0 vs 1,2,3,4)
  - Multi distribution prediction (0, 1, 2, 3, 4)

Imbalanced data

Task	Percentage of labels
bin_maj [0, 1]	65 %, 34 %
bin_one [0, 1]	46 %, 53 %
bin_all [0, 1]	83 %, 16 %
multi_maj [0, 1, 2, 3, 4]	67%, 3%, 15%, 10%, 1%

Training samples: 7984

Unique words: 32,362

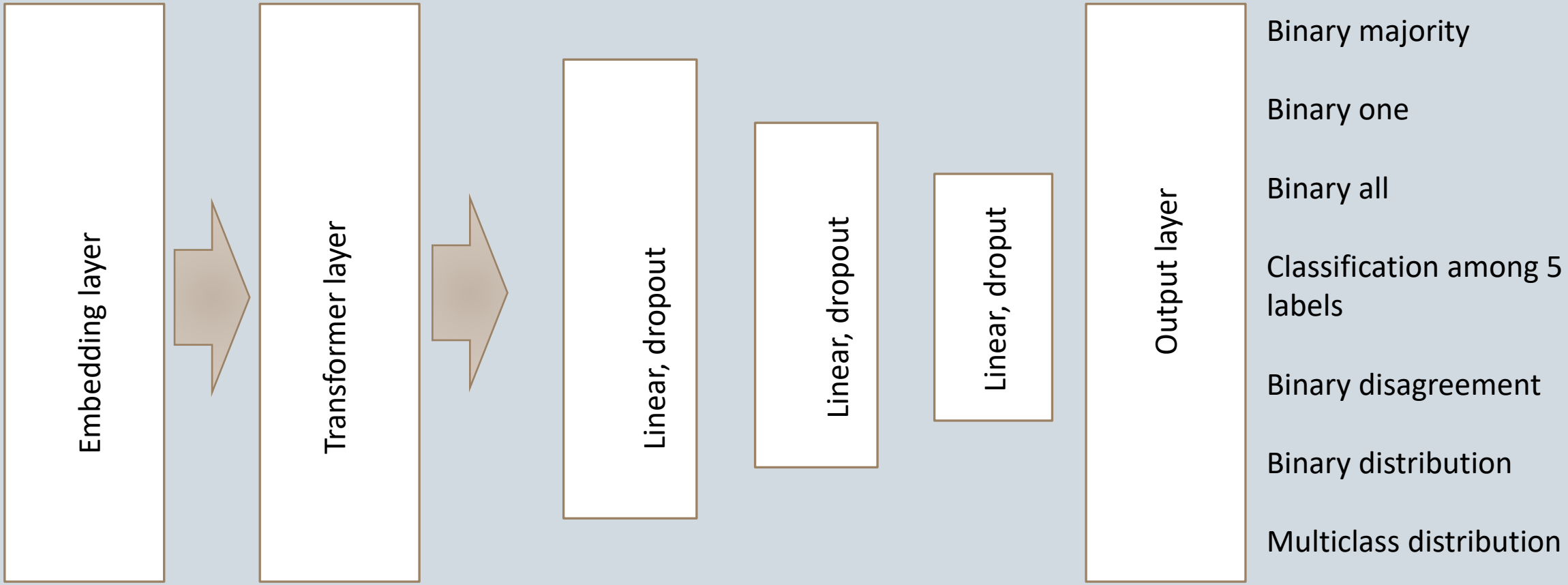
Sample size: 173 words (longest), 0 words (shortest), Avg: 32.86

Model setup

maxLen = 80

Loaded 'cc.de.300.vec’ embeddings (default 300 dimensions)

Network Architecture

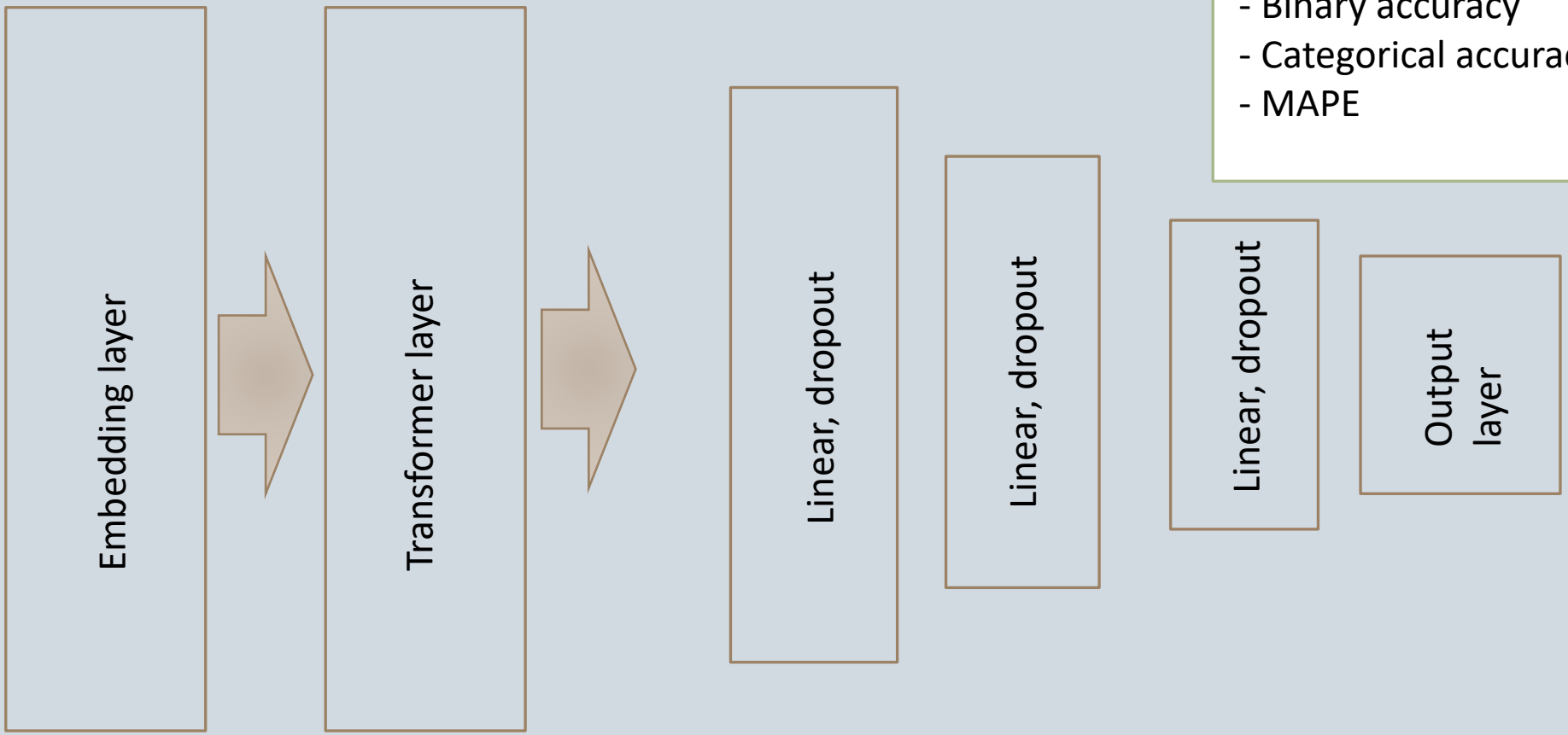


40 Batch size, 5 epochs, 0.2 validation split, shuffle=true

Network Architecture

Loss:
- Binary focal cross entropy
- Categorical focal cross entropy
- MSE

Metric:
- Binary accuracy
- Categorical accuracy
- MAPE



Embedding layer	Transformer layer	Linear, dropout	Linear, dropout	Linear, dropout	Output layer
Untrainable 300 size	Two attention heads 30 size for FF layer 0.1 dropout rate	Leaky_relu 0.3 dropout rate			Sigmoid for binary classification Softmax for multiclass classification Linear for distributions

Results

Team	Model	Rank	Score	MultiMaj F1	BinMaj F1	BinOne F1	BinAll F1	DisagreeBin F1
LMR	deSeDector	6	0.476	0.273	0.584	0.521	0.536	0.46

Team	Model	Rank	Score	JS Dist Multi	JS Dist Bin
LMR	deSeDector	5	0.388	0.426	0.349

Conclusion

We proposed a transformer based multi-task learning architecture to perform multiple classification/distribution tasks as one architecture

The network architecture is simple and light-weight

The loss functions suited for imbalanced data improved accuracy

The approach yielded good results and needs to be investigated further

Limitations and Future work

The feature representation of tasks differ despite using same data for similar tasks.

- In future we will explore separation of common layers and task specific layers
- Can also split into 2 or 3 multi-task learning models by combining highly similar tasks for the same model

A modest architectural was used to cut on use of hardware resources and training time.

- A bigger architecture should used to learn better nuanced features across all tasks

Our approach to address the skewed representation can further be improved through a better technique

Thank you