BOUN-TABI@SMM4H’22: Text-to-Text Adverse Drug Event Extraction with Data Balancing and Prompting

Gökçe Uludoğan and Zeynep Yirmibeşoğlu
Department of Computer Engineering, Boğaziçi University

- **Viewed ADE classification and extraction as sequence-to-sequence problems → two systems based on the T5 model**
- **Observed a significant class imbalance → Over- and undersampling**
- **Prompting for ADE extraction**
- **Models outperformed the current state-of-the-art method**

Official Results

Task 1a: ADE Classification

Model	Precision	Recall	F1
Our model	**0.688**	0.625	**0.655**
Mean	0.646	0.497	0.562
Magge et al. (2021)	0.61	0.64	0.63

Task 1b: ADE Extraction

	Partial			Strict		
Model	P	R	F1	P	R	F1
Our model	0.507	**0.549**	**0.527**	**0.384**	**0.412**	**0.398**
Mean	**0.539**	0.517	**0.527**	0.344	0.339	0.341
Magge et al. (2021)	0.53	0.38	0.44	-	-	-

Scan the QR code to go to the **repository**

Methodology

1. Model

Text-to-Text Transfer Transformer (T5) model (Raffel et. al, 2019)

- Task-denoting prefix: **<<assert ade>>** or **<<ner ade>>**
- Fine-tuned separately for classification and extraction

2. Data Balancing

To eliminate the class imbalance

- Over- and undersampling
- 1:1 and 2:1 (noADE:ADE) ratios

3. Prompting

Templates applied to input and output text

- Probability of text can be modelled
- Useful in low-resource scenarios
- Used prompting for extraction
- Three templates

Input	Output
Is there a negative drug effect in : [X]	[Y] is a negative drug effect. There isn’t a negative drug effect.
Did the patient suffer from a side effect? [X]	Yes, the patient suffered from [Y]. No, the patient didn’t suffer from a side effect.
[X] Did the patient suffer from a side effect?	Yes, the patient suffered from [Y]. No, the patient didn’t suffer from a side effect.

4. Ensemble Modeling

To compensate for the strengths and weaknesses of the models

- Majority voting
- Chose the span predicted by at least half of the models
- Combined predictions of different models by taking intersection

Validation Results

Task 1a: ADE Classification

Model	Precision	Recall	F1
Raw data	0.75	0.69	0.72
Balanced data (1:1 ratio)	0.75	0.86	0.80
Balanced data (2:1 ratio)	0.73	0.86	0.79

Task 1b: ADE Extraction

Model	Partial F1	Strict F1
Raw data	0.605	0.481
Balanced data (1:1 ratio)	0.612	**0.503**
Balanced data (2:1 ratio)	0.639	0.482
Prompt/T1	0.636	0.424
Prompt/T2	**0.662**	0.408
Prompt/T3	0.638	0.393
Ensemble	0.657	0.500

References
1. *Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. J. Mach. Learn. Res., 21(140), 1-67.*
2. *Magge, A., Tutubalina, E., Miftahutdinov, Z., Alimova, I., Dirkson, A., Verberne, S., ... & Gonzalez-Hernandez, G. (2021). DeepADEMiner: a deep learning pharmacovigilance pipeline for extraction and normalization of adverse drug event mentions on Twitter. Journal of the American Medical Informatics Association, 28(10), 2184-2192.*

gokce.uludogan@boun.edu.tr
zeynep.yirmibesoglu@boun.edu.tr