NLP4DL Workshop - EGC 2022 (Blois, France)

Fine-tuning Pre-trained Transformer Language Models for Biomedical Event Trigger Detection

Laura Zanella and Yannick Toussaint

LORIA (Université de Lorraine, CNRS, Inria)

Café TAL

October 3, 2022

Introduction

Introduction

- Biomedical event extraction is a complex information extraction task that helps to identify key information from large sets of textual data for further applications;
    - e.g. pathway curation, study of biomolecular mechanisms of infectious diseases, drug labeling and epigenetic changes.

Introduction

- Biomedical event extraction is a complex information extraction task that helps to identify key information from large sets of textual data for further applications;
    - e.g. pathway curation, study of biomolecular mechanisms of infectious diseases, drug labeling and epigenetic changes.

- A biomedical event contains an event trigger and one or more arguments;
    - Event triggers generally refer to nouns or verbs that express a circumstance, process or eventuality.
    - Arguments refer to biomedical entities or other events.

Biomedical event extraction: Definition

‘[...] O3-cytokine associations were mediated by aMT6s.’

Biomedical event extraction: Definition

‘[...] O3-cytokine associations were mediated by aMT6s.’

Biomedical event extraction: Definition

Event 1 - *Binding*

- **Trigger word**: ‘associations’
- **Trigger category**: Binding
- **Argument word**: ‘O3’
- **Argument category**: Simple_chemical
- **Role**: Theme

Biomedical event extraction: Definition

Event 1 - *Binding*

- **Trigger word**: ‘associations’
- **Trigger category**: Binding
- **Argument word**: ‘O3’
- **Argument category**: Simple_chemical
- **Role**: Theme

Event 2 - *Regulation*

- **Trigger word**: ‘mediated’
- **Trigger category**: Regulation
- **Argument word 1**: *Event 1*
- **Role 1**: Theme
- **Argument word 2**: ‘aMT6s’
- **Argument category 2**: G_or_G_P
- **Role 2**: Cause

Biomedical event trigger detection: Definition

- The extraction of biomedical events can be divided into two main sub-tasks; event trigger detection and argument identification.

Biomedical event trigger detection: Definition

- The extraction of biomedical events can be divided into two main sub-tasks; event trigger detection and argument identification.

- Event trigger detection identifies and classifies the event triggers into a set of predefined categories of events.
    - It has a critical role in building events, since the triggers are the targets that allow the construction of an event.
    - More than 60 % of biomedical event extraction errors occur in this sub-task [8].

Event trigger detection is challenging!

- The same event can be represented in the form of different expressions;
  - '[...] the mechanism that activates$_{+Reg}$ infiltrating macrophages [...]'
  - '[...] a limited role in post-ischemic macrophage activation$_{+Reg}$.'
  - '[...] antibodies that activated$_{?}$ inflammatory cytokine expression.'

Event trigger detection is challenging!

- The same event can be represented in the form of different expressions;
  - '[...] the mechanism that activates$_{+Reg}$ infiltrating macrophages [...]'
  - '[...] a limited role in post-ischemic macrophage activation$_{+Reg}$.'
  - '[...] antibodies that activated$_{?}$ inflammatory cytokine expression.'

- They can be represented as single words or multi-words;
  - '[...] the neuroprotective efficacy$_{-Reg}$ of epigallocatechin-3-gallate (EGCG) [...]'
  - '[...] TLR4/NF-B pathway in LPS+A-induced$_{+Reg}$ rat microglia [...]'

Event trigger detection is challenging!

- They can present in-domain language or not;
  - '[...] Noncanonical Inflammasome via TLR4/NF-B Pathway$_{Pathway}$.'
  - '[...] and Neurotoxicity$_{Carcinogenesis}$ by Suppressing the Activation of Inflammasome'.
  - 'EGCG attenuates$_{_Reg}$ microglial inflammation [...]'

Event trigger detection is challenging!

- They can present in-domain language or not;
  - '[...] Noncanonical Inflammasome via TLR4/NF-B Pathway$_{Pathway}$.'
  - '[...] and Neurotoxicity$_{Carcinogenesis}$ by Suppressing the Activation of Inflammasome'.
  - 'EGCG attenuates$_{_Reg}$ microglial inflammation [...]'

- The same word can represent a different event according to the context (or the manual annotation?);
  - 'Anti-Survival and Pro-Apoptotic Effects$_{Reg}$ of 6-Shogaol [...]'
  - '[...] has anticancer effects$_{_Reg}$ on many types of tumors'.

Biomedical event trigger detection: Techniques

- Neural networks have been widely adopted for event trigger detection since they do not require the design of functions or use additional tools for their training.

- Models pre-trained on transformers architectures are commonly used for solving NLP tasks due to their positive achievements in performance.

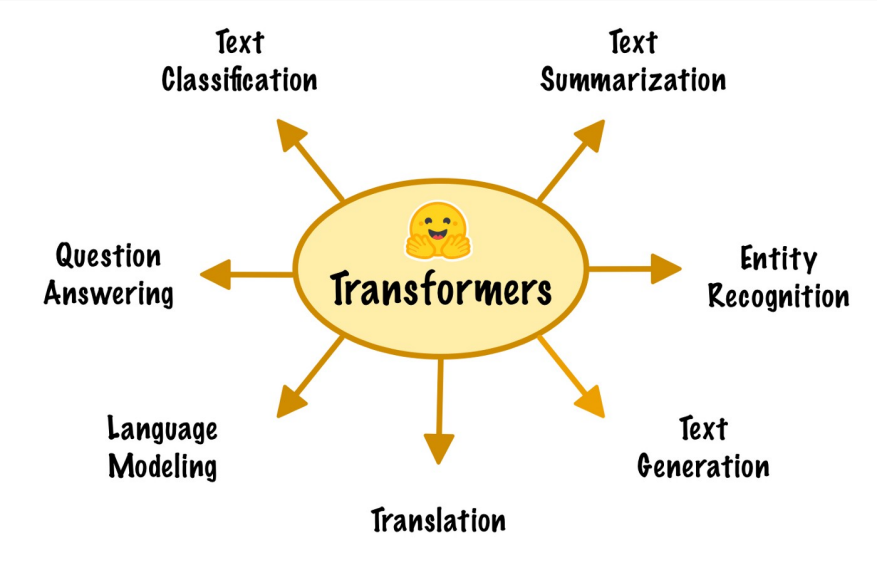
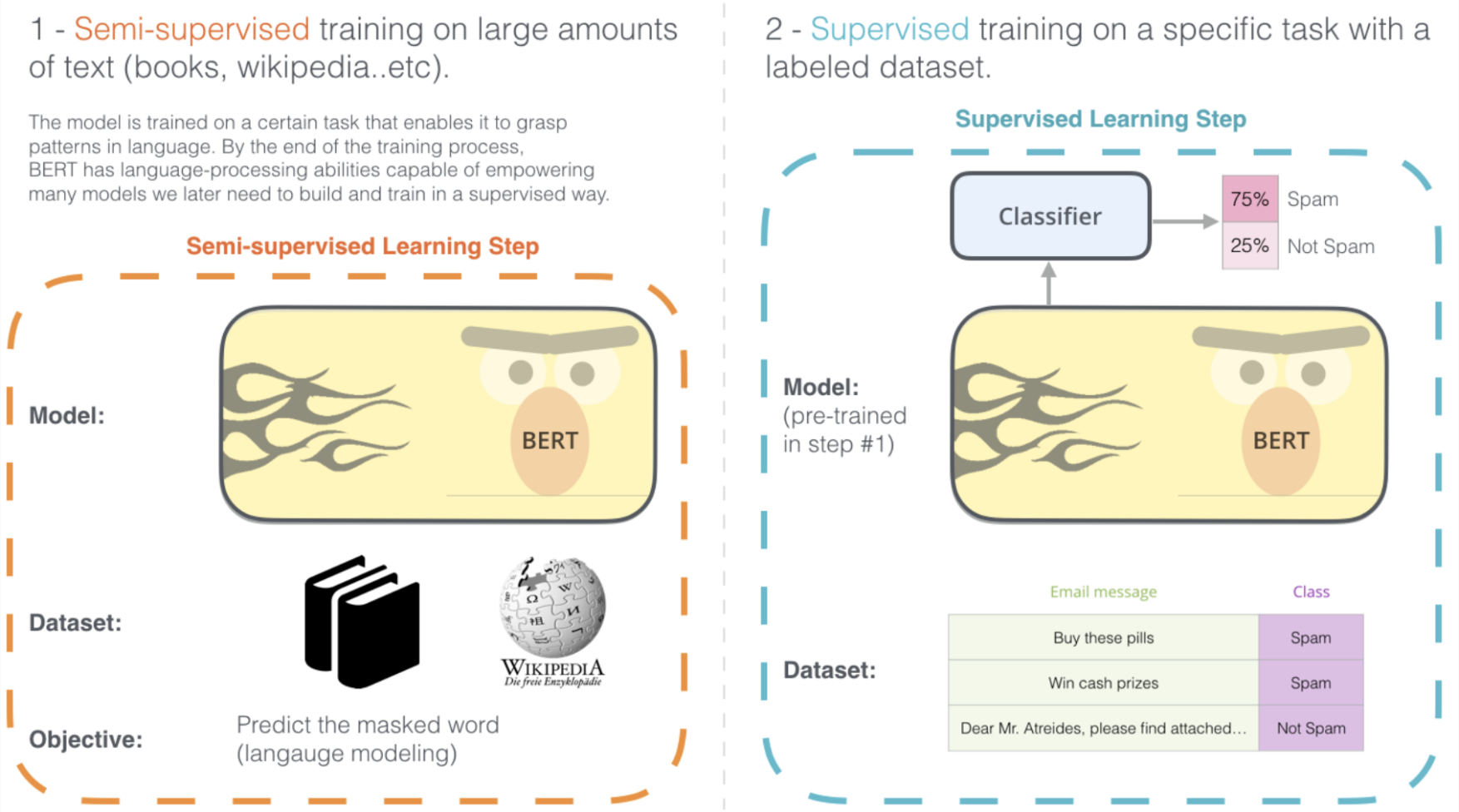


Figure obtained from: https://towardsdatascience.com/transformers-implementing-nlp-models-in-3-lines/-of-code-475639c3611d

Bidirectional Encoder Representations from Transformers



- BERT variants in the biomedical domain: BioBERT, PubMedBERT, BioMedRoBERTa, among others.

Figure obtained from `https://jalammar.github.io/illustrated-bert/`.

Work proposal and Contributions

This work compares five transformer language models (BERT, BioBERT, SciBERT, PubMedBERT and BioMedRoBERTa) for detecting biomedical event triggers using seven merged biomedical datasets to identify which model is the most appropriate for tackling this task;

Work proposal and Contributions

This work compares five transformer language models (BERT, BioBERT, SciBERT, PubMedBERT and BioMedRoBERTa) for detecting biomedical event triggers using seven merged biomedical datasets to identify which model is the most appropriate for tackling this task;

1. Identify whether using a transformer model pre-trained on the biomedical domain language presents advantages in performance.

2. Analyze whether using different biomedical corpus together for the models' training can improve event triggers detection.

Methods

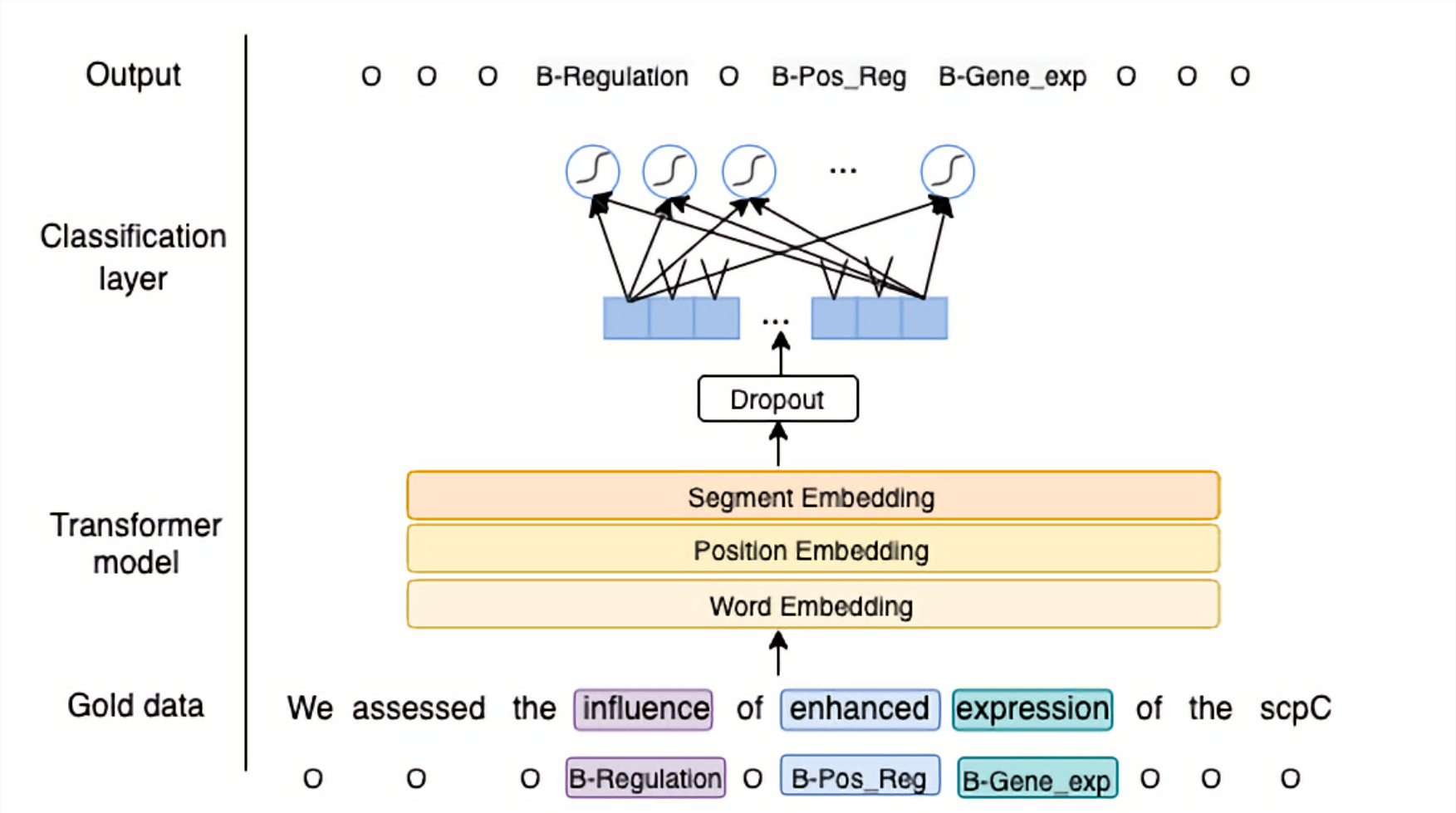
Detection of biomedical event triggers

The detection of biomedical event triggers can be considered as a multi-class problem;

IOB (Inside-Outside-Beginning) annotation of event triggers

‘PTHrP	**drives**	breast	tumor	**initial**	**progression**	and	**metastasis** ...’
‘O’	*‘B-Regulation’*	*‘O’*	*‘O’*	*‘B-Development’*	*‘I-Development’*	*‘O’*	*‘B-Metastasis’*

Model proposal



Experimental settings and Evaluation

Statistics of the corpus

Dataset	No. Triggers	No. Events	Documents	Train/Dev/Test
CG 2013	9,790	17,248	PubMed abstracts	300/100/200
EPI 2011	2,035	2,453	PubMed abstracts	600/200/400
GENIA 2011	10,210	13,560	MEDLINE abstracts	1,000 (total)
GENIA 2013	4,676	6,016	PMC full-text	34 (total)
ID 2011	2,155	2,779	PMC full-text	15/5/10
PC 2013	6,220	8,121	PubMed abstracts	260/90/175
MLEE	5,554	6,677	PubMed abstracts	131/44/87

- Total train and test sets: 19,855 and 4,964 sentences.
- Total trigger classes: 58.

Pre-trained language models

Model	Pre-training	Corpus
BERT	from scratch	WikiPedia + BookCorpus
BioBERT	from BERT	PubMed
SciBERT	from scratch	PMC* + Semantic scholar
PubMedBERT	from scratch	PMC* + PubMed
BioMedRoBERTa	from BERT	Semantic scholar

- Each model was fine-tuned for 10, 30 and 100 epochs.
- Precision, Recall and F1-score were measured for evaluation.

*PMC = PubMed Central

Evaluation of pre-trained models fine-tuned for trigger detection

Model	10 epochs			30 epochs			100 epochs		
	P	R	F1	P	R	F1	P	R	F1
BERT	**0.57**	**0.67**	**0.62**	**0.60**	**0.68**	**0.64**	0.62	0.68	0.65
BioBERT	0.51	0.61	0.55	0.57	0.59	0.58	**0.72**	**0.75**	**0.73**
SciBERT	0.59	0.64	0.61	0.61	0.65	0.63	0.70	0.70	0.70
PubMedBERT	0.49	0.61	0.54	0.58	0.66	0.61	0.58	0.62	0.60
BioMedRoBERTa	0.48	0.49	0.47	0.52	0.52	0.51	0.55	0.50	0.52

Category grained evaluation (BioBERT - 100 epochs)

Trigger category	P	R	F1	Support	Trigger category	P	R	F1	Support
Amino_acid_catabolism	1.00	1.00	1.00	1	Entity	0.63	0.74	0.68	398
Glycolysis	1.00	0.90	0.95	10	Degradation	0.68	0.68	0.68	19
Acetylation	0.86	0.99	0.92	82	Transcription	0.65	0.70	0.67	175
Phosphorylation	0.89	0.94	0.91	207	Synthesis	1.00	0.50	0.67	2
Deglycosylation	0.83	1.00	0.91	5	Conversion	0.55	0.75	0.64	28
Process	0.84	0.96	0.90	136	Regulation	0.66	0.57	0.61	556
Deacetylation	0.81	1.00	0.90	13	Blood_vessel_development	0.52	0.72	0.60	18
Metastasis	0.84	0.92	0.88	53	Transport	0.62	0.54	0.59	42
Methylation	0.85	0.90	0.87	73	Planned_process	0.65	0.54	0.59	104
Demethylation	0.75	1.00	0.86	3	Metabolism	0.57	0.57	0.57	7
Ubiquitination	0.82	0.90	0.86	67	Cell_death	0.56	0.58	0.57	43
Gene_expression	0.82	0.88	0.85	754	Growth	0.50	0.67	0.57	3
Hydroxylation	0.82	0.85	0.84	27	DNA_demethylation	0.40	1.00	0.57	2
Glycosylation	0.81	0.84	0.82	67	DNA_domain_or_region	0.57	0.57	0.57	7
DNA_methylation	0.82	0.82	0.82	77	Development	0.49	0.54	0.51	39
Cell_differentiation	0.92	0.73	0.81	15	Dephosphorylation	0.33	1.00	0.50	1
Carcinogenesis	0.78	0.81	0.79	31	Deubiquitination	1.00	0.33	0.50	3
Activation	0.78	0.80	0.79	65	Inactivation	0.44	0.53	0.48	15
Protein_catabolism	0.70	0.87	0.78	30	Catalysis	0.38	0.56	0.45	16
Pathway	0.79	0.76	0.78	168	Breakdown	0.40	0.50	0.44	4
Cell_proliferation	0.77	0.73	0.75	37	Mutation	0.45	0.41	0.43	32
Binding	0.72	0.79	0.75	434	Protein_processing	0.25	1.00	0.40	1
Negative_regulation	0.71	0.79	0.75	586	Anaphora	0.23	0.14	0.18	49
Localization	0.71	0.77	0.74	164	Protein_domain_or_region	0.00	0.00	0.00	5
Infection	1.00	0.56	0.71	9	Cell_division	0.00	0.00	0.00	2
Cell_transformation	0.76	0.67	0.71	39	Catabolism	0.00	0.00	0.00	5
Positive_regulation	0.72	0.68	0.70	1,276	Remodeling	0.00	0.00	0.00	1
Dissociation	0.64	0.78	0.70	9	Translation	0.00	0.00	0.00	2
Death	0.69	0.69	0.69	16	Dehydroxylation	0.00	0.00	0.00	1

Examples of ambiguous event triggers in gold data

‘When the dickkopf protein (Dkk) binds to Kremen and LRP5, this last **undergoes internalization and therefore becomes unable to bind Wnt**; this leads to degradation of beta-catenin and to inhibition of bone formation’.

Examples of ambiguous event triggers in gold data

'When the dickkopf protein (Dkk) binds to Kremen and LRP5, this last **undergoes internalization and therefore becomes unable to bind Wnt**; this leads to degradation of beta-catenin and to inhibition of bone formation'.

(annotation from gold data)

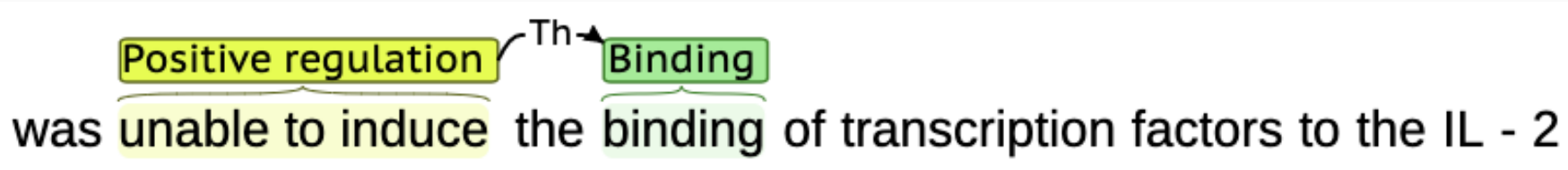
- The model incorrectly classified ***becomes*** as 'Regulation'.

Examples of ambiguous event triggers in gold data

'On the other hand, stimulation of T cells with mAb 9.3 increased the level of intracellular Ca2+ and triggered the activation of p56 (lck) and c-Raf-1, but **was unable to induce the binding of transcription factors to the IL-2** promoter.'

Examples of ambiguous event triggers in gold data

'On the other hand, stimulation of T cells with mAb 9.3 increased the level of intracellular Ca2+ and triggered the activation of p56 (lck) and c-Raf-1, but **was unable to induce the binding of transcription factors to the IL-2** promoter.'



(annotation from gold data)

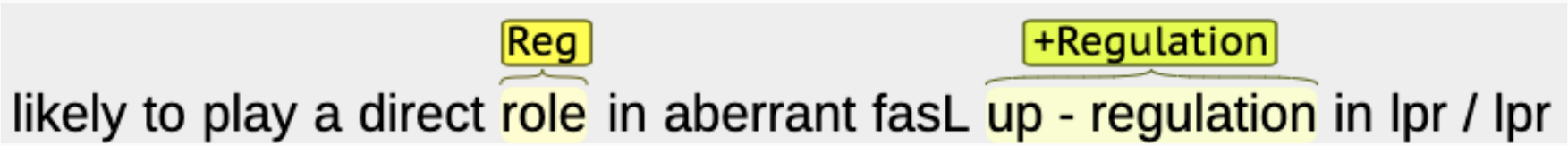
- The model incorrectly classified ***unable to induce*** as 'Negative regulation'.

Examples of ambiguous event triggers in gold data

‘Thus, Egr-2, in addition to Egr-3, regulates FasL expression in activated normal T cells, and Egr-2 is **likely to play a direct role in aberrant fasL up-regulation in lpr/lpr** and gld/gld CD4(-)CD8(-)T cells.’

Examples of ambiguous event triggers in gold data

'Thus, Egr-2, in addition to Egr-3, regulates FasL expression in activated normal T cells, and Egr-2 is **likely to play a direct role in aberrant fasL up-regulation in lpr/lpr** and gld/gld CD4(-)CD8(-)T cells.'



(annotation from gold data)

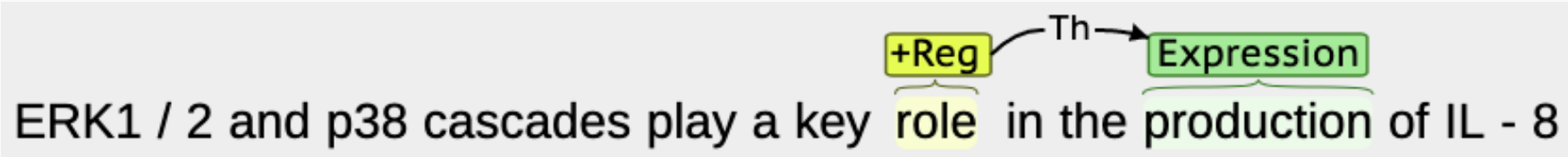
- The model correctly classified ***role*** as 'Regulation'.

Examples of ambiguous event triggers in gold data

‘By using PD-98059 and SB203580, two potent and selective inhibitors of MEK1 and p38, respectively, we have demonstrated that both **ERK1/2 and p38 cascades play a key role in the production of IL-8** by monocytes and PMN stimulated with bacterial fractions.’

Examples of ambiguous event triggers in gold data

'By using PD-98059 and SB203580, two potent and selective inhibitors of MEK1 and p38, respectively, we have demonstrated that both **ERK1/2 and p38 cascades play a key role in the production of IL-8** by monocytes and PMN stimulated with bacterial fractions.'



(annotation from gold data)

- The model incorrectly classified ***role*** as 'Regulation'.

Conclusions and Future work

Conclusions

- Biomedical data presents very specialized language that can be difficult to model for event trigger detection.

Conclusions

- Biomedical data presents very specialized language that can be difficult to model for event trigger detection.
- BioBERT presents the highest performance in detecting biomedical event triggers when it is trained during 100 epochs.
    - This result suggests that a model pre-trained on biomedical data that starts its pre-training from BERT weights is the best strategy for biomedical event trigger detection.

Conclusions

- Biomedical data presents very specialized language that can be difficult to model for event trigger detection.
- BioBERT presents the highest performance in detecting biomedical event triggers when it is trained during 100 epochs.
    - This result suggests that a model pre-trained on biomedical data that starts its pre-training from BERT weights is the best strategy for biomedical event trigger detection.
- Using different corpus merged as input data can enrich event detection, since they provide more trigger categories and samples to train the model.

Conclusions

- Biomedical data presents very specialized language that can be difficult to model for event trigger detection.
- BioBERT presents the highest performance in detecting biomedical event triggers when it is trained during 100 epochs.
    - This result suggests that a model pre-trained on biomedical data that starts its pre-training from BERT weights is the best strategy for biomedical event trigger detection.
- Using different corpus merged as input data can enrich event detection, since they provide more trigger categories and samples to train the model.
- However, the categories of triggers with high number of samples do not necessarily present high performance.
    - These results suggest that the triggers samples may present ambiguities, making it difficult for the model to achieve the generalization, even if the number of samples is relatively significant.

Future work

For the next steps:

- to enrich the information given to the model by adding extra features, as the Parts-Of-Speech [4] or the syntactic dependency path [8], to reduce ambiguities;

- to merge the trigger categories with the lowest support to other categories with similar events to reduce the imbalance of the data.

Fine-tuning Pre-trained Transformer Language Models for Biomedical Event Trigger Detection

Laura Zanella and Yannick Toussaint

LORIA (Université de Lorraine, CNRS, Inria)

October 3, 2022

References I

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova.
Bert: Pre-training of deep bidirectional transformers for language understanding.
arXiv preprint arXiv:1810.04805, 2018.

[2] A. N. Jagannatha and H. Yu.
Bidirectional rnn for medical event detection in electronic health records.
In *Proceedings of the conference. Association for Computational Linguistics. North American Chapter. Meeting*, volume 2016, page 473. NIH Public Access, 2016.

[3] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang.
Biobert: a pre-trained biomedical language representation model for biomedical text mining.
Bioinformatics, 36(4):1234–1240, 2020.

[4] B. Portelli, E. Lenzi, E. Chersoni, G. Serra, and E. Santus.
Bert prescriptions to avoid unwanted headaches: A comparison of transformer architectures for adverse drug event detection.
In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1740–1747, 2021.

References II

[5] P. V. Rahul, S. K. Sahu, and A. Anand.
Biomedical event trigger identification using bidirectional recurrent neural network based models.
arXiv preprint arXiv:1705.09516, 2017.

[6] A. Ramponi, R. van der Goot, R. Lombardo, and B. Plank.
Biomedical event extraction as sequence labeling.
In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5357–5367, 2020.

[7] H.-L. Trieu, T. T. Tran, K. N. Duong, A. Nguyen, M. Miwa, and S. Ananiadou.
Deepeventmine: end-to-end neural nested event extraction from biomedical texts.
Bioinformatics, 36(19):4910–4917, 2020.

[8] J. Wang, J. Zhang, Y. An, H. Lin, Z. Yang, Y. Zhang, and Y. Sun.
Biomedical event trigger detection by dependency-based word embedding.
BMC medical genomics, 9(2):123–133, 2016.