# Zero-shot and Few-shot classification of biomedical articles in context of the COVID-19 pandemic

1

Simon Lupart

Grenoble INP – Ensimag

Laboratoire d'Informatique et Systèmes – Naver Labs Europe

5 Avril – 19 Septembre 2021

### Contents



- Introduction
  - Problematic
  - Biomedical domain



- Architectures
  - Zero-shot architecture
  - Learning the hierarchy
  - Zero-shot input



Results

- Results in zero/few-shot
- Results of the probing tasks

Oiscussion and Limits

### Contents

(1)

#### Introduction

- Problematic
- Biomedical domain



- Architectures
  - Zero-shot architecture
  - Learning the hierarchy
  - Zero-shot input



Results

- Results in zero/few-shot
- Results of the probing tasks

Oiscussion and Limits

### Context of the pandemic

- Creation of the Bibliovid association to support scientific monitoring
  - Team of doctors from the Grenoble University Hospital
- Need for automatic annotation tools. The difficulty comes from the biomedical domain which has evolved significantly with the pandemic:
  - Appearance of **new concepts**, on which there is little or no data.



### Context of the pandemic

- Creation of the Bibliovid association to support scientific monitoring
  - Team of doctors from the Grenoble University Hospital
- Need for automatic annotation tools. The difficulty comes from the biomedical domain which has evolved significantly with the pandemic:
  - Appearance of **new concepts**, on which there is little or no data.



Zero-shot learning is a problem setup in machine learning, where at test time, a learner observes samples from **classes that were not observed during training**. Respectively in few-shot, the model will have observed a small number of documents with these classes.



Introduction – Biomedical domain

## Medline – National Library of Medicine

Medline is a biomedical dataset:

- Millions of annotated biomedical articles
- Around **29,933 classes** (29,933 MeSH)
- Structured as a hierarchy
- MeSH have multiple positons in the hierarchy

MeSH	COVID-19	Bronchitis
Tree Number	C01.748.214	C01.748.099
All MeSH	Ø	Ø
Diseases Category		
Infections	-01	-01
Respiratory Tract Infections	748	748
	214	099 0 ' ``

Introduction – Biomedical domain

### Medline – National Library of Medicine

Medline is a biomedical dataset:

- Millions of annotated biomedical articles
- Around 29,933 classes (29,933 MeSH)
- Structured as a hierarchy
- MeSH have multiple positons in the hierarchy





### Data and MeSH

We work on a subset of documents from Medline

- 27,321 COVID-19 related articles
- with 8,140 MeSH descriptors
- 747 held-out MeSH for zero-shot
- Documents annotated with 13 MeSH (average)



### Data and MeSH

We work on a subset of documents from Medline

- 27,321 COVID-19 related articles
- with 8,140 MeSH descriptors
- 747 held-out MeSH for zero-shot
- Documents annotated with 13 MeSH (average)



#### **Objectives and Motivations**

- Compare zero-shot and few-shot classification results
- 2 Encode hierarchical information about MeSH to be able to perform zero-shot better

### Contents



- Introduction
  - Problematic
  - Biomedical domain

#### **2** Architectures

- Zero-shot architecture
- Learning the hierarchy
- Zero-shot input



Results

- Results in zero/few-shot
- Results of the probing tasks

Oiscussion and Limits

Architectures – Zero-shot architecture

#### Learning the semantic

- Classic classification architecture do not adapt to zero-shot, because the output is static.
- Open architectures enable zero-shot, because they allow the model to learn the semantics of the classes.

#### Examples (Learning the semantic)

- Classes similarities: « Breast Cancer » / « Kidney Cancer »
- Description of the classes: « Infection: Invasion of the host organism by microorganisms or their toxins or by parasites that can cause pathological conditions or diseases. »



### Models

- Two architectures fine-tuned on BioBERT:
  - **BERT-1-block:** with the [SEP] token and a dense classification layer
  - **BERT-2-blocks**: with two separate inputs, and a projection or an attention mecanism



### Decoder

- Prediction of the hierarchy with a decoder block that generate the Tree Number of the given MeSH.
- Multi-Tasks Learning:
  - Common trunc: the BioBERT block
  - Two branches: classification and decoder
- Learning a mutual representation benefic for the two tasks.



#### Multi-Tasks coefficients

$$loss_{total} = \sum_{i=0}^{n} c_i \times loss_i + regularization$$

$$loss_{total} = \frac{1}{2\sigma_1^2} \times loss_1 + \frac{1}{2\sigma_2^2} \times loss_2 + \log(\sigma_1\sigma_2)$$

Architectures – Zero-shot input

### Transformation of the dataset

The open architecture requires pairs ([MeSH, Abstract], boolean)

Example with 4 classes: « Diagnostic », « Therapeutic », « Prognostic » et « Recommendations »					
• Classic model:	<i>Input:</i> Abstract_1		<i>Output:</i> (True, False, False, False)		
• Open architecture:	<i>Input:</i> (« Diagnostic », (« Therapeutic », (« Prognostic », (« Recommendations »,	Abstract_1) Abstract_1) Abstract_1) Abstract_1)	<i>Output:</i> True False False False		

Scaling problem with the 9,000 MeSH descriptors. We need to select which negatives we want to add.

Architectures – Zero-shot input

### Transformation of the dataset (2)

Choice of the pairs ([MeSH, Abstract], boolean)

First we add the annotated MeSH to an abstract as positives

#### **Balanced**:

- As many negatives as positives
- Given a MeSH, this MeSH should appear as many time as positive than as negative
- (Never add ancestors as negatives)

#### Train/Test configurations

- Training: Balanced
- Test: Balanced and Siblings

#### **2** Siblings:

- Add all the ancestors as positives
- Add all the siblings as negatives



### Contents



- Introduction
  - Problematic
  - Biomedical domain



- Architectures
  - Zero-shot architecture
  - Learning the hierarchy
  - Zero-shot input

#### **3** Results

- Results in zero/few-shot
- Results of the probing tasks

Oiscussion and Limits

### Results zero-shot

• Results with respect to the test dataset in zero-shot and few-shot.



### Results zero-shot

- Despite the lower f1-score on both balanced and zsc balanced test dataset, the BERT-1block MTL has a overall better precision
- This precision is benefic in setting where you have difficult negatives (i.e the siblings setting)



### Results zero-shot

- Despite the lower f1-score on both balanced and zsc balanced test dataset, the BERT-1block MTL has a overall better precision
- This precision is benefic in setting where you have difficult negatives (i.e the siblings setting)



## Results by frequency

 Results with respect to the number of occurrences of the MeSH in the train set (in siblings settings):

#### Interpretations

- The more the MeSH appears, the better are the results
- 2 MTL model seems better in low ressources setups



# Probing Tasks

#### Definition (Probing model)

Auxiliary models of **diagnosis**, that we train on a main model. The goal is to extract **internal representations** of the main model, and then try to predict sub-information from those representations. The objective is to assess whether an information is implicitly included in the internal representation of our model or not.

Prediction of the length of the shortest path MeSH to MeSH

Probe	Distance Error (std)
BERT without finetuning (ft)	1.494 (-)
Random	2.597 (-)
BERT-1-block STL	1.462(0.053)
BERT-1-block MTL	$\underline{1.323}$ (0.013)
$\operatorname{BERT-2-blocks+proj}\operatorname{STL}$	$1.490\ (0.054)$
BERT-2-blocks+proj MTL	$1.353\ (0.014)$

#### Prediction if common ancestors



### Contents



- Introduction
  - Problematic
  - Biomedical domain



- Architectures
  - Zero-shot architecture
  - Learning the hierarchy
  - Zero-shot input



Results

- Results in zero/few-shot
- Results of the probing tasks

#### Oiscussion and Limits



#### Discussion and Limits

### Limits – Multi-tasks Learning

- The prediction of the hierarchy task is more complex than the classification task
- Not enough iterations to enable the model to learn the hierarchy



#### Interpretations

• The secondary task can be seen as a **regularization task**, that makes the model more robust.

#### Discussion and Limits

### Limits – MeSH Hierarchy

#### Example: Hierarchy of the MeSH « COVID-19 »

In the annotation « COVID-19 » can be annotated without « Infection ».

- Try to add all the ancestors, to make the secondary task easier for the model.
- It would increase the number of inputs, but it's possible.

All MeSH Categories <u>Diseases Category</u> <u>Respiratory Tract Diseases</u> <u>Lung Diseases</u> <u>Pneumonia</u> <u>Pneumonia, Viral</u> **COVID-19** 

All MeSH Categories <u>Diseases Category</u> <u>Respiratory Tract Diseases</u> <u>Respiratory Tract Infections</u> **COVID-19** 

All MeSH Categories <u>Diseases Category</u> <u>Respiratory Tract Diseases</u> <u>Respiratory Tract Infections</u> <u>Pneumonia</u> <u>Pneumonia</u>, Viral **COVID-19**  All MeSH Categories <u>Diseases Category</u> <u>Infections</u> <u>Respiratory Tract Infections</u> COVID-19

All MeSH Categories

Diseases Category <u>Infections</u> <u>Respiratory Tract Infections</u> <u>Pneumonia</u> <u>Pneumonia, Viral</u> **COVID-19** 

All MeSH Categories Diseases Category Infections Virus Diseases Pneumonia, Viral COVID-19

#### All MeSH Categories

Diseases Category Infections Virus Diseases <u>RNA Virus Infections</u> <u>Nidovirales Infections</u> <u>Coronaviridae Infections</u> Coronavirus Infections

COVID-19

24

### Limits – Deepness

Results with respect to the deepness of the MeSH

#### Interpretations

- ① Deeper MeSH are not necessarly more complex.
- Precision of the MTL model is better on some deeper MeSH.



### Conclusion

#### Zero-shot

Learning the semantic of the classes is possible, and this enable models to adapt to zero-shot settings with new classes.

#### Comparison MTL and STL

In the current settings, it is difficult to say that the MTL model is really better than the STL one, but still we see that he has a **better representation of the hierarchy**.

#### Limits

**Complex Task**, we also lack knowledge about the MeSH hierarchy, and how is the annotation of Medline articles made.

# Questions

Discussion and Limits

### LitCovid and Bibliovid

Transfer to new dataset: LitCovid and Bibliovid

Litcovid	F1-score	Precision	Recall
Baseline IsIn	0.329	0.520	0.241
Baseline Cosine Similarity	0.308	0.228	0.471
BERT-1-block STL	0.512	0.444	0.604
BERT-1-block MTL	0.465	0.401	0.553
BERT-2-blocks+proj STL	0.478	0.394	0.609
BERT-2-blocks+proj MTL	0.480	0.530	0.438
BERT-2-blocks+att $STL$	0.379	0.408	0.353
BERT-2-blocks+att MTL	0.452	0.347	0.649