Deep Learning for Anomaly Detection and Cause Identification

Quentin Fournier

quentin.fournier@polymtl.ca

École Polytechnique de Montréal
Laboratoire DORSAL

How to leverage the event fields?

Why Use the Event Fields?

- The information is already available...
- ...or it can be made available at a low cost.
- The more information is considered, the better will be the model[1].
- In particular neural networks strive with more data.

[1] There are some caveats, for example, models may be subject to the curse of dimensionality.

Which Event Field to Consider?

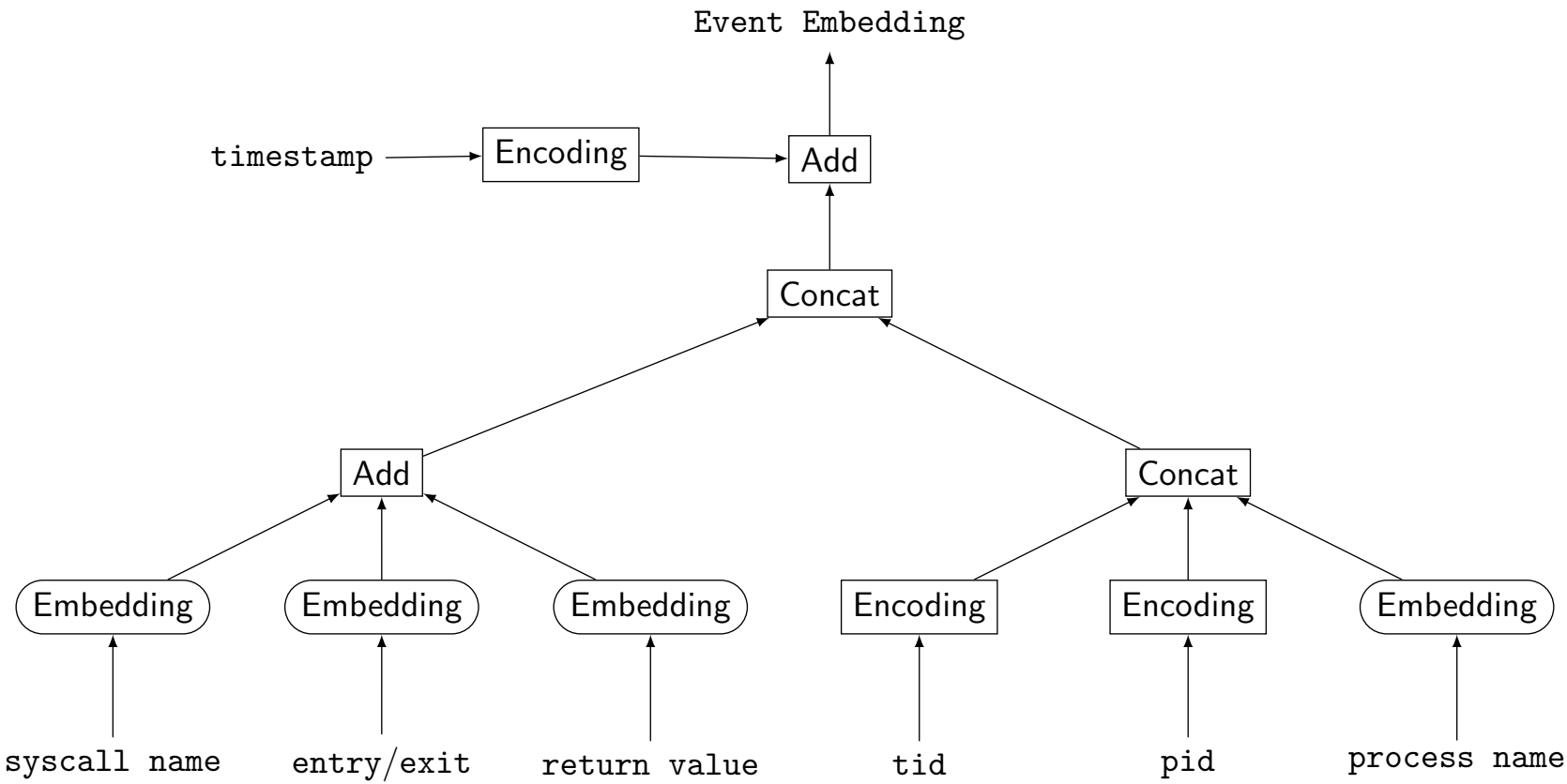
- Let us focus on the system calls.
- Some fields are in virtually every event:
  - `system call name`
  - `timestamp`
  - `entry/exit`
  - `return value`
  - `process name`
  - `process id (pid)`
  - `thread id (tid)`

How to Use the Event Fields?

- Models take as input a vectorial representation of the event.
- This representation should take into account the fields.
- The simplest representation is one-hot encoding:
  - `event a` $\rightarrow \{1, 0, 0, ..., 0\}$
  - `event b` $\rightarrow \{0, 1, 0, ..., 0\}$
  - ...
- Not practical... The representation must be learned.

How to Use the Event Fields?

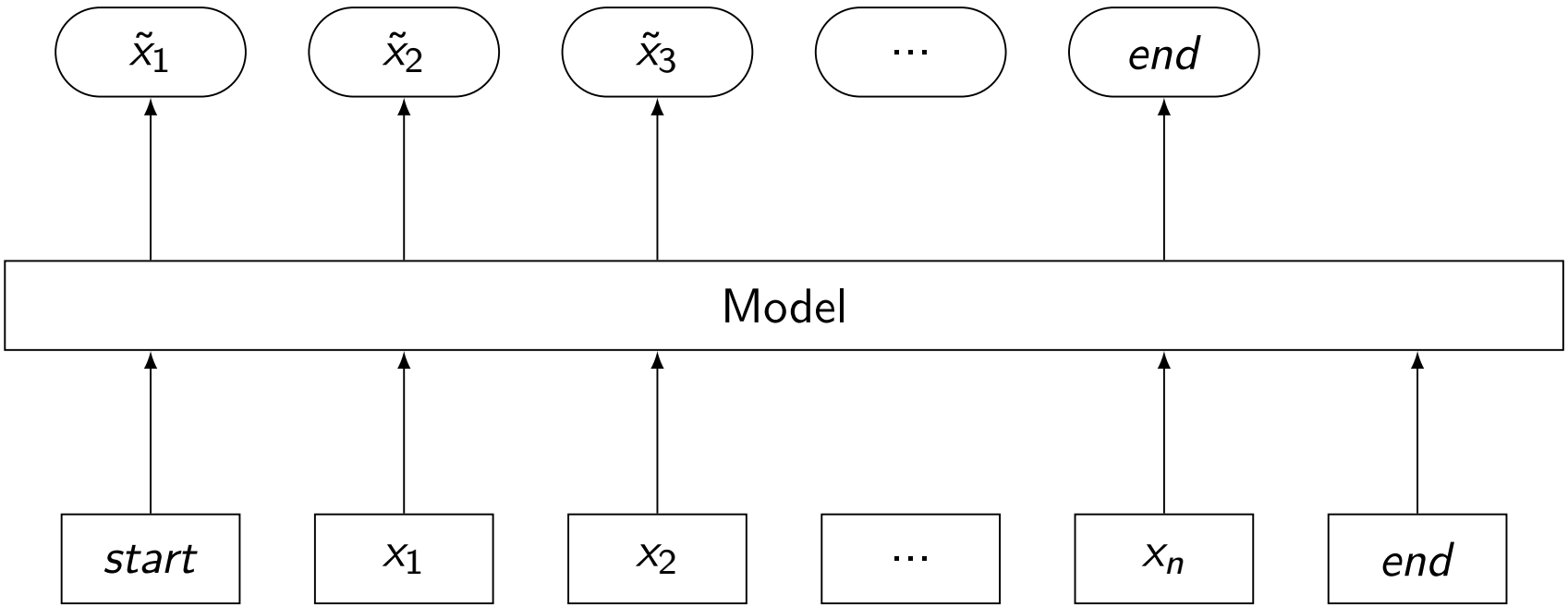
$$Encoding(x) = \left[sin\left(\frac{x}{1}\right), \ldots, sin\left(\frac{x}{n}\right) \right]$$



LM Objective

- Let us consider the task of Language Modelling (LM).
- Predict the next event based on the previous ones.
- Do not require labels.
- Allow computing the likelihood of a sequence.
- Sequences with a low probability may be anomalies.

LM Objective - Example



Experiments

- The model is a neural network.
- Sequences correspond to web requests.
- Inputs x_i are the vectorial representation of system calls.

Results - LSTM

Model	Cross-entropy		Accuracy		Training time
	Train	Test	Train	Test	
2-gram			31.5	31.5	0h29
3-gram			43.1	43.2	0h29
4-gram			46.7	46.8	0h31
{`system call name`}	0.074	0.073	97.3	97.4	0h44
Previous + {`timestamp`}	0.076	0.074	97.3	97.3	0h41
Previous + {`entry/exit, success/failure`}	0.066	0.065	97.6	97.7	0h38
Previous + {`process name, pid, tid`}	0.061	**0.059**	97.8	**97.9**	0h52

Results - Transformer

Model	Cross-entropy		Accuracy		Training time
	Train	Test	Train	Test	
2-gram			31.5	31.5	0h29
3-gram			43.1	43.2	0h29
4-gram			46.7	46.8	0h31
{system call name}	0.100	0.084	96.6	97.1	0h53
Previous + {timestamp}	0.165	0.117	94.1	96.0	1h24
Previous + {entry/exit, success/failure}	0.097	0.081	96.7	97.2	0h50
Previous + {process name, pid, tid}	0.076	**0.065**	97.3	**97.7**	1h08

- Timestamps degrade the model... their representation is inappropriate.

How to leverage the huge amount of data available?

Available Data

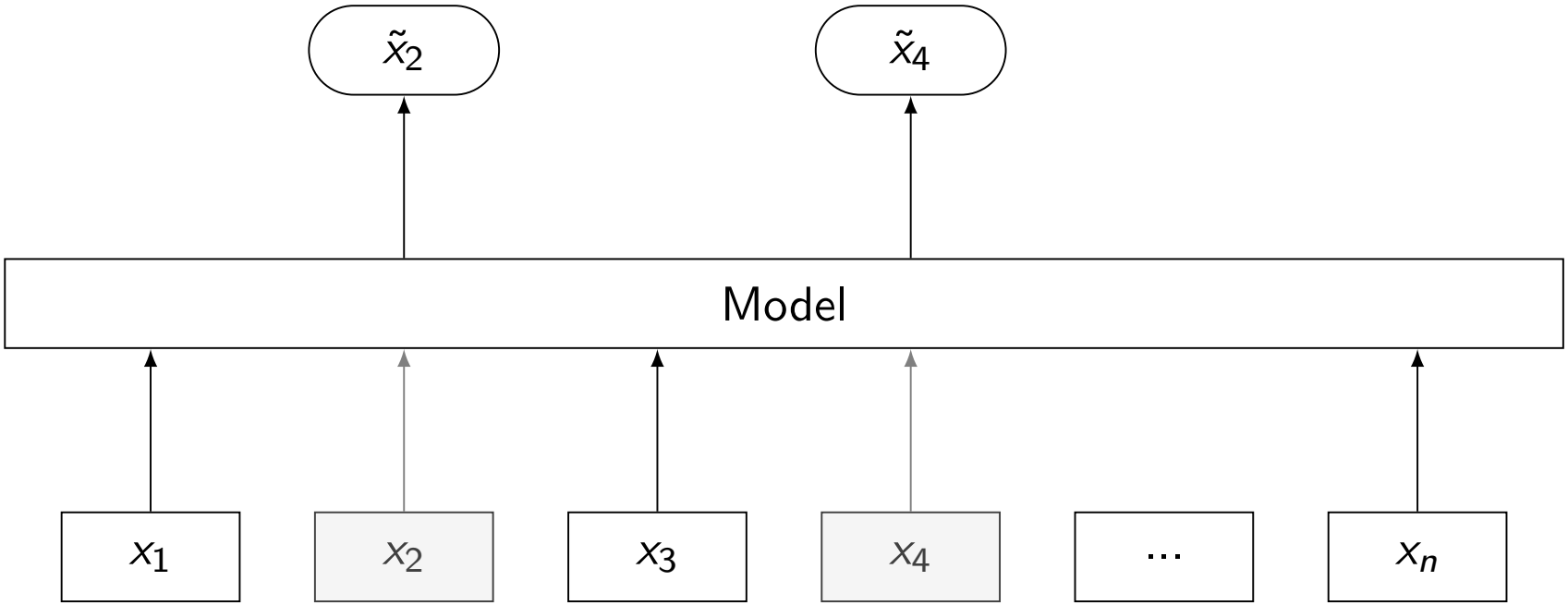
- Huge amount of data available, or easily obtainable.
- Unlabelled and boring...
- Yet there is the data underlying structure[2].
- Some models benefit from learning this structure.
- How to learn this structure from many unlabelled examples?

[2] Underlying structure may be understood as the data distribution, or the set of syntactic and semantic rules.

MLM Objective

- Language Model is one approach...
- But it is directional, only the previous events are considered.
- Masked Language Model (MLM) address this limitation.
- Mask some of the input and learn to predict them.
- Learn to reconstruct the data, hence its underlying structure.

MLM Objective - Example

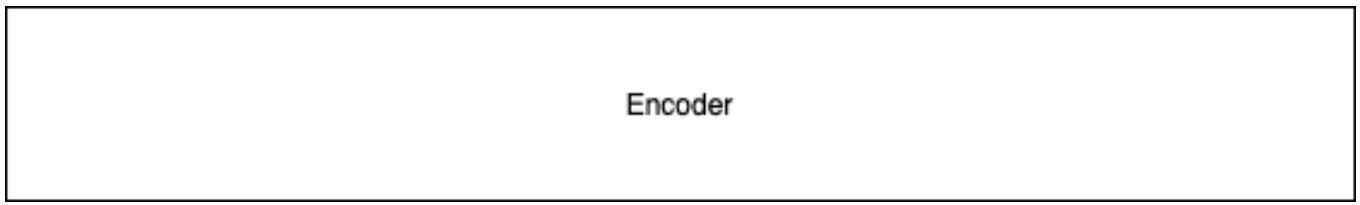


How to Use MLM

- As an unsupervised pre-training step
- Learn the structure of the data, then use the model for another task

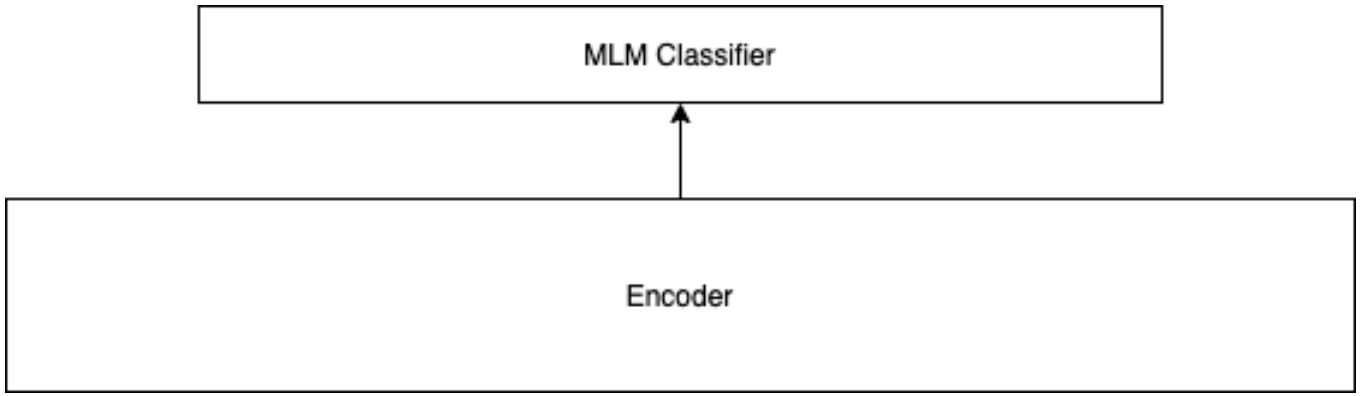
How to Use MLM

1. Create an encoder which takes as input the vectorial representation of events:



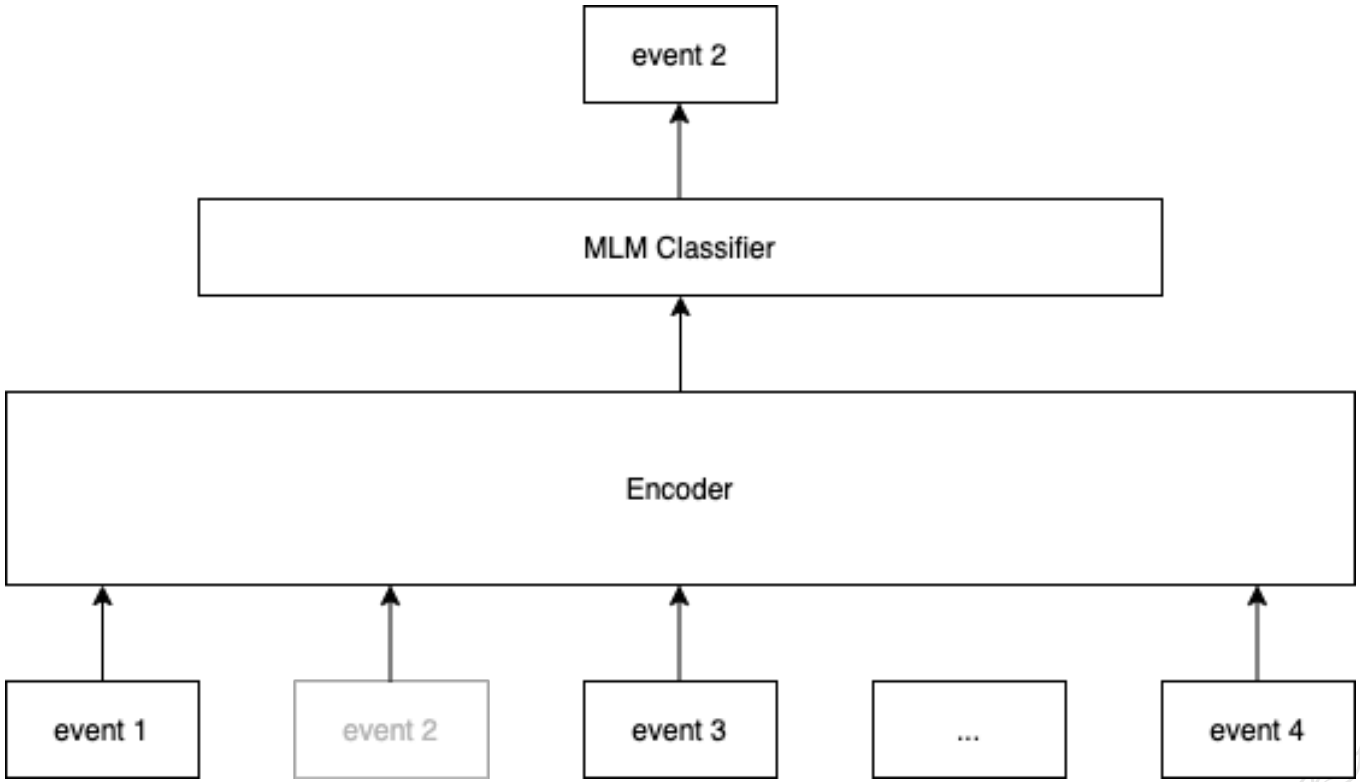
How to Use MLM

2. Add the classifier for the Mask Language Modelling (MLM):



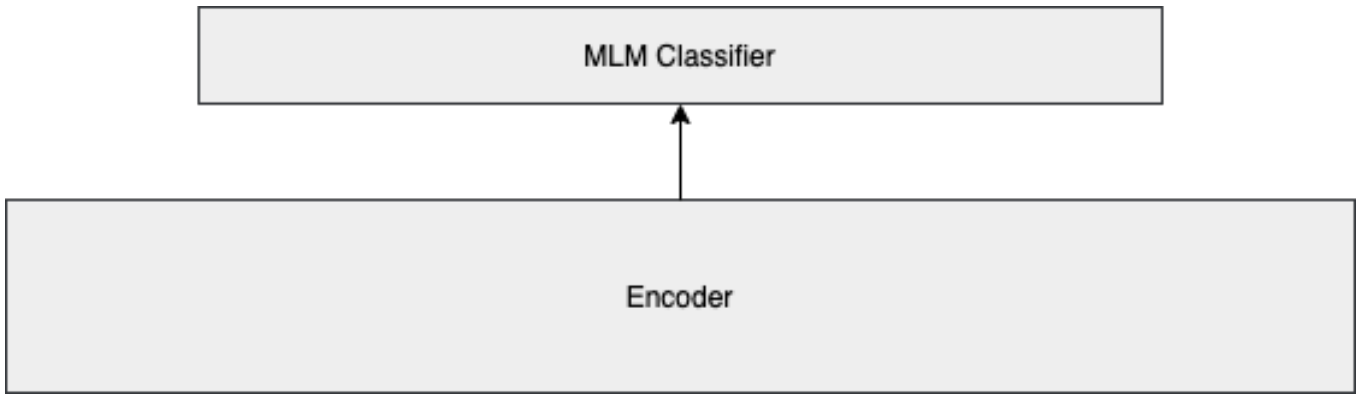
How to Use MLM

3. Train the encoder and the classifier using MLM:



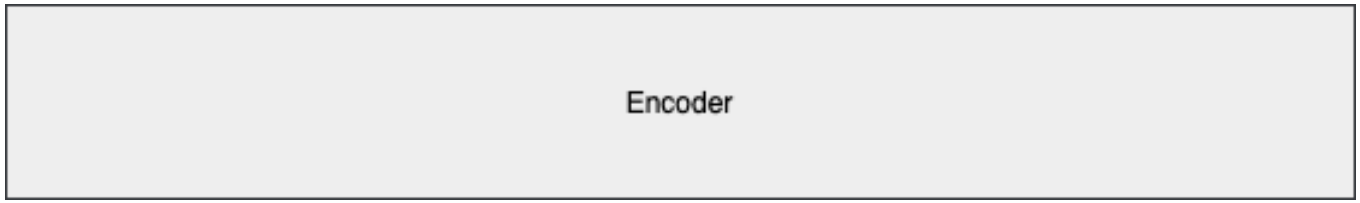
How to Use MLM

4. Now both the encoder and the classifier are trained on MLM:



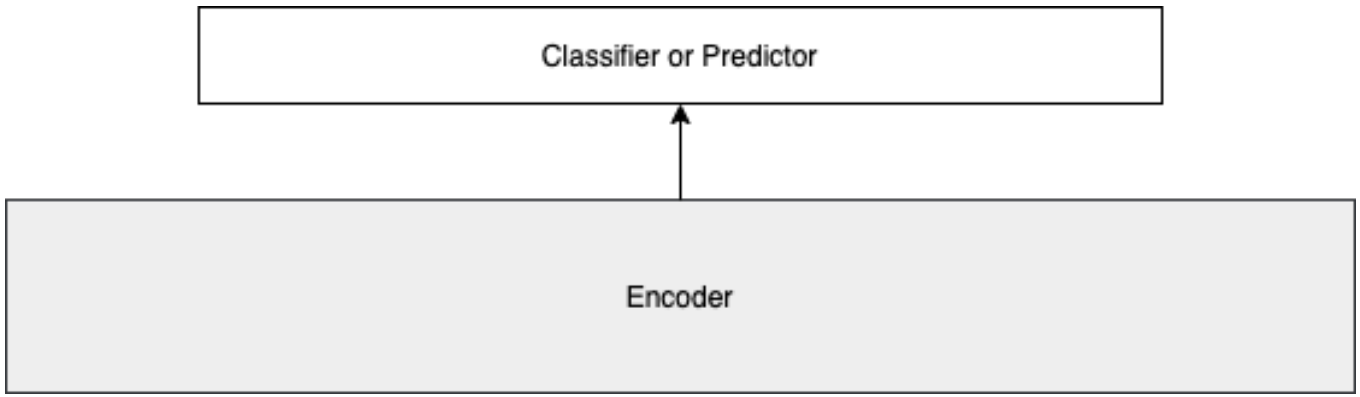
How to Use MLM

5. The classifier is removed since we only need the encoder. It is said the encoder has been pre-trained.



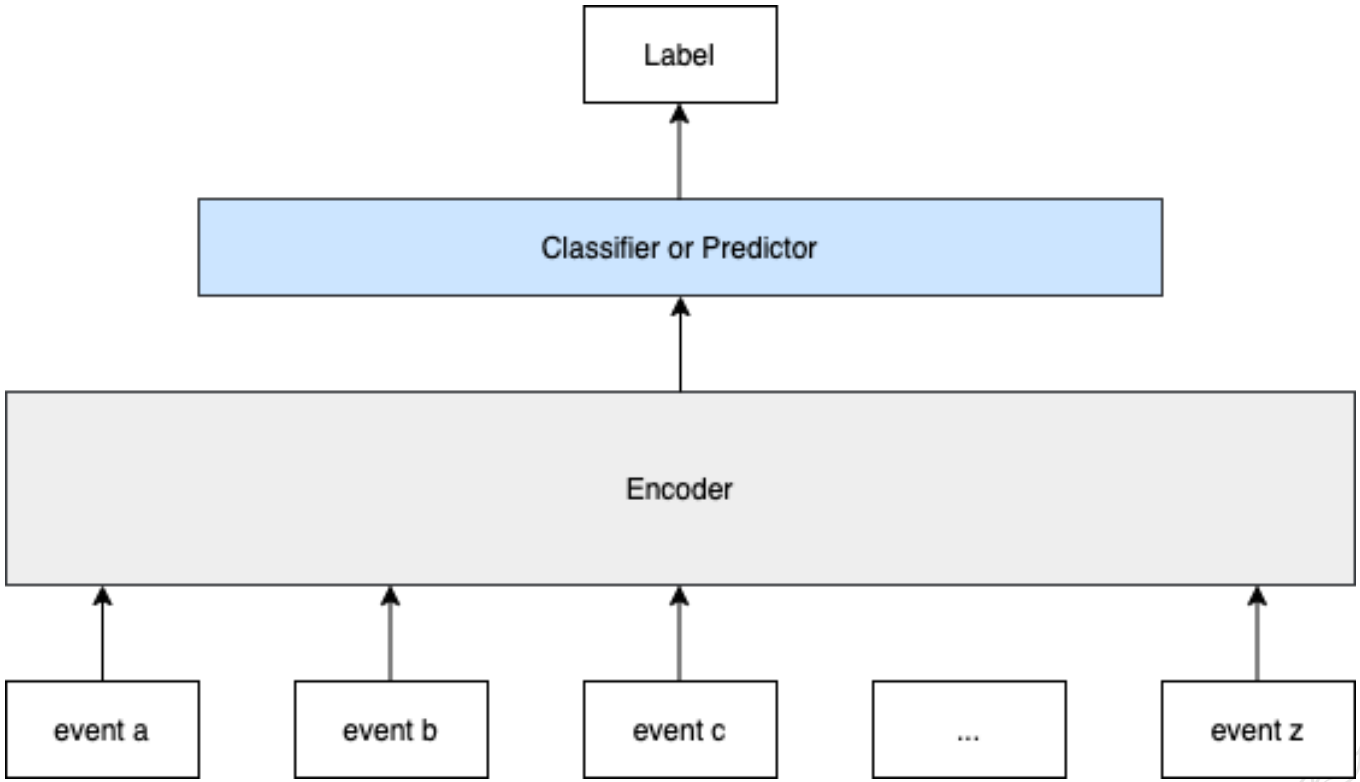
How to Use MLM

6. Add the classifier or the predictor for the final task:



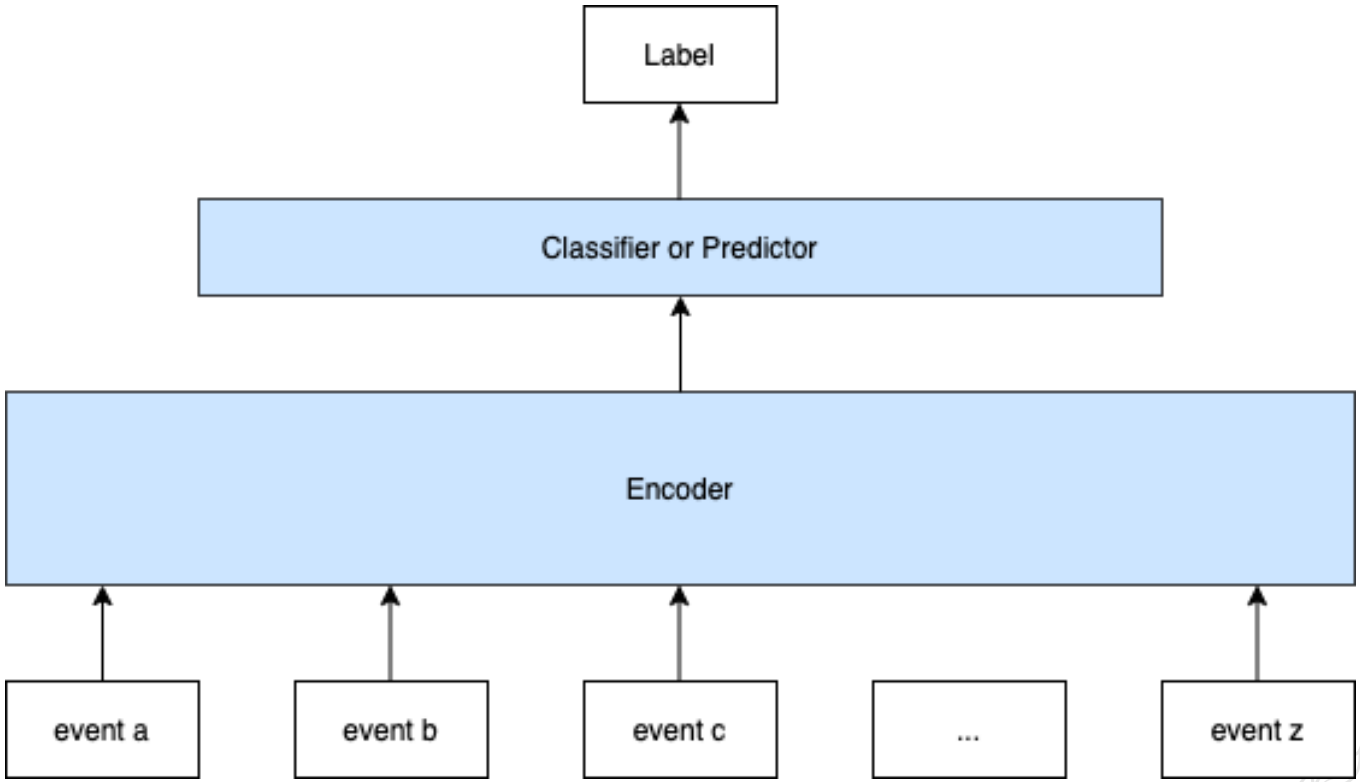
How to Use MLM

7. Train only the classifier or decoder:



How to Use MLM

8. Fine-tune the whole model to get the best performance:



Preliminary Results - Transformer

- No preliminary results for the impact of pre-training.
- Ciena is collecting labelled data.
- Early results indicate that event fields seem to have a large impact on the pre-training.

Future Works

- Evaluate the impact of pre-training with MLM.
- Better quantify the impact of event fields.
- Compute the sequence likelihood to detect anomalies.
- Reduce the complexity of the models used.

Thank You