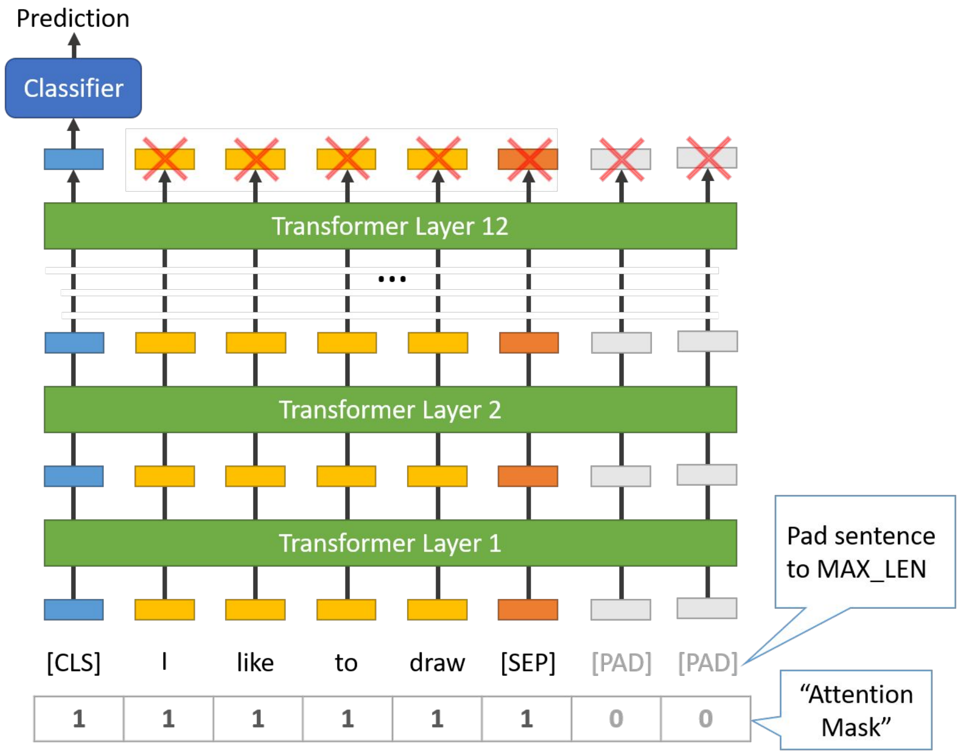
CM Meeting

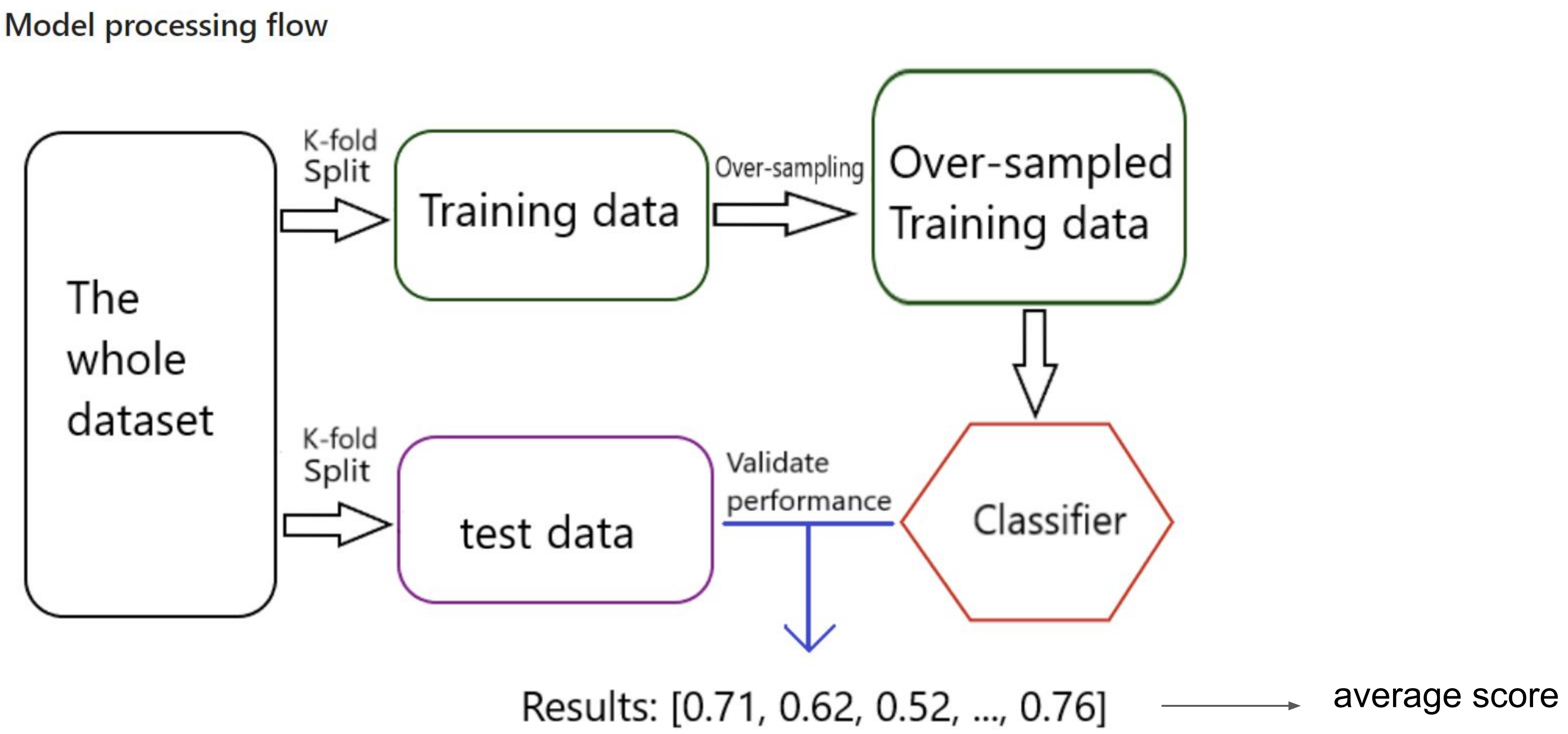
-- 16/11/2021 Shaomu

Current methods for causal relation detection

1. **N-gram** models with pre-trained word embedding
2. Bert embedding + Logistic regression
3. Bert sequence classification model (Fine-tune the model)



Current structure of experiment for causal relation detection

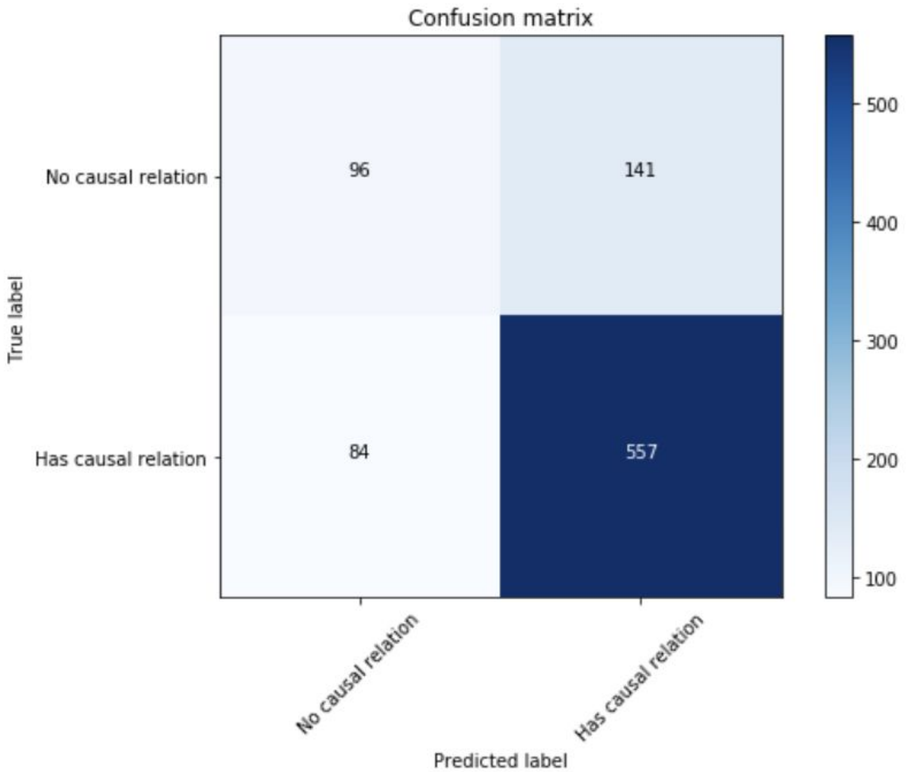


uni-gram models with pre-trained word embedding

classification report on test set is:

	precision	recall	f1-score	support
class 0	0.53	0.41	0.46	237
class 1	0.80	0.87	0.83	641
accuracy			0.74	878
macro avg	0.67	0.64	0.65	878
weighted avg	0.73	0.74	0.73	878

confusion matrix on test set is:



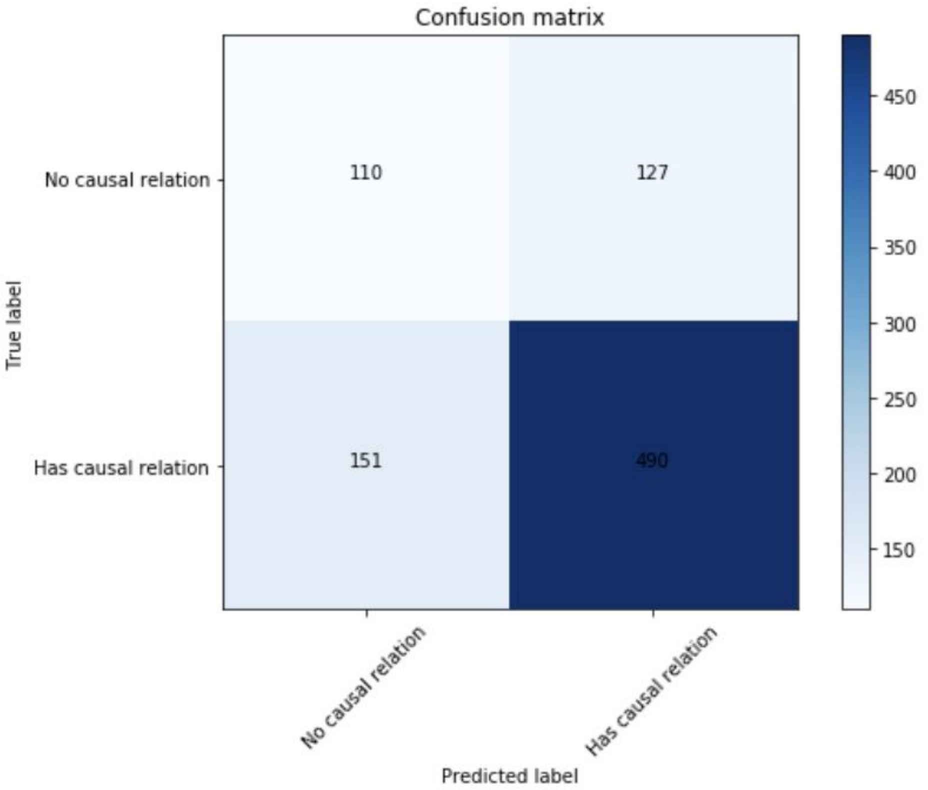
Bert embedding + Logistic regression

classification report on test set is:

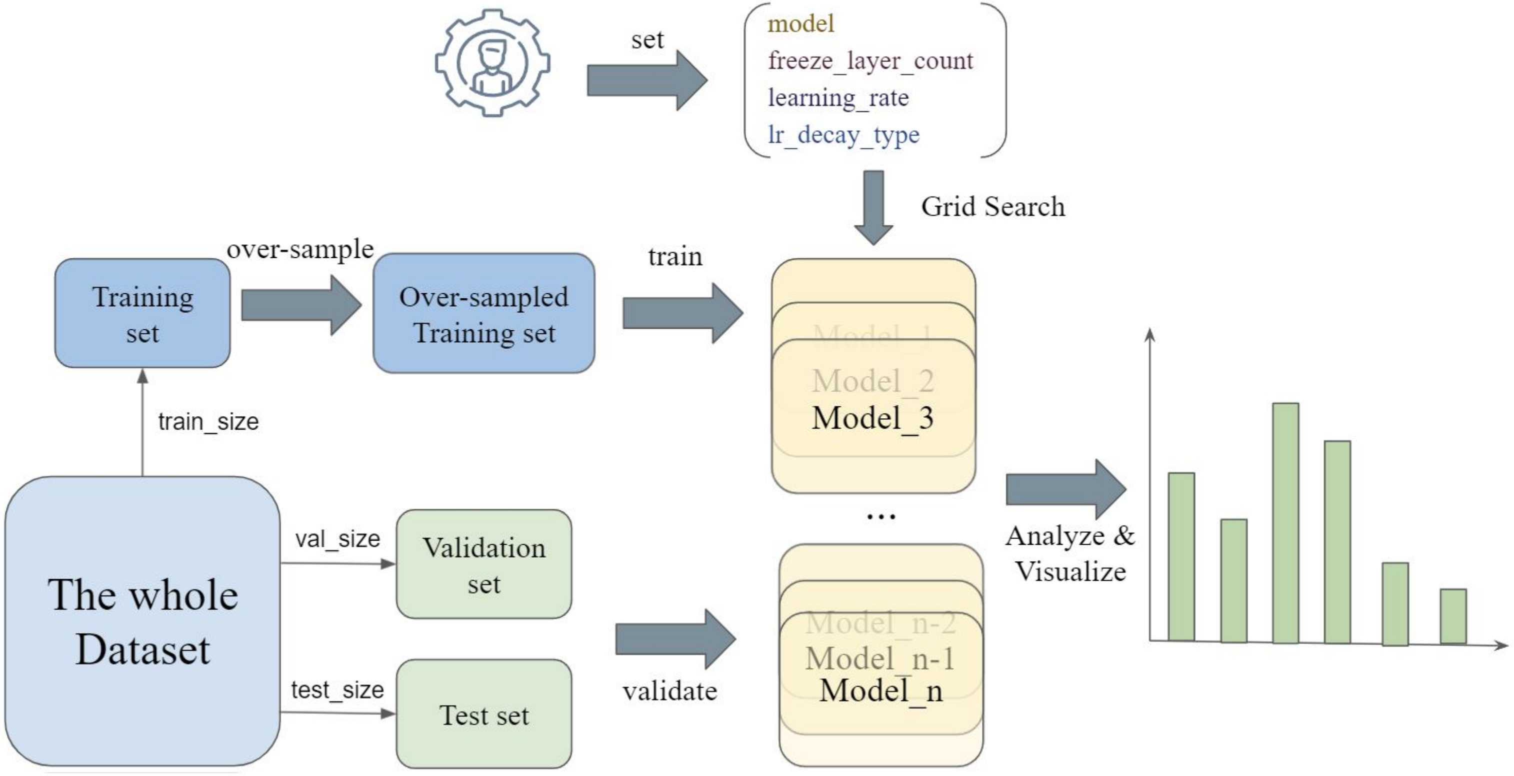
	precision	recall	f1-score	support
class 0	0.42	0.46	0.44	237
class 1	0.79	0.76	0.78	641
accuracy			0.68	878
macro avg	0.61	0.61	0.61	878
weighted avg	0.69	0.68	0.69	878

confusion matrix on test set is:

confusion matrix on test set is:



Bert sequence classification model (Fine-tune the model) -- Hyper param tuning

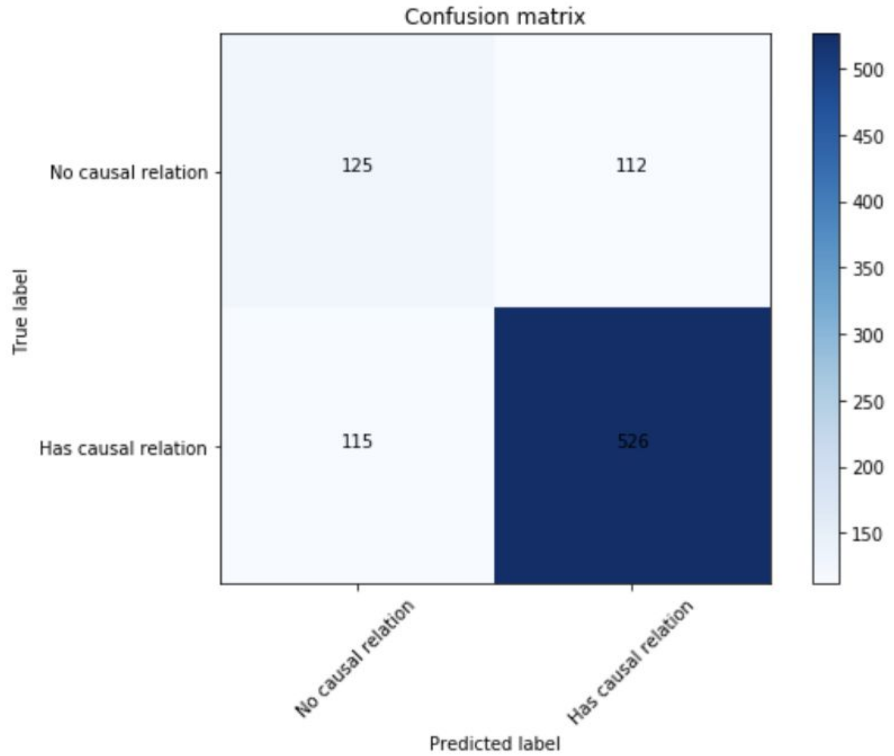


Bert sequence classification model (Fine-tune the model)

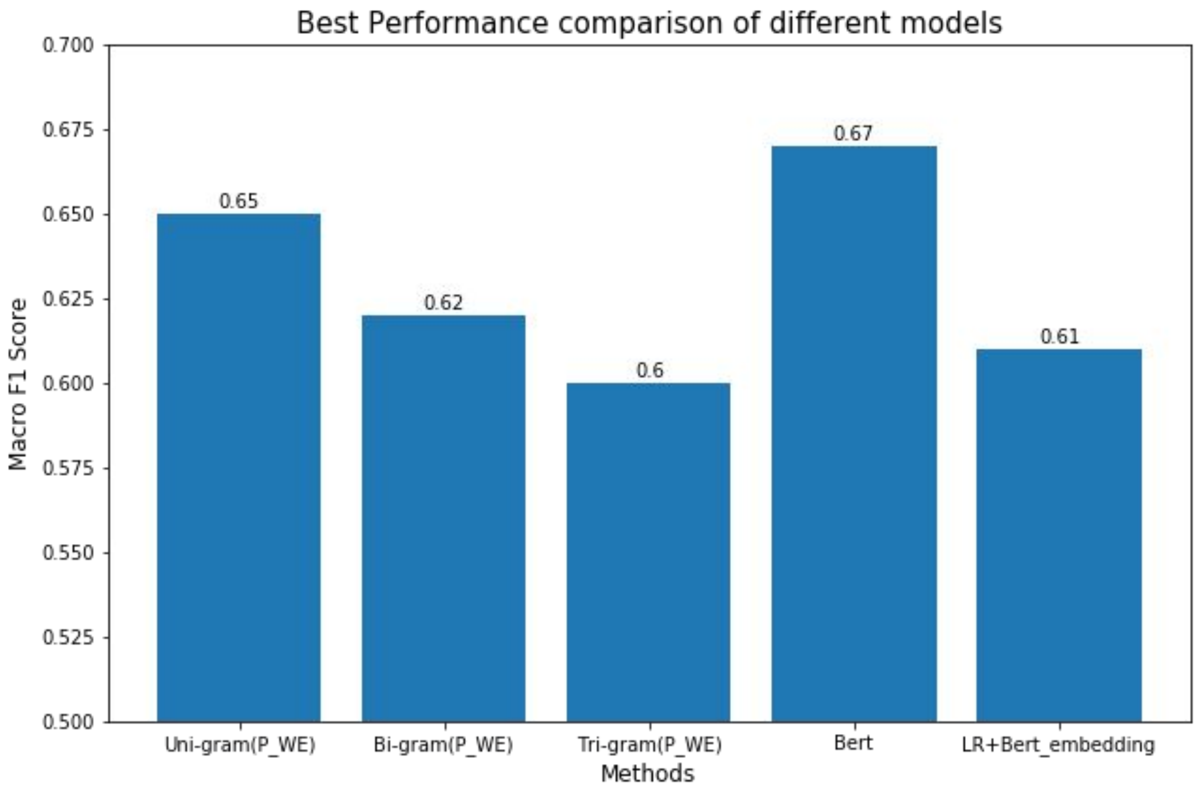
classification report on test set is:

	precision	recall	f1-score	support
class 0	0.52	0.53	0.52	237
class 1	0.82	0.82	0.82	641
accuracy			0.74	878
macro avg	0.67	0.67	0.67	878
weighted avg	0.74	0.74	0.74	878

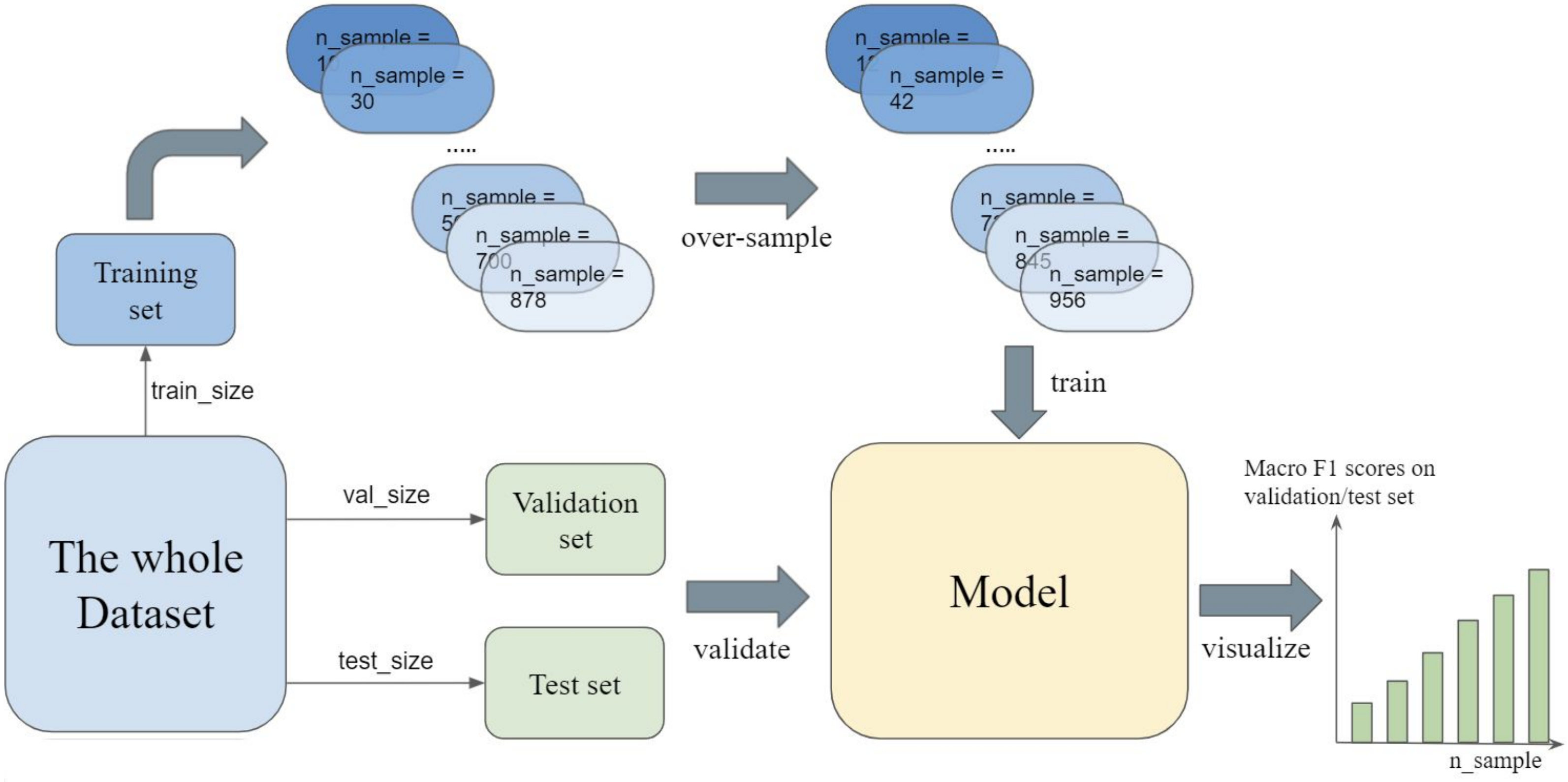
confusion matrix on test set is:



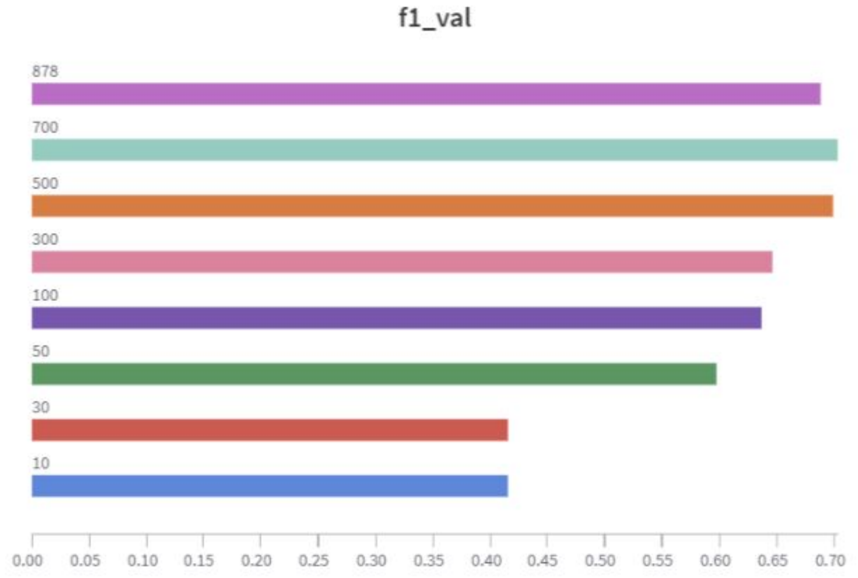
Current results for causal relation detection

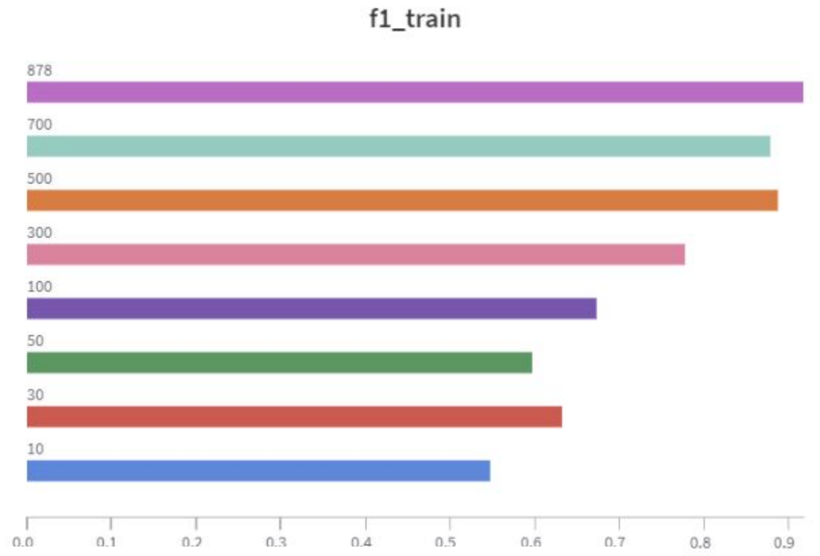


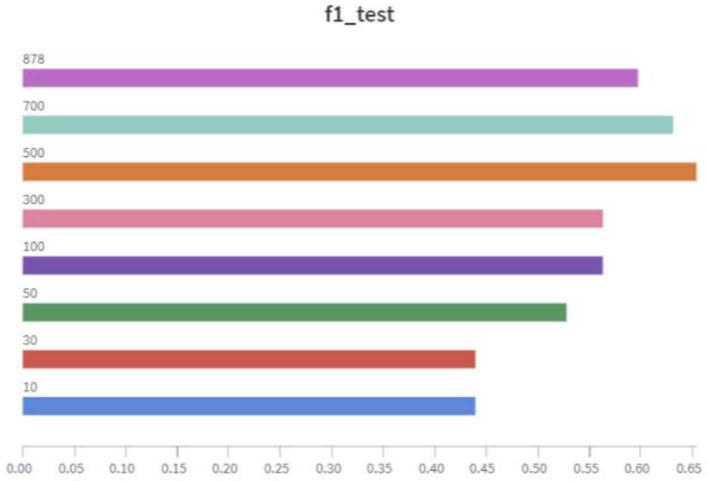
Sensitivity analysis



Sensitivity analysis







Existing datasets about causal relation detection

Dataset:

1. **[Relation Classification] SemEval-2010** Task 8: Multi-Way Classification of Semantic Relations Between Pairs of Nominals, see: https://arxiv.org/pdf/1911.10422.pdf **(#dataset statistics: 10000 totally)**

   Cause-Effect (CE). An event or object yields an effect. Example: those cancers were caused by radiation exposures **(1134 totally)**

   Instrument-Agency (IA). An agent uses an instrument. Example: phone operator

   Product-Producer (PP). A producer causes a product to exist. Example: a factory manufactures suits

   Content-Container (CC). An object is physically stored in a delineated area of space. Example: a bottle full of honey was weighed

   Entity-Origin (EO). An entity is coming or is derived from an origin (e.g., position or material). Example: letters from foreign countries

   Entity-Destination (ED). An entity is moving towards a destination. Eg. the boy went to bed

   Component-Whole (CW). An object is a component of a larger whole. Example: my apartment has a large kitchen

   Member-Collection (MC). A member forms a nonfunctional part of a collection. Example: there are many trees in the forest

   Communication-Topic (CT). An act of communication, written or spoken, is about a topic. Example: the lecture was about semantics

Existing datasets about causal relation detection

Datasets:

1. **[ADE]** Adverse Drug Effect (ADE) Corpus Dataset (*biomedical text*), see: https://huggingface.co/datasets/ade_corpus_v2
   (#dataset statistics: 11223 totally, Cause-Effect: 6720)

2. **[Causal-TimeBank]** Annotating Causality in the TempEval-3 Corpus, see: https://aclanthology.org/W14-0702/
   #dataset statistics:
   - 6,811 EVENTs (only instantiated events by MAKEINSTANCE tag of TimeML)
   - 5,118 temporal links
   - 171 **causal signals**
   - 318 **causal links**

3. **[Event StoryLine]** The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction
   #dataset statistics:
   - 258 documents concerning calamity events
   - 112 **Cause-Effect**

Existing solutions similar to causal relation detection task

Paper:

1. **[R-Bert]** Enriching Pre-trained Language Model with Entity Information for **Relation Classification**, see https://arxiv.org/pdf/1905.08284.pdf, https://github.com/monologg/R-BERT, https://github.com/wang-h/bert-relation-classification

2. **[Skeleton-Aware BERT]** An Extensible Framework of Leveraging Syntactic Skeleton for Semantic **Relation Classification**, see: https://dl-acm-org.proxy.library.uu.nl/doi/pdf/10.1145/3402885

3. **[Entity Attention Bi-LSTM]** Semantic **Relation Classification** via Bidirectional LSTM Networks with Entity-aware Attention using Latent Entity Typing, see: https://arxiv.org/pdf/1901.08163.pdf, https://github.com/roomylee/entity-aware-relation-classification

4. **[C-Bert]** Causal-BERT : Language models for **causality detection** between events expressed in text, see: https://arxiv.org/pdf/2012.05453.pdf

5. **[k-CNN]** Knowledge-oriented convolutional neural network for **causal relation extraction** from natural language texts, see: https://www.sciencedirect.com/science/article/pii/S0957417418305177

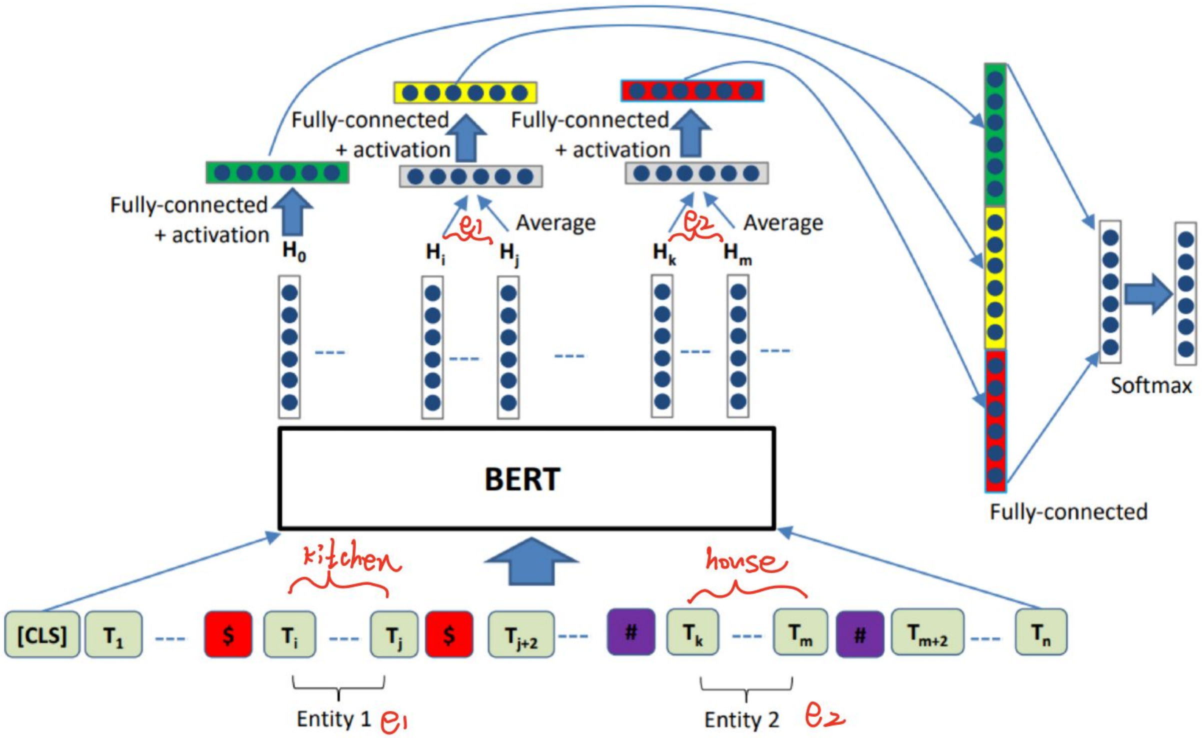
Existing solutions similar to causal relation detection task

Some of them are using the entity as one of their features, then to determine if the enetity pair has the causal relation, which is different from our approach.

with entity feature: Most of the <e1> taste <\e1> of strong onions comes from the <e2> smell <\e2>.

without entity feature: Most of the taste of strong onions comes from the smell.

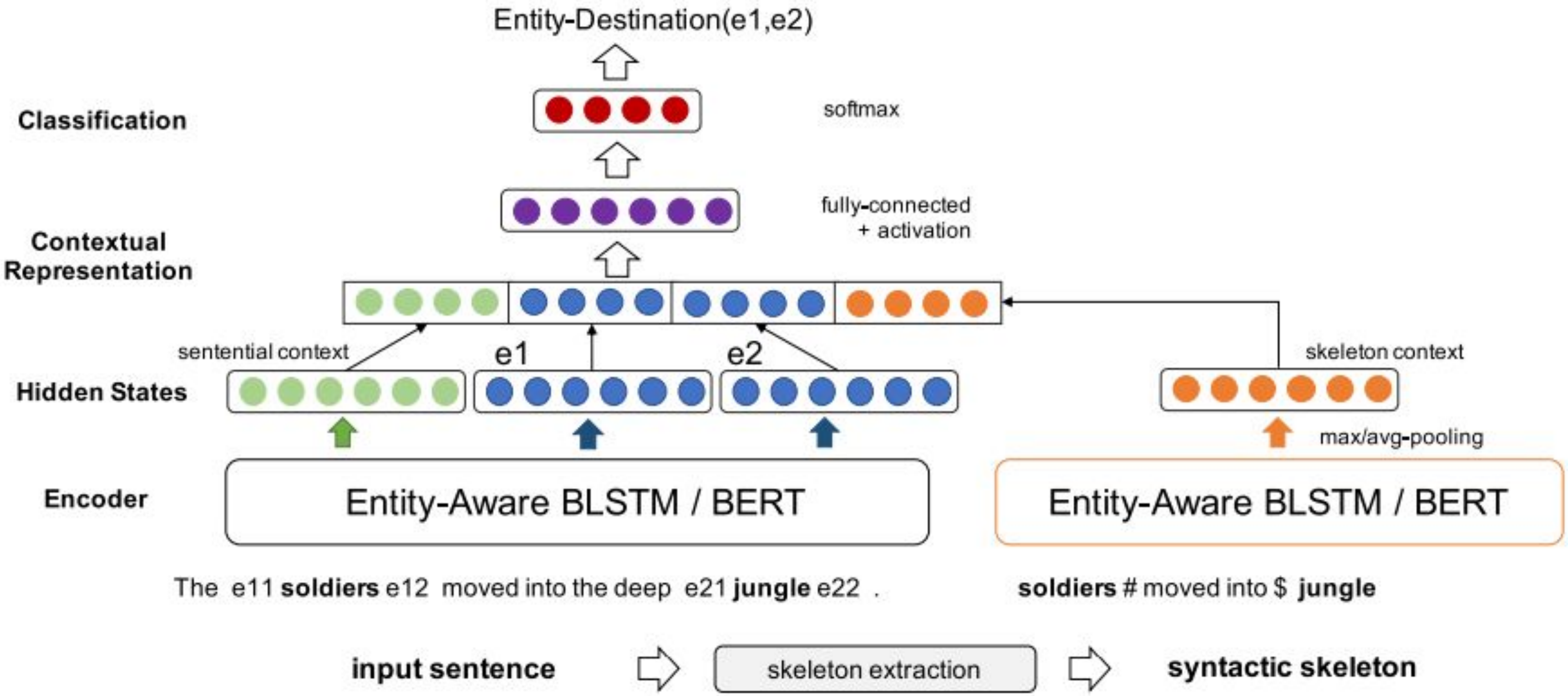
R-bert Model structure



Method	F1	Description
R-bert-no-sep no-ent	81.09	discard the special separate tokens & hidden vector output of the two entities
R-bert-no-sep	87.98	only discard the special separate tokens
R-bert-no-ent	87.99	only discard the hidden vector output
R-bert	89.25	keep everything

[CLS] The $ kitchen $ is the last renovated part of the # house#. “

Skeleton-Aware BERT Model structure



Causal-BERT : Language models for causality detection between events expressed in text

	Semeval 2007	Semeval 2010	ADE
C-BERT	93.78	97.68	97.10
Event Aware C-ERT	94.94	98.35	97.85

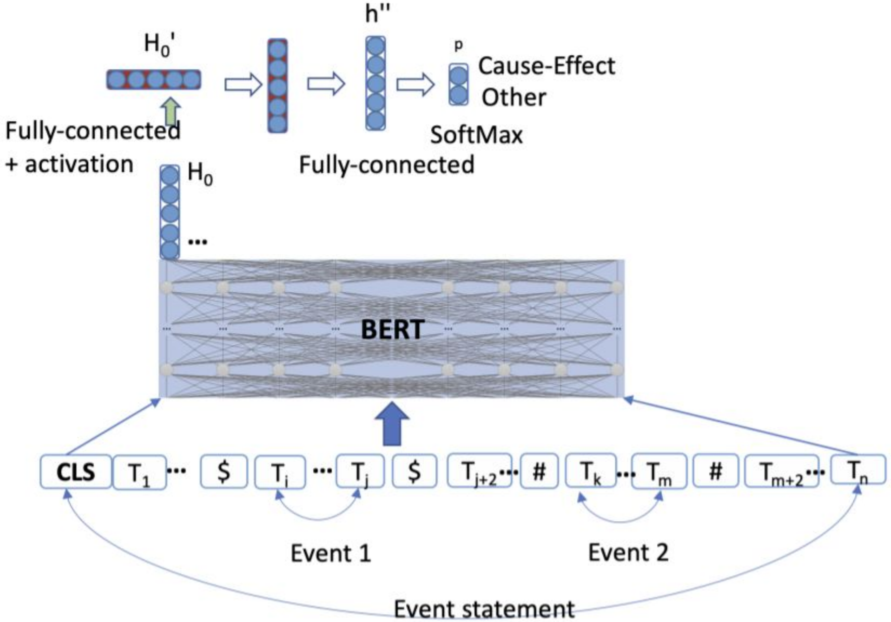


Fig. 2. C-BERT

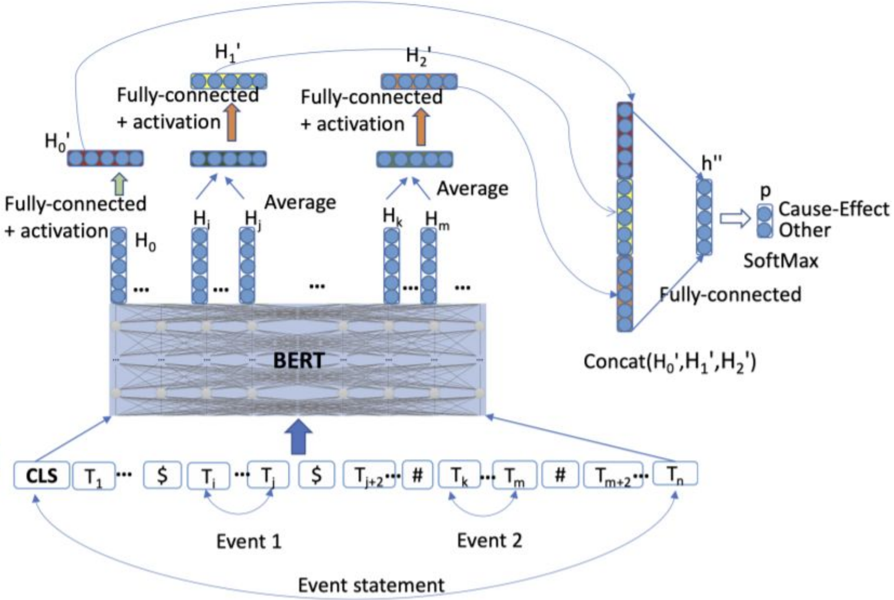


Fig. 3. Event aware C-BERT

Existing solutions’ results for Relation Classification

Method	F1 score on **SemEval 2010**
Entity Attention Bi-LSTM	85.2
R-Bert	89.25
Skeleton-Aware BERT	89.94 (avg) 90.36 (best)

Note:
All three methods are aimed at **semantic relation classification** (containing approx. 10% causal relation)

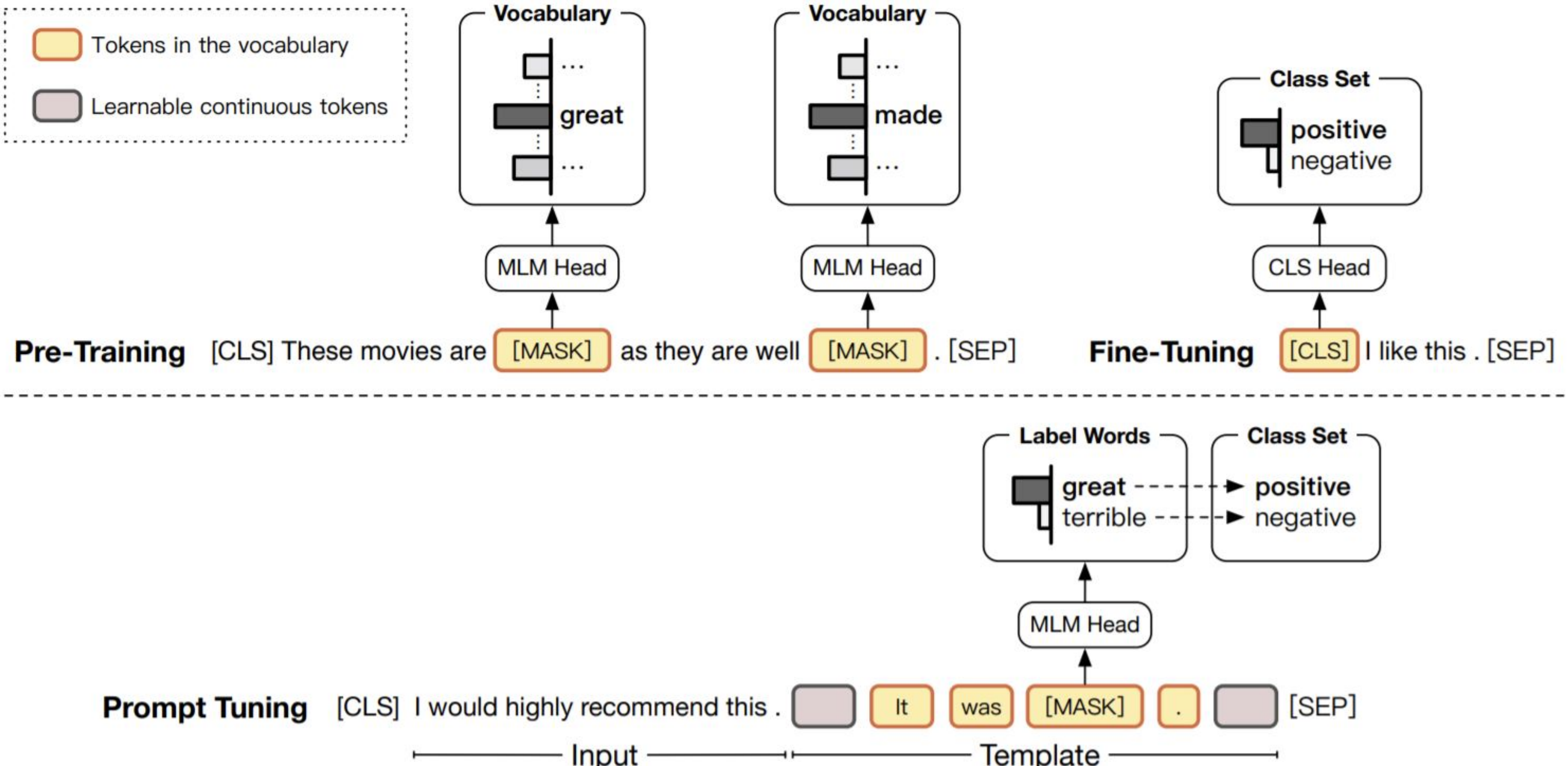
Existing solutions' results for Causal Relation Detection

Method	F1 score on **SemEval2010**	F1 score on **Causal-TB**	F1 score on **Event-SL**	F1 score on **SemEval2007**	F1 score on **ADE**
K-CNN	91.82	76.29	81.25	-	-
C-Bert	98.35	-	-	95.31	97.85

Note:
Both methods are aimed at **Causal Relation Detection**
(only containing causal relation)

Existing solutions to improve transformer encoder

1. **[Prompt tuning]** OpenPrompt: An Open-source Framework for Prompt-learning, see: https://github.com/thunlp/OpenPrompt, https://arxiv.org/abs/2111.01998 [easy to implement]



Existing solutions to improve transformer encoder

1. **[Data augmentation]** Data Augmentation Approaches in Natural Language Processing: A Survey, see: https://arxiv.org/abs/2110.01852

2. **[Avoid overfitting for transformer]** I believe there're tons of paper…

Summary

1. We could do causal relation task on other dataset first: to see whether we will have performance that related to those SOTA model.
   a. If so, at least we know the problem could be our task is harder [might need expert knowledge];
   b. If not, implement some existing method to check whats wrong with our model.

2. Add special tokens and hidden vector output of the two entities to our model [easy]

3. Use other dataset to train our model as well

4. Try SOTA methods in the field of relation detection/ causal relation detection. [depend on method]

5. Improve Tansformer encoder [depend on method]