ULB UNIVERSITÉ LIBRE DE BRUXELLES

ULB BRUFACE BRUSSELS FACULTY OF ENGINEERING VUB

VUB VRIJE UNIVERSITEIT BRUSSEL

Towards neural architectures for Fact Extraction and VERification

Scientific Fact Verification

Academic year: 2020-2021

In order to be awarded the Degree of Master of Science in Electrical Engineering major in Information and Communication Technology System

Student: Boya Zhang

Promotor: Prof. Dr. Ir. Nikolaos Deligiannis
Copromotor: Dr. Ir. Giannis Bekoulis

CONTENT

INTRODUCTION

- **Motivation:** Provide ordinary people with a method to identify the veracity of scientific claims on the internet.
- **Goal:** Classify the SCIFACT claims into three classes with three steps of procedure.

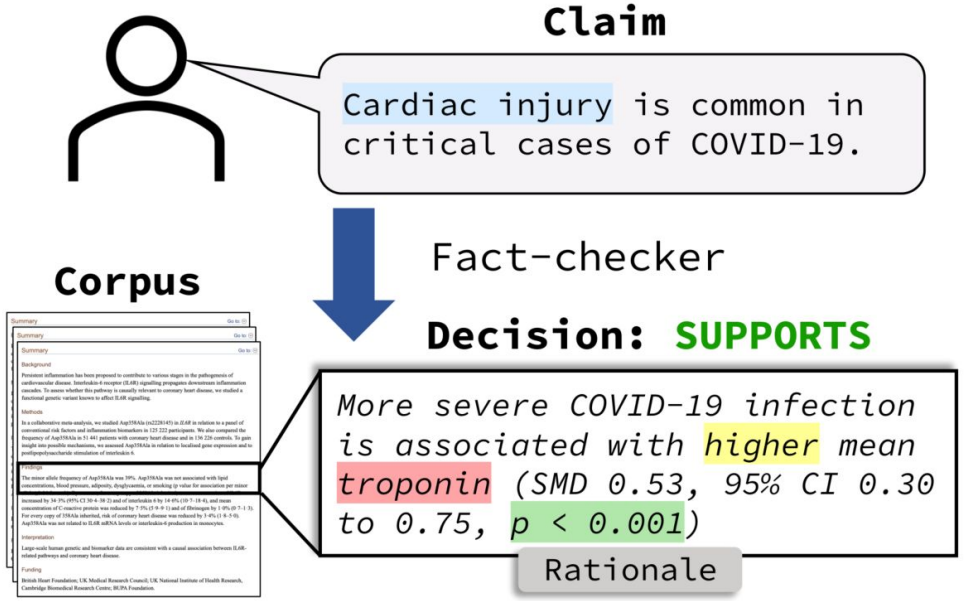
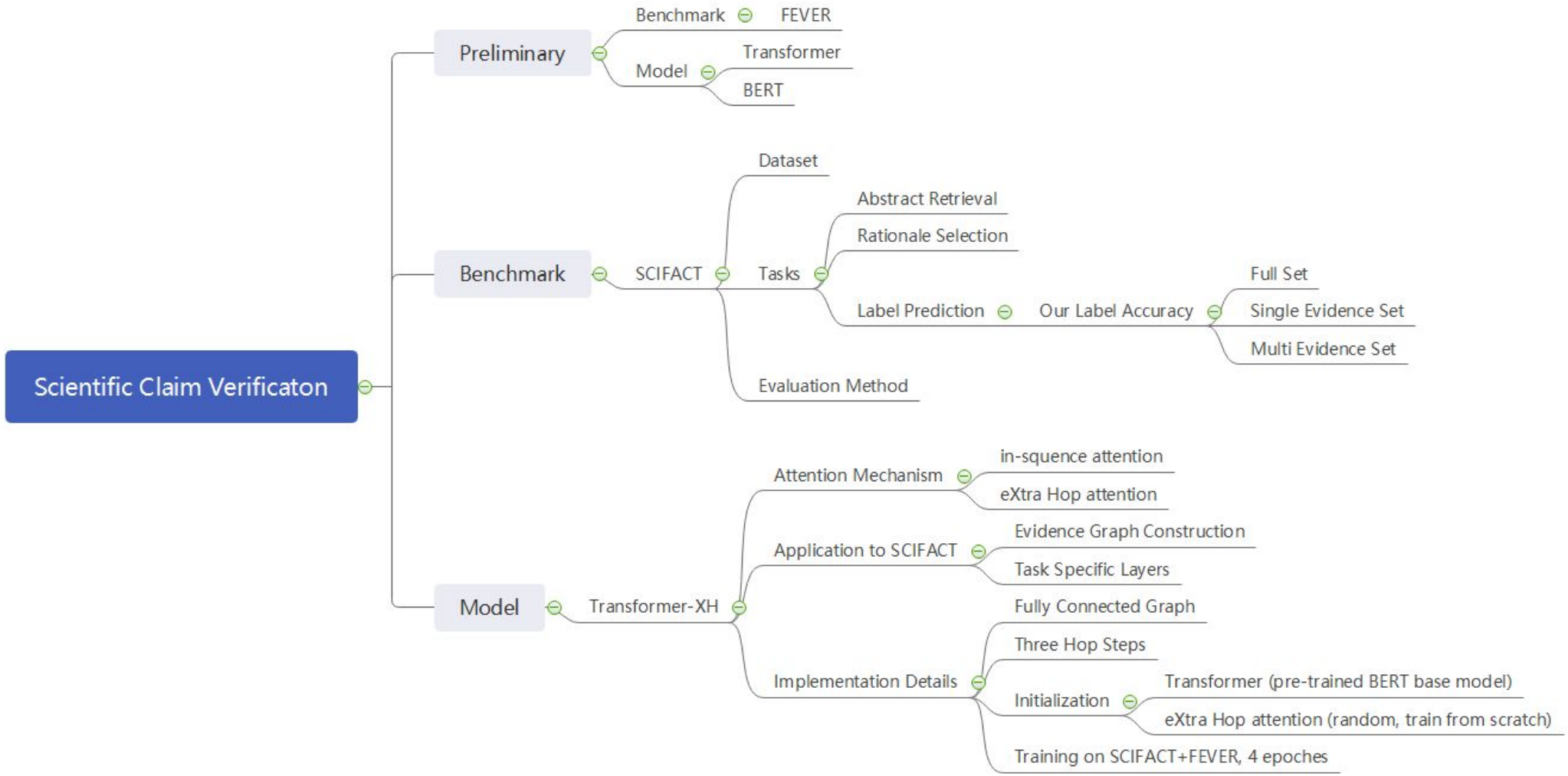


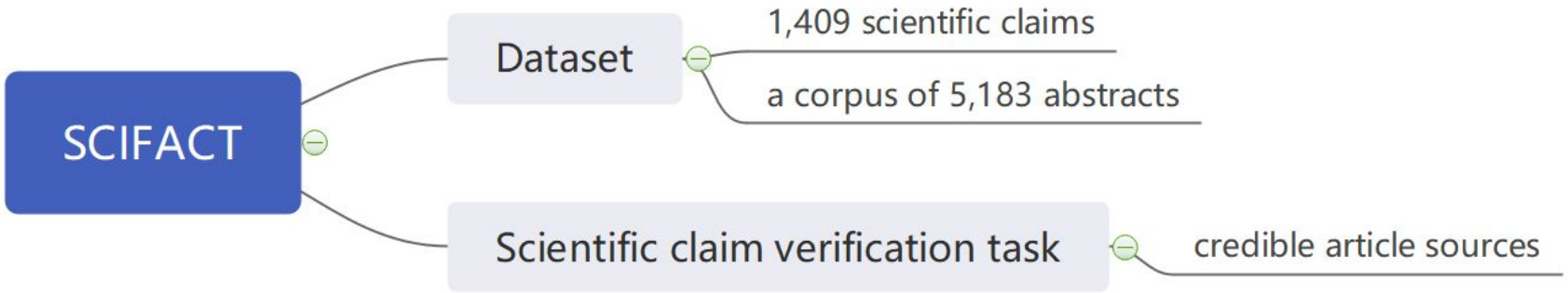
Figure: An example of scientific claim verification (Wadden et al., 2020).

INTRODUCTION



BENCHMARK

SCIFACT



BENCHMARK

Dataset

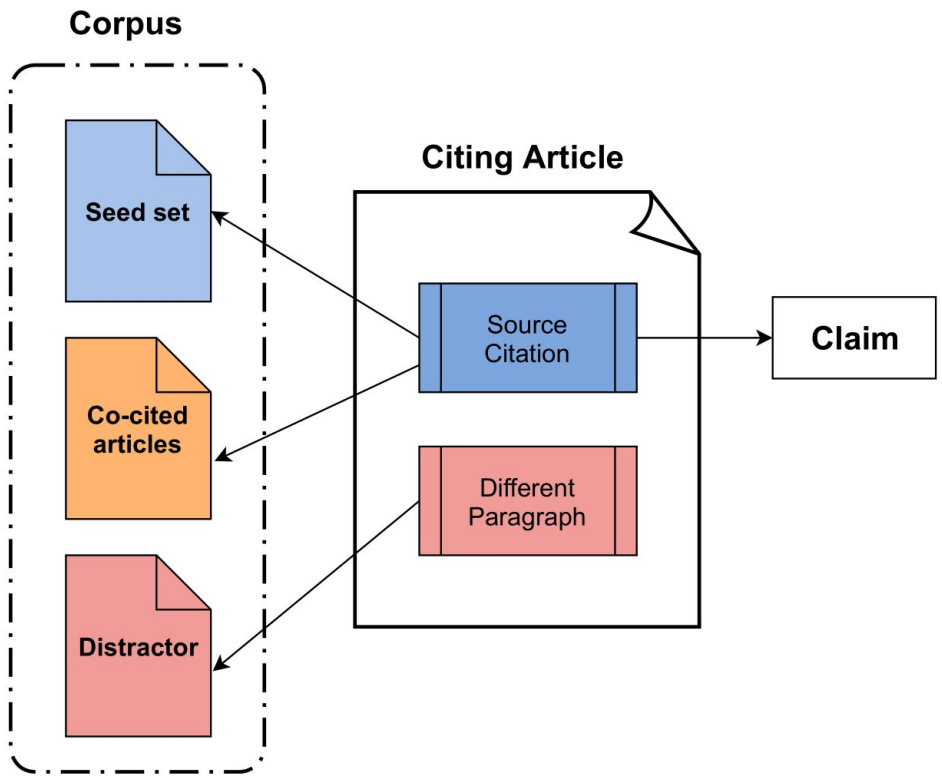


Figure: Corpus Creation.

Set	SUPPORTS	REFUTES	NO INFO	Total
Training	332	304	173	809
Dev	124	112	64	300
Test	100	100	100	300
Total	556	516	337	1409

Table: Training, Dev and Test Set Sizes for Each Classes of SciFact Dataset.

BENCHMARK

Dataset

Claim: 1/2000 in UK have abnormal PrP positivity. **Evidence: [SUPPORT]** **[abstract/ Prevalent abnormal prion protein in human appendixes after bovine spongiform encephalopathy epizootic: large scale survey]** RESULTS Of the 32,441 appendix samples 16 were positive for abnormal PrP, indicating an overall prevalence of 493 per million population (95% confidence interval 282 to 801 per million). **Label:** SUPPORTS
Claim: ALDH1 expression is associated with better breast cancer outcomes. **Evidence: [CONTRADICT]** **[abstract/ ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome]** In a series of 577 breast carcinomas, expression of ALDH1 detected by immunostaining correlated with poor prognosis. **Label:** REFUTES
Claim: 0-dimensional biomaterials show inductive properties. **Label:** NO INFO

Figure: Three Examples from SCIFACT Dataset.

BENCHMARK

Pipeline

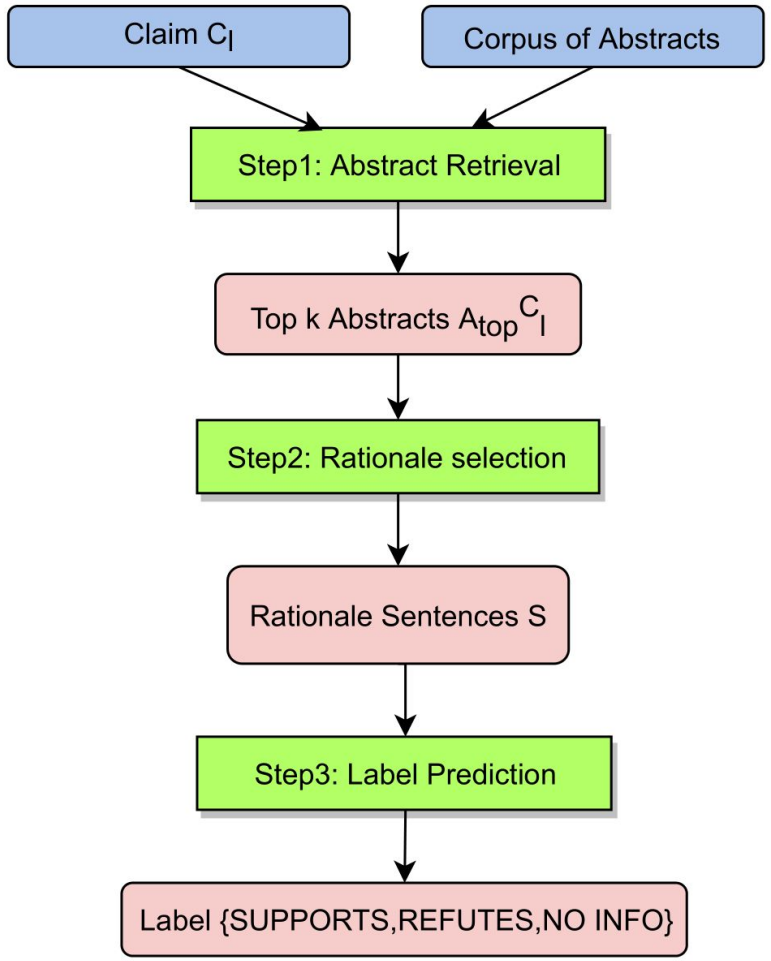
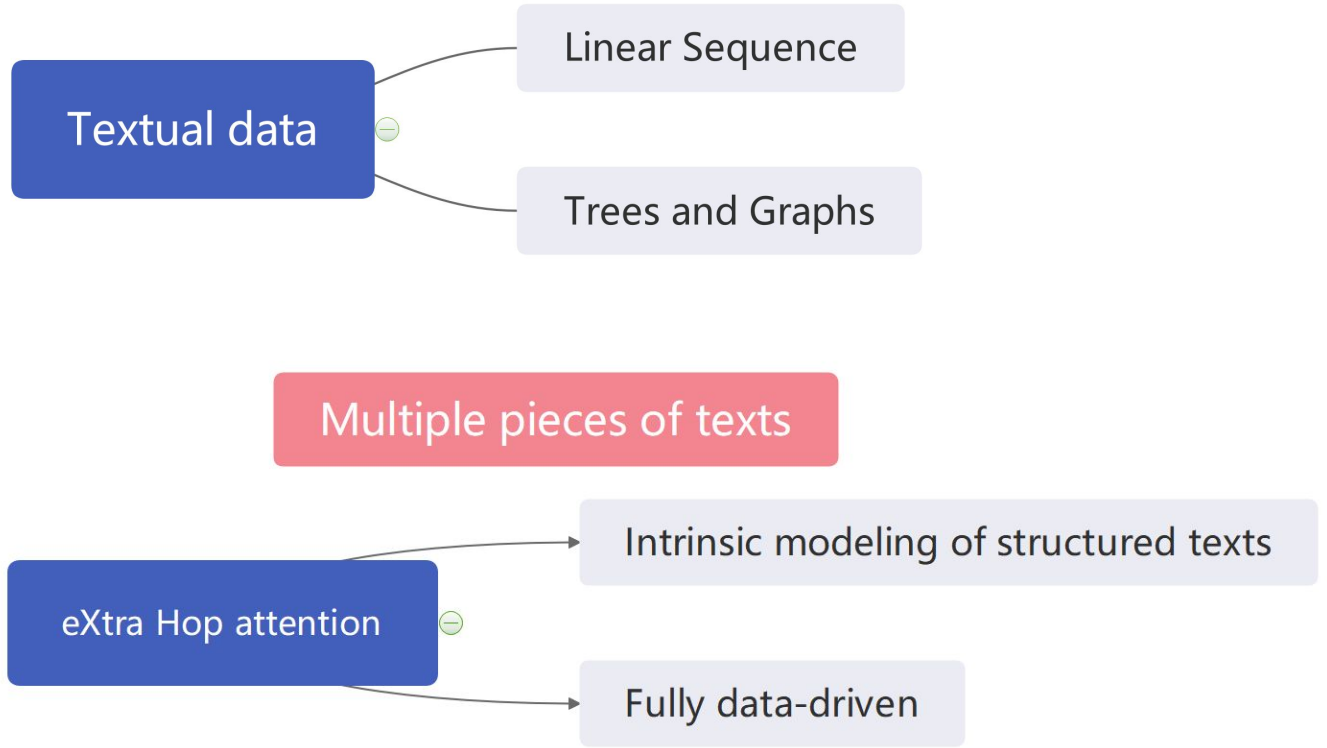


Figure: The Baseline System of SciFact Dataset.

MODEL

Transformer-XH



MODEL

Transformer-XH

Transformer ⟹ **Transformer-XL** ⟹ **Transformer-XH**

- Attention calculated over all token pairs
- Hard to scale to long text sequences

- Breakdown longer texts
- Propagate information between adjacent text segments
- Cannot deal with text segments organized in nontrivial structures

- Link structured text sequence with eXtra Hop attention
- Propagate information along graph edges
- Enable information sharing between connected text sequence

MODEL

Attention Mechanism

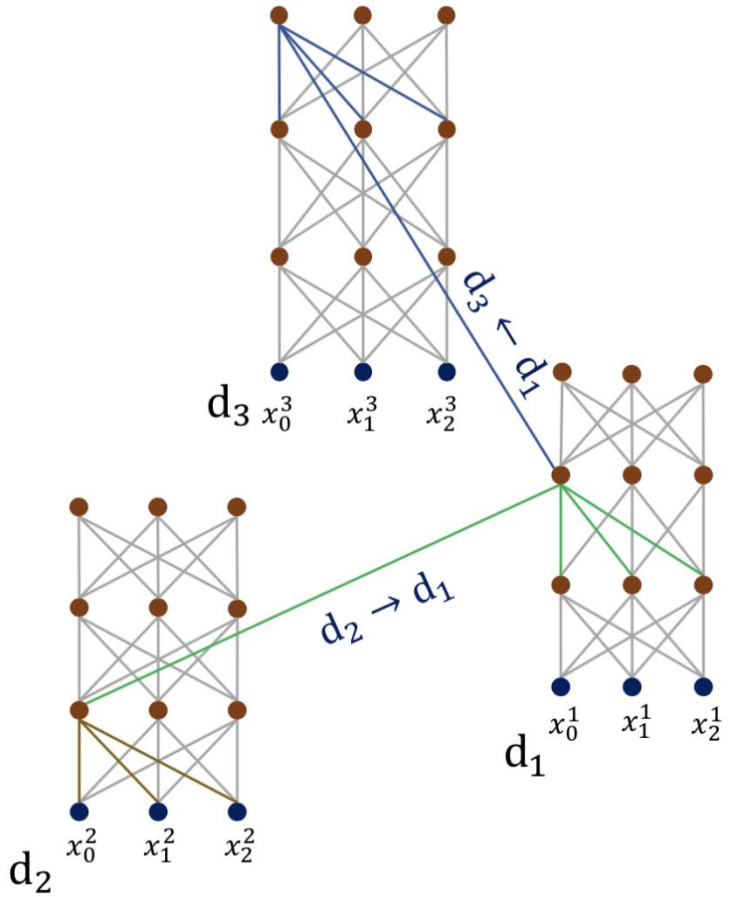


Figure: Hop Attentions on the Path from node d2 to d1 to d3(Zhao et al., 2020).

MODEL

Application to SCIFACT

- **Evidence Graph Construction**
- **Transformer-XH on Evidence Graph**
  1. Global representation of the graph
  2. Task specific layers
     1. Fact prediction per node ([CLS])
     2. Importance of each node in the graph
     ✔ Final prediction

MODEL

Implementation Details

- **Graph Structures:** Fully connected graph
- **Parameters Initialization**
  - Transformer part: pre-trained BERT base model
  - eXtra Hop attention part: random, train from scratch
- **Hop steps:** 3
  - Transformer-XH reaches its peak performance with three hops in the ablation study (Zhao et al.,2020).
- **Training set:** SCIFACT+FEVER, trained for 4 epochs

RESULTS

Label Prediction, model trained on FEVER

- Label Accuracy on FEVER dev set: 78.03
- Label Accuracy on SCIFACT dev set: 38.0

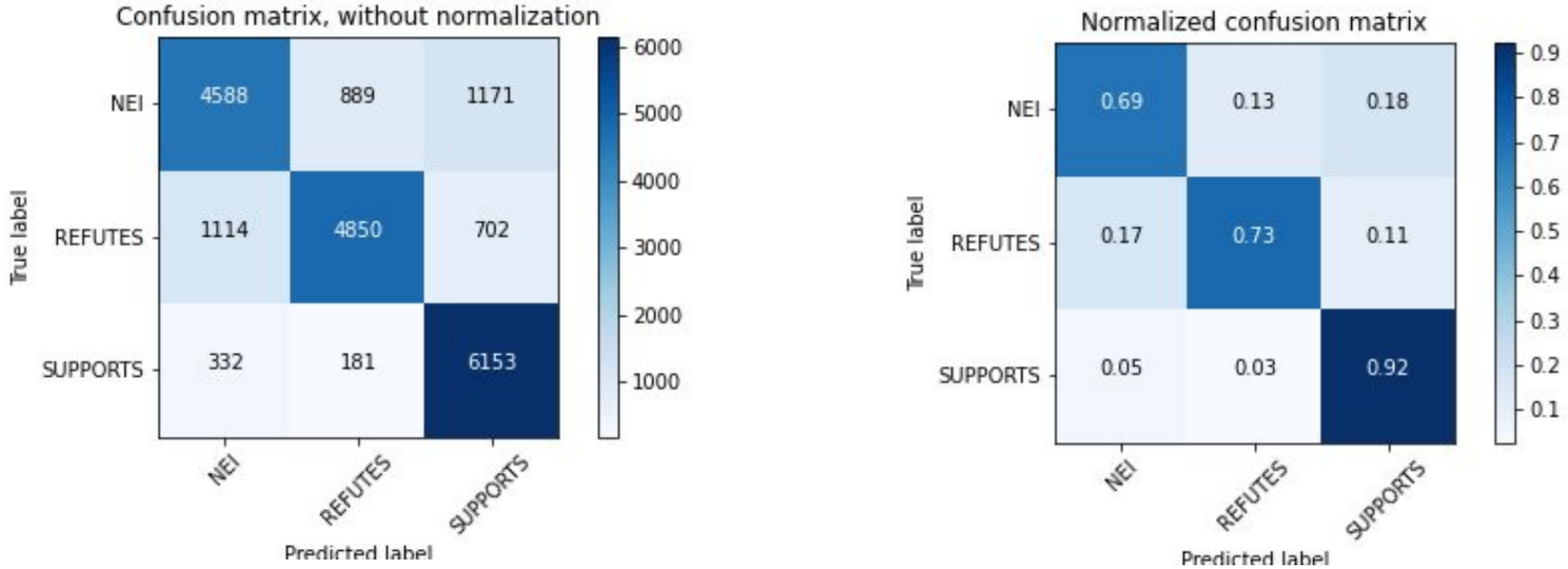


Figure: Confusion Matrix of FEVER Dev Set Label Prediction.

RESULTS

Label Prediction

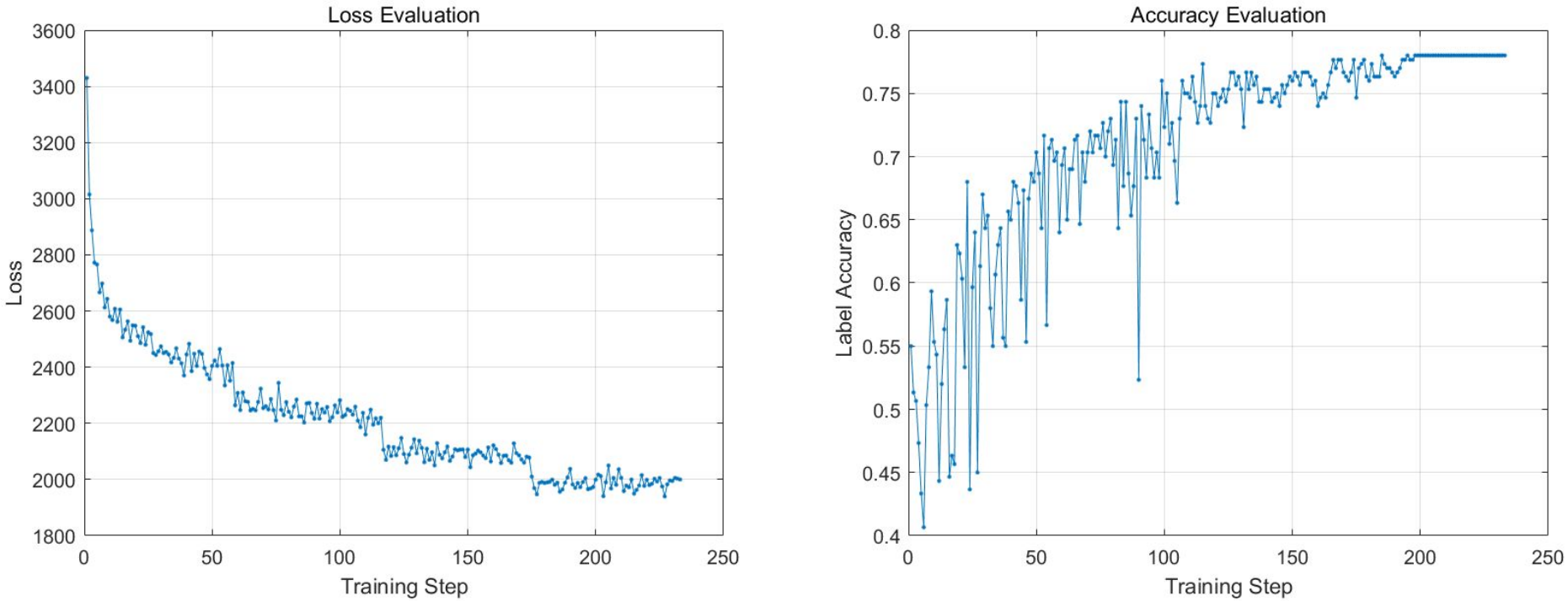


Figure: Training Process of Transformer-XH on SCIFACT Task.

RESULTS

Label Prediction

Model	Full	Single Evidence	Multi Evidence
Transformer-XH	78.0	68.75	82.35

Table: Label Accuracy on SCIFACT Dev Claims.

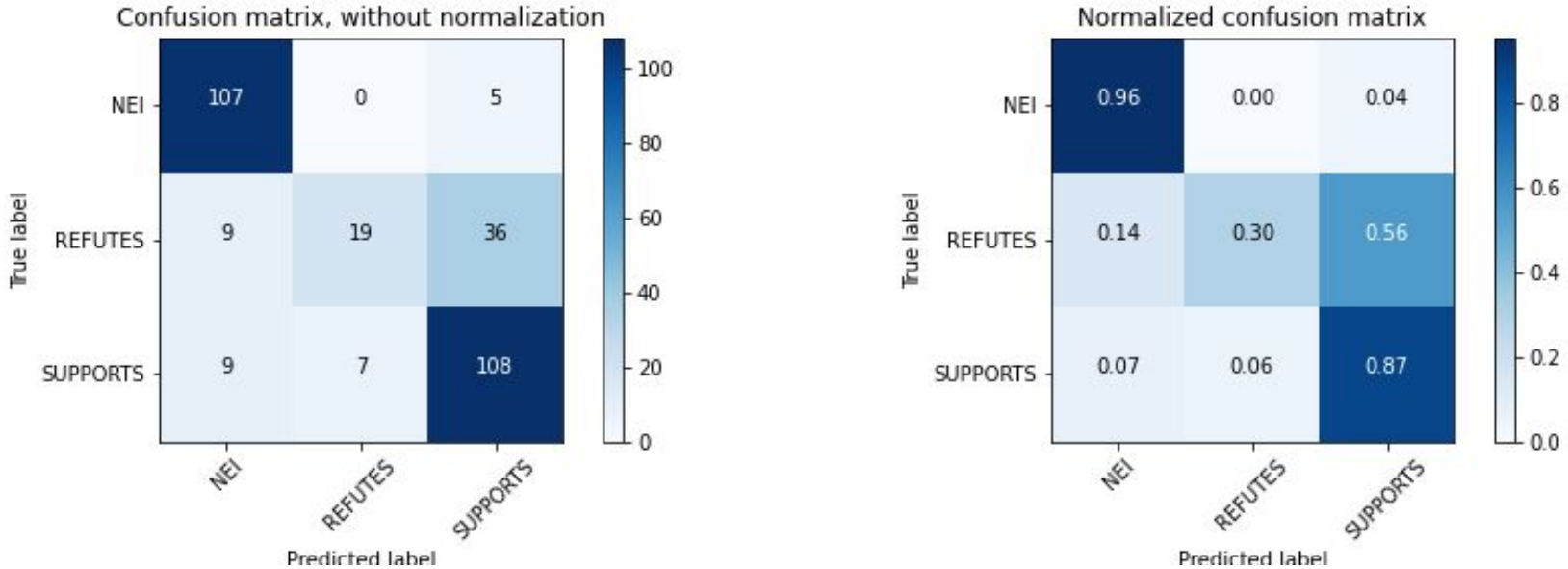


Figure: Confusion Matrix for Label Accuracy on the SCIFACT Development Set.

RESULTS

Label Prediction

Model	Full	Single Evidence	Multi Evidence
Transformer-XH	78.0	68.75	82.35

Table: Label Accuracy on SCIFACT Dev Claims.

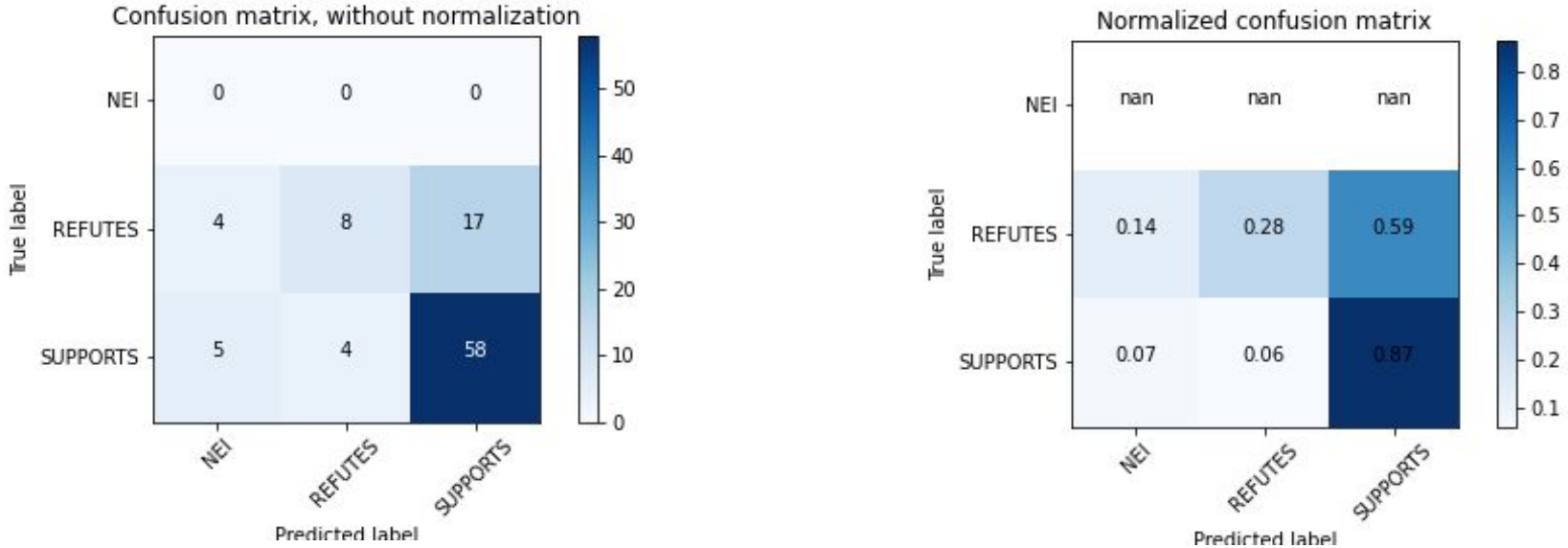


Figure: Confusion Matrix for Label Accuracy on the Single Evidence Claims.

RESULTS

Label Prediction

Model	Full	Single Evidence	Multi Evidence
Transformer-XH	78.0	68.75	82.35

Table: Label Accuracy on SCIFACT Dev Claims.

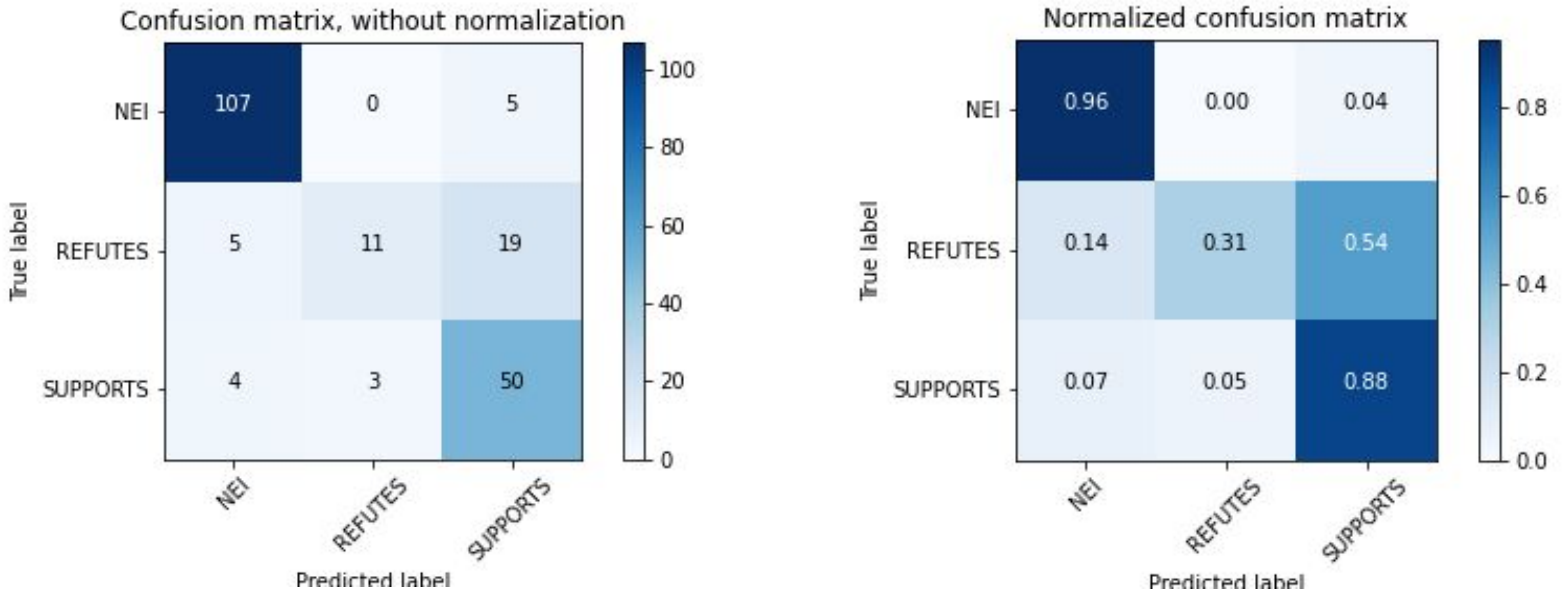


Figure: Confusion Matrix for Label Accuracy on the Multi Evidence Claims.

RESULTS

Label Prediction

Training Data	Model	Label Prediction
SCIFACT	BioMedRoBERTa	71.7
SCIFACT	RoBERTa-base	62.9
SCIFACT	SCIBERT	69.2
SCIFACT	RoBERTa-large	75.7
UKP Snopes	RoBERTa-large	71.3
FEVER	RoBERTa-large	67.6
FEVER+SCIFACT	RoBERTa-large	**81.9**
FEVER+SCIFACT	Transformer-XH(Full)	**78.0**
FEVER+SCIFACT	Transformer-XH(Single Evidence)	68.75
FEVER+SCIFACT	Transformer-XH(Multi Evidence)	**82.35**

Table: Comparison on Different Training Dataset and Models for Label Prediction.

RESULTS

Abstract Retrieval

- TF-IDF

Number of Abstracts	Hit One	Hit All
1	0.7467	0.7333
3	**0.8467**	**0.8333**
5	0.89	0.87
10	0.9267	0.91
20	0.95	0.94
50	0.9767	0.9633
100	0.9867	0.98

Table: Number of Abstracts and Hit Scores for Abstract Retrieval.

RESULTS

Rationale Selection

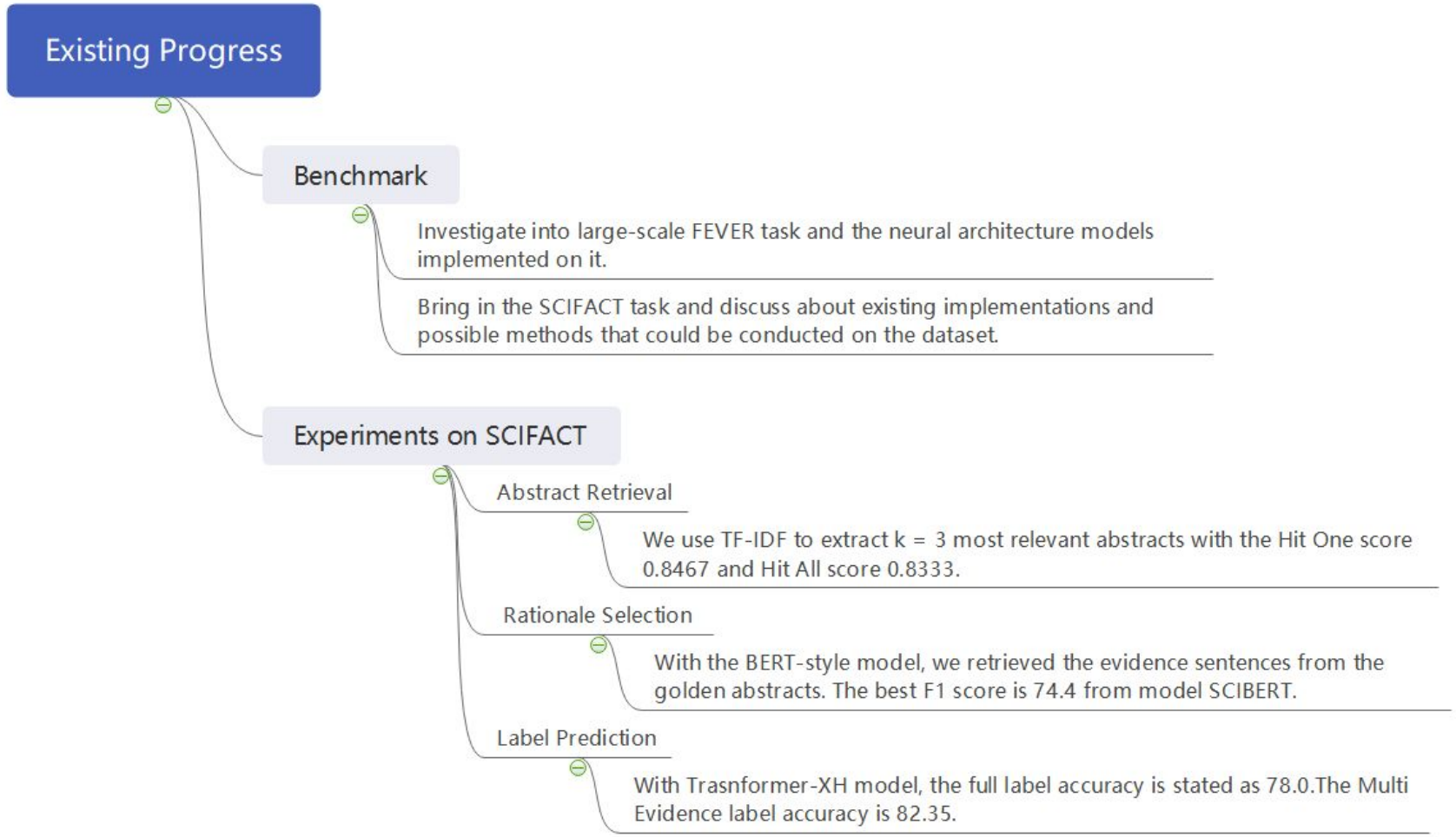
- BERT

		Rationale Selection		
Training Data	Model	P	R	F1
SCIFACT	BioMedRoBERTa	75.3	69.9	72.5
FEVER	RoBERTa-large	41.5	57.9	48.4
UKP Snopes	RoBERTa-large	42.5	62.3	50.5
FEVER+SCIFACT	RoBERTa-large	72.4	67.2	69.7
SCIFACT	RoBERTa-base	76.1	66.1	70.8
SCIFACT	RoBERTa-large	**73.7**	**70.5**	**72.1**
SCIFACT	SCIBERT	**74.5**	**74.3**	**74.4**

Table: Results for Rationale Selection.

CONCLUSION

Existing Progress



CONCLUSION

Future Work

- Improve the work on abstract retrieval with mention or key word-based method in combination with TF-IDF.
- Improve the work on rationale selection with continuous in-domain training method.
- Conduct pipeline evaluation.
- Study on verifying scientific claims with daily basis corpus can bring huge application value.

THANK YOU

Question?

ULB UNIVERSITÉ LIBRE DE BRUXELLES

ULB BRUFACE BRUSSELS FACULTY OF ENGINEERING VUB

VUB VRIJE UNIVERSITEIT BRUSSEL