Master Thesis Final Presentation

Multilingual and Cross-Modal Embedded Representation for Tweets Aggregation

Houcem Ben Makhlouf

Jury Members:
Prof. Dr. techn. Wolfgang Nejdl
Prof. Dr. Ralph Ewerth

Supervisors:
Dr. Erick Elejalde
Sergej Wildemann, M.Sc.

Contents

Introduction

Motivation

- Twitter is good source of information

- Twitter data is rich:
  - Reflects the user behaviour
  - Understand the public opinion

- Twitter raw data is challenging to deal with

→ Data representation learning can be used for better understanding of the raw data

Problem Definition

- Twitter Data representation learning
  - Deep learning based model

- Recent work
  - Concentrate on few modalities
  - Rarely combine multilingual and multimodal
  - Absence of rich multilingual multimodal datasets

- → Multilingual and multimodal representation of tweets by leveraging a topic classification task.

Dataset

Collection Process

- Data collection from news outlet
  - Tweet text with interactions(replies, quotes)
  - Based on three languages(English, German, French)

Language	Tweets	Replies	Quotes
English	28.94	28.47	38.56
French	28.17	25.05	21.65
German	37.20	42.15	26.03
Other	5.69	4.33	13.76

Annotation Process

- Hashtag Classification
  - Preliminary manual inspection of hashtags
  - Co-occurrence matrix

Subset	Tweets	Replies	Quotes
MCMTRA_1	36,049	55,436	32,737
MCMTRA_2	13,056	52,123	26,767

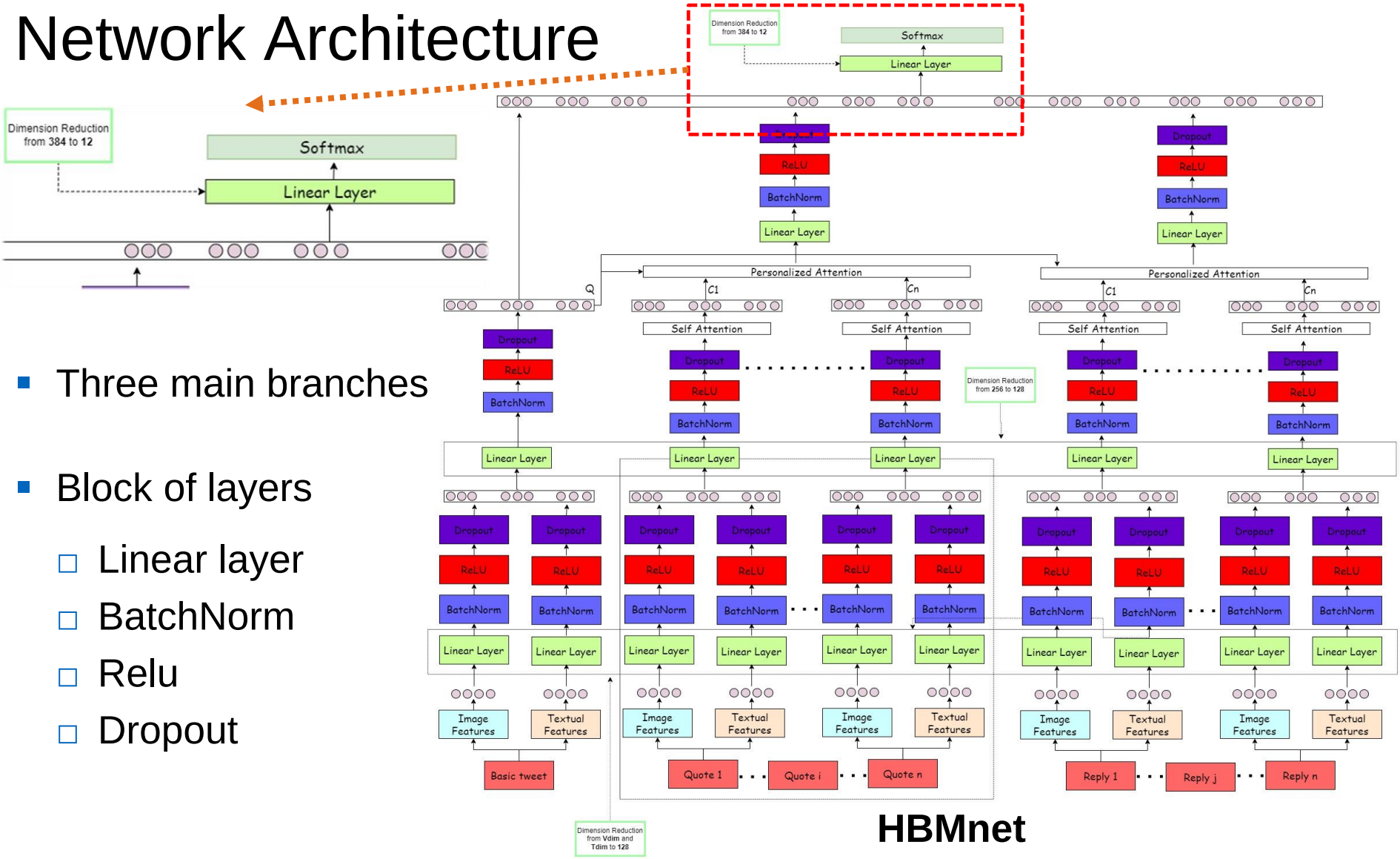
Proposed Method

Proposed Method

- Deep NN taking tweet, replies and quotes as input
  - Visual and text embeddings(VSE)
- Based on pre-trained models for embeddings extraction
  - XLM-RoBERTa and ResNet-50 for text and image respectively
- Attention mechanisms to learn intra and inter modal features
  - Intra-modal: within each reply and quote
  - Inter-modal: between tweet and corresponding replies and quotes

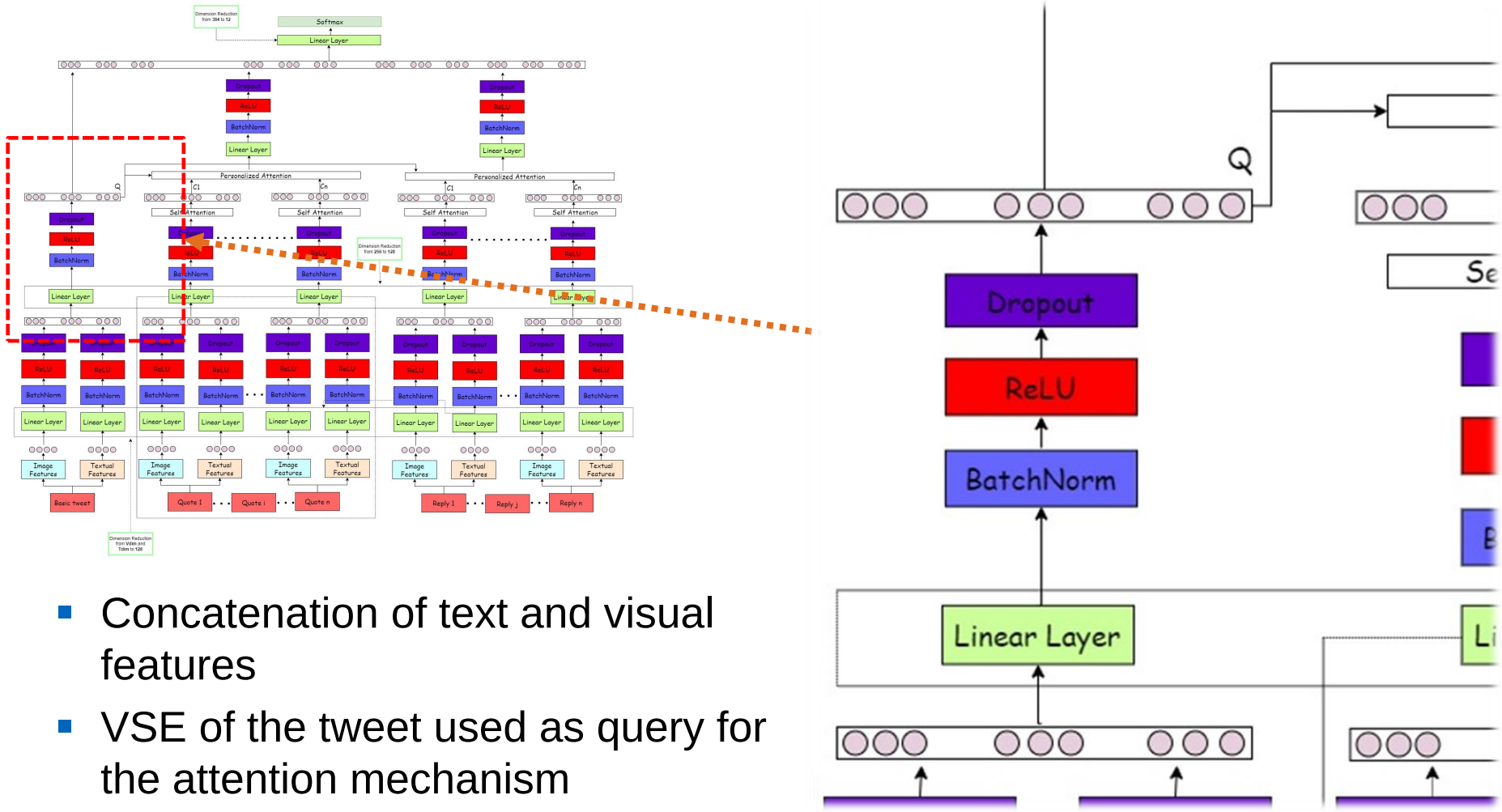
Network Architecture

- Three main branches
- Block of layers
  - Linear layer
  - BatchNorm
  - Relu
  - Dropout



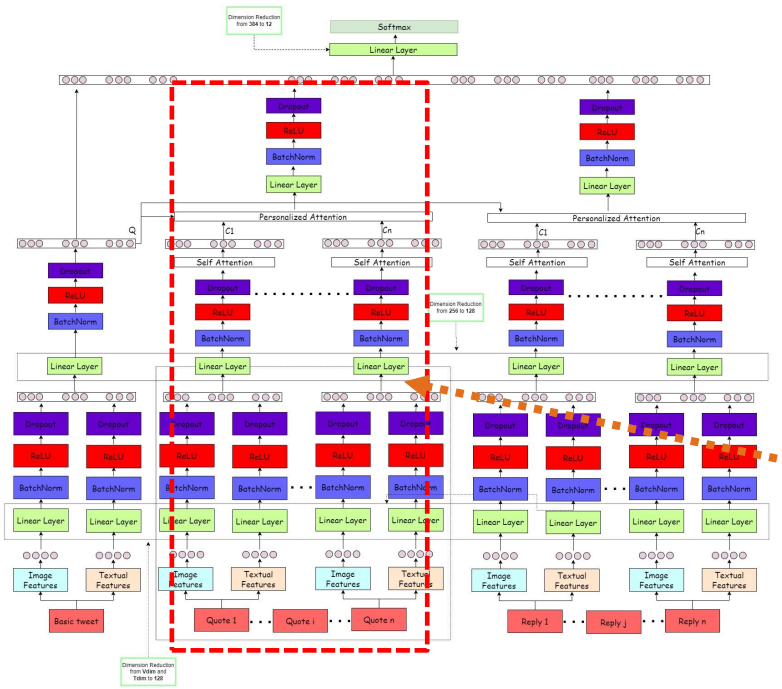
HBMnet

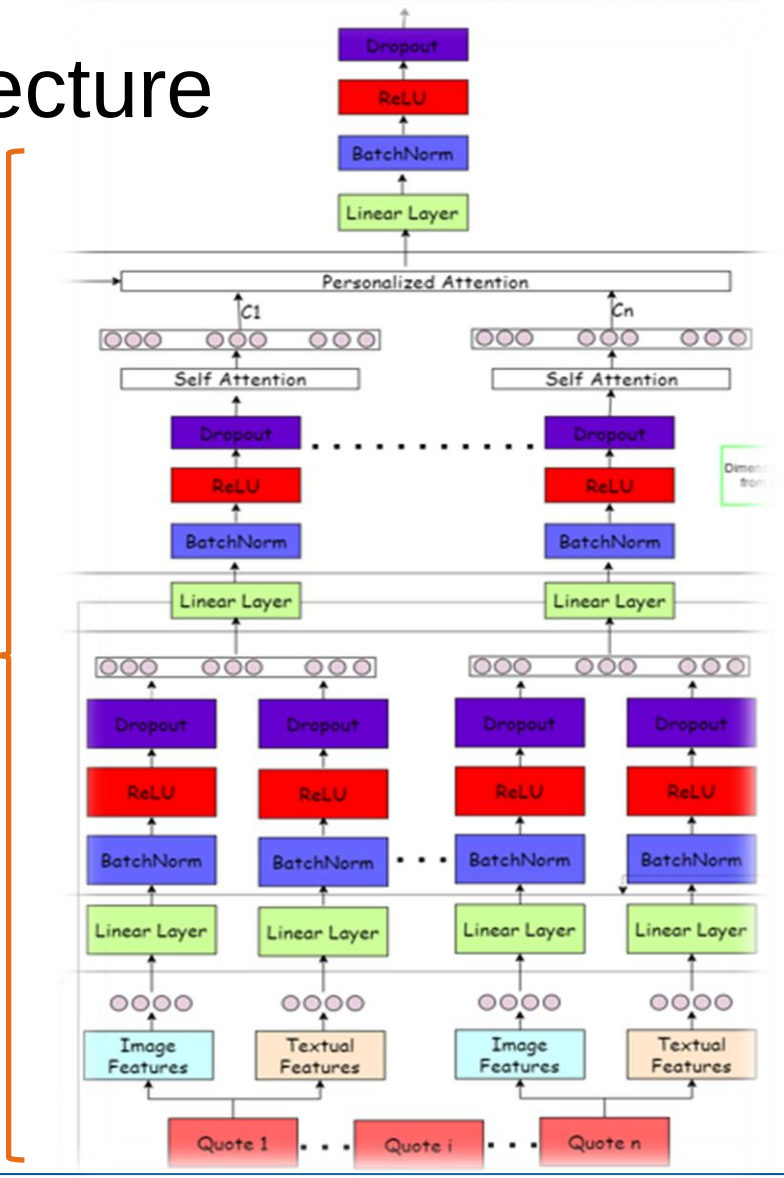
HBMnet: Network Architecture



- Concatenation of text and visual features
- VSE of the tweet used as query for the attention mechanism

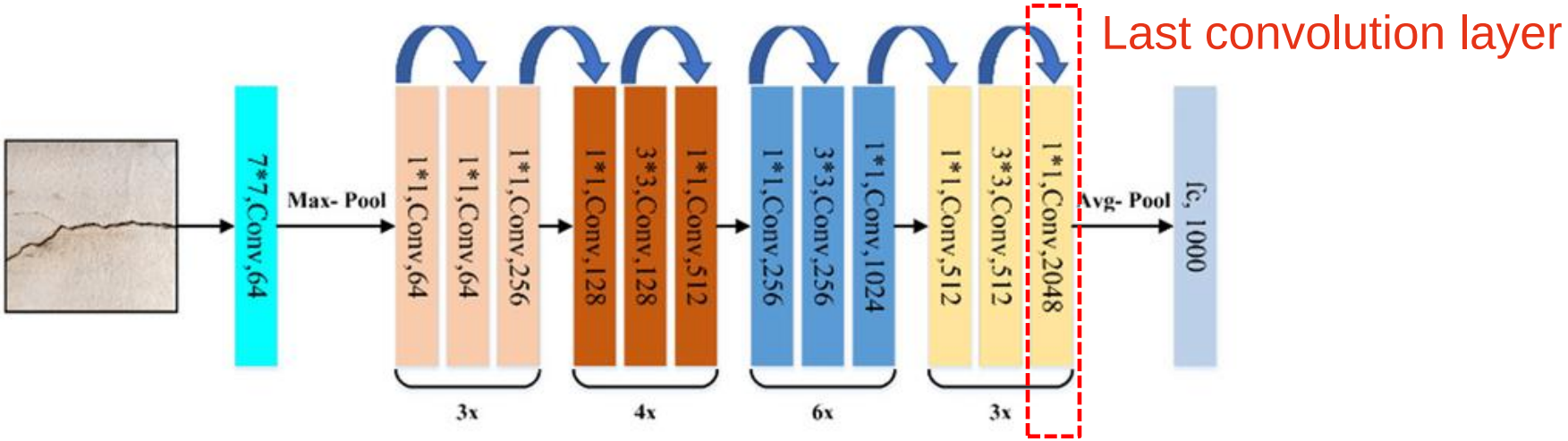
HBMnet: Network Architecture





- Self-attention to capture intra-modal features
- Replies and quotes used as context for the attention mechanism

Visual Tweet Representation



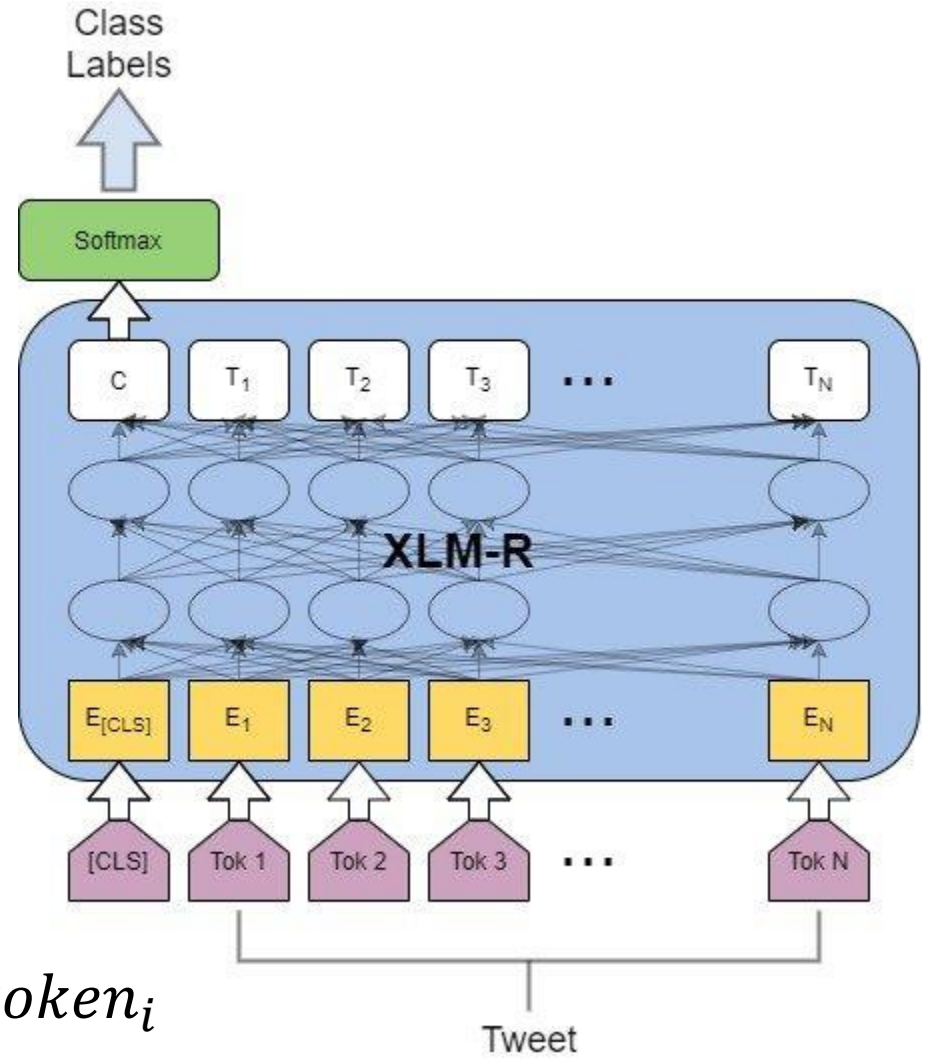
- Object features extraction with ResNet-50
- final convolutional layer outputs 2048 feature maps each of size 7 x 7

→ 2048-dimensional vector after pooling with a global average.

Text Tweet Representation

- Pre-processing of tweet text
- Meaning of words within one tweet sentence are equally important
- The average of the last four layers is the embedding taken in consideration

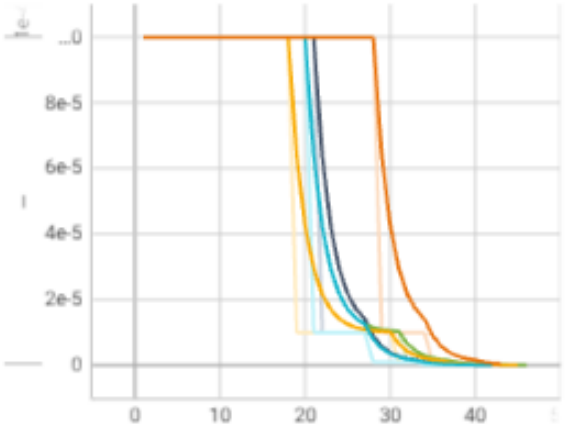
→ 768-dimensional vector for every tweet after averaging over word embeddings



Sentence embedding= $\frac{1}{n}\sum_{i=1}^{n} token_i$

Training

- 1000 max number of epochs with early stop
  - Training is stopped if the performance is not increased within the last 20 epochs
- Adam optimizer
- batch size = 128



- Initial learning rate: 1×10^{-4}
  - Multiplied by a factor of 0.1 if the validation loss does not decrease
  - Metrics quantity is updated every 5 epochs

Experiments And Results

Evaluation Metrics

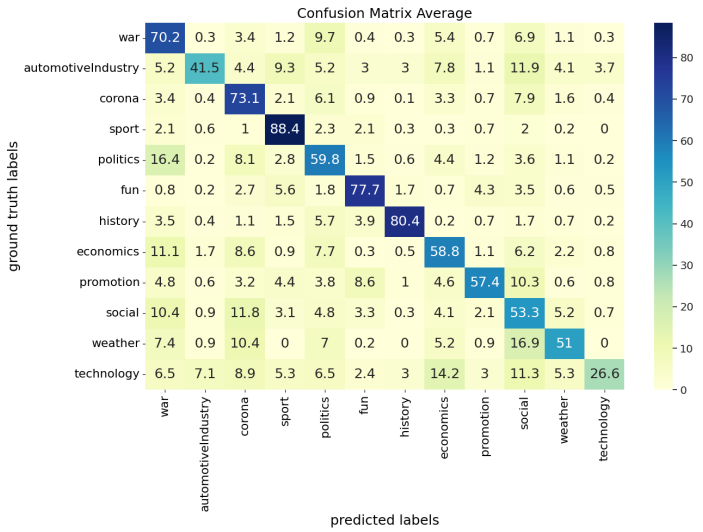
- Accuracy score

$$\frac{TP + TN}{TP + FN + TN + FP}$$

- Weighted F1 score

$$\frac{2 \times (precision \times recall)}{(precision + recall)}$$

- confusion matrix

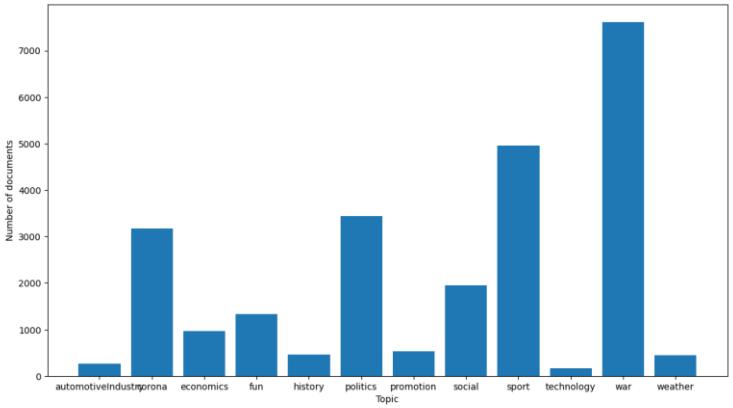


Initial Document Classification

Method	Model	MCMTRA_1		MCMTRA_2	
		ACC	F1	ACC	F1
Multi-label	hbmnet_baseline_t	62.13	69.73	61.83	70.33
	hbmnet_baseline_tv	**63.95**	**70.22**	**62.12**	**70.62**
Multi-class	hbmnet_baseline_t	64.52	68.82	63.83	74.53
	hbmnet_baseline_tv	**64.67**	**69.02**	**64.12**	**74.56**

Data imbalance Handling

- Data imbalance was detected in:
  - Diagonal of the confusion matrix filled with zeros
- weighted random sampler provided by PyTorch
  - Generates weights for each class based on the supplied number of samples

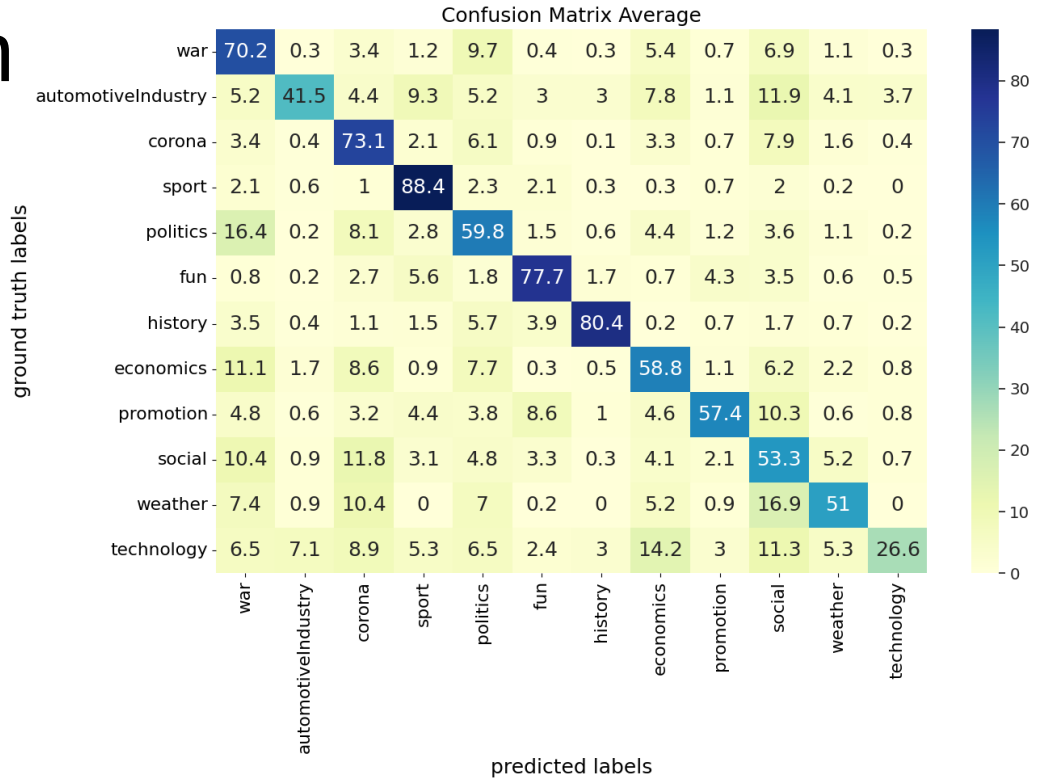


Multimodal Document Classification

Method	Model	MCMTRA_1		MCMTRA_2	
		ACC	F1	ACC	F1
Multi-class	hbmnet	65.13	68.63	66.23	69.81
	hbmnet_w_likes	64.23	67.35	66.26	69.36
	hbmnet_w_self_attention	66.44	**69.85**	68.76	**70.76**
	hbmnet_w_attention	**66.65**	69.56	**68.89**	70.36

Results Combination

- Best proposed model after averaging over 5 folds
- Impact of data imbalance still persists



Subdataset	Model	ACC	F1
MCMTRA_2	hbmnet_attention_likes	68.33	70.09
	hbmnet_w_double_attention	**70.42**	**73.46**

Multilingual document classification

- Evaluation for the multilingual document classification task
- Three topics(”corona”, ”politics” and ”sport”)
- Three languages(German, French and English)
  - Train for German language
  - Test for French and English languages

Metric	Valset German	Testset French	Testset English
Accuracy	86.09	85.64	73.64
F1	86.22	85.02	75.41

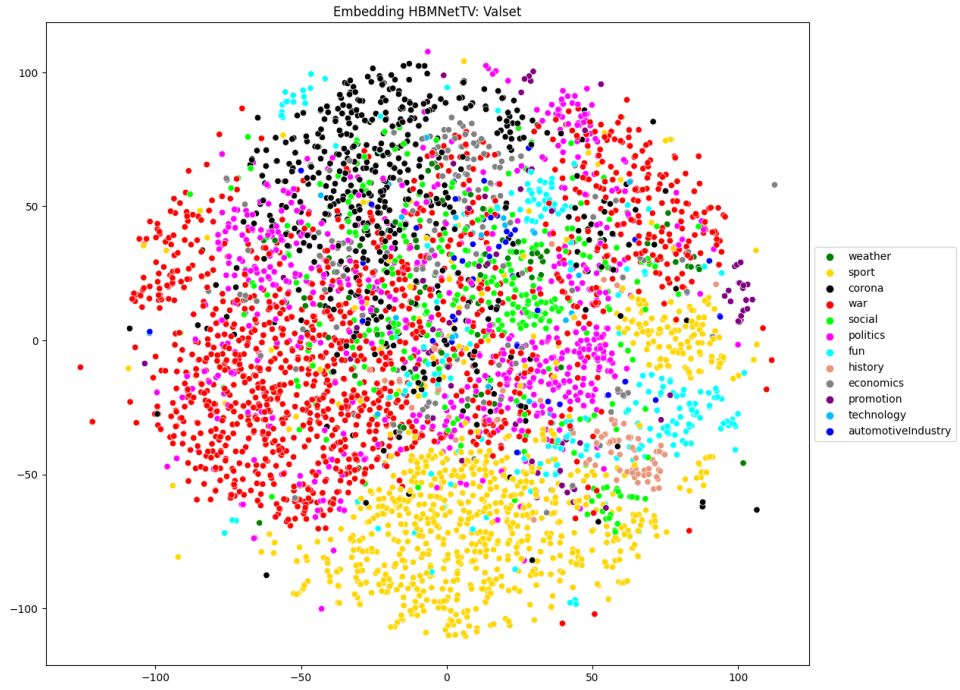
Multilingual document classification

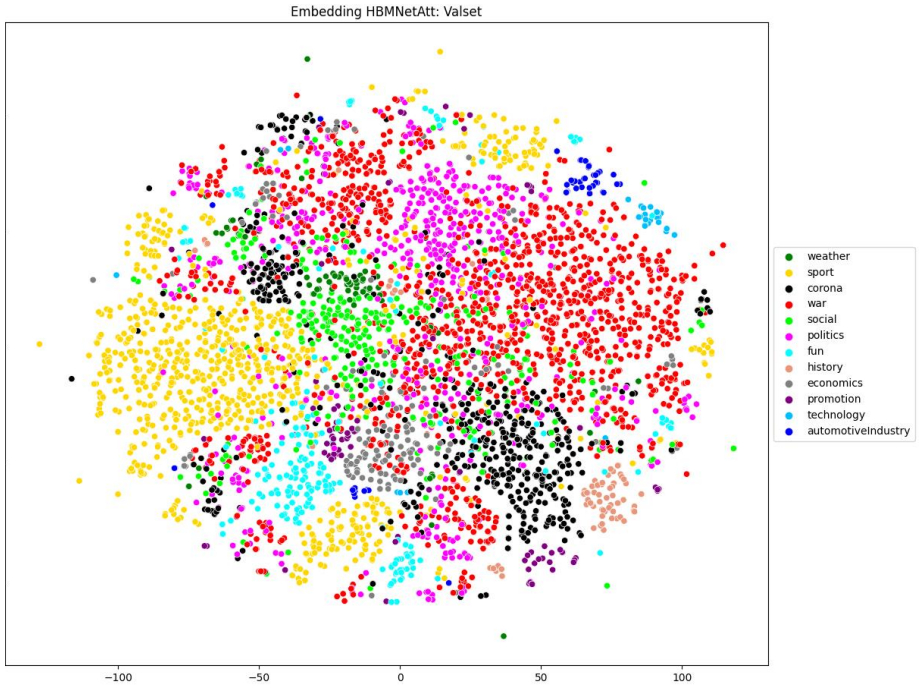
Test with English language

Test with French language

- High-dimensional embedding vectors using t-SNE
  - Each high-dimensional data point is projected in a two-dimensional space

Multilingual Multimodal Document Representation





bimodal baseline

HBMnet

- High-level document representations of bimodal baseline model and best proposed model

Conclusion

Conclusion

- Deep Learning-based Method for data representation learning by leveraging a topic classification task
- Based on state-of-the-art encoders to extract text and vision features
- Usage of tweet with corresponding interactions
- Build dataset “MCMTRA” to assess the effectiveness of the proposed approach
- Using multimodalities(replies, quotes and images) in multilingual manner improves the tweet representation

Future Work

- Labelling process of the data
- More investigation of topics
  - Some topics can be merged
  - The classification task can concentrate on pattern detection
- More techniques for data imbalance handling
- Pre-processing
- automatic hyperparameter tuning

Thank you!

Questions