Leveraging Transformer Models for Multilingual Hierarchical Classification of Articles

A Study of Unified mBERT and XLM-R with Attention Networks

Team 28

Pavithra Raghavan, Takitazwar Parthib, Nila Maitra Chaity, Sabikunnahar Talukder Pyaasa, Isra Zaman

1 Problem

The task focuses on identifying narratives, classifying them, and analyzing entity roles to support analysts studying target-specific disinformation.

Objective

- **Task**: Assign accurate narrative and sub narrative labels to news articles using a two-level hierarchical taxonomy.
- **Goal**: Improve classification accuracy across diverse, multilingual news content.

Challenges:

- Multiple interrelated multi-label, multi-class dataset.
- Multilingual dataset.
- Fine-grained hierarchical classification.

Research Questions:

RQ1: How effectively can mBERT and XLM-R models perform in multi-label, multi-class narrative classification within a two-level taxonomy of narrative and sub narrative labels?

RQ2: To what extent does the proposed model generalize across datasets in different languages, and how does its adaptability affect performance in multilingual contexts?

2 Data

Data Overview and Visualizations

- **Source:** News articles provided in five languages (e.g., English, Russian, etc.).[4]
- **Annotations:** Articles labeled with multiple narratives and sub-narratives.
  - **Hierarchical Taxonomy:**
    - Main Narrative - 3 (2 main classes + fallback case)
    - Sub-Narrative - 20
    - Sub-Sub-Narrative - 69

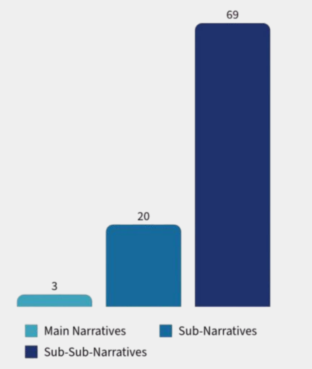

Fig 1: Number of Unique Classes

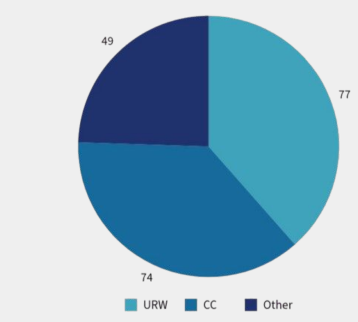

Fig 2: Number of Article/Narrative Class

Correlation Between Narrative Levels:

- **Main Narratives & Sub-narratives:** Main narratives (e.g., "Climate Change") connect to specific sub-narratives (e.g., "Green policies as geopolitical tools").
- **Interconnected Propagation:** Sub-narratives align with broader themes (e.g., "Hidden plots" under "Climate Change").
- **Insight Potential:** Analyzing correlations uncovers how narratives propagate and evolve.

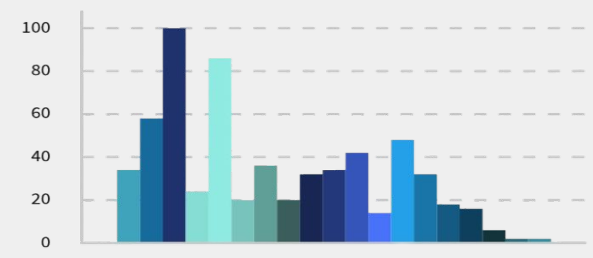

Fig 3: Number of Article per Sub Narrative Class

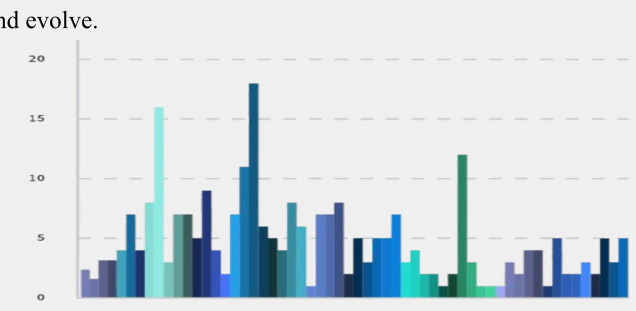

Fig 4: Number of Article per Sub Narrative Class

Class Imbalance Observations:

- **Disproportionate Representation:** Largest class has 123 articles; smallest class has only 2 articles.
- **Impact:** Skews model performance, favoring majority classes and underperforming on minority ones.

Strategies used: Class Weights, SMOTE (Synthetic Minority Oversampling Technique)

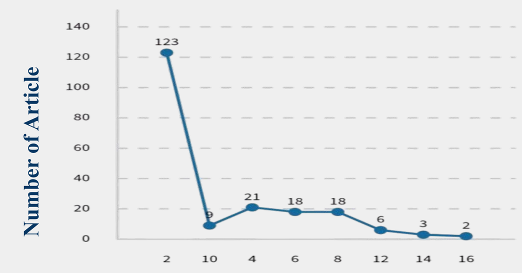

Fig 5: Class Frequency vs Number of Articles

3 Approach

Model: Unified mBert

Data Preprocessing

- **Dual Encoders** tokenizer extracts rich language features. [1]
- MultiLabelBinarizer encodes multi-label data for training and decodes predictions.

Hierarchical Classification Framework:

- **Flattened** hierarchical labels into a single vector representation.
- **Reconstructed** hierarchical relationships from predicted class values.

Transformer-based Model:

- 24 transformer layers, 768 hidden units, fused via a 1024-dimensional ReLU layer.
- Includes a dropout layer (rate 0.3) and a fully connected classifier with sigmoid activation.
- **Fusion layer** combines encoder outputs; **final classifier** produces multi-label output.

Training and Evaluation:

- **Used AdamW** optimizer and **Binary Cross-Entropy** with Logits Loss.
- Evaluation Metrics: **F1-score**, Precision, and Recall, Accuracy.

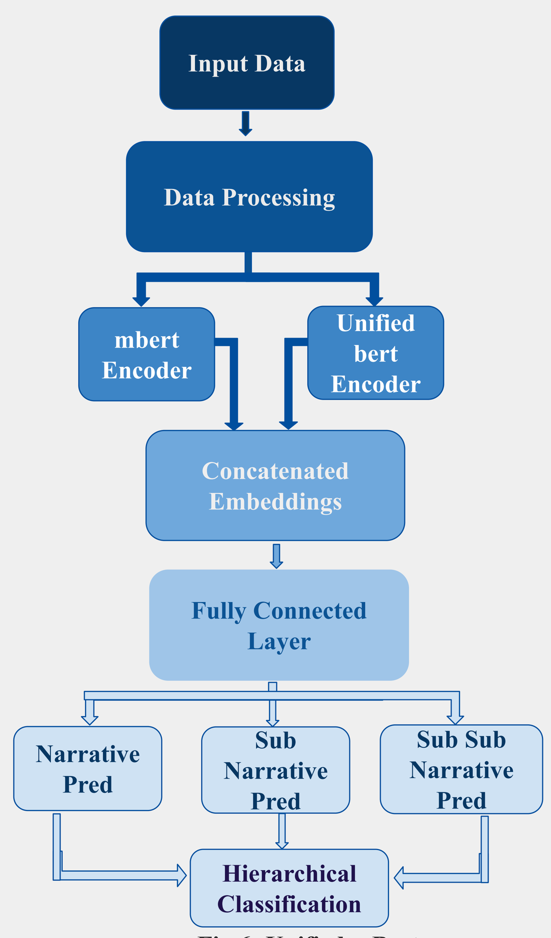

Fig 6: Unified mBert

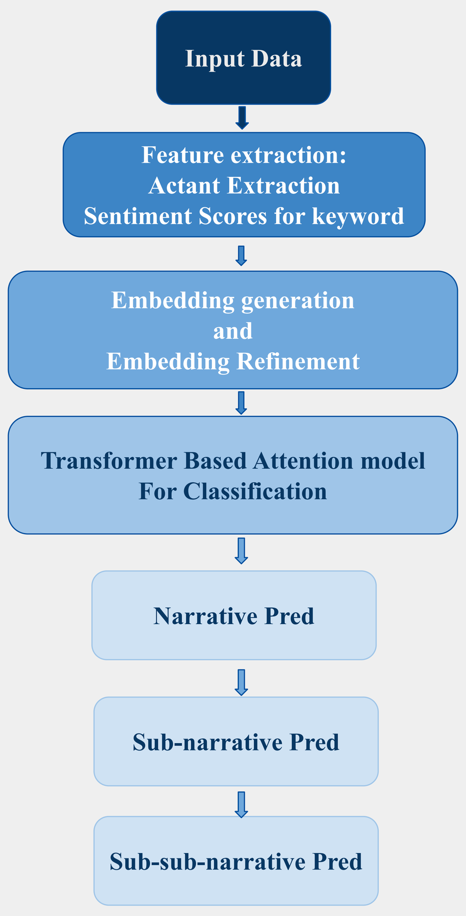

Fig 7: XLM-R model

Model: XLM-R embedded Transformer Based Attention Model

Data Preprocessing:

- Extracted top 10 keywords from Level 1 classes and computed sentiment scores.
- Tokenized articles and actants [3] using **XLM-R embeddings[2]** for feature representation.
  a. actants used: subject, object, helper, and opponent

Transformer Based Attention Model:

- Multi-layer transformer with **16-head attention**, **dropout (30%)**, and **L2 regularization**.
- Positional encoding and **batch normalization** were included for stable and efficient learning.

Training and Evaluation:

- Hierarchical training where each level's outputs informed the next level (Narrative → Level 1 → Level 2).
- Early stopping and learning rate decay ensured efficient convergence.
- Metrics computed: **Accuracy**, **F1-Score (Macro/Weighted)**, and **Balanced Accuracy**.

4 Result

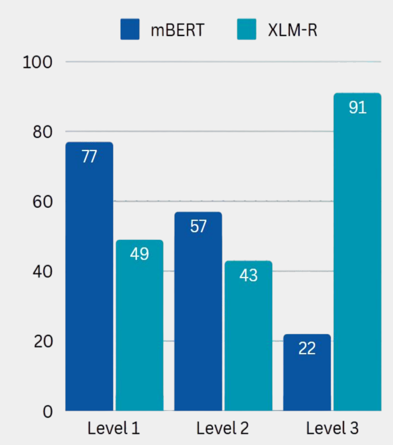

Fig 8: Overall Accuracy

Level 1 - Narrative
Level 2 - Sub-narrative
Level 3 - Sub-sub-narrative

	Narrative	Sub-Narrative	Sub-Sub-Narrative
English mBert	71%	38%	18%
English XLM-R	87%	26%	33%
Russian mBert	78%	35%	13%
Russian XLM-R	80%	20%	19%

Fig 9: F1 scores

5 Analysis

Predicted Values	Actual Values: Positive	Actual Values: Negative
Positive	57	38
Negative	82	177

Fig 10:mBert model Confusion Matrix

Predicted Values	Actual Values: Positive	Actual Values: Negative
Positive	105	59
Negative	46	248

Fig 11: XLM-R model Confusion Matrix

	True Values	Predicted Values
0	1, 7, 38	1, 2, 42
1	0, 4, 18	0, 15, 51
2	0, 13, 8	0, 16, 29
3	1, 10, 33	1, 10, 15
4	1, 6, 36	1, 6, 36
5	0, 17, 37	0, 17, 37

Fig 12: mBert Sample Hierarchical Predictions

	True Values	Predicted Values
0	2, 8, 22	2, 8 , 11
1	0, 0, 4	0, 0, 4
2	1, 1, 1	1, 1, 1
3	2, 7, 9	2, 10, 9
4	2, 8, 22	2, 8, 11
5	0, 0, 3	0, 0, 4

Fig 13: XLM-R Sample Hierarchical Predictions

- Captured some hierarchical relationships.
- Struggled with accurate contextual predictions, indicating need for better handling of dependencies between levels.
- Performance across classes showed inconsistencies, with the model favoring certain classes, highlighting challenges with imbalanced data distributions.
- Issues with Unified mBert model:
  - **Thresholding:** Default threshold might be suboptimal.
  - **Language-Specific Accuracy:** Lower due to inadequate domain-specific fine-tuning.
- Issues with XLM-R embedded model:
  - **Accuracy Increases:** Measures overall performance, not class-specific.
  - **F1 Score Decreases:** Penalizes class imbalance.

6 Conclusion/Future Work

- The models effectively classify narratives into multiple levels.
- Performance can improve for complex and detailed narrative elements.
- Results emphasize the need for refinement in handling diverse languages and structures.

Future work:

- To leverage language-specific adapters, alignment techniques, and enhanced feature extraction to improve multilingual adaptation.
- Data Augmentation to solve Class-imbalance.
- Fine-tuning the model for unsupervised learning and domain-specific narratives.

7 Reference

[1] J. Libovický, R. Rosa, and A. Fraser, “How Language-Neutral is Multilingual BERT?,” *arXiv.org*, Nov. 08, 2019. https://arxiv.org/abs/1911.03310
[2] Z. Chi *et al.*, “XLM-E: Cross-lingual Language Model Pre-training via ELECTRA,” *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Jan. 2022, doi: 10.18653/v1/2022.acl-long.427.
[3] D. Herman, “Existentialist roots of narrative actants,” *Studies in 20th & 21st Century Literature*, vol. 24, no. 2, Jun. 2000, doi: 10.4148/2334-4415.1484.
[4] “Call for task proposals,” *SemEval-2025*. https://semeval.github.io/SemEval2025/cft.html

UNIVERSITÄT BONN INFORMATIK CAISA Lab