Evidence Detection

MANCHESTER 1824
The University of Manchester

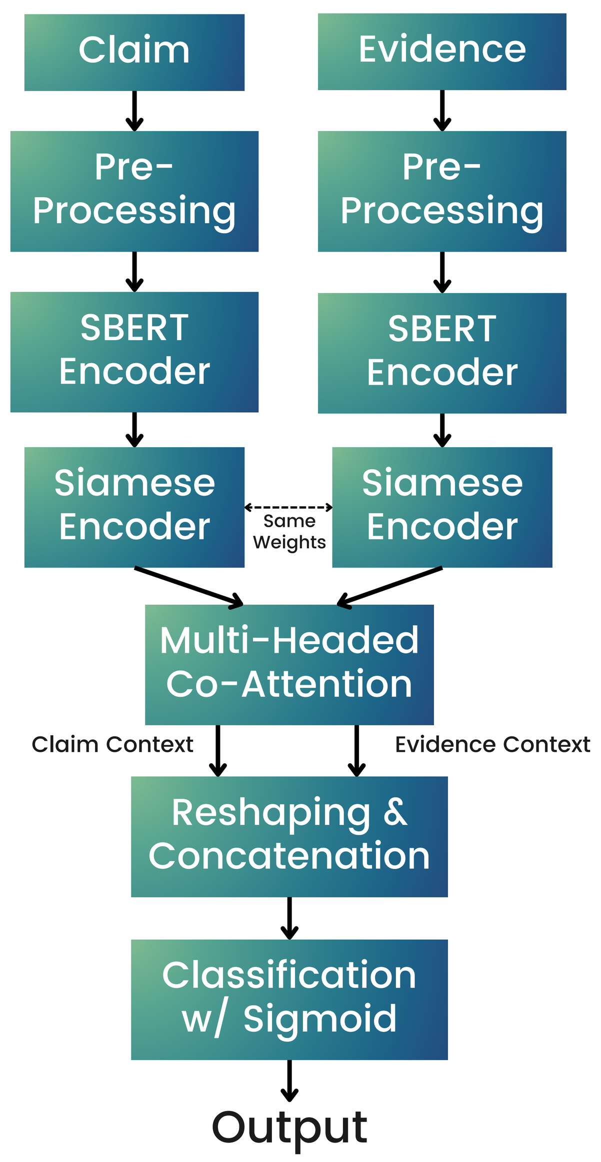
By: Dhruv Sharma & Tuan Chuong Goh

Introduction

Evidence Detection (ED) is an important area of NLU which analyses whether a piece of evidence is related to a claim [1]. ED supports fact-checking systems, academic research validation, and information verification pipelines.

ED in NLU is a classification task which determines whether a piece of text (evidence) substantiates, contradicts, or is relevant to a given claim. This fundamental capability enables automated verification systems, improves information retrieval accuracy, and helps combat misinformation by establishing connections between assertions and their supporting evidence [2,3].

Siamese Model with Multi-Headed Co-Attention



Input Processing:

- **Pre-processing** - Remove citation markers [ref] from text.
- **SBERT Embeddings** - Convert claim and evidence texts into 384-dimensional vectors.

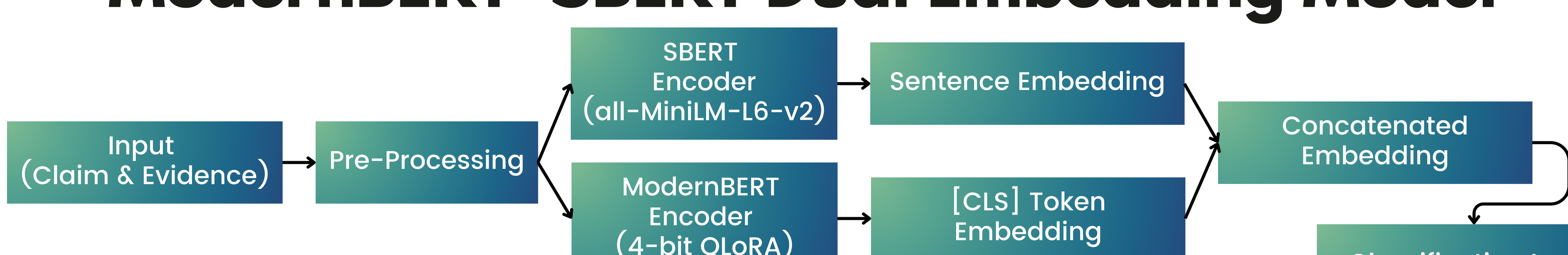
Relationship Modelling:

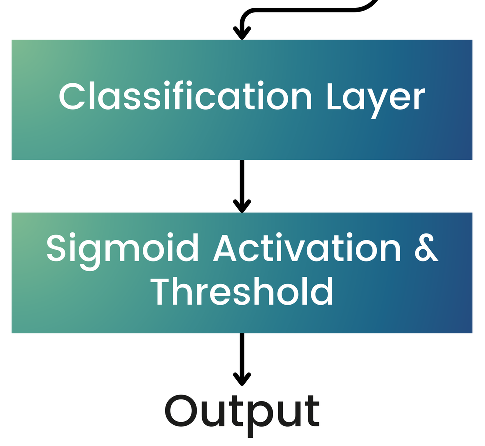
- **Shared Siamese Encoder** - Processes claim and evidence using identical layers with shared weight to create a representation in a unified feature space.
- **Multi-Headed Co-Attention** - Enable bi-directional attention where claims attend to relevant evidence elements and vice versa, with multiple heads to capture different relationship patterns simultaneously.

Classification:

- **Feature Fusion** - Concatenates and reshapes the claim and evidence contexts for the decision layer to make an informed choice
- **Decision Layers** - Processes combined features through dense layers with dropout and batch normalization
- **Output** - Produces predictions using Sigmoid Activation Function and then thresholded to get the binary results.

ModernBERT+SBERT Dual Embedding Model





Input Processing:

- **Pre-processing** - Removes citation markers ([ref]), normalises text using Unidecode for accents, eliminates irregular spacing around punctuation, and standardises whitespace
- **Dual Embedding Approach:**
  - SBERT Encoder (all-MiniLM-L6-v2) - 384-dimensional sentence embeddings
  - ModernBERT Encoder (4-bit QLoRA) - Contextual [CLS] token embeddings

Relationship Modeling:

- **Separate Embedding Paths** - SBERT embeddings are produced separately for claim and evidence, then averaged
- **Feature Integration** - [CLS] token embeddings from ModernBERT capture contextual understanding
- **Combined Representation** - Both embedding types are concatenated to form a unified feature space

Classification:

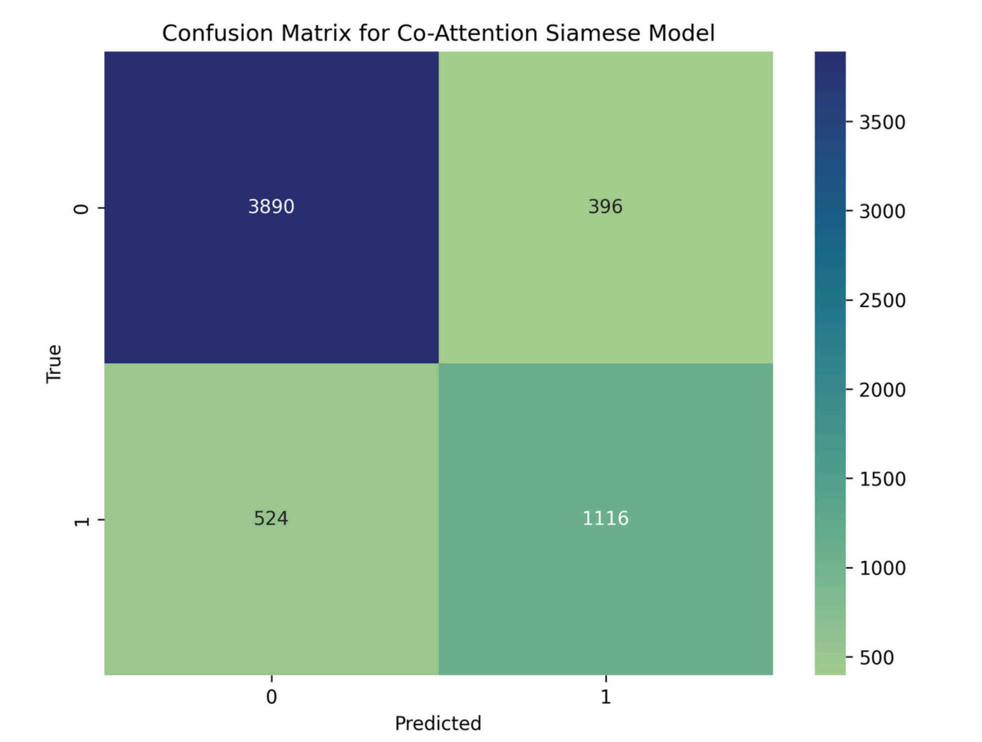
- **Dense Neural Network** - Processes the combined embeddings through layers
- **Output** - Produces predictions using Sigmoid Activation Function for claim-evidence relevance using a threshold

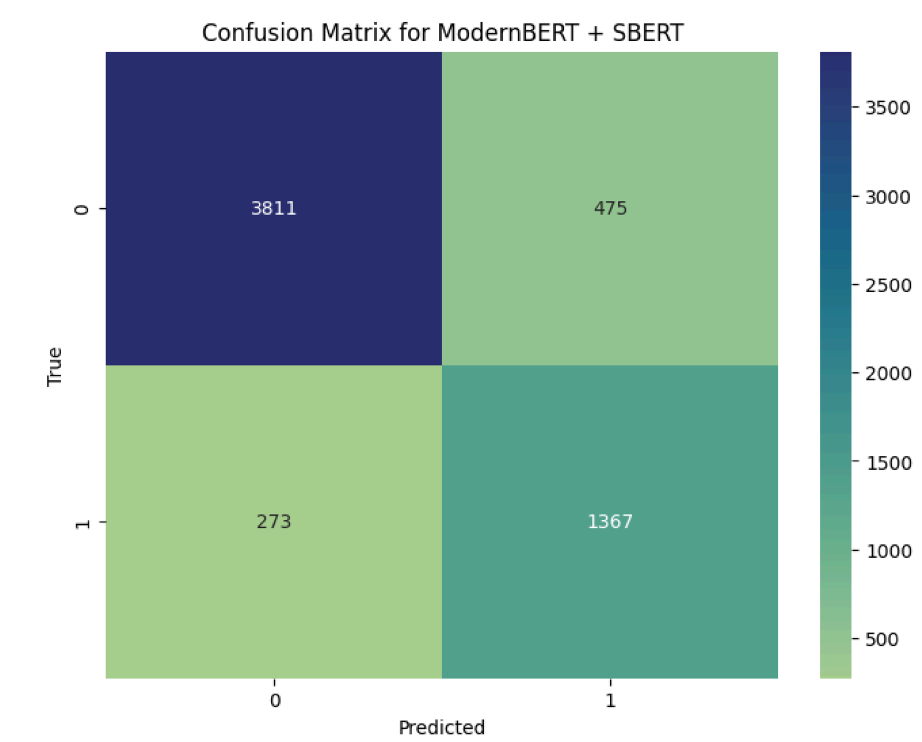
Model Enhancements:

- **Data Augmentation** - Synonym replacement for improved generalization
- **Complementary Embeddings** - Combines semantic similarity with contextual understanding
- **Efficient Implementation** - Memory-optimized with 4-bit quantization and Low-Rank Adaptation

Results & Analysis

Model / Metric	Accuracy	Macro Precision	Macro Recall	Macro F1-Score	Weighted Precision	Weighted Recall	Weighted F1-Score	MCC
Co-Attention Siamese Model	84.4%	80.9%	79.4%	80.1%	84.2%	84.5%	84.3%	0.60
Transformer Model	87.4%	83.8%	86.1%	84.8%	88.0%	87.4%	87.6%	0.699





- Our ModernBERT+SBERT dual embedding model achieved excellent performance with 87.4% accuracy and a weighted F1-score of 87.6%.
- The ModernBERT+SBERT model outperformed the standard BERT baseline (87.09% weighted F1-score) by 0.51 percentage points while maintaining comparable accuracy but improving memory efficiency through 4-bit quantization.
- Similarly, our Co-Attention Siamese model (84.3% weighted F1-score) significantly surpassed the LSTM baseline (78.36% weighted F1-score) by 5.94 percentage points, demonstrating the effectiveness of the co-attention mechanism for claim-evidence relationship modelling.
- Analysis of the confusion matrices reveals that the ModernBERT+SBERT approach reduced false negatives compared to the Co-Attention Siamese model (273 vs 524), indicating better sensitivity in identifying relevant evidence.
- The MCC scores further highlight the improvement: ModernBERT+SBERT (0.699) vs BERT baseline (0.6748) and Co-Attention Siamese (0.60) vs LSTM baseline (0.4674).
- Both approaches showed particular strength in precision metrics, with ModernBERT+SBERT achieving 0.88 weighted precision compared to Co-Attention Siamese's 0.842.

Conclusion

- Our Co-Attention Siamese model is a powerful architecture for evidence detection, with multi-headed co-attention mechanisms enabling a nuanced understanding of claim-evidence relationships.
- The ModernBERT+SBERT dual embedding approach leverages complementary strengths of semantic similarity and contextual understanding, creating a unified representation that captures complex textual relationships.
- Synonym replacement augmentation in the ModernBERT+SBERT model enhances its robustness to linguistic variations, making it particularly effective at identifying evidence despite variations in expression, though at a tradeoff having different precision and recall.
- Future work could explore ensemble methods combining both architectures' strengths or investigate domain-specific adaptations.
- It would be better to utilise the ModernBERT model for more robust results, but if inference time is more important, the Siamese Model is a great alternative.

References:

[1] J. Thorne, A. Vlachos, C. Christodoulopoulos, and A. Mittal, “FEVER: a large-scale dataset for Fact Extraction and VERification,” arXiv.org, Mar. 14, 2018. https://arxiv.org/abs/1803.05355

[2] K. Singh, P. Reisert, N. Inoue, P. Kavumba, and K. Inui, “Improving Evidence Detection by Leveraging Warrants,” Proceedings of the Second Workshop on Fact Extraction and VERification (FEVER), pp. 57–62, Jan. 2019, doi: 10.18653/v1/d19-6610.

[3] R. Aly et al., “FEVEROUS: Fact Extraction and VERification Over Unstructured and Structured information,” arXiv.org, Jun. 10, 2021. https://arxiv.org/abs/2106.05707