**Technical Explanation: Hindi Emotion Classification using BioBERT, BlueBERT & MultiBERT**

By Abhiraj Ghose

E23CSEU0014

Project Overview

This project implements a comprehensive emotion classification system for Hindi emotional poetry using three transformer-based language models: BioBERT, BlueBERT, and MultiBERT. The research investigates the effectiveness of domain-specific biomedical models versus multilingual models for cross-lingual emotion analysis in Hindi text.

Technical Architecture

Model Implementations

The project implements six distinct model variants across three transformer architectures:

1. MultiBERT Models: Built on `bert-base-multilingual-cased`, designed for multilingual understanding

2. BioBERT Models: Based on `dmis-lab/biobert-base-cased-v1.1`, pre-trained on biomedical literature

3. BlueBERT Models: Using `bionlp/bluebert-base-uncased-ms`, specialized for clinical text processing

Each model is implemented with both basic configurations and enhanced versions incorporating Hindi emotional embeddings. The core architecture consists of:

- Base Transformer Layer: Pre-trained BERT variants with 12 transformer layers

- Classification Head: Linear layer with dropout (0.3-0.6) for regularization

- Optional Feature Fusion: Integration of Hindi emotional embeddings (100-dimensional)

Hindi Emotional Embeddings System

A key innovation is the development of a comprehensive Hindi emotional vocabulary system containing 200+ terms across six categories:

- Core Emotions: दर्द (pain), खुशी (happiness), प्रेम (love), डर (fear)

- Intensity Modifiers: बहुत (very), तीव्र (intense), गहरा (deep)

- Poetry Terms: कविता (poetry), गजल (ghazal), शेर (couplet)

- Psychological States: मन (mind), आत्मा (soul), भावना (emotion)

- Relationship Terms: रिश्ता (relationship), दोस्ती (friendship)

- Life Concepts: जिंदगी (life), समय (time), मौत (death)

The embeddings are computed using random initialization with uniform distribution [-0.1, 0.1] and averaged when multiple emotional terms are present in the text.

Dataset and Preprocessing

Dataset Characteristics

- Size: 240 balanced samples (80 each: Negative, Neutral, Positive)

- Source: Hindi emotional poetry with diverse linguistic expressions

- Quality: 91.7% consistency rate through manual validation

- Language: Pure Hindi text with Devanagari script

Data Processing Pipeline

1. Text Preprocessing: Unicode normalization and whitespace handling

2. Label Encoding: Systematic emotion-to-integer mapping

3. Stratified Splitting: 70% training, 20% testing, 10% validation

4. Tokenization: Model-specific tokenizers with max length 128-512 tokens

5. Feature Extraction: Optional Hindi emotional embeddings integration

Training Methodology

Hyperparameter Optimization

The training employs carefully tuned hyperparameters based on empirical analysis:

- Learning Rates:

  - MultiBERT: 1e-5 (reduced from 2e-5 for stability)

  - BioBERT: 1e-5 (optimized for biomedical domain)

  - BlueBERT: 5e-6 (significantly reduced to prevent class collapse)

- Batch Size: 8 (memory-optimized for available hardware)

- Epochs: 6-10 (conservative strategy to prevent overfitting)

- Sequence Length: 128 tokens (optimal for Hindi poetry)

Regularization and Stability Features

To address overfitting and training instability:

- Enhanced Dropout: 0.5-0.6 (increased from baseline 0.3)

- Weight Decay: L2 regularization (0.01-0.025)

- Gradient Clipping: Maximum norm 0.5-1.0

- Early Stopping: Patience of 3-4 epochs with validation monitoring

- Learning Rate Scheduling: ReduceLROnPlateau with factor 0.3-0.5

Performance Analysis

Model Performance Comparison

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | F1-Score | AUC-ROC | Key Insight |
| MultiBERT (Basic) | 65.31% | 0.6548 | 0.8347 | Best overall performance |
| MultiBERT + Hindi | 61.22% | 0.6105 | 0.7849 | Hindi features showed mixed results |
| BlueBERT + Hindi | 52.08% | 0.5232 | 0.6999 | Domain mismatch evident |
| BioBERT + Hindi | 50.00% | 0.5082 | 0.7116 | Medical domain less effective |
| BlueBERT (Basic) | 35.42% | 0.2619 | 0.6100 | Baseline performance |
| BioBERT (Basic) | 33.33% | 0.3102 | 0.4915 | Random guessing level |

Critical Findings

1. Language Appropriateness > Domain Specificity: MultiBERT's multilingual capabilities significantly outperformed biomedical domain-specific models for Hindi text processing.

2. Domain Mismatch Effect: BioBERT and BlueBERT, despite sophisticated architectures, showed poor performance due to English biomedical pre-training conflicting with Hindi emotional content.

3. Feature Engineering Impact: Hindi emotional embeddings provided modest improvements (2-6%) for domain-specific models but showed diminishing returns for MultiBERT.

4. Class-wise Performance Patterns:

   - Negative emotions: Most accurately classified (78.57% precision)

   - Positive emotions: Moderate performance (66.67% precision)

   - Neutral emotions: Most challenging (52.94% precision)

Technical Implementation Details

Training Infrastructure

- Hardware: CPU-based training with 4.5-minute total execution time

- Framework: PyTorch with Transformers library

- Optimization: AdamW optimizer with linear warmup scheduling

- Monitoring: Real-time loss tracking with early stopping mechanisms

Evaluation Metrics

Comprehensive evaluation using multiple metrics:

- Primary: Accuracy and F1-score (macro/weighted)

- Statistical: Matthews Correlation Coefficient (MCC)

- Probabilistic: AUC-ROC for multi-class classification

- Per-class: Precision, Recall, F1-score for each emotion

Visualization and Analysis

- Confusion Matrices: Heat-map visualization of classification patterns

- Training Curves: Loss and accuracy progression with validation monitoring

- Performance Comparison: Radar charts and bar plots for model comparison

Conclusions and Technical Insights

The project successfully demonstrates that language-appropriate models significantly outperform domain-specific models when there's a mismatch between pre-training domain and target application. The 65.31% accuracy achieved by MultiBERT represents a substantial improvement over random baseline (33.33%) and validates the importance of multilingual pre-training for cross-lingual emotion analysis.

The comprehensive regularization strategy effectively prevented overfitting, while the Hindi emotional embeddings system, though providing marginal improvements, offers a foundation for future feature engineering approaches. The systematic evaluation framework provides reproducible benchmarks for future research in Hindi emotion classification.

This work establishes a robust technical foundation for emotion analysis in low-resource languages and demonstrates the critical importance of model-task alignment in transfer learning applications.