Fake News Detection Using Deep Learning and Fuzzy Logic

Authors:

Abdullah Al Masood
Department of CSE, RUET

Md. Rabiul Islam
Department of CSE, RUET

Md. Faysal Ahamed
Department of ECE, RUET

1. Introduction
2. Problem Statement
3. Related Work
4. Research Questions
5. Objectives
6. Outcomes
7. Dataset
8. Methodology
9. Results

Introduction

- **Digital Age Transformation:** The internet, particularly social media, has become a primary source of news and information.
- **Rise of Fake News:** Misinformation threatens social order, undermines media credibility, and impacts democratic processes.
- **Rapid Spread and Influence:** Fake news often spreads faster and has greater impact than genuine news.
- **Need for Detection Tools:** Urgent need for reliable methods to detect and mitigate fake news.

Problem Statement

How can we develop advanced techniques to enhance fake news detection by leveraging deep learning models, incorporating methods to effectively address uncertainty, complexity, and variability in diverse datasets?

Related Work (1)

Title: *An Enhanced Fake News Detection System With Fuzzy Deep Learning (2024)*

Authors: Cheng Xu, M-Tahar Kechadi

Methods: Developed FDHN model integrating fuzzy logic with deep learning; evaluated on LIAR and enhanced LIAR2 datasets.

Contributions: Achieved state-of-the-art results; improved dataset quality; addressed uncertainty in detection.

Limitations: Lacks full utilization of attention mechanisms, potentially limiting focus on specific content.

Difference: Our work enhances fuzzy logic with attention mechanisms to improve focus and uncertainty handling.

Related Work (2)

Title: *Fake News Detection on Social Media Using a Natural Language Inference Approach (2022)*

Authors: Fariba Sadeghi, Amir Jalaly Bidgoly, Hossein Amirkhani

Methods: Used Natural Language Inference (NLI) with related news from reputable sources; introduced FNID dataset.

Contributions: Demonstrated NLI utility; enhanced accuracy with external information; provided new dataset.

Limitations: Relies on availability of accurate related news; computationally intensive; scaling challenges.

Difference: Our approach integrates fuzzy logic and attention, reducing reliance on external sources.

Related Work (3)

Title: *Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach (2021)*

Authors: Jamal Abdul Nasir, Osama Subhani Khan, Iraklis Varlamis

Methods: Combined CNN and RNN for classification; evaluated on ISO and FA-KES datasets.

Contributions: Outperformed non-hybrid methods; explored generalization; analyzed dataset variability.

Limitations: Generalization issues; requires large labeled datasets; potential overfitting.

Difference: Our model uses fuzzy logic and attention to enhance generalization and reduce data dependency.

Research Questions

- How can fuzzy logic be optimized within deep learning frameworks to improve fake news detection accuracy?
- Can attention mechanisms enhance the handling of uncertainty in fact-checking processes?
- How effective is the integration of fuzzy logic and attention mechanisms in addressing dataset variability?

Objectives

- Develop a deep learning model integrating optimized fuzzy logic and attention mechanisms.
- Enhance detection accuracy by focusing on critical textual features and managing uncertainty.
- Evaluate the model on diverse datasets to ensure robustness and generalization.

Outcomes

- **Outcomes:** Improved fake news detection accuracy; robust model with better generalization across datasets.
- **Impacts:** Enhanced media credibility; support for democratic processes; reduced spread of misinformation.

LIAR and LIAR2 Datasets

- **LIAR Dataset (2017):** Benchmark with 12,823 human-labeled U.S. political statements, classified into six truthfulness categories (True, Mostly True, Half True, Mostly False, False, Pants on Fire). Includes basic contextual data (speaker, subject, context). Limited by sparse contextual features, labeling errors, and lack of detailed speaker or justification texts.
- **LIAR2 Dataset:** Enhanced version with 23,000 statements. Improvements include detailed speaker descriptions, justification texts, temporal data, corrected labeling errors, and enriched contextual/numerical features.
- **Benefits:** Improved model accuracy, better handling of nuanced patterns, and mitigation of LIARs limitations.

Research Workflow

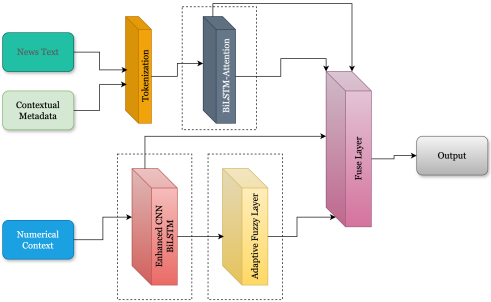

Figure 1: Proposed Research Workflow

Model Architecture Components

Textual Data Processing

- **Tokenization**: Uses **BERT's tokenizer** to convert news statements into token IDs, handling subword tokenization and padding.
- **Embedding Layer**: Maps tokens to dense vector representations.
- **BiLSTM with Attention**: Applies a **Bidirectional LSTM (BiLSTM)** to capture sequential dependencies, followed by a self-attention mechanism to focus on relevant tokens.
- **Output**: Produces a feature vector for truthfulness classification.

Model Architecture Components

Contextual Metadata Processing

- **Input**: Processes metadata fields (subject, speaker, job title, state, party affiliation, context) using **BERT's tokenizer**.
- **Embedding and BiLSTM**: Concatenates tokenized metadata, embeds it, and applies a **BiLSTM** with self-attention to extract contextual features.
- **Output**: Generates a feature vector capturing metadata context.

Numerical Metadata Processing

- **Input**: Utilizes numerical features (counts of previous truth assessments).
- **Enhanced CNN-BiLSTM**: Employs a **1D CNN** with multiple kernel sizes to extract local patterns, followed by a **BiLSTM** with attention to model dependencies.

Model Architecture Components

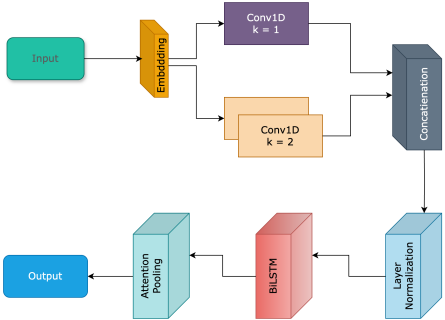

Figure 2: Enhanced CNN-BiLSTM Architecture for Numerical Metadata

Model Architecture Components

- **Fuzzy Logic**: Integrates an **Adaptive Fuzzy Layer** with Gaussian membership functions and feature attention to handle uncertainty.
- **Adaptive Fuzzy Layer Details**:
  - Uses Gaussian membership functions with learnable mean (`membership_miu`) and standard deviation (`membership_sigma`).
  - Feature attention via learnable weights (`feature_weights`) with sigmoid activation for soft feature selection.
  - Output: Feature vector from averaged weighted memberships for classification.
- **Output**: Produces a feature vector for classification.

Model Architecture Components

Feature Fusion and Classification

- **Concatenation**: Combines feature vectors from:
  - News Text Module (BiLSTM output).
  - Text Metadata Module (BiLSTM output).
  - Numerical Metadata Module (CNN-BiLSTM and Fuzzy Layer output).
- **Fusion Layer**: Feeds concatenated features into a fully connected layer with dropout for robust integration.
- **Classification**: Computes probabilities for six truthfulness categories (pants-fire, false, barely-true, half-true, mostly-true, true) using a multi-class classification loss function.
- **Training**: Optimizes the model to minimize validation loss, saving the best-performing model.

Hyperparameters and Architectural Details

Table 1: Hyperparameters and Architectural Details

Parameter	Value
Random Seed	42
Vocabulary Size	30,522
Embedding Dimension	256
BiLSTM Hidden Dimension	128
BiLSTM Layers	1
CNN Filters (Enhanced CNNBiLSTM)	32
CNN Kernel Sizes	1, 2
Dropout Rate	0.3
Batch Size	32
Learning Rate	0.001
Max Epochs	100
Number of Runs	10

Result and Analysis

Comparative Performance of Proposed Model

Dataset	Model	Test Loss	Accuracy (%)	F1 Macro (%)	F1 Micro (%)
LIAR	FDHN (Original)	0.3425	46.50	47.00	46.50
	FDHN (Reproduced)	0.3556	44.28	44.19	44.28
	Proposed (BA + ECB + AF)	**0.3394**	**46.49**	**46.50**	**46.49**
LIAR2	FDHN	0.2145	70.95	70.17	70.95
	Proposed (BA + ECB + AF)	**0.2041**	**72.47**	**72.01**	**72.47**

Table 2: Comparison of FDHN and Proposed Model on LIAR and LIAR2 datasets

Training and Validation Loss Trend

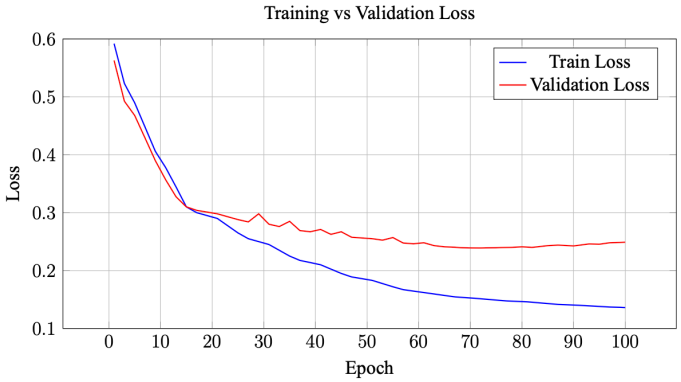

Figure 3: Training and Validation Loss on LIAR2 Dataset

Confusion Matrix on LIAR2 (Proposed Model)

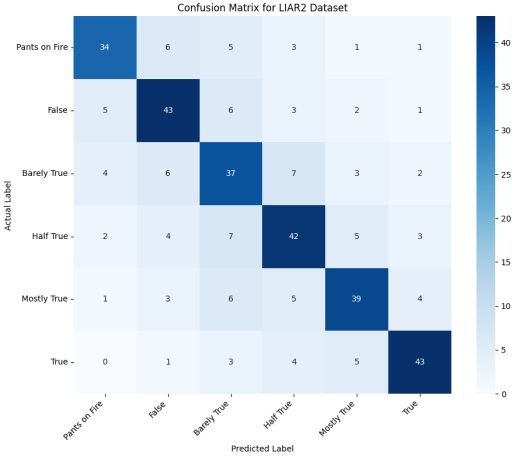

Figure 4: Confusion Matrix on LIAR2 Dataset

Component-wise Performance on LIAR2

Configuration	Test Loss	Accuracy (%)	F1 Macro (%)	F1 Micro (%)
FDHN (Baseline)	0.2145	70.95	70.17	70.95
BA	0.2076	72.13	71.12	72.13
BA + ECB	0.2054	71.99	71.06	71.99
BA + ECB + AF	**0.2041**	**72.47**	**72.01**	**72.47**

Table 3: Component-wise Performance Improvement on LIAR2 Dataset[1]

[1] BA: BiLSTM-Attention; ECB: Enhanced CNN-BiLSTM; AF: Adaptive Fuzzy Layer

Component Analysis: BiLSTM-Attention

- **Purpose:** Replace baseline TextCNN with BiLSTM-Attention to capture long-range dependencies in claims and justifications.
- **Design Rationale:**
  - **BiLSTM:** Captures bidirectional sequential context, critical for nuanced reasoning.
  - **Attention Layer:** Highlights key tokens, amplifying important evidence in short or ambiguous claims.
- **Performance Insights:**
  - +0.08% accuracy on LIAR; +1.18% on LIAR2.
  - Particularly effective for dispersed evidence in justification paragraphs.
- **Interpretability:** Attention weights provide transparent rationale, aiding fact-checker trust.

Component Analysis: Enhanced CNNBiLSTM

- **Purpose:** Improve metadata handling by refining numerical feature processing.
- **Design Enhancements:**
  - **Multi-scale CNN:** Kernels of size 1 and 2 capture granular and relational patterns.
  - **Layer Normalization:** Stabilizes learning across heterogeneous metadata.
  - **BiLSTM:** Models sequential dependencies in speaker credibility histories.
- **Performance Insights:**
  - +0.94% accuracy on LIAR; -0.14% on LIAR2.
  - Stronger impact with richer metadata (speaker history, justification).
- **Interpretability:** Highlights influential temporal patterns, aiding deception trend detection.

Component Analysis: Adaptive Fuzzy Layer

- **Purpose:** Introduce fuzzy logic to model uncertainty and adapt feature weights dynamically.
- **Core Innovations:**
  - **Learnable Membership Functions:** Adaptively capture feature distributions.
  - **Feature-wise Attention:** Emphasize predictive features, suppress noisy inputs.
  - **Dynamic Weighting:** Adjusts per class and dataset richness.
- **Performance Insights:**
  - +1.19% accuracy on LIAR (46.49% vs. 45.30%); +0.48% on LIAR2.
  - Significant reduction in test loss, improving robustness and generalization.
- **Adaptability:** Learns class-specific rules, making predictions more reliable.

Future Research Directions

- **Multimodal Expansion:** Integrate images, audio, and video with text using cross-modal attention or graph neural networks for better misinformation detection.
- **Domain Generalization:** Extend beyond political claims to domains like health, finance, and science to ensure broad applicability.
- **Real-Time Detection:** Optimize for low-latency deployment using pruning, quantization, and concept drift adaptation for dynamic online settings.
- **Explainable AI (XAI):** Employ SHAP, LIME, and visual heatmaps to enhance interpretability and user trust, especially for journalists and fact-checkers.
- **Language Diversity:** Expand to multilingual and low-resource settings via mBERT, cross-lingual transfer, and data augmentation.
- **Hybrid Knowledge Integration:** Leverage external knowledge bases (e.g., Wikidata, Snopes) using entity linking or graph embeddings for fact-grounded predictions.

References (1/2)

Xu, C., & Kechadi, M. T.
An Enhanced Fake News Detection System With Fuzzy Deep Learning.
IEEE Access, vol. 12, pp. 1–15, 2024.

Sadeghi, F., Bidgoly, A. J., & Amirkhani, H.
Fake News Detection on Social Media Using a Natural Language Inference Approach.
Information Processing & Management, vol. 59, no. 2, pp. 102763, 2022.

Nasir, J. A., Khan, O. S., & Varlamis, I.
Fake News Detection: A Hybrid CNN-RNN Based Deep Learning Approach.
International Journal of Information Management Data Insights, vol. 1, no. 1, pp. 100007, 2021.

References (2/2)

Hochreiter, S., & Schmidhuber, J.
Long Short-Term Memory.
Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

Zadeh, L. A.
Fuzzy Sets.
Information and Control, vol. 8, no. 3, pp. 338–353, 1965.

Wang, W. Y.
"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection.
Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 422–426, 2017.

Thank you!