BRAUDE
College of Engineering, karmiel

Authorship Attribution in Arabic Texts Using Deep Impostors Method

Nuwar Dabbah & Nour Dabbah

Research Project 26-1-R-15

Supervisors: Prof. Zeev Volkovich & Dr. Renata Avros

Authorship Attribution

Definition

Authorship attribution is the task of identifying the author of a given text by analyzing distinctive linguistic and stylistic writing patterns rather than topical content.

Intrinsic Approaches

Rely solely on internal stylistic features such as vocabulary usage, syntax, and structural patterns within the text itself.

Extrinsic Approaches

Compare questioned texts against reference texts from known authors to identify stylistic similarity and distinguish authorship.

Authorship Attribution in Arabic

The Challenge

- *Many texts have unclear or disputed authorship due to historical preservation and copying practices.*
- *The Arabic language presents high linguistic and stylistic variability.*
- *Existing computational tools for Arabic authorship analysis remain limited.*

Why Does It Matter ?

- *Many historically important texts have uncertain or disputed origins.*
- *Reliable authorship analysis supports literary and historical scholarship.*
- *Data driven approaches are needed to complement traditional methods.*
- *Addressing this gap advances research in Arabic Natural Language Processing.*

Research Objectives

Primary Goal *: Develop a computational research framework for authorship attribution in Arabic literary texts using the Deep Impostors methodology.*

Research Objectives

01

Framework Investigation

Investigate the suitability of the Deep Impostors framework for Arabic authorship recognition.

02

Stylistic Analysis

Analyze stylistic patterns in the selected Arabic texts across different authors and historical periods.

03

Linguistic Adaptation

Address linguistic variability in Arabic through adaptation of data driven methods and Deep Impostor one.

Limitations of Existing Solutions

> *Research on Arabic authorship attribution commonly applies traditional machine learning models and transformer based approaches such as BERT and AraBERT.*
> *These methods focus on extracting linguistic and semantic features and have shown promising results on modern Arabic texts.*

Closed Set Attribution

Most approaches rely on closed set attribution.

Modern Text Focus

Limited suitability for classical Arabic texts.

Document Level Analysis

Difficulty capturing stylistic variation across long documents.

Generic Approaches

Insufficient focus on author specific analysis.

Proposed Research Framework

1. The research proposes an authorship attribution framework based on the Deep Impostors methodology.

2. The framework adopts an extrinsic, impostor-based approach to model authorial style.

3. Texts are analyzed at the segment level rather than as a single document.

4. Stylistic behavior is represented through signal-level analysis.

5. The framework is designed to be robust to linguistic and stylistic variability in Arabic.

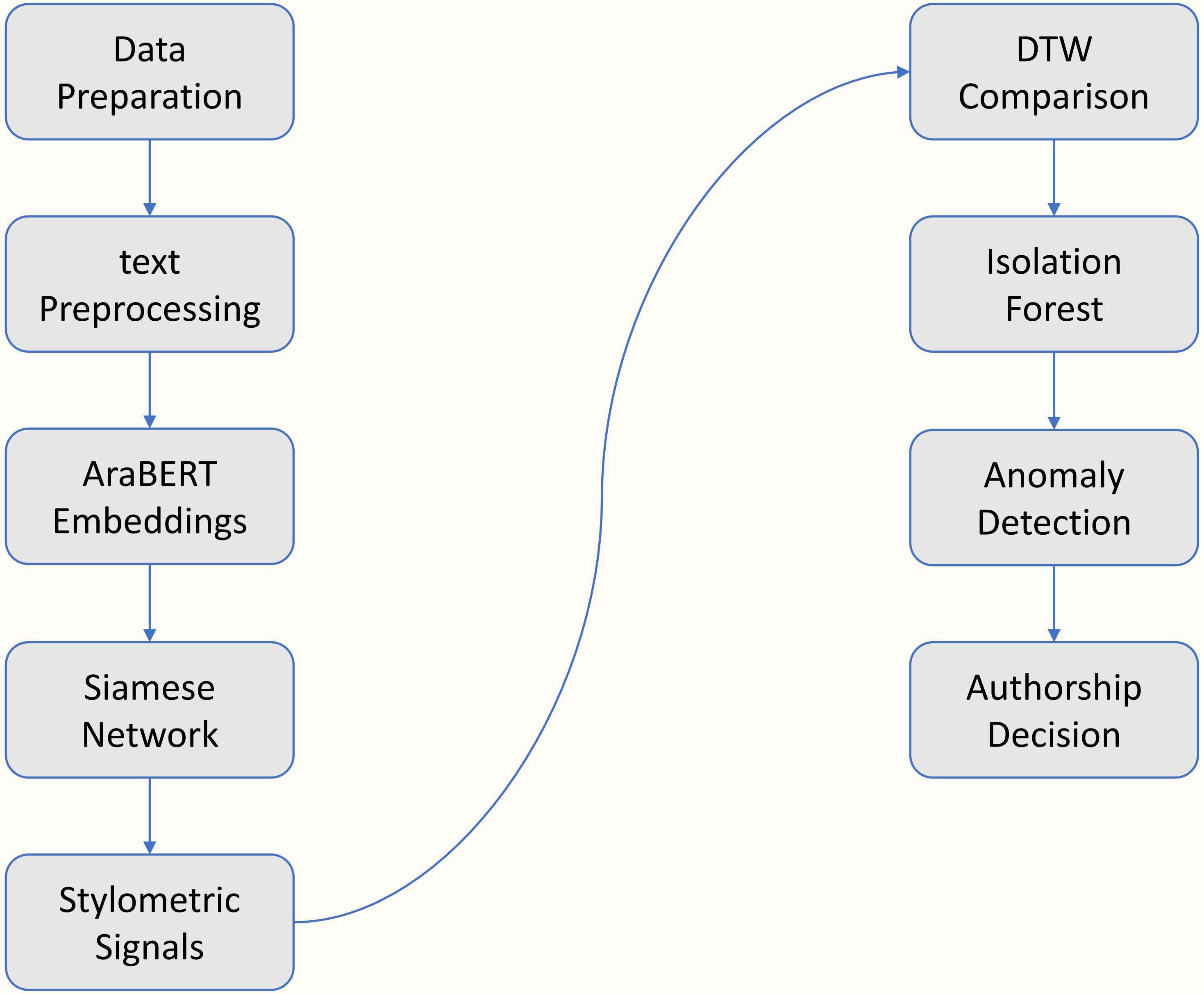
Deep Impostors Approach

- *Authorship attribution is performed through relative comparison, not direct classification.*
- *A questioned text is compared against texts created by multiple impostor authors, evaluating stylistic similarity, rather than absolute scores.*
- *Texts must be homogenized according to the topics, aiming just compare the writing styles.*

Deep Impostors Framework

1. *Divide texts into batches of L words*
2. *Select impostor pairs (Zi, Zj) from different authors*
3. *Train binary classifier to distinguish between impostors*
4. *Apply classifier to test document → binary sequence {0,1,1,...}*
5. *Aggregate results into real valued signal components*

System Architecture Overview



Arabic Text Preprocessing

Step	Process	Example Output
Step 1	Original Text	إنَّ العلمَ نورٌ، والجهلُ ظلامٌ يعمي القلوبَ
Step 2	Tokenization	["إنَّ", "العلمَ", "نورٌ", "،", "والجهلُ", "ظلامٌ", "يعمي", "القلوبَ"]
Step 3	Normalization	["ان", "العلم", "نور", "والجهل", "ظلام", "يعمي", "القلوب"]
Step 4	Optional: Stopword Filtering	["العلم", "نور", "الجهل", "ظلام", "يعمي", "القلوب"]
Step 5	Optional: Light Morphological Processing	["علم", "نور", "جهل", "ظلام", "يعمي", "قلب"]

Training Process

Training Loop Overview

- *During training, we construct pairs of impostor authors rather than using the target author directly.*
- *Each pair represents a binary classification task between two impostor styles.*

Batch Construction

- *Impostor pairs are divided into fixed size batches.*
- *Each batch contains segment level embeddings extracted using AraBERT.*

Balanced Sampling

- *We apply balanced sampling to ensure that all impostor classes are equally represented in each batch.*
- *This prevents bias toward dominant authors and stabilizes the learning process.*

Purpose

- *This training strategy prepares reliable representations for the post training stage.*
- *The learned embeddings are later used in the Siamese / Deep Impostors framework to improve stylistic discrimination and overall model accuracy.*

Embedding Representation using AraBERT

- *Arabic texts are segmented into fixed length overlapping chunks to handle model input constraints.*
- *Each chunk is encoded using AraBERT contextual embeddings, which generate representations sensitive to surrounding context.*
- *Texts' tokenization enables robust handling of rare words and Arabic morphological variation.*
- *Embeddings are extracted at the segment level, allowing to catch stylistic behavior to be analyzed.*
- *These segment level embeddings serve as the input for impostor based classification and signal construction.*

Siamese Network

- *A Siamese Network is a neural network architecture that takes two inputs and passes them through identical networks with shared weights.*
- *The outputs are compared using a similarity function, such as contrastive loss, to determine how similar or dissimilar the inputs are.*
- *Advantages of a Siamese Network: Efficiency, flexibility, and a focus on relationships, making it ideal for comparison-based tasks.*
- *The Siamese Network employs CNN and BiLSTM layers to leverage both local and global features of the text.*

The CNN layer captures local patterns and stylistic features, helping extract nuances from text embeddings through convolutional operations.

The BiLSTM layer takes these local features and processes them bidirectionally, modeling sequential dependencies and capturing long-term contextual relationships.

Stylometric Signal Construction & DTW

Stylometric Signal Construction

- *Segment level classification outputs are ordered sequentially along the document.*
- *Local stylistic decisions are aggregated to form a one dimensional stylometric signal.*
- *The signal represents how the writing style evolves across the text.*
- *This representation captures stylistic consistency and variation over document segments.*

Dynamic Time Warping (DTW)

- *DTW is used to compare stylometric signals from different texts.*
- *It aligns signals of different lengths and structures.*
- *Enables comparison despite variations in document order or writing pace.*
- *Provides a robust similarity measure between stylistic patterns.*

Outcome

- *Set of anomaly scores computed for each text resting upon pairwise DTW distances.*

Calculation of Random Forest Scores

- *After computing DTW distances between pairs of text signals, we use these distances to build a feature space.*
- *Each comparison (target vs. impostor text) is represented by its DTW based similarity profile.*
- *We feed these features into a Random Forest classifier, a robust ensemble machine learning model.*

What the Random Forest Does:

it analyzes multiple decision trees to estimate the probability that two texts were written by the same author. It handles noisy features and avoids overfitting, important for real-world, messy Arabic texts.

Why This Step Matters:

Converts raw DTW distances into a meaningful authorship similarity score. Helps bridge the gap between raw signal similarity and higher-level authorship prediction.

Summarize Clustering and Detection

Final Verification Through Dual Methods

Method 1: Isolation Forest Principle

Anomaly Score Calculation:

- *Score ≈ 1: Strong outlier (impostor text).*
- *Score ≈ 0.5: Normal pattern (authentic text).*

Method 2: K-Means Clustering (k=2)

- *Group documents by stylistic signal similarity.*
- *Validate Isolation Forest results.*
- *High confidence when both methods agree.*

Decision Logic:

- *Authentic texts: Hard to isolate, cluster together.*
- *Impostor texts: Easy to isolate, form separate cluster.*

Summary & Expected Contributions

- *Presented a research framework for authorship attribution in Arabic literary texts.*
- *Integrated contextual embeddings with impostor based stylometric analysis.*
- *Proposed a multi stage methodology combining signal construction, anomaly detection, and clustering.*
- *Addressed key challenges related to linguistic variability in Arabic.*
- *Provided a foundation for future experimental evaluation and validation.*

Evaluation Plan

Test No.	Test Subject	Expected Result
1	Validate tokenization and normalization: Test Arabic text splitting and standardization.	Accurate and consistent segmentation and formatting of text across the dataset.
2	Verify stopword removal and stemming: Ensure reduction of uninformative content while preserving key lexical features.	Text retains core meaning and stylistic elements, reducing dimensionality.
3	Assess AraBERT embeddings: Evaluate the semantic and stylistic representation of text segments.	High quality embeddings that differentiate writing styles effectively.
4	Evaluate similarity scoring with Siamese Network: Test how well the model distinguishes between similar and dissimilar text pairs.	Lower distance for same author pairs; higher distance for different authors.
5	Test robustness to noisy input: Introduce variations like typos or missing chunks.	Minimal accuracy drop; model remains stable under imperfect input.
6	Validate signal representation and DTW alignment: Examine the numeric representation of style and its distance comparison.	Signal distances reflect meaningful stylistic similarity or difference.
7	Evaluate clustering: Group texts based on similarity using Isolation Forest and k means.	Clear and interpretable clustering that aligns with author boundaries.
8	Run end to end system test: Assess the framework's full pipeline from raw input to classification.	Consistent and accurate authorship attribution across multiple text cases.

THANK YOU !