Detecting Girl Impersonators on social media Through Writing style/patterns Using Deep Learning

الكشف عن منتحلي شخصية الفتيات على وسائل التواصل الاجتماعي من خلال أسلوب/أنماط الكتابة باستخدام التعلم العميق

Written by:

E. Mohammad Omar Mahairi

mohammad_omar_234296@svuonline.org

Supervised by:

D. Bassel Alkhatib

t_balkhatib@svuonline.org

Keywords

***RNNs**: Recurrent Neural Networks, **LSTM**: Long Short-Term Memory, **Glove**: Global Vector for Word Representation, **SNSs**: Social Networking Sites.*

Introduction

In recent years, social media has become a platform where people freely express themselves, communicate, connect, and share their sentiments with others. However, these platforms might be infiltrated by impostors trying to pass off as other genders to spread disinformation, manipulate others, steal money, or pose as a security threat. Hence, gender impersonation becomes a serious social media trust issue. As a result, it will be ambitious to regulate abusive content on social media purely through user profiles and prior labels because of this potential threat.

There are various ways for an impostor to behave like a different gender in an online environment or behaviorally while deceiving others. One way an impostor can perpetrate this impersonation

behavior is by potentially using a different writing style while communicating with others, or by pretending the identity of a girl by sharing the same profile information and mimic her behavior by posting pictures and videos of her daily life as if it were her.

Purpose of the Study

The purpose of this research is to use deep learning methods and techniques for identifying and mitigating girl impersonators on social media platforms by using Recurrent Neural Network (RNNs) for feature extraction, (Glove) for word embedding and long short-term memory (LSTM) for text and sentiment analyzing. We will be applying these methods on real datasets by integrating with social media platforms to test its accuracy.

Background and Motivation

Due to its scalability and popularity, social networking sites (SNSs) have become integral parts of our daily lives, connecting millions of users worldwide. These platforms facilitate communication, information sharing, and social interactions. However, everything has pros and cons, and when we come across these platforms, the cons are more than pros. There are several challenges related to privacy, security and authenticity, Including:

Anomaly Detection in SNSs:

Anomalies, such as fake profiles, spam, and malicious activities, can harm users and the platform itself. Traditional rule-based approaches struggle to keep up with the evolving tactics of malicious actors (this is where Deep Learning comes in). Deep learning techniques offer promise due to their ability to learn complex patterns from large-scale data.

Gender Detection:

Understanding gender's orientation and interests on SNSs is crucial for personalized services, marketing, and content recommendation. Deep learning models can analyze textual content, images, and behavioral patterns to infer gender.

Typing Patterns and Fake Profiles:

Keystrokes, mouse clicks, and touch strokes exhibit unique patterns for each user. Challenges include handling noisy data, scalability, and ensuring robustness against adversarial attacks.

In this research we will try to address these challenges by using Deep Learning algorithms because of their accuracy and robustness.

Research Questions

There are couple of questions that come to mind when we deal with such big problem:

1. How can I know if a profile on Facebook is fake or suspicious?
2. How can I determine the gender of the real person (the person behind the scenes) through the writing style?
3. How can we distinguish between real published images and images taken from sites such as Pinterest or Google images?

In this research, we will try to answer these questions by showing real examples and testing the algorithms on real datasets.

Research Methodology

1. **Data Collection and Preprocessing:**

   Preprocess the datasets combined from social media (posts, comments, etc.) by:

   - Tokenizing text (splitting into words or sub word units).
   - Removing stop words, special characters, and URLs.
   - Handling misspellings and abbreviations.
   - Balancing the dataset (if needed).

2. **Feature Extraction:**
   - Utilizing Glove (Global Vectors for Word Representation) embeddings:
   - Converting each post/comment into a fixed-length vector by averaging the Glove vectors of its constituent words.

3. **Model Architecture:**

   - Using Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks
   - Designing an architecture that combines Glove embeddings with RNN/LSTM layers:

     - Input: Glove vectors for each post/comment.
     - RNN/LSTM layers to learn sequential patterns.
     - Output layer for binary classification (fake vs. genuine).

4. **Training and Evaluation:**

   - Splitting the dataset into training, validation, and test sets.
   - Training the model using backpropagation and gradient descent:
   - Optimizing hyperparameters (learning rate, batch size, etc.).

- Monitoring loss and accuracy during training.

5. **Evaluate the model on the test set:**

- Metrics: Accuracy, precision, recall, F1-score.
- Considering cross-validation for robustness.

6. **Fine-Tuning and Interpretability:**

- Experimenting with different RNN/LSTM architectures (stacked LSTMs, bidirectional LSTMs, etc.).
- Visualizing attention weights to understand which parts of the text contribute most to predictions.
- Investigating misclassified instances to identify model weaknesses.

7. **Deployment and Real-World Testing:**

- Deploying the trained model in a controlled environment (e.g., a simulated chatbot).
- Evaluating its performance on real-world SNS data:
- Monitoring false positives/negatives.
- Addressing any biases or ethical concerns.

Plan of Work

Month 1: Project Setup and Data Collection:

- Defining project goals and objectives.
- Setting up development environment (Python, TensorFlow, etc.).
- Collecting and preprocessing SNS data (posts, comments, etc.).

Months 2-3: Feature Extraction and Glove Embeddings:

- Implementing Glove-based feature extraction.
- Validating the quality of the embeddings.

Months 4-5: RNN and LSTM Model Development:

- Designing the RNN/LSTM architecture.
- Training initial models using Glove features.
- Optimizing hyperparameters (learning rate, batch size, etc.).

Months 6-7: Model Evaluation and Fine-Tuning:

- Evaluating model performance (accuracy, precision, recall).
- Experimenting with different RNN/LSTM configurations.

- Visualizing attention weights and interpreting model decisions.

Months 8-9: Ethical Considerations and Bias Analysis:

- Addressing privacy concerns related to gender profiling.
- Investigating potential biases in the model.
- Refining the model to minimize biases.

Months 10-11: Deployment and Real-World Testing:

- Deploying the model in a controlled environment (e.g., chatbot).
- Collecting real-world SNS data for testing.
- Monitoring false positives/negatives.

Month 12: Finalize Research, Write Thesis, and Present Findings:

- Summarizing findings.
- Writing thesis.
- Preparing for presentations.

Bibliography

1. "Fake profile recognition using profanity and gender identification on online social networks".
2. "Ensemble fake profile detection using machine learning (ML)".
3. "The Doppelgänger Bot Attack: Exploring Identity Impersonation in Online Social Networks".
4. “How to detect and report a fake social media account”.
5. “Social media impersonation: What is it? How to stop it”.
6. “Social Media Impersonation: How To Identify, Prevent, & Respond”.