

Authorship Attribution for Social Media Forensics

ABSTRACT:

The veil of anonymity provided by smartphones with pre-paid SIM cards, public Wi-Fi hotspots, and distributed networks like Tor has drastically complicated the task of identifying users of social media during forensic investigations. In some cases, the text of a single posted message will be the only clue to an author's identity. How can we accurately predict who that author might be when the message may never exceed 140 characters on a service like Twitter? For the past 50 years, linguists, computer scientists and scholars of the humanities have been jointly developing automated methods to identify authors based on the style of their writing. All authors possess peculiarities of habit that influence the form and content of their written works. These characteristics can often be quantified and measured using machine learning algorithms. In this article, we provide a comprehensive review of the methods of authorship attribution that can be applied to the problem of social media forensics.

INTRODUCTION:

Forensic authorship attribution is the process of inferring something about the characteristics of an author from the form and content of their writing present in a collection of evidence. The emergence of social media as a primary mode of communication has challenged the traditional assumption that a forensic investigation will have access to long-form writing (i.e., letters and emails). In this

Technofist,

YES Complex, 19/3&4, 2nd Floor, Dinnur Main Road, R.T.Nagar, Bangalore-560032

Ph:080-40969981, Website:www.technofist.com. E-mail:technofist.projects@gmail.com

article, we frame the problem as a computational pipeline, in which features are extracted from very small samples of text, and scalable supervised learning is deployed to train author-specific models and make predictions about unknown samples.

TECHNOFIST