

Can Anonymous Posters on Medical Forums be Reidentified?

Bobicev Victoria, Sokolova Marina, El Emam Khaled, Jafer Yasser, Dewar Brian, Jonker Elizabeth, Matwin Stan

<https://doi.org/10.2196/jmir.2514>

Abstract

Background: Participants in medical forums often reveal personal health information about themselves in their online postings. To feel comfortable revealing sensitive personal health information, some participants may hide their identity by posting anonymously. They can do this by using fake identities, nicknames, or pseudonyms that cannot readily be traced back to them. However, individual writing styles have unique features and it may be possible to determine the true identity of an anonymous user through author attribution analysis. Although there has been previous work on the authorship attribution problem, there has been a dearth of research on automated authorship attribution on medical forums. The focus of the paper is to demonstrate that character-based author attribution works better than word-based methods in medical forums. **Objective:** The goal was to build a system that accurately attributes authorship of messages posted on medical forums. The Authorship Attributor system uses text analysis techniques to crawl medical forums and automatically correlate messages written by the same authors. Authorship Attributor processes unstructured texts regardless of the document type, context, and content. **Methods:** The messages were labeled by nicknames of the forum participants. We evaluated the system's performance through its accuracy on 6000 messages gathered from 2 medical forums on an in vitro fertilization (IVF) support website. **Results:** Given 2 lists of candidate authors (30 and 50 candidates, respectively), we obtained an F score accuracy in detecting authors of 75% to 80% on messages containing 100 to 150 words on average, and 97.9% on longer messages containing at least 300 words. **Conclusions:** Authorship can be successfully detected in short free-form messages posted on medical forums. This raises a concern about the meaningfulness of anonymous posting on such medical forums. Authorship attribution tools can be used to warn consumers wishing to post anonymously about the likelihood of their identity being determined.