Skysong Innovations
Advancing the Arizona State University Knowledge Enterprise

1475 N. Scottsdale Road, Suite 200
Scottsdale, AZ 85287-3538
Phone: 480 884 1996 Fax: 480 884 1984

Case ID:M21-053P^
Published: 9/8/2021

Inventors

Ahmadreza Mosallanezhad

Ghazaleh Beigi

Huan Liu

Contact

Shen Yan
shen.yan@skysonginnovations.com

Deep Reinforcement-Learning-Based Text Anonymization for Attribute Privacy

Background Social media users generate a tremendous amount of data in forms such as profile information, network connections, and online posts. Online vendors use this data to understand user preferences and further predict their future needs. However, because user-generated data is rich in content, the data can be used by malicious attackers to infer users' sensitive information. Recent research has shown that textual data alone may contain sufficient information about users' private attributes that they do not intend to disclose such as age, gender, location, political views, and sexual orientation.

Anonymizing textual information comes at the cost of utility loss for future applications, and hence, a privacy-utility balance must be established. Recent success of reinforcement learning (RL) highlights an alternative methodology in which the feedback of attackers and utility can be included in a reward function, enabling control of the privacy-utility balance. Furthermore, an RL software agent can perturb parts of an embedded text for preserving both utility and privacy, instead of retraining an embedding as is common in adversarial learning. Invention Description Researchers at Arizona State University have developed a novel Reinforcement Learning-based Text Anonymizer, namely, RLTA, composed of two main components: 1) an attention-based task-aware text representation learner that extracts latent embedding representation of the original text's content for a given task, and 2) a deep reinforcement-learning-based privacy and utility preserver that converts the problem of text anonymization to a one-player game in which the agent's goal is to learn the optimal strategy for text embedding manipulation to satisfy both privacy and utility. The Deep Q Learning algorithm is then used to train the agent capable of changing the text embedding with respect to the received feedback from the privacy and utility subcomponents. Potential Applications • Preserving privacy of sensitive user information in textual data • Social media Related Publication: Deep Reinforcement Learning-based Text Anonymization against Private-Attribute InferenceFaculty Profile of Professor Huan Liu