

Safeguarding Patients Data for Reasonable Health Care Cost

Srikanth Gampa and Rajiv Bagai

Department of Electrical Engineering and Computer Science, Wichita State University

Health care providers including hospitals are required by law to publicly release their patient data, removing explicitly identifiable attributes, such as name, social security number, and address. The release of this data helps research laboratories, and federal and state governments (the state of Kansas in this case) in tasks, such as analyzing geographical movement of diseases, disease eradication, and drug discovery. For a long time, organizations believed the data to be adequately protecting patient privacy as long as all explicitly identifying attributes were removed from it. Motivated insurance companies could perform analyses to decipher people's medical history and raise premiums of those with sensitive medical history, thereby raising the overall health care cost of the society. Several sophisticated data anonymization concepts have since been proposed by the research community, of which t-closeness is a leading one. The currently available t-closeness algorithm is capable of handling only one sensitive attribute, such as a patient's diagnosed disease. We extend the state-of-the-art by building an algorithm that is capable of achieving t-closeness in the presence of multiple sensitive attributes. Our algorithm achieves significant user anonymity and, if adopted to anonymize the data before releasing, will contribute towards lowering the overall health care cost of the society.