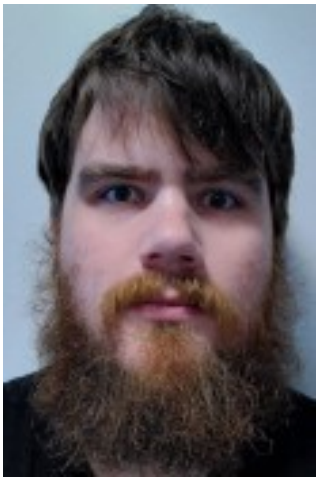
Entwicklung eines Frameworks zur Anonymisierung von personenbezogenen zeitkontinuierlichen Daten

(Masterarbeit)



FREIMUT HAMMER

Motivation

Um bei medizinischen Studien den wissenschaftlichen Standards zu entsprechen ist es notwendig, Patientendaten zu veröffentlichen. Hierfür werden diese anonymisiert, um die Patientenrechte zu wahren. Viele Parameter der Studiendaten sind zeitkontinuierliche Signale, z. B. EKG Verläufe. Für diese Art von Daten wird noch eine Methode zur Anonymisierung benötigt.

Stand der Technik

Für diskrete oder gar kategoriale Daten bieten sich k-anonymity und t-closeness zur Sicherstellung der Anonymität an. K-Anonymity ist ein generalisierender Anonymisierungsalgorithmus, bei dem mindestens k verschiedene Quasi-Identifier Ausprägungen zusammen gruppiert werden. T-closeness erweitert die Anonymität um die Forderung nach einem Höchstabstand der Verteilung der sensitiven Attribute in der Äquivalenzklasse von der Verteilung in der Population. Mit Hilfe von CASTLE (**C**ontinuously **A**nonymizing **ST**reaming data via adaptive c**L**ust**E**ring)[1] und SABRE (a **S**ensitive **A**ttribute **B**ucketization and **RE**distribution framework for t-closeness)[2] können diese Konzepte auf diskrete Datenströme angewendet werden.

Zielsetzung

Das Ziel dieser Masterarbeit ist es, ein Werkzeug zur Anonymisierung einer Vielzahl von Daten zu entwickeln. Insbesondere sollen dabei medizinische Daten, wie z. B. EKG anonymisiert werden können, da diese zeitkontinuierlich sind und meist eine hohe Datendichte haben. Der Schwerpunkt liegt hierbei auf der Erweiterung der k-anonymity für zeitkontinuierliche Daten. Die Qualität der anonymisierten Daten soll durch verschiedene Metriken beurteilt werden können. Zusätzlich soll die Anonymisierung um t-closeness erweitert werden können, wenn die Verteilung in der Population gegeben ist.

Geplante Vorgehensweise

Für die Umsetzung müssen geeignete Metriken für Abstand, Cluster-Qualität und Informationsverlust definiert werden. Außerdem müssen die Daten anhand geeigneter Heuristiken in Abschnitte geteilt werden, welche dann unabhängig voneinander verarbeitet werden.

Zuerst soll eine Möglichkeit zur Anonymisierung einzelner Abschnitte mittels k-anonymity geschaffen werden. Diese wird dann schrittweise um Funktionen wie das Aufteilen der Daten in Abschnitte, t-closeness und verschiedene Parameter zur Reduktion und Quantifizierung des Informationsverlustes erweitert.

Das Framework soll über eine geeignete grafische Benutzeroberfläche mit entsprechender Visualisierung der Daten bedienbar sein.

1 J. Cao et al. “Castle: Continuously anonymizing data streams” 2010
2 J. Cao et al. “SABRE: a Sensitive Attribute Bucketization and REdistribution framework for t-closeness” 2011

Betreuer: Mateusz Buglowski
12.11.2021

Informatik 11
Embedded Software
RWTH AACHEN UNIVERSITY