CONFERENCE OF EUROPEAN STATISTICIANS

Expert Meeting on Statistical Data Confidentiality

1-3 December 2021, Poland

Some of entity resolution and the impacts on personal privacy.

Rebecca C. Steorts (Duke University)
beka@stat.duke.edu

Abstract

Whether the goal is to estimate the number of people that live in a congressional district, to estimate the number of individuals that have died in an armed conflict, or to disambiguate individual authors using bibliographic data, all these applications have a common theme—integrating information from multiple sources. Before such questions can be answered, databases must be cleaned and integrated in a systematic and accurate way, commonly known as structured entity resolution (record linkage or de-duplication). In this article, we review motivational applications and seminal papers that have led to the growth of this area. We review modern probabilistic and Bayesian methods in statistics, computer science, machine learning, database management, economics, political science, and other disciplines that used throughout industry and academia in applications such as human rights, official statistics, medicine, citation networks, among others. Finally, we discuss current research topics of practical importance, such as implications to privacy.