Título: Local dampening: Differential privacy for non-numeric queries via local sensitivity

Data: 25/05/2021

Horário: 13h

Local: Videoconferência

Resumo:

Differential privacy is the state-of-the-art formal definition for data release under strong privacy guarantees. A variety of mechanisms has been proposed in the literature for releasing the output of numeric queries (e.g., the Laplace mechanism and smooth sensitivity mechanism). Those mechanisms guarantee different privacy by adding noise to the true query's output. The amount of noise added is calibrated by the notions of global sensitivity and local sensitivity of the query that measure the impact of the addition or removal of an individual on the query's

output. Mechanisms that use local sensitivity add less noise and, consequently, have a more accurate answer. However, although there has been some work on generic mechanisms for releasing the output of non-numeric queries using global sensitivity (e.g., the Exponential mechanism), the literature lacks generic mechanisms for releasing the output of non-numeric queries using local sensitivity to reduce the noise in the query's output. In this work, we remedy this shortcoming and present the local dampening mechanism. We adapt the notion of local sensitivity for the non-numeric setting and leverage it to design a generic non-numeric mechanism. We provide theoretical comparisons to the exponential mechanism and show under which conditions the local dampening mechanism is more accurate than the exponential mechanism. We illustrate the effectiveness of the local dampening mechanism by applying it to three diverse problems: (i) median selection. We report the median element in the database; (ii) Influential node analysis. Given an influence metric, we release the top-k most influential nodes while preserving the privacy of the relationship between nodes in the network; (iii) Decision tree induction. We provide a private adaptation to the ID3 algorithm to build decision trees from a given tabular dataset. Experimental evaluation shows that we can reduce the error for median selection application up to 18, reduce the use of privacy budget by 2 to 4 orders of magnitude for influential node analysis application and increase accuracy up to 12% for decision tree induction when compared to global sensitivity based approaches. We illustrate the effectiveness of the local dampening mechanism by applying it to three diverse problems: (i) median selection. We report the median element in the database; (ii) Influential node analysis. Given an influence metric, we release the top-k most influential nodes while preserving the privacy of the relationship between nodes in the network; (iii) Decision tree induction. We provide a private adaptation to the ID3 algorithm to build decision trees from a given tabular dataset. Experimental evaluation shows that we can reduce the error for median selection application up to 18, reduce the use of privacy budget by 2 to 4 orders of magnitude for influential node analysis application and increase accuracy up to 12% for decision tree induction when compared to global sensitivity based approaches.

Banca examinadora:

- Javam de Castro Machado (Orientador)
- Divesh Srivastava (AT&T Labs Research - EUA)
- Agma Juci Machado Traina (USP-São Carlos)
- Altigran Soares da Silva (UFAM)
- José Soares Andrade Junior (UFC)
- João Paulo Pordeus Gomes (MDCC/UFC)