Efficient and Practical Approach for Private Record Linkage

Reviewed by Liyue Fan

Summary:
This paper proposes a privacy-preserving record linkage protocol without the need of a trusted third party. The matched vectors are within $\delta$ Euclidean distance. The first phase transforms the original/private vectors into complex space and performs a blocking step to generate likely-linked pairs. The second phase uses a secure method to compute the Euclidean distance between vectors to produce the final matched pairs. As they have shown, the first phase requires $O(t + m\log m + n\log n)$ operations where $m$ and $n$ are input size and $t$ is the number of likely-linked pairs. The second phase is shown to be secure with very little information learned by data holders.

Strength:
1. Two-Party protocol without cryptography.
2. Proposed an efficient linear search blocking scheme.
3. Proposed a private scheme to compute the distance between vectors.

Weakness:
1. Assume that original records are transformed into a vector space using (Monica Scannapieco, 2007). It is not a practical assumption and limits the application of this work.
2. The blocking step generates a lot false positives. The best precision can be achieved is around 20%.

Discussion:
1. Is the first DFT coefficient a good representation of original vector, in terms of utility and security?
2. It might be interesting to see how to potentially breach the privacy or reconstruct vectors from the communication in the second phase, which the authors sidestepped.
Works Cited