Διαφορική Ιδιωτικότητα (Differential Privacy)

Κυριάκος Κρητικός

Τμήμα Μηχανικών Πληροφοριακών & Επικοινωνιακών Συστημάτων

Ορισμός

- Αποτελεί σύστημα για τον δημόσιο διαμοιρασμό πληροφορίας για ένα σύνολο δεδομένων περιγράφοντας μοτίβα ομάδων στο σύνολο δεδομένων χωρίς να αποκαλύπτει πληροφορία ατόμων σε αυτό
  - Αν το αποτέλεσμα την εκτέλεσης μιας μικρής αντικατάστασης σε μια βάση δεδομένων είναι αρκετά μικρό, τα αποτελέσματα επερωτήσεων δεν μπορούν να χρησιμοποιηθούν για την εξαγωγή πληροφορίας για ένα άτομο – αυτό επομένως παρέχει ιδιωτικότητα
- Άλλος ορισμός: περιορισμός στους αλγορίθμους δημοσίευσης συντιθέμενων/συνολικών πληροφοριών για μια στατιστική βάση δεδομένων που περιορίζει την αποκάλυψη της ιδιωτικής πληροφορίας των εγγραφών
  - Ένας αλγόριθμος θεωρείται διαφορικά ιδιωτικός εφόσον ένας παρατηρητής που βλέπει το αποτέλεσμά του δεν μπορεί να διακρίνει αν η πληροφορία για ένα άτομο χρησιμοποιήθηκε στον υπολογισμό
  - Οι αλγόριθμοι αυτοί είναι ικανοί να αντιστέκονται σε επιθέσεις προσδιορισμού & επαναπροσδιορισμού (identification & re-identification/linkage)

ε-Διαφορική Ιδιωτικότητα

- Αποτελεί έναν μαθηματικό ορισμό της απώλειας ιδιωτικότητας που σχετίζεται με οποιαδήποτε έκδοση δεδομένων από μια στατιστική βάση δεδομένων
  - Προφανώς, η ιδιωτικότητα ενός ατόμου δεν παραβιάζεται από μια στατιστική έκδοση δεδομένων αν τα δεδομένα του ατόμου αυτού δεν υπάρχουν στην βάση
  - Επομένως, σκοπός της διαφορικής ιδιωτικότητας είναι να δοθεί σε κάθε άτομο το ίδιο επίπεδο ιδιωτικότητας που θα ήταν το αποτέλεσμα της διαγραφής των δεδομένων του (από την βάση)
    - Με άλλα λόγια, οι στατιστικές συναρτήσεις που εκτελούνται πάνω από την βάση δεν θα πρέπει να εξαρτώνται σε μεγάλο βαθμό από τα δεδομένα ενός ατόμου

ε-Διαφορική Ιδιωτικότητα

- Συνεπώς, το πόσο ένα άτομο συνεισφέρει στο αποτέλεσμα μιας επερώτησης εξαρτάται εν μέρει από το αν τα δεδομένα του ατόμου εμπλέκονται στην επερώτηση
  - Αν η βάση περιέχει τα δεδομένα από ένα άτομο, τα δεδομένα του συνεισφέρουν κατά 100%
  - Αν η βάση περιέχει δεδομένα από 100 άτομα, τότε τα δεδομένα κάθε ατόμου συνεισφέρουν κατά 1%
  - Άρα, αν μια επερώτηση γίνεται στα δεδομένα ολοένα λιγότερων ατόμων, περισσότερος θόρυβος θα πρέπει να προστεθεί στο αποτέλεσμα της επερώτησης ώστε να παραχθεί το ίδιο επίπεδο ιδιωτικότητας

ε-Διαφορική Ιδιωτικότητα

- Ορισμός
  - Έστω πως *ε* είναι ένας θετικός πραγματικός αριθμός και *A* ένας τυχαιοποιημένος αλγόριθμος που λαμβάνει ένα σύνολο δεδομένων ως είσοδο. Έστω *im A* να προσδιορίζει την εικόνα του Α. Ο αλγόριθμος *A* είναι ε-διαφορικά ιδιωτικός αν για κάθε ζεύγος συνόλων δεδομένων *D1* & *D2* που διαφέρουν σε ένα μόνο στοιχείο (πχ. τα δεδομένα ενός ατόμου) και για όλα τα υποσύνολα S της εικόνας του *A* ισχύει το εξής:

$$\Pr[A(D_1) \in S] \leq \exp(\varepsilon) \cdot \Pr[A(D_2) \in S]$$

Ευαισθησία

- Έστω πως *d* είναι ένας θετικός ακέραιος, *D* μια συλλογή από σύνολα δεδομένων και *f*: $D \rightarrow R^d$ είναι μια συνάρτηση. Η ευαισθησία της συνάρτησης *f*, *Δf* ορίζεται ως: $\Delta f = \max||f(D1)-f(D2)||_1$, όπου το μέγιστο υπολογίζεται για όλα τα ζευγάρια *D1* & *D2* που ανήκουν στην *D* και διαφέρουν ως προς ένα μόνο στοιχείο ενώ ||·||1 σημαίνει την l_1 νόρμα.
- Το ζητούμενο είναι να έχουμε συναρτήσεις/αλγορίθμους διαφορικής ιδιωτικότητας που να έχουν χαμηλή ευαισθησία

Μηχανισμός Laplace

- Προσθέτει Laplace θόρυβο στο αποτέλεσμα μιας επερώτησης
- Έστω πως η συνάρτηση εξόδου της *A* είναι μια συνάρτηση πραγματικών τιμών, *f* είναι η αρχική συνάρτηση που εφαρμόζεται στα δεδομένα και $T_A(x) = f(x) + Y$ όπου Y ~ Lap(λ). Η $T_A(x)$ μπορεί να θεωρηθεί ως μια συνεχής τυχαία μεταβλητή για την οποία ισχύει ότι:
  - pdf($T_{A,\ D1}(x) = t$) / pdf($T_{A,\ D2}(x) = t$) = noise($t-f(D1)$ / noise($t-f(D2)$)
  - Και άρα: $e^{|f(D1)-f(D2)|\ /\ \lambda} <= e^{\Delta(f)/\lambda}$
  - Ο τελευταίος εκθέτης Δ(f)/λ θεωρείται ο παράγοντας ιδιωτικότητας ε
- Σημειώνεται πως είναι δυνατόν η χρήση διαφορετικών ειδών θορύβου (πχ. Gaussian) αλλά μπορεί να χρειαστούν έναν πιο χαλαρό ορισμό της διαφορικής ιδιωτικότητας

Τυχαίες Απαντήσεις

- Η απάντηση σε μια απλή ερώτηση (Ναι ή Όχι) ακολουθεί την εξής διαδικασία:
  - Ρίψη νομίσματος
    - Αν έχουμε κορώνα, τότε ξανά ρίψη του νομίσματος και επιστροφή ειλικρινούς απάντησης
    - Αν έχουμε γράμματα, τότε ξανά ρίψη του νομίσματος
      - Αν έχουμε κορώνα, τότε δώσε μια πιθανή απάντηση (πχ. Ναι)
      - Διαφορετικά, δώσε μια άλλη πιθανή απάντηση (πχ. Όχι)
      - Δεν είναι σίγουρο πια από τις απαντήσεις είναι η σωστή - ειλικρινής
- Μια επερώτηση μπορεί να θεωρηθεί ως μια σύνθεση απλών ερωτήσεων
  - Οπότε ο μηχανισμός αυτός μπορεί να επεκταθεί και σε ολόκληρες επερωτήσεις
- Η εμπιστευτικότητα εξασφαλίζεται από τον ανασκευασμό (refutability) των απαντήσεων

Σταθερές Μετατροπές

- Μετατροπές που μπορούν να εφαρμοστούν σε ένα σύνολο δεδομένων που επιτρέπουν τον υπολογισμό διαφορικής ιδιωτικότητας (differential privacy computation)
- Μια μετατροπή *T* είναι *c*-σταθερή αν η απόσταση Hamming μεταξύ *T*(*A*) & *T*(*B*) είναι το πολύ *c*-φορές την απόσταση Hamming μεταξύ των *A* & *B* για οποιεσδήποτε βάσεις δεδομένων *A*, *B*