差分プライバシのプライバシ費用による一般化k-匿名化の評価手法の提案

明治大学　総合数理学部

堀込光

研究背景

- **一般化k-匿名化は代表的な匿名加工手法の一つである.**
  →PWSCUP2018[1]では一般化手法が採用

- **k-匿名化におけるkなどの明確な規定はなく加工をどこまで不明確である.**

[1]濱田 浩気，他，“PWSCUP2018:匿名加工再識別コ ンテストの設計 履歴データの一般化・再識別 ”, コ ンピュータセキュリティシンポジウム (CSS2018)， pp.935-940，2018.

実験目的

- **差分プライバシの観点から一般化k-匿名化の安全性を評価する手法の提案する**

差分プライバシ

- **データのYesと回答した人数と，データからある個人を削除したデータのYesと回答した人数に差がない**

→ある個人がいてもいなくても人数に差がない，その個人のプライバシは保証される

Twitterやってますか？

江口	Yes
梶間	No
草野	Yes
柴山	Yes
平山	No
堀込	No
山﨑	Yes
高松	Yes
合計	5

江口	Yes
梶間	No
草野	Yes
柴山	Yes
平山	No
堀込	No
山﨑	Yes
合計	4

例えば...

- 確率pで嘘の回答をする

江口	Yes
梶間	No
草野	Yes
柴山	Yes
平山	No
堀込	No
山﨑	Yes
高松	Yes
合計	5

江口	Yes
梶間	No
草野	Yes
柴山	Yes
平山	No
堀込	No
山﨑	Yes
合計	4

江口	No
梶間	No
草野	No
柴山	Yes
平山	No
堀込	Yes
山﨑	Yes
高松	Yes
合計	4

江口	Yes
梶間	Yes
草野	Yes
柴山	No
平山	No
堀込	No
山﨑	Yes
合計	4

k-匿名

- **データベースDの任意のレコード$t \in D$に対して，同一の順識別子の値の組がk以上存在しているとき，データベースDはk-匿名であるという．**

年齢	職業	年収
26歳	サラリーマン	≤50K
31歳	サラリーマン	>50K
45歳	地方公務員	≤50K
55歳	国家公務員	≤50K
19歳	サラリーマン	≤50K

2-匿名化

年齢	職業	学歴
10～30代	サラリーマン	≤50K
10～30代	サラリーマン	>50K
40～50代	公務員	≤50K
40～50代	公務員	≤50K
10～30代	サラリーマン	≤50K

提案手法：サンプリング後に一般化k-匿名化

- **ランダムサンプリングを行ったデータベースにk-匿名化を行うことでプライバシ費用を計算する.**
  - **確率的な操作を行うため**

D

年齢	職業	年収
10代	サラリーマン	≤50K
20代	サラリーマン	>50K
20代	地方公務員	≤50K
40代	サラリーマン	>50K
60代	地方公務員	≤50K
50代	国家公務員	≤50K
20代	地方公務員	≤50K
20代	サラリーマン	≤50K

サンプリング →

年齢	職業	年収
10代	サラリーマン	≤50K
20代	サラリーマン	>50K
60代	地方公務員	≤50K
50代	国家公務員	≤50K
20代	サラリーマン	≤50K

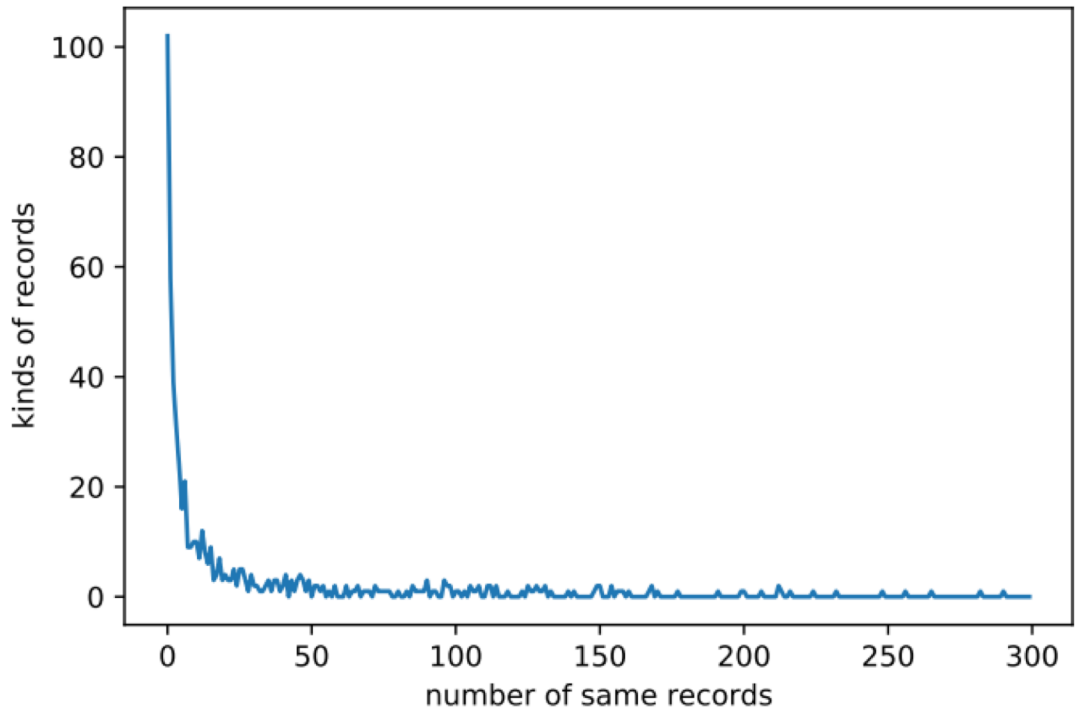
k-匿名化 →

S

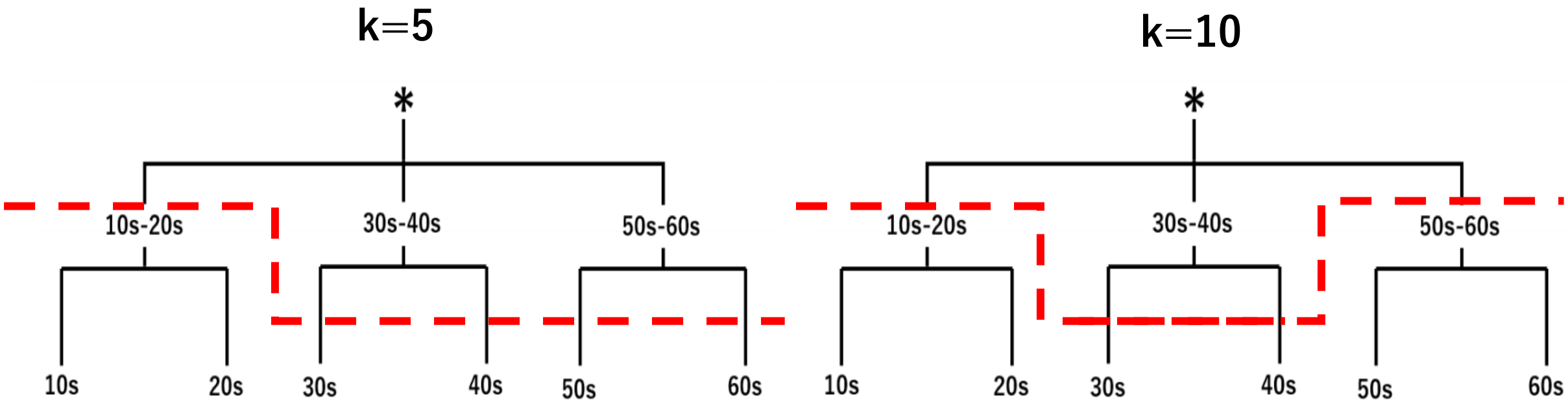
年齢	職業	年収
10～20代	サラリーマン	≤50K
10～20代	サラリーマン	>50K
50～60代	公務員	≤50K
50～60代	公務員	≤50K
10～20代	サラリーマン	≤50K

実験：使用データ(Adult Data set)

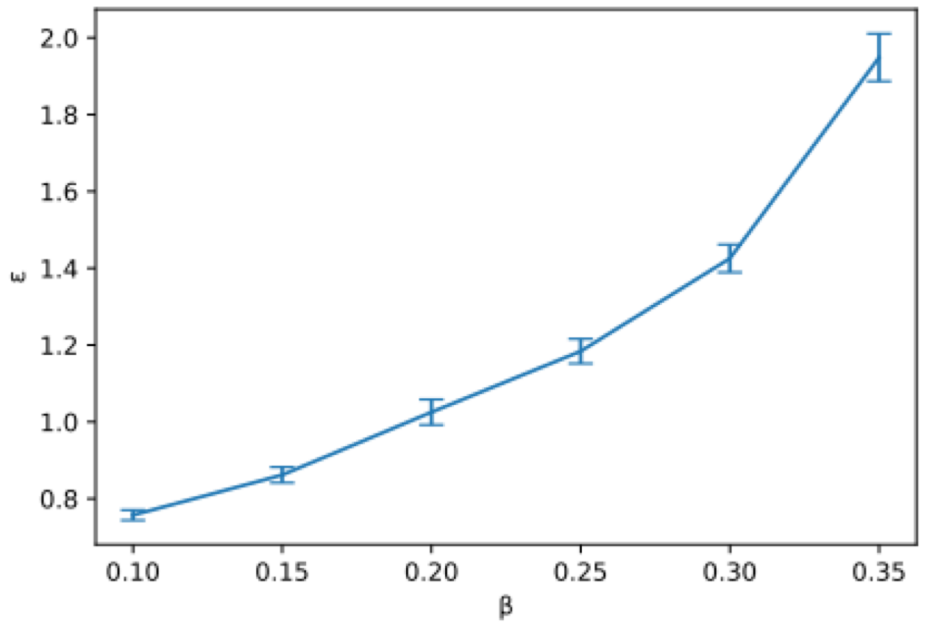
- 欠損値を削除した45,233レコードで実験を行う
- 属性：Age(年齢)，Workclass(職業)，Education(学歴)を使用する
- Age属性に関しては，一の位を切り捨てた値を離散値として扱う．



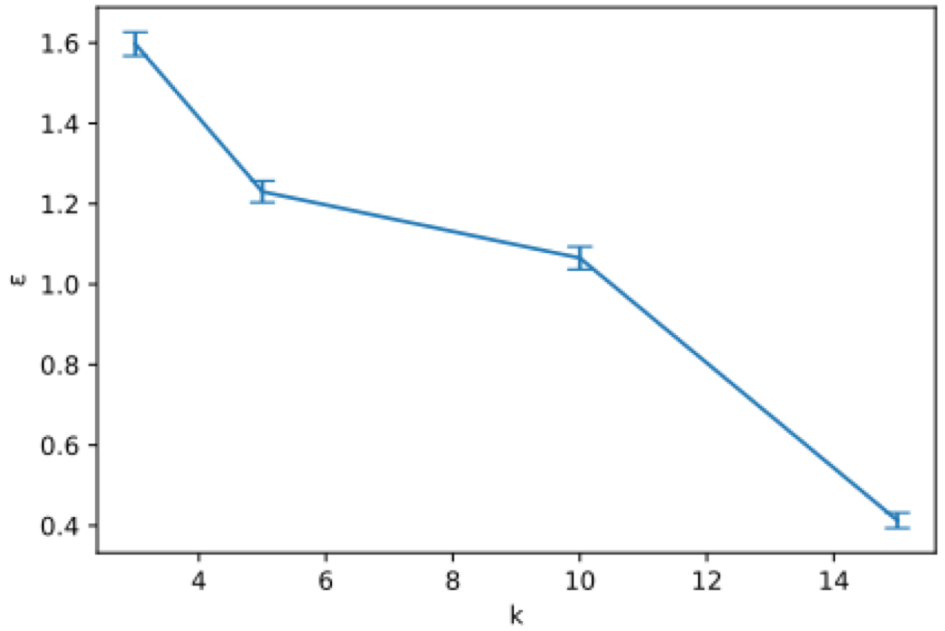
実験結果(1)
匿名化指標kによる一般木の変化



実験結果（2）



k=10に固定した際のサンプリング率βによるεの変化



サンプリング率$\beta = 0.2$に固定した際のkによるεの変化

考察

Twitterをやっているかの例では,

$$p = \frac{1}{1 + e^{\varepsilon}}$$

で嘘の回答をすると差分プライバシを満たす.

この時, $\beta = 0.2$, $k = 10$の際, $\varepsilon = 1.05$を上記の式に代入すると,

$p = 0.26$となる.

26%の割合で嘘の回答をするデータと同様の安全性と言える.

まとめと課題

- k-匿名化の安全性を差分プライバシの観点から評価する手法を提案した.

- 提案手法を数式化すること
- また，本実験で得られたデータをもとに他の手法との比較として，有用性の調査を行っていきたい.