

1F-03

秘匿暗号処理の一考察

和田 平司 三角田 秀実

A Consideration of Anonymization and Encryption

Heiji Wada Hidemi Misumida

(所属なし)

キーワード 匿名加工 暗号処理

あらまし あるデータ表または表を匿名化する際に色々な加工法がある。データ表や表を暗号化するときも同様である。

我々は、独自に秘匿化の前処理に一誤り訂正符号を用いることで秘匿化を行うことができたので報告する。

1. はじめに

パーソナルデータ利活用の為の匿名化技術には削除法や仮名化、k-匿名化、l-多様化、差分プライバシー化などがあげられる。

そこで、我々は匿名化を行う場合、ZIP コードや ID、会員 NO.のコードに一誤り訂正符号化法を用いることで匿名化の前処理を図ることができた。ZIP コードのある桁に一誤りを発生させて、k=2 匿名化を k=4 匿名化に簡単に行うことができた。この前処理は ZIP コード又は年齢の数値に一誤りを発生させることにより、k-匿名化を増している。レコード特定に効果があり、属性特定にはそれほど効果はないと考えられる。

2. 本論

2-1) 一般にパーソナルデータは氏名の替りに ZIP コードや ID、会員 NO.を用いることが多い。表 1 にオリジナルのパーソナルデータ表を示す。

ZIP コードの数はバイナリー値 5bits を 1 桁とする。

然も k=1 の匿名化である。よって、ZIP コード全体で 25bits のバイナリー値で、4 桁として、補注の項目を一誤り訂正符号の検査点数部として扱い、

そこに記入する(書き込みマスクする、*の箇所へ)

とすると、ZIP コードをハミング符号の (25,20) 符号に割り当てることになる。

ZIPコード	年 齢	補 注	国 籍	病 名
1305	28	*	カナダ	心 臓
1305	21	*	イギリス	感 染
1315	11	*	日 本	風 邪
1315	16	*	中 国	風 邪
1325	11	*	シソガボール	ガ ン
1325	15	*	日 本	ガ ン
1355	22	*	ロシア	感 染
1355	28	*	アメリカ	心 臓

表 1 パーソナルデータ (加工なしの場合)

ZIP コードを符号長としてデータの 1bit 割付けには色々あるが、本研究の場合は 5bits で 1 桁を表す値とする。

2-2) 表 1 のパーソナルデータ (加工なし) の ZIP コードの bit 列に一誤りを発生させ、すなわち、4 桁の ZIP コードをハミング符号の (n,k) 符号と考える。しかし、一誤りは必ずしもすべての ZIP コードのデータに発生させなくても良い。

2-3) 例えば、表 1 の ZIP コードの 2 桁目に 1bit の誤りを故意に発生させる。

この場合、2 桁目の bit 列のどの位置に誤りを発生させるかは 2 通りの方法がある。

- 1) 故意に誤り位置を指定する。
- 2) ランダムに誤り位置を指定する。

表 2 の場合、1) の故意に誤り位置を指定した場合の値を示す。

ZIP コード	年 齢	補 注	国 籍	病 名
1345	28	*	カナダ	心 臓
1345	21	*	イギリス	感 染
1305	11	*	日 本	風 邪
1305	16	*	中 国	風 邪
1305	11	*	シンガポール	ガ ン
1305	15	*	日 本	ガ ン
1345	22	*	ロシア	感 染
1345	28	*	アメリカ	心 臓

表 2 ZIP コードの 2 桁目に一誤りが

発生した場合

2-4) 表 2 の ZIP コードを順番に並び替え、年齢を一般化し、11-20 と 21-30 とする匿名化ができる。

表 3 は $l=2$ 多様化の場合の秘匿化を行ったものである。

$k=4$ 匿名化 $l=2$ 多様性

ZIP コード	年 齢	補 注	国 籍	病 名
1305	11-20	*	アジア	ガ ン
1305	11-20	*	アジア	ガ ン
1305	11-20	*	アジア	風 邪
1305	11-20	*	アジア	風 邪
1345	21-30	*	*	感 染
1345	21-30	*	*	心 臓
1345	21-30	*	*	心 臓
1345	21-30	*	*	感 染

表 3 匿名化を行った場合

3. 考察

3-1) パーソナルデータの前処理を行うと、 $k=2$ 匿名化を $k=4$ 匿名化にすることができる。

3-2) ZIP コードは誤りの変換が行われているので特定することが難しくなる。

3-3) n 誤り訂正符号を用いると、色々応用が広がる。ZIP コード及び年齢を含めて、加工が可能になる。

4. 結論

我々は秘匿暗号処理の前処理として、一誤り訂正符号（ハミング符号）を用いて、ZIP コード等に故意に発生させる。ことによって、秘匿化を増すことができた。 $k=2$ 匿名化を $k=4$ 匿名化にできた。

今回我々は秘匿暗号処理の加工を行うのに一誤りを卓上で故意に反転させたものを用いて ZIP コードの秘匿化を行った。

その結果を表 3 に示すことができた。残された課題としてプログラムにてアルゴリズムを導いて一誤りを ZIP コードに発生させることである。

参考文献

(1) "一誤り訂正符号の提案"

和田、愛谷、丸山、松本

電子通信学会論文誌 81/4Vol.J64 NO.4
P.338～P.339

(2) "個人データの活用とプライバシー保護"

電子情報通信学会誌 Vol.98 NO.3
P.178～P.217

(3) "匿名加工とプライバシーの保護"

情報処理学会誌 Vol.59 NO.5
P.438～P.460