KNOM Review '16-01 Vol.19 No.01
https://doi.org/10.22670/knom.2016.19.1.22

엔트로피와 클러스터링을 이용한 익명화 기법 연구

이 다 은◆, 홍 충 선 *

A Study on Anonymity Scheme Using Entorpy and Clustering

Da Eun Lee◆, Choong Seon Hong*

요 약

개인 정보를 다루는 서비스에서 개인 정보의 유출을 방지하기 위해 익명화 등을 이용해 데이터를 비식별화 한다. 하지만 개인 정보의 유출을 방지하기 위해 익명화 된 데이터는 복구가 불가능 하다. 따라서 익명화 과정에서 서비스에 필요한 정보의 손실을 줄여 데이터의 유용성을 높이고 기존의 개인 정보 보호 기능을 수행 할 수 있는 기법이 필요하다. 따라서 본 논문에서는 개인 정보 보호와 데이터의 유용성을 높이기 위해 클러스터링과 엔트로피를 이용한 익명화 기법에 대해 제안하였다. 서비스에서 활용 될 데이터의 유용성을 높이기 위해 k-means 클러스터링을 이용한 준식별자의 클러스터링 기법과 다른 데이터와의 연계성을 줄이기 위해 엔트로피를 이용한 k-anonymity 기법을 제안 하였다.

Key Words : Anonymity, k-Anonymity, Entropy, Clustering, k-Means

ABSTRACT

In the service that deals with personal information, data is de-identified to prevent leak of personal information such as anonymity. However, the anonymized data can not be recovered. So, we need new anonymity scheme that increases utility and reduces linkage between data. Thus, in this paper, we propose anonymity scheme using entropy and clustering. Using entropy, we propose to reduce linkage between data. And using k-means clustering, we propose to increase utility of data.

※이 논문은 2016년도 정부(미래창조과학부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (R0126-16-1009, ICBMS 플랫폼 간 정보 모델 연동 및 서비스 매쉬업을 위한 스마트 중재 기술 개발)
◆ First Author : Kyung Hee University Department of Computer Science and Engineering, daeunlee@khu.ac.kr
* Corresponding Author : Kyung Hee University Department of Computer Science and Engineering, cshong@khu.ac.kr
논문번호 : KNOM2016-01-09, Received April 28, 2016; Revised June 3, 2016; Accepted July 7, 2016

Ⅰ. 서 론

최근 사물인터넷, 클라우드, 빅데이터, 모바일 등 다양한 환경을 이용한 서비스가 증가하고 있다. 이러한 서비스는 사용자에게 다양한 서비스를 제공하기 위해 개인 정보를 포함한 데이터를 다룬다. 여기서 개인 정보는 개인은 식별할 수 있는 정보로 이름 주소, 나이, 주민등록번호 등을 포함하는 정보를 말한다[1].

표 1 개인 정보의 예
Table 1 Example Personal Data

Type	Examples
Identity	Name, Phone number, Resident registration number, etc.
Economy	Income, Asset, Bank account, Credit rating, etc.
Social	Phone number, Address, E-mail, etc.
Body	Health, Height, Weight, etc.
Biometrics	Fingerprint, Iris, DNA, etc.

개인 정보가 포함된 데이터는 기업, 정부, 학교와 같은 많은 기관에서 관련 연구 및 서비스 제공 등의 다양한 목적으로 수집되고 공유된다. 하지만 개인 정보가 포함된 데이터는 다양한 보안 위협에 의해 노출 될 수 있다. 개인 정보가 포함된 데이터에서 개인의 신원을 파악할 수 있는 이름, 주민등록번호 등의 식별자를 삭제하거나 변경하여 개인 정보가 노출되는 것을 방지한다. 하지만 식별자를 삭제하거나 변경하더라도, 다른 데이터와 연계시키면 개인 정보가 노출 될 수 있다.

개인 정보를 포함하고 있는 데이터에서 개인을 식별할 수 없도록 개인 식별요소를 삭제 또는 대체하거나 다른 데이터와 쉽게 결합하지 못하도록 하여 특정 개인을 알아볼 수 없도록 하는 것을 비식별화라고 한다. 비식별화 방법에는 익명화가 있다[2]. 익명화는 개인 식별정보를 제거하는 방법으로 익명화 된 개인 정보는 복원할 수 없다.

비식별화 과정에서 개인 정보가 노출되는 것을 방지하기 위해 데이터의 개인 식별 요소를 많이 제거할수록 정보의 손실이 증가하여 데이터의 정확성은 낮아진다. 따라서 기존에 존재하는 익명화에 관한 연구는 대부분 개인 정보 보호 만 고려하여 개인 식별 요소를 제거하는 방법이나 전체 테이블의 정보 손실률을 줄이는 방법으로 연구가 진행되어 왔다[3][4][5]. 하지만 개인 정보 보호 수준을 높이게 되면 많은 정보 손실이 발생한다. 또한 전체 데이터의 정보 손실을 줄였다고 하더라도 서비스마다 활용되는 데이터가 다르기 때문에 서비스에 유용성이 높은 데이터를 제공했다고 볼 수 없다. 따라서 개인 정보 보호와 데이터의 유용성을 높이는 연구가 필요하다.

따라서 본 논문에서는 개인 정보 보호와 유용성을 높인 익명화 기법을 제안한다. 새로운 익명화 기법을 제안하기에 앞서 기존의 익명화 기법인 k-anonymity에 대해 분석하고 entropy와 k-means 클러스터링 알고리즘을 이용하여 유용성을 높인 익명화 기법을 제안한다.

본 논문은 다음과 같은 구성으로 이루어진다. 제 2장에서는 제안사항에 기반이 되는 k-anonymity, entropy, k-means 알고리즘에 대해 설명한다. 제 3장에서는 본 논문의 제안사항을 데이터 클러스터링과 엔트로피를 이용한 익명화 두 부분으로 나누어서 설명한다. 제 4장에서는 제안 사항에 대한 성능 평가를 실시한다. 마지막으로 제 5장에서 결론을 맺는다.

Ⅱ. 관련 연구

이 장에서는 제안 사항을 설명하기에 앞서 제안 사항에 기반이 되는 익명화와 클러스터링, 엔트로피에 대해 설명한다.

1. 익명화(Anonymity)

익명화는 데이터에 포함된 개인 식별 정보를 삭제하거나 알아볼 수 없는 형태로 변환하는 것을 말한다. 일반적으로 익명화는 데이터 테이블을 식별자, 준식별자, 민감정보로 구분한다. 이 후 테이블의 식별자를 삭제하고 준식별자를 치환하여 익명화를 수행한다. 익명화의 가장 대표적인 알고리즘은 k-anonymity이다.

1.1 k-Anonymity

k-anonymity는 데이터들 간의 연관성을 줄이기

위해 고안된 익명화 기술이다. k-anonymity를 만족한다는 것은 주어진 데이터 집합에서 준식별자 속성값들이 동일한 레코드가 적어도 k개 존재한다는 것을 의미한다[6][7].

그림 1은 k=3을 만족하도록 익명화한 데이터 테이블 예시이다. 그림 1의 위쪽 테이블은 익명화하기 전 원본 테이블이다. 아래 테이블은 k=3을 만족하도록 익명화 한 결과를 나타내는 테이블이다. 익명화 결과 테이블에서 레코드 1~3과 4~6의 준식별자의 속성값들이 동일하다는 것을 확인 할 수 있다.

	Non-Sensitive			Sensitive
	Age	Gender	Zip Code	Disease
1	23	F	43711	Flu
2	27	F	43711	Flu
3	24	F	43711	Flu
4	32	M	43715	Fracture
5	38	F	43712	AIDS
6	33	M	43711	Hepatitis

	Non-Sensitive			Sensitive
	Age	Gender	Zip Code	Disease
1	20 – 29	F	43711	Flu
2	20 – 29	F	43711	Flu
3	20 – 29	F	43711	Flu
4	30 – 39	*	4371*	Fracture
5	30 – 39	*	4371*	AIDS
6	30 – 39	*	4371*	Hepatitis

그림 1. k-anonymity를 만족하는 데이터 테이블 예(k=3)
Fig. 1 An Example k-Anonymity (k=3)

2. Entropy

어떤 확률변수의 불확실성의 정도를 측정하는 것을 엔트로피라고 한다. 불확실성이 높아질수록 정보의 양은 더 많아지고 엔트로피는 커진다. 예를 들어 동전 던지기의 경우 H, T 두 가지 경우만 나타내므로 엔트로피는 1이다. 하지만 특정한 면이 나올 확률이 더 높은 경우 엔트로피는 1보다 작아진다. 특정한 면을 예측해서 맞출 확률이 더 높아졌기 때문에 엔트로피가 더 작아진 것이다. 표본 공간이 총 N개의 서로 다른 값으로 이루어진 확률 질량 함수 $p_i = 1/N$에 대해 엔트로피 H는 아래의 수식과 같이 정의 된다[8].

$$H = -\sum_{i=1}^{N} p_i \cdot \log_2 p_i$$

3. Clustering

클러스터링이란 수많은 데이터를 비슷한 특징을 가지는 클러스터로 군집화 시키는 것을 말한다. 그림 2은 클러스터링의 예시를 나타낸 것이다.

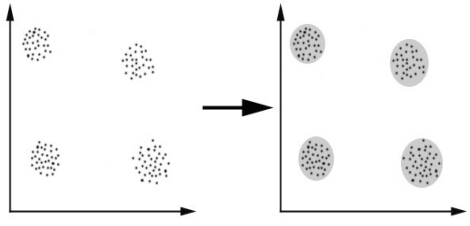

그림 2. Clustering 예시
Fig. 2 An Example Clustering

3.1 k-means Clustering

k-means 클러스터링은 주어진 데이터를 k개의 클러스터로 군집화 시키는 알고리즘이다. k-means 클러스터링은 각 클러스터와 거리 사이의 분산을 최소화하는 방식으로 동작한다. k-means 클러스터링은 자율 학습의 일종으로 레이블이 달려 있지 않은 입력 데이터에 레이블을 달아주는 역할을 수행한다. k-means 알고리즘은 μ_i가 집합 S_i의 중심점이라 할 때 각 집합의 중심점에서 집합 내 오브젝트간 거리의 제곱 합을 최소로 하는 집합 S를 찾아 클러스터링 한다. 최소 거리를 이용한 k-means 알고리즘에 대한 수식은 다음과 같이 나타낼 수 있다[9].

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} ||x - \mu_i||^2$$

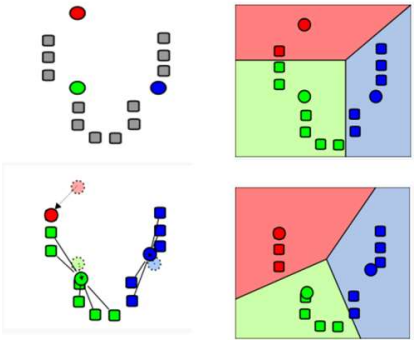

그림 3. k-means Clustering 과정
Fig. 3 k-means Clustering Process

Ⅲ. 제안 사항

이번 장에서는 앞서 설명한 k-anonymity와 엔트로피, k-means clustering을 이용한 익명화 알고리즘을 제안한다. 그림 4는 제안사항의 구조도를 나타낸 그림이다. 제안사항을 간단히 설명하자면 첫 번째로 서비스로부터 얻은 사용자의 개인정보가 담겨있는 테이블을 서비스에서 활용하고자 하는 준식별자 속성값에 대해 k-means 알고리즘을 이용하여 클러스터링을 한다. 두 번째로 클러스터링 한 테이블을 엔트로피를 이용한 k-anonymity를 사용하여 익명화 한다.

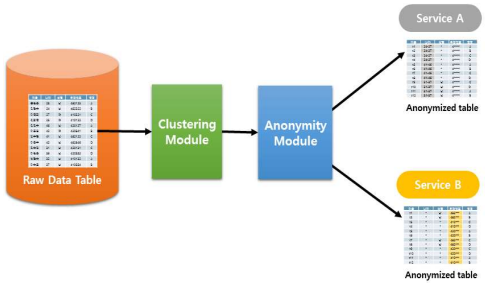

그림 4. 제안 사항 구조도
Fig. 4 Architecture

1. k-means를 이용한 데이터 클러스터링

본 논문에서 사용 될 데이터 테이블의 구조는 준식별자와 민감 정보로 이루어져 있다. 이번 세션에서는 각 서비스에서 필요로 하는 준식별자의 속성값의 기준으로 k-means 알고리즘을 이용하여 클러스터링하는 과정에 대해 설명한다.

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	25	F	43711	Flu
2	22	F	43711	Fracture
3	31	M	43712	Fracture
4	38	F	43713	AIDS
5	45	F	43713	Flu
6	49	M	43723	Cancer
7	52	M	43727	Cancer
8	58	M	43721	Flu

그림 5. 데이터 테이블 예
Fig. 5 An Example Data Table

그림 5는 데이터 테이블의 예시 이다. 데이터 테이블에서 서비스에서 필요로 하는 각각의 준식별자에 대한 속성값들을 k-means 알고리즘을 이용해 클러스터링 한다. 그림 6, 7, 8은 그림 5의 데이터 테이블의 준식별자 Age, Gender, Zip Code에 대한 속성값을 클러스터링 한 결과를 나타낸다.

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	25	F	43711	Flu
2	22	F	43711	Fracture
3	31	M	43712	Fracture
4	38	F	43713	AIDS
5	45	F	43713	Flu
6	49	M	43723	Cancer
7	52	M	43727	Cancer
8	58	M	43721	Flu

그림 6. 클러스터링 결과(Zip Code)
Fig. 6 Result of Clustering (Zip Code)

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	25	F	43711	Flu
2	22	F	43711	Fracture
3	31	M	43712	Fracture
4	38	F	43713	AIDS
5	45	F	43713	Flu
6	49	M	43723	Cancer
7	52	M	43727	Cancer
8	58	M	43721	Flu

그림 7. 클러스터링 결과(Gender)
Fig. 7 Result of Clustering (Gender)

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	25	F	43711	Flu
2	22	F	43711	Fracture
3	31	M	43712	Fracture
4	38	F	43713	AIDS
5	45	F	43713	Flu
6	49	M	43723	Cancer
7	52	M	43727	Cancer
8	58	M	43721	Flu

그림 8. 클러스터링 결과 (Age)
Fig. 8 Result of Clustering (Age)

2. 엔트로피를 이용한 k-anonymity

공격자가 익명화 된 데이터 테이블과 다른 데이터 테이블을 연계하여 원본 데이터를 유추 할 수 없도록 하기 위해서는 데이터 테이블의 가장 높은 엔트로피 값을 구하여 익명화를 하여야 한다[10]. 따라서 이번 세션에서는 엔트로피를 이용하여 최적의 익명화 값 k를 예측하는 알고리즘을 설명한다.

2.1 알고리즘

다양한 서비스로부터 들어오는 정보 레코드는 이전의 레코드들을 기반으로 하는 데이터 테이블 확률 r_i 값을 갖는다. r_i은 다음과 같이 정의 한다.

$$r_i = \frac{information\,record}{whole\,record}\; i = 1,2,...,n$$

r_i은 다음과 같이 정규화한 확률 p_i를 갖는다.

$$p_i = \frac{r_i}{\sum_{i=1}^{n} r_i}$$

정규화한 확률 p_i를 이용하여 엔트로피 H를 계산한다.

$$H = -\sum_{i=1}^{N} p_i \bullet \log_2 p_i$$

엔트로피 H를 이용하여 최적의 익명화 값 k를 구한다.

$$K = \arg\max H_i$$

표 2는 앞서 설명한 엔트로피를 이용한 k값을 예측하는 알고리즘을 수도 코드로 나타낸 것이다. 아래의 알고리즘을 통해 구한 k값을 이용하여 전체 데이터 테이블에 대해 k-anonymity를 만족하도록 익명화한다.

표 2 엔트로피를 이용한 k 값 예측 알고리즘
Table 2 Pseudo Code for Calculating k Using Entropy

Algorithm : Pseudo code for an optimal k using entropy	
Input : information record IR number of set N **Output :** an optimal K	
1 :	//Compute r_i
2 :	**for** (i=1; i<=N; i++) **do**
3 :	$r_i = \frac{IR_i}{i}$;
4 :	**end**
5 :	//compute p_i
6 :	**for** (i=1; i<=N; i++) **do**
7 :	**for** (j=1; j<=i; j++) **do**
8 :	$p_i = \frac{r_i}{\sum_{j=1}^{i} r_j}$;
9 :	**end**
10 :	**end**
11 :	//Compute optimal K

12 : **for** (i=1; i<=N; i++) **do**

13 : $$H_i = -\sum_{i=1}^{N} p_i \log_2 p_i;$$

14 : **end**

15 : output arg max H_i

3. 엔트로피와 k-means 클러스터링을 이용한 k-anonymity

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	25	F	43711	Flu
2	22	F	43711	Fracture
3	31	M	43712	Fracture
4	38	F	43713	AIDS
5	45	F	43713	Flu
6	49	M	43723	Cancer
7	52	M	43727	Cancer
8	58	M	43721	Flu

	Quasi-identifier			Sensitive
	Age	Gender	Zip Code	Disease
1	2*	F	43711	Flu
2	2*	F	43711	Fracture
3	*	*	4371*	Fracture
4	*	*	4371*	AIDS
5	*	*	4371*	Flu
6	*	M	4372*	Cancer
7	*	M	4372*	Cancer
8	*	M	4372*	Flu

그림 9. 익명화 결과 (k=2, Clustering=Zip Code)
Fig. 9 Result of k-Anonymity (k=2, Clustering=Zip Code)

이번 세션에서는 3.1장에서 클러스터링 한 데이터 테이블과 3.2장에서 구한 k값을 이용하여 데이터 테이블의 익명화를 데이터 예시를 통해 설명한다.

그림 9는 앞서 설명한 엔트로피를 이용한 k-anonymity의 예시를 나타낸 그림이다. 그림의 위에 있는 테이블은 익명화하기 전 k-means 클러스터링을 이용하여 준식별자의 속성값으로 클러스터링 한 결과 테이블이다. 그림 7은 준식별자 Zip Code 값으로 클러스터링 하였다. 엔트로피를 이용하여 k 값을 구하기 위해 클러스터링과정에서 각 클러스터의 전 테이블의 레코드를 기반으로 한 r_i를 구한다. r_i를 이용하여 정규화한 확률 p_i를 구한다. p_i를 이용해 테이블의 엔트로피를 구한다. 알고리즘을 이용하여 예제 테이블의 엔트로피를 계산하여 k=2를 구할 수 있다. 엔트로피를 이용하여 구한 k=2를 만족하도록 테이블을 익명화 한다.

Ⅳ. 성능 평가

이번 장에서는 3장에서 제안한 엔트로피와 클러스터링을 이용한 익명화 기법에 대한 성능 평가 결과를 분석한다. 성능 평가는 제안 사항인 각각의 준식별자를 클러스터링 한 결과에서 익명화 한 데이터 테이블의 정보 손실률과 일반적으로 무작위로 익명화한 k-Anonymity의 정보 손실률[11]을 비교하였다.

1. 정보 손실률

이번 세션에서 성능 평가에 앞서 성능 평가를 위한 정보 손실률을 구하는 방법을 설명한다. 정보 손실률은 속성값이 숫자인 것과 문자인 것으로 나누어서 계산 한다.

가장 먼저 속성값이 문자인 경우 정보 손실률은 정보를 식별 할 수 있으면 0%, 정보를 식별 할 수 없으면 100%로 계산한다. 해당 준식별자의 정보 손실률은 각 레코드의 정보 손실률의 평균값이다.

속성값이 숫자인 경우에는 해당 준식별자의 범위를 기준으로 익명화로 손실된 속성값의 범위에 대한 백분율로 계산하였다. 예를 들어 준식별자인 나이에서 나이의 범위가 1~100이라 할 때 25살인 사람의 익명화한 데이터 결과가 2* 이면 정보 손실률은 (10/100)*100으로 10%가 된다. 또한 나이가 30인 사람의 익명화 결과가 * 이면 정보 손실률은 (100/100)*100으로 정보 손실률은 100%가 된다[11].

2. 성능 평가 환경

본 논문의 성능평가 환경은 다음과 같다. 구현 환경은 Windows 7 운영체제에서 자바(JDK 1.8.0_91)를 사용하였다. 데이터 테이블은 50*5이며 50개의 레코드와 5개의 속성으로 구성되어 있다. 속성은 1개의 식별자와 3개의 준식별자, 1개의 민감 정보로 구성되어 있다. k-means 클러스터링은 오렌지[12]를 사용하였고 k-anonymity는 ARX[13] 라이브러리를 사용하였다.

표 3 구현 환경
Table 3 Evaluation Environment

Item	Contents
Operating System	Windows 7
Language	Java(JDK 1.8.0_91)
Table Size	50*5
Total Record	50
Total Attribute	5
Num. of Identifier	1
Num. of Quasi-Identifier	3
Num. of Sensitive Attribute	1
k-Means	Orange 3
k-Anonymity	ARX

3. 성능 평가 결과

그림 10, 11, 12은 제안 사항인 준식별자를 클러스터링하여 k-anonymity로 익명화한 데이터와 일반적인 k-anonymity로 익명화한 데이터의 정보 손실률을 비교한 그래프이다.

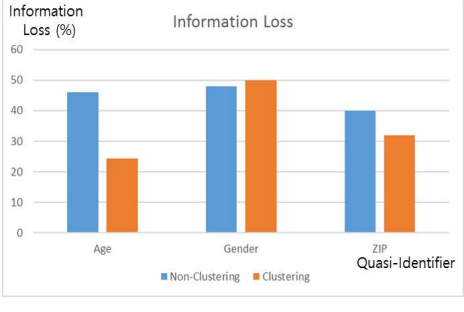

그림 10. 정보 손실률 (Clustering = Age)
Fig. 10 Information Loss of Age Clustering

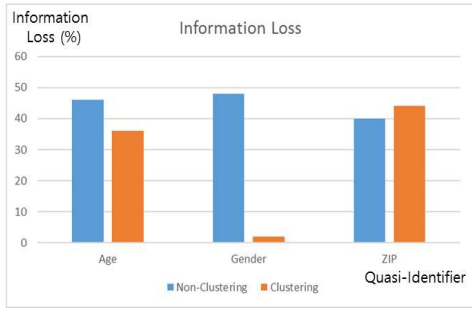

그림 11. 정보 손실률 (Clustering = Gender)
Fig. 11 Information Loss of Gender Clustering

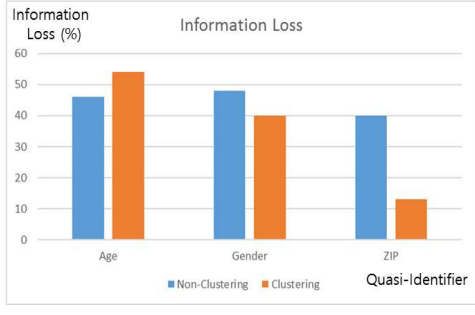

그림 12. 정보 손실률 (Clustering = Zip Code)
Fig. 12 Information Loss of Zip Code Clustering

그림 10은 준식별자 Age에 대해서 클러스터링 한 후 익명화 한 데이터와 일반적인 익명화 방법으로 익명화 한 데이터의 정보 손실률을 나타낸 그래프이다. 그래프에서 준식별자 Age의 정보 손실률이 클러스터링 한 후 익명화 한 데이터가 클러스터링을 하지 않고 익명화 한 데이터 보다 더 적은 정보 손실률을 보이는 것을 알 수 있다. 하지만 Gender에서는 클러스터링 하지 않은 익명화 데이터 보다 더 높은 정보 손실률을 보이는 것을 알 수 있다.

그림 11은 준식별자 Gender에 대해서 클러스터링 한 후 익명화 한 데이터와 일반적인 익명화 방법으로 익명화 한 데이터의 정보 손실률을 나타낸 그래프이다. 그래프에서 준식별자 Gender의 정보 손실률이 클러스터링 한 후 익명화 한 데이터가 클러스터링을 하지 않고 익명화 한 데이터 보다 더 적은 정보 손실률을 보이는 것을 알 수 있다. 하지만 Zip Code에서는 클러스터링 하지 않은 익명화 데이터 보다 더 높은 정보 손실률을 보이는 것을 알 수 있다.

그림 12는 준식별자 Zip Code에 대해서 클러스

터링 한 후 익명화 한 데이터와 일반적인 익명화 방법으로 익명화 한 데이터의 정보 손실률을 나타낸 그래프이다. 그래프에서 준식별자 Zip Code의 정보 손실률이 클러스터링 한 후 익명화 한 데이터가 클러스터링을 하지 않고 익명화 한 데이터 보다 더 적은 정보 손실률을 보이는 것을 알 수 있다. 하지만 Age에서는 클러스터링 하지 않은 익명화 데이터 보다 더 높은 정보 손실률을 보이는 것을 알 수 있다.

클러스터링 한 준식별자를 제외 한 다른 준식별자 모두가 클러스터링 하지 않고 익명화 한 데이터 보다 더 낮은 정보 손실률을 보이지는 않다. 하지만 각각의 준식별자로 클러스터링하여 익명화 한 데이터와 클러스터링을 하지 않고 익명화한 데이터의 정보 손실률을 비교한 그래프에서 각각의 클러스터링 한 준식별자의 데이터는 클러스터링 하지 않고 익명화 한 데이터 보다 낮은 손실률을 보인다. 이는 서비스에서 활용하고자 하는 준식별자의 데이터의 유용성이 클러스터링 하지 않고 익명화 한 데이터 보다 클러스터링 된 데이터를 익명화 한 데이터가 더 높다고 볼 수 있다.

V. 결 론

개인 정보의 노출을 방지하기 위해 데이터의 익명화를 이용하여 비식별화 된 데이터는 복구가 불가능 하다. 따라서 익명화 과정에서 서비스에 필요한 정보의 손실을 줄여 데이터의 유용성을 높이고 기존의 개인 정보 보호 기능을 수행 할 수 있는 기법이 필요하다.

따라서 본 논문에서는 개인 정보 보호와 데이터의 유용성을 높이기 위해 클러스터링과 엔트로피를 이용한 익명화 기법에 대해 제안하였다. 서비스에서 활용 될 데이터의 유용성을 높이기 위해 k-means 클러스터링을 이용한 준식별자의 클러스터링을 제안하였다. 다른 데이터와의 연계성을 줄이기 위해 엔트로피를 이용하여 k-Anonymity의 k값을 계산하는 기법을 제안하였다.

제안 사항의 성능 평가 결과 익명화 과정에서 클러스터링을 한 익명화 결과가 모든 속성 값에서 일반적으로 익명화 한 결과 보다 모든 준식별자에서 더 적은 정보 손실률을 보이지는 않았다. 하지만 서비스에서 필요로 하는 준식별자의 정보 손실률은 해당 준식별자를 클러스터링 한 익명화 결과가 일반적으로 익명화 한 결과 보다 더 적은 정보 손실률을 보였다. 이는 일반적인 익명화 데이터 보다 유용성이 높은 데이터를 서비스에 제공 할 수 있다는 것을 의미한다.

References

[1] KISA. [Online]. Available: https://guide.kisa.or.kr/guide/privacy.jsp

[2] IHE IT Infrastructure Technical Committee, “IHE IT Infrastructure Handbook : De-Identification,” pp1-41, June 2014.

[3] Yonghong Xie, Qing He, Dezheng, Xiaojing Hu, “Medical Ethics Privacy Protection Based on Combininbg Dstributed Randomization With k-anonymity,” 2015 8th International Congress on Image Signal Processing, October 2015.

[4] Yonglei Yao, Laurence T. Yang, Neal N. Xiong, “Anonymity-Based Privacy-Preserving Data Reporting for Participatory Sensing,” IEEE INTERNET OF THINGS JOURNAL, vol. 2, issue no. 5, pp381-390, October 2015.

[5] David F. Nettleton, Julian Salas, “A data driven anonymization system for information rich online social network graphs,” Expert Systems With Applications vol. 55, pp87-105, 2016.

[6] Chikwang Hwang, Choong Seon Hong, Jongwon Choe, “A Study on Service-based Secure Anonymization for Data Utility Enhancement,” Journal of KIISE, vol. 42, issue no. 5, pp681-689, May 2015.

[7] Latanya Sweeney “K-ANONYMITY : A MODEL FOR PROTECTING PRIVACY,” International Journal on Uncertainty Fuzziness and Knowledge-based System, vol. 10, issue no. 5, pp557-770, 2002.

[8] WIKIPEDIA. [Online]. Available : https://en.wikipedia.org/wiki/entropy

[9] WIKIPEDIA. [Online]. Available : https://en.wikipedia.org/wiki/k-means

[10] Da Eun Lee, YoungKi Kim, Choong Seon Hong, “Anonyropy : An Efficient Anonymization Scheme Using Entropy in Smarat Mediator for Mashup Service,” The 18th Asia-Pacific Network operations and

Management Symposium, October 2016
[11] Xiangwen Liu, Quingquin Xie, Liangmin Wang, "A Personalized Extend (a,k)-Anonymity Model," 2015 Third International Conference on Advanced Cloud and Big Data, November 2015
[12] Orange, [Online], Available : http://orange.biolab.si/
[13] ARX, [Online], Available : http://arx.deidentifier.org/
Commun. 2010 (KICS ICC 2010),

이 다 은 (Da Eun Lee)

2015년 : 경희대학교 컴퓨터공학과(공학사)
2017년 : 경희대학교 컴퓨터공학과(공학석사)
<관심분야> IoT, 정보 보호, 익명화

홍 충 선 (Choong Seon Hong)

1983년 : 경희대학교 전자공학과(공학사)
1985년 : 경희대학교 전자공학과(공학석사)
1997년 : Keio University, Department of Information and Computer Science(공학박사)
1988~1999년 : 한국통신통신망연구소 수석연구원/네트워킹 연구실장
1999년~현재 : 경희대학교 컴퓨터공학과 교수
<관심분야> 인터넷 서비스 및 망 관리구조, 미래인터넷, IP mobility, Sensor Networks, Network Security