World Wide Web (2018) 21:1773–1785
https://doi.org/10.1007/s11280-018-0604-z

The purpose driven privacy preservation for accelerometer-based activity recognition

Soumia Menasria[1] · Jianxin Wang[1] · Mingming Lu[1]

Received: 12 October 2017 / Revised: 16 February 2018 / Accepted: 25 May 2018 /
Published online: 3 September 2018

Abstract Accelerometer-based activity recognition (AAR) attracted a lot of attentions due to the wide spread of smartphones with energy-efficiency. However, since accelerometer data contains individual characteristics; AAR might raise privacy concerns. Although numerous privacy preservation approaches, such as "privacy filtering, differential privacy, and inferential privacy", have been proposed to conceal sensitive information, unfortunately they cannot address the privacy problem associated with AAR. In this paper, we report our efforts to control the use of the AAR while preserving the privacy. To achieve this task, our method leverages a connection to agglomerative information bottleneck, through which the amount of disclosed data can be compressed so that irrelevant private information can be reduced, and a connection to general privacy statistical inference framework, where both of the privacy leakage and utility accuracy are considered as mutual information. Our experimental results have shown that the proposed solution can greatly reduce privacy leakage while maintaining a relative good utility.

Keywords Accelerometer · Activity recognition · Agglomerative information bottleneck · Statistical inference · Mutual information · Privacy-utility tradeoff

This article belongs to the Topical Collection: *Special Issue on Deep Mining Big Social Data*
Guest Editors: Xiaofeng Zhu, Gerard Sanroma, Jilian Zhang, and Brent C. Munsell

✉ Mingming Lu
mingminglu@csu.edu.cn

Soumia Menasria
menasriasoumia@csu.edu.cn

Jianxin Wang
jxwang@csu.edu.cn

[1] School of Information Science and Engineering, Central South University, Changsha, 410083, China

1 Introduction

The widespread of smartphones, equipped with diverse and sophisticated sensors, especially accelerometer sensors, offer a great opportunity for better understanding human context [2, 3]. However, this also raised new privacy concerns [11, 18]. Although numerous existing works have considered the issues of context privacy [5, 22], especially location privacy [12], only a few works have paid attentions to context privacy issues related to accelerometers [8, 10]. Especially, to the best of our knowledge, none of the existing works have studied the privacy issues of AAR.

Human activity, as a key part of human context, has attracted increasingly attentions from both research and industry communities [14]. Among them, AAR has received much more attention, because it does not require user's intervention and it is much more energy-efficient [31]. Though, accelerometer seems innocuous at the first glance, it actually can leak non-negligible user context information, such as user identity [11]. Works done on this [3, 28, 30], have shown that the accelerometer data for AAR can be utilized to recognize users with accuracy over 99%.

Existing common privacy-preserving approaches, such as privacy filtering [6], differential privacy [13], cannot address the privacy problem associated with AAR due to the following reasons: 1) in the case that a utility-privacy trade-off exists, simply filtering out privacy data might compromise the data utility; 2) differential privacy cannot resist the statistical inference attack because the properties of target protected by differential privacy can be inferred by statistical learning [23, 24]. Although a novel inference privacy preserving approach, called IpShield [27], has been proposed recently, it cannot address the privacy problem associated with AAR either, because it assumed privacy leakage to be a boolean variable without considering the possibility of partial privacy leakage.

The proposed solution is motivated by our experiment study on a large-scale accelerometer-data set for human activity recognition [3]. More specifically, the utility (the AAR accuracy) and the privacy (the identity accuracy) adapt differently to information loss in terms of sampling rate reduction, as shown in Figure 1, from which, it can be observed that the accuracies of both activity recognition and identity recognition reduce along with the increment of sampling interval. However, the reduction rate of identity recognition is much faster than that of the activity recognition. The phenomenon shown in Figure 1 actually reflects a general utility-privacy trade-off [1, 7], which opens a door for designing a novel approach to maintain relatively acceptable utility while minimizing the risk of privacy leaking.

Although Mirco et al. proposed SensorSift [17], which could balance utility and privacy by transforming raw sensor data into a sifted representation, and utilized existing state-of-art machine learning algorithms to verify the effect of the sifted representation, the advance of machine learning algorithms might still be able to infer unexpected privacy information. Thus, it is necessary to formalize the utility privacy trade-off theoretically. Therefore, we came up with a new purpose (utility) driven privacy preserving framework; which extends the information bottleneck framework [19, 20] by introducing privacy.Due to the proposed framework, we suggested two Hierarchical Agglomerative Clustering (HAC) founded on algorithms to balance the utility and privacy trade-off. Experimental study has shown that the proposed algorithms can achieve a high utility with a low privacy leakage.

The contributions of this work can be summarized as following: 1) to the best of our knowledge, we first propose the inference privacy problem associated with smartphone accelerometer sensing; 2) we extend an information theory built on framework information

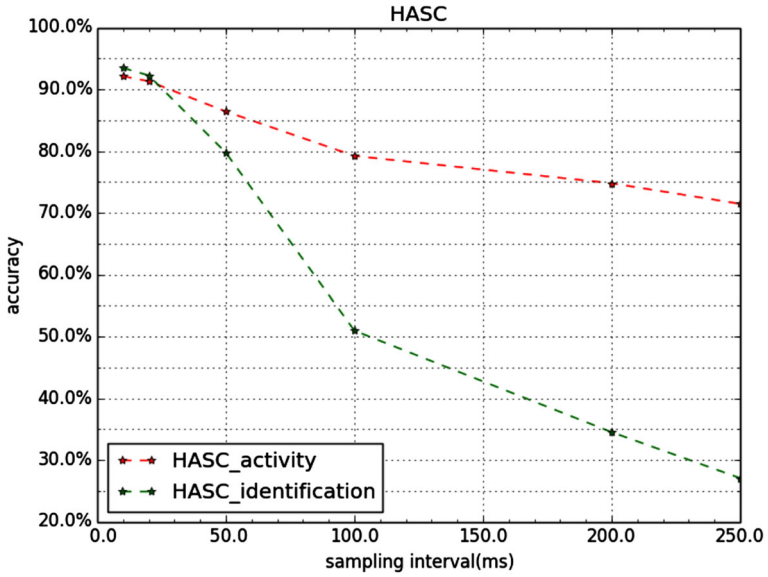

Figure 1 An illustrative example

bottleneck by taking privacy leakage into account, which can formally quantify inferential privacy leakage; 3) we design two information bottleneck based algorithms to achieve a good balance between utility and privacy.

The rest of this paper is organized as follows. The related work is depicted in Section 2. In Section 3, the problem is formulated. In Section 4, two HAC-based clustering algorithms are provided. In Section 5, the experimental evaluation settings and results are presented. Conclusions and future works are described in Section 6.

2 Related works

Several works [1, 7, 17, 29] that have been put into debate to analyze the tradeoff between utility and privacy. Among them: Differential privacy [13, 16, 23] provided a robust formal guarantees about the anonymity of the participants in a clean database. Despite its wide usage, differential privacy could not guarantee zero private information leakage and therefore might lead to privacy leakage due to statistical inference [24].

Therefore, inferential privacy [19, 26], which depends on a general privacy statistical inference framework [9], have been proposed to address the privacy leakage issue that cannot be addressed by differential privacy. Inferential privacy captured the privacy threat scenario incurred by releasing data with certain utility purpose to a potential malicious adversary utilizing statistical inference attacks, where the privacy leakage has been measured in term of inference cost gained under a distortion constraint. Inferential privacy enforced data to be converted before releasing through a probabilistic privacy mapping, which hide privacy information as much as possible.

Besides the privacy models, the metric for privacy/utility is the other important factor [15] to evaluate the soundness and robustness of the privacy preserving works. Among the proposed metrics, mutual information is a relatively good one, as it can accurately reflect

the privacy/utility information contained in the data. For example, [25] and Ali et al. [4] utilized mutual information to model the privacy leakage in the differential privacy model and the inferential privacy model, respectively. Moreover, Ali et al. [4] established a connection between their proposed model and the information bottleneck model, which well modeled the tradeoff between learning accuracy and learning complexity, i.e., the length of the compressible data. In our work, we extend information bottleneck model by introducing the privacy variable and utilize mutual information to model the privacy.

3 Problem statement

3.1 Setting

The privacy-preserving problem discussed in this work can be formalized through a purpose driven privacy framework, as shown in Figure 2, where the random variable $X \in \mathcal{X}$ represents the original accelerometer data (public or observable data), the correlated random variable $S \in \mathcal{S}$ denotes the private (personal) information, such as a user's identity, random variable $Y \in \mathcal{Y}$ indicates the set of recognized activities (utility), and random variable $T \in \mathcal{T}$ shows the compressed version of the original data X. It has to note that, although for the AAR utility-privacy trade-off problem, the original data $\mathcal{X}$ may include the private information S, Figure 2 illustrates a more general case, where $\mathcal{S}$ has an overlap with $\mathcal{X}$ but is not exactly in it. In the purpose-driven privacy framework, a user's accelerometer data may disclose to some applications for the utility of recognized activity.

The correlation between X and S is computed by the joint probability $P_{S,X}$. Due to this correlation, privacy attackers can utilize various machine learning algorithms to infer users' privacy. To decrease the inference threat on users' private data, is necessary to transform the disclosed data X to a filtered data T before disclosing under a probabilistic mapping $P_{T|X}$, such an amount of information contained in $\mathcal{T}$ with respect to X, is just enough for the intended purpose, Y, through a conditional distribution $P_{Y|T}$. The potential information leakage about S is minimized under the utility constraint. In the purpose-driven privacy preserving framework, both the utility and privacy can be estimated through mutual information, where the mutual information $I(T; Y)$ and $I(T; S)$ represent the utility and privacy, respectively.

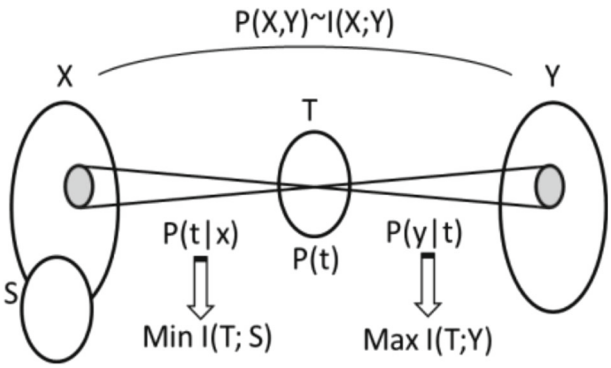

Figure 2 Purpose driven privacy

The compressed representation T is generated by passing X through a probabilistic mapping, called *privacy mapping*. Formally, this mapping can be characterized by a conditional distribution $P_{T|X}$. The goal of this mapping is to reduce the inference threat on private data S by minimizing the dependency between the private variable S and T, measured by $I(T;S)$, while preserving the utility through a conditional distribution $P_{Y|T}$ by maintaining the dependency between relevant variable Y and T, measured by $I(T;Y)$. The basic trade-off is between minimizing the compression information while maximizing the relevant-information. An illustration of this idea is given in Figure 2.

3.2 Problem formulation

The problem described previously can be formulated as a constrained optimization problem as shown in (1), where the optimization object is to maximize the intended utility $I(T;Y)$, and the optimization constraint is to let the privacy leakage below some predefined threshold ε, among all feasible privacy mapping $P_{T|X}$.

$$\max_{P_{T|X}:I(T;S)\leq\varepsilon} I(T;Y) \tag{1}$$

A common way to address a constrained optimization problem is to convert into an unconstrained optimization problem through Lagrange multiplier, because it is relatively easy to solve. The constrained optimization problem formulated in (1) can be converted into an unconstrained optimization problem as shown in (2).

$$max\mathcal{L}(T,Y,S,\lambda) = max(I(T;Y) - \lambda I(T;S)) \tag{2}$$

The lagrange multiplier λ operates as a control parameter, which adjusts the trade-off between the amount of utility, $I(T;Y)$, and the amount of privacy leakage, $I(T;S)$. The value of λ has a significant impact on the optimization results: a larger value implies a preference to reduce privacy leakage, while a smaller value denotes a preference to increase utility.

However, the drawback of this unconstrained formulation is that the lagrange multiplier λ is hard to be estimated in some applications. After thoroughly examining the relationship in terms of information among the original data X, the intended variable Y, and the private data S, we identify a potential priority-based on optimization scheme, which partitions the original data X into four components and compresses them according to their relatively importance to Y and S, respectively.

Figure 3 illustrates the mutual information among X, Y, and S. So what needs to be noticed here is that $H(X)$, the information of X, is actually consists of four components, namely, $I(X;Y|S)$, $I(S;Y;X)$, $I(X;S|Y)$, and $H(X|Y,S)$. Moreover, the contributions of those four information components to the reduction of privacy leakage and the improvement of application utility vary. Among the four components of $H(X)$; $I(X;S|Y)$ is the most desirable component to be reduced. The underlying reasons lie in that, on one hand, it does not contribute to the utility, as the information related to utility is exclusively contained in $I(X;Y)$, which consists of $I(X;Y|S)$ and $I(S;Y;X)$. On the other hand, the reduction of $I(X;S|Y)$ can greatly reduce the privacy leakage because it contains private information about S. $H(X|Y,S)$ is the second one to be reduced, because it also has nothing to do with the utility, even though it does not contain private information. $I(S;Y;X)$ is the third one to be reduced, because, although its reduction can reduce privacy leakage, meanwhile it reduce utility. $I(X;Y|S)$ is the least to be reduced, as it contains no private information, and its reduction can compromise the utility.

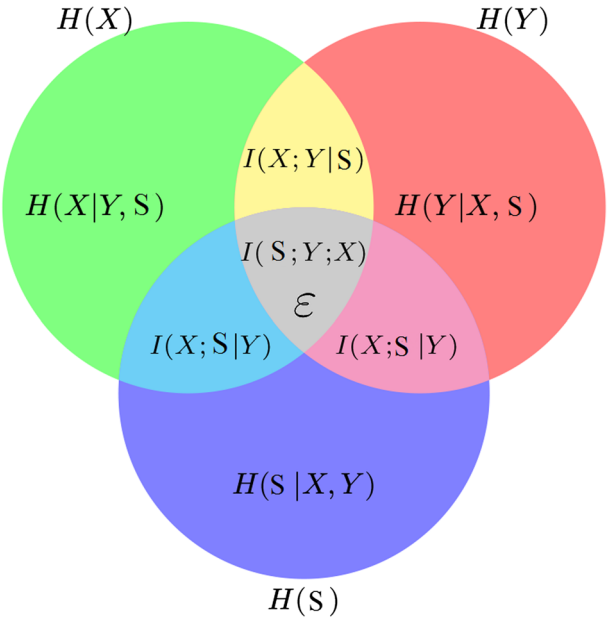

Figure 3 Information theoretic measures for three variables S, Y, and X.

4 Algorithm for privacy preserving

We adopt the HAC [21] founded on clustering scheme to address the suggested purpose driven privacy protection problem. HAC is a clustering mechanism, which initializes each data point as one cluster, and merges two clusters at a time developed on some predefined metric until certain threshold has been satisfied. In order to decide which pair should be merged, we proposed two metrics based on information theoretic framework: the first one based on the lagrange multiplier as shown in (2), and the second one based on the priority-based scheme described in Section 3.2. The idea is to minimize the loss of mutual information between successive clustering and to reduce the leakage of private data.

To gage utility and privacy information loss due to clustering, we adopt the *Jensen-Shannon (JS) divergence* base metric proposed in Proposition 1 in [21], where *JS divergence (JSD)* is used to estimate the distance between the two conditional probability $P(d|t_i)$ and $P(d|t_j)$ as shown in Figure 4. Here, $d \in D$ represents either $y \in Y$ or $s \in S$. Thus, the information loss, due to the merge of two clusters t_i and t_j, can be formulated as (3).

$$I_{ij}(T; D) = (p(t_i) + p(t_j)) * JS_{\Pi_2}[p(d|t_i), p(d|t_j)] \tag{3}$$

In (3), Π presents the merge prior distribution of t_i in the merged subset.

4.1 Lagrange multiplier based algorithm

Our first algorithm focuses on the tradeoff between utility and privacy, using unconstraint optimization through Lagrange multiplier $I_{ij}(T; Y) - \lambda I_{ij}(T; S)$ as metric, to select the clusters to be merged as shown in Algorithm 1.

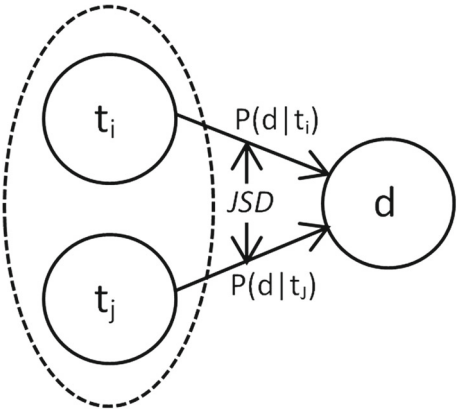

Figure 4 The cost of merging illustration

Algorithm 1 Lagrange multiplier based algorithm

Input: $S = |S|$, $N = |X|$, $M = |Y|$
Output: T_m: m-partition of X into clusters, for every $1 \leq m \leq N$
Initialization:
Construct $T \equiv X$ **for** *every* $i, j = 1..N$ **,** $i < j$ **do**
  $I_{ij}(T; Y) = (p(t_i) + p(t_j)) * JS_{\Pi_2}[p(y|t_i), p(y|t_j)]$
  $I_{ij}(T; S) = (p(t_i) + p(t_j)) * JS_{\Pi_2}[p(s|t_i), p(s|t_j)]$
Loop:
for $t = 1..N - 1$ **do**
  Find $\{\alpha, \beta\} = argmin_{i\ j}\{I_{ij}(T; Y) - \lambda I_{ij}(T; S)\}$
  Merge $\{t_\alpha, t_\beta\} \rightarrow \acute{t}$ Update $T = \{T - \{t_\alpha, t_\beta\}\} \cup \{\acute{t}\}$
  (T is now a new $(N - t)$ partition of X with $N - t$ clusters)
  Update $I_{ij}(T; S)$, $I_{ij}(T; Y)$ costs and pointers w,r,t, $\acute{t}$ (only for couples contained t_α or t_β)

To find the best pair to merge, in each iteration, the selecting metric, $I_{ij}(T; Y) - \lambda I_{ij}(T; S)$, is maintained up to date for each pair of clusters. Based on the maintained metrics, in each iteration, the pair with the minimum metric will be selected for merge so that the combined information associated with remaining clusters will be maximized. Once the merge finishes, both the cluster set and the corresponding utility and privacy will be updated accordingly.

4.2 Privacy constraint based algorithm

As mentioned previously, both of the proposed solutions are founded on HAC clustering scheme. Thus, the second solution differs from the first one only in the selection metric. Instead of using Langrange multiplier to trade off utility and privacy, the second one (Algorithm 2) tries to minimize the privacy leakage first. By defining $\varepsilon = I(S; Y; X)$, as long as the privacy leakage $I(T; S)$ is larger than ε, Algorithm 2 can iteratively merges the pair of clusters with the maximal privacy information loss until no further privacy information can be reduced without compromising the utility $I(T, Y)$. In this case, the merging process will continue by merging the pair of clusters with minimal utility information loss.

Algorithm 2 Privacy constraint based algorithm

Input: ε, $S = |S|$, $N = |X|$, $M = |Y|$
Output: T_m: m-partition of X into clusters, for every $1 \leq m \leq N$
Initialization:
Construct $T \equiv X$
for *every* $i, j = 1..N$ **,** $i < j$ **do**
  $I_{ij}(T; Y) = (p(t_i) + p(t_j)) * JS_{\Pi_2}[p(y|t_i), p(y|t_j)]$
  $I_{ij}(T; S) = (p(t_i) + p(t_j)) * JS_{\Pi_2}[p(s|t_i), p(s|t_j)]$
Loop:
for $t = 1..N - 1$ **do**
  Find $\{\alpha,\beta\}$:
  if $I(T; S) > \varepsilon$ **then**
    $\{\alpha, \beta\} = argmax_{i,j}\{I_{ij}(T; S)\}$;
  else
    $\{\alpha, \beta\} = argmin_{i,j}\{I_{ij}(T; Y)\}$;
  Merge $\{t_\alpha, t_\beta\} \rightarrow \acute{t}$
  Update $T = \{T - \{t_\alpha, t_\beta\}\} \cup \{\acute{t}\}$
  (T is now a new $(N - t)$ partition of X with $N - t$ clusters)
  Update $I_{ij}(T; S)$, $I_{ij}(T; Y)$ costs and pointers w,r,t, $\acute{t}$ (only for couples contained t_α or t_β)

5 Experimental evaluation

5.1 Data set

The data adopted in this evaluation is Human Activity Sensing Consortium (HASC) (http://hasc.jp/hc2011/download-en.html) data set, which collected mobile phone sensor data from 107 users, who placed their mobile phone in different positions. The sensor data include gyroscope data, GPS data, accelerometer data, and etc. Those sensor data have been label with the six corresponding human activity: stay, walk, jog, skip, stUp (stairs up), and stDown (stairs down). To simplify the evaluation, we select only the accelerometer data when the mobile phones are put on the waist. The sampling rate for the accelerometer data is 100 Hz. It was performed by all subjects in a controllable lab environment.

In our experiment, Agglomerative Information Bottleneck (aIB) [21], which is the closest related work as mentioned in Section 2, is used to compare with the proposed algorithms.

5.2 Experiments

5.2.1 Setting λ

In this experiments we tend to study the effect of the tradeoff factor λ on utility and privacy leakage accuracies. We changed the value of λ from the extreme point, where $\lambda = 0$ to $\lambda = 5$ as shown in Figure 5. The optimal value of λ obtained is then applied to evaluation data.

Figure 5 illustrates that in the extreme point of $\lambda = 0$, the privacy leakage is high, while utility is in its highest accuracy. The utility accuracy start to decrease from $\lambda = 2$ and it

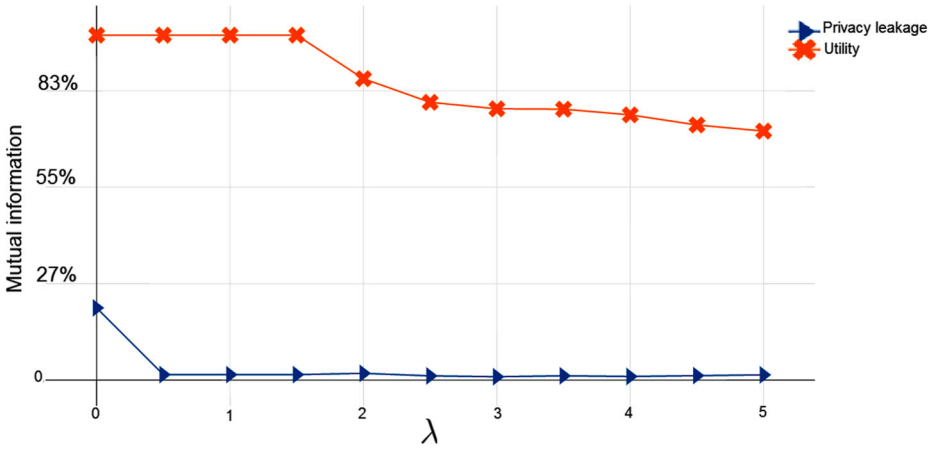

Figure 5 Results of utility and privacy leakage with different values of λ

is continue decreasing with the increment of λ value, while the privacy leakage decreases when $\lambda = 0.5$.

As shown in Figure 5, the optimal tradeoff value of λ is in the interval [0.5, 1.5].

To evaluate the proposed algorithms, we plot the mutual information variation along with partition size, as shown in Figure 6. From Figure 6, it can be observed that aIB algorithm and Algorithm 1 achieve the best utility (AAR accuracy), while Algorithm 1 shows the lowest privacy leakage accuracy. Algorithm 2 illustrates that the accuracies of both physical activity recognition and identity recognition increases along the increment of partition size, and the privacy leakage using this algorithm is less than aIB algorithm, while the utility accuracy with aIB is greater than Algorithm 2. It can be concluded that Algorithm 1 is the best among the three algorithms, as it can maximize the utility and minimize the privacy leakage.

5.2.2 Cluster selection

HAC seeks to build a hierarchy of clusters, starts from many clusters, and pair of clusters are merged as one and move up the hierarchy until reaching only one cluster. In the end one

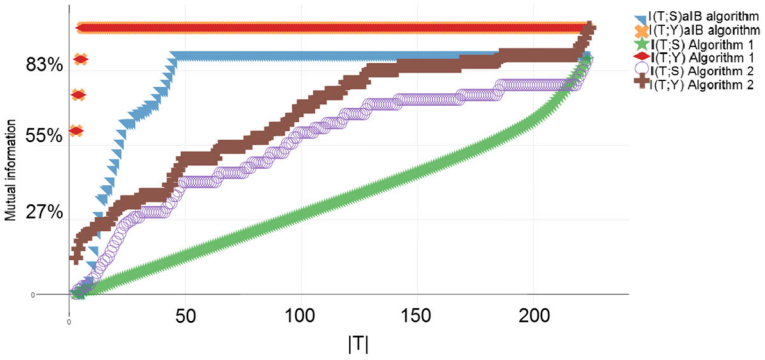

Figure 6 The mutual information variation along with partition size

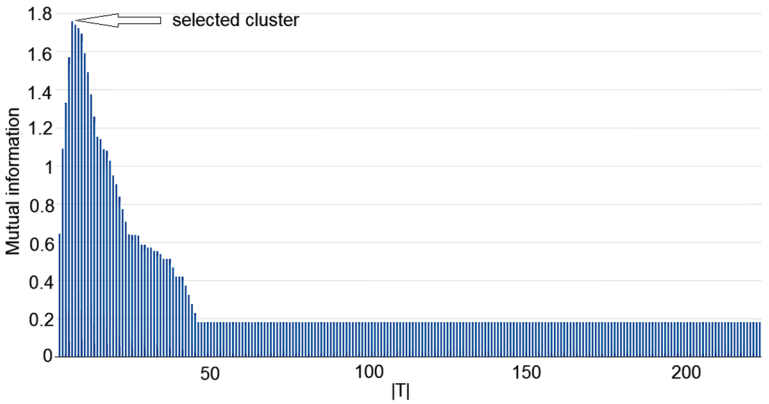

Figure 7 Chosen model for aIB algorithm

of those clusters must be chosen. The selected cluster represents a compact representation, which contains the most informative subset with minimal amount of data with respect to both utility and privacy. As mentioned before, $I(T;Y)$ denotes the utility accuracy, while $I(T;S)$ denotes the privacy leakage accuracy. The distance $I(T;Y) - I(T;S)$ between utility accuracy and privacy leakage accuracy shows the disparity between them. A small disparity denotes that utility and privacy leakage accuracies are close to each other and both of them are high or low. A large disparity denotes that utility and privacy leakage accuracies are far from each other, and it shows a high utility accuracy and small privacy leakage accuracy. The goal of cluster selection is to find the cluster that has low privacy leakage and high utility accuracy, which presents a large disparity between them. In this work, we propose to use the difference between utility and privacy leakage accuracies as criterion to find the best cluster that represents the data set, that contains the minimum average amount of private information while utility accuracy is high as possible, the chosen models shown in Figures 7, 8 and 9.

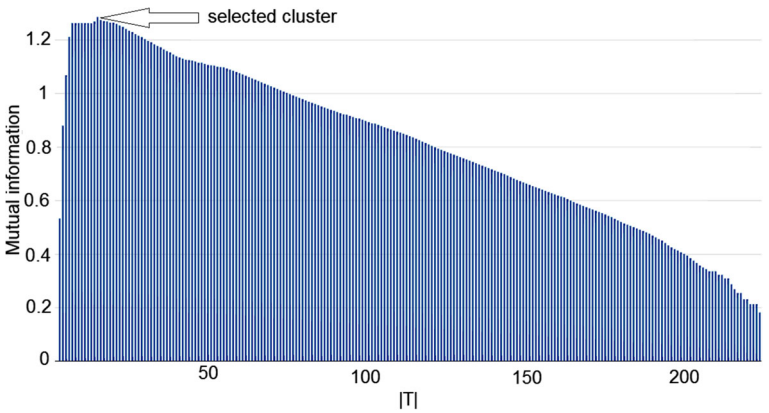

Figure 8 Chosen model for Algorithm (1)

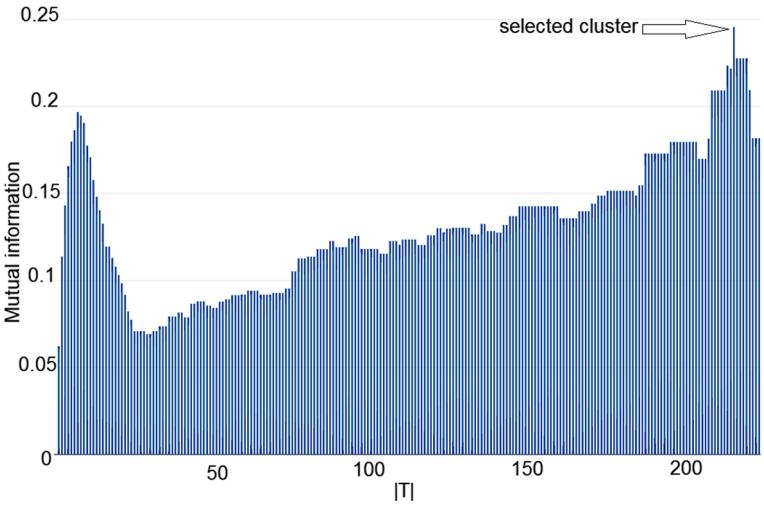

Figure 9 Chosen model for Algorithm (2)

5.2.3 Accuracy of both identity recognition and activity Recognition for the selected cluster

In Figure 10, we plot the accuracy of physical activity recognition and identity recognition together of the selected cluster and compare the different algorithms. We extract 13 features, namely: average, Root Mean Square (RMS), variance, standard deviation (STD), peak frequency, the average of the absolute deviation (AAD), median, energy, maximum value, minimum value, 70th percentile value, 80th percentile value, 90th percentile value. Using RandomForest (RF) classificatory we obtain the results illustrated in this Figure 10, where the red bar represents identity recognition accuracy, and the blue bar represents physical acitivity recognition accuracy. We observe that for a 100% utility accuracy, the average

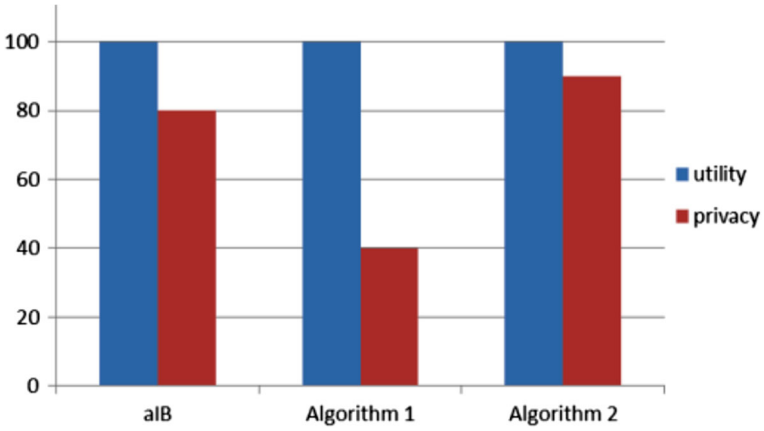

Figure 10 Accuracy of utility and privacy for the chosen partition

amount of privacy leakage is less than 40% using Algorithm 1. While it is larger than 40% for aIB algorithm and Algorithm 2.

6 Conclusion and future work

In this work, we studied the privacy-utility trade-off, which can keep the physical activity recognition accuracy and identity recognition (privacy concerns) balanced from the data contributors. Based on explanatory sensing data, this work proposed two new mechanisms for physical activity recognition under privacy constraint founded on the HAC, which extend the agglomerative information bottleneck algorithm. Intensive experiments have illustrated the effectiveness of the proposed algorithms. In future work, we will investigate the opportunity to use Generative Adversarial Network, in particular, we will adopt an adversarial multicriteria learning strategy by integrating shared knowledge from multiple segmentation criteria to extract the criteria-invariant features and the criteria-specific features to achieve a good balance between utility and privacy.

References

1. Kishore, A.R., Latha, T.S., Niveditha, K.: Theoretic information on utility-privacy exchange in databases, Internation Journal of Engineering Research and Applications (IJERA), pp. 15–17 (2015)
2. Khan, A., Mellor, S., Berlin, E., Thompson, R., McNaney, R., Olivier, P., Ploz, T.: Byond activity recognition: Skill assessment from accelerometer data, UBICOMP (2015)
3. Bayat, A., Pomplun, M., Tran, D.A.: A study on human activity recognition using accelerometer data from smartphones. In: The 11th international conference on mobile systems and pervasive computing (MobiSPC) (2014)
4. Makhdoumi, A., Salamatian, S., Fawaz, N., Medard, M.: From the information Bottelenck to the privacy funnel, ArXiv e-prints (2014)
5. Braun, A., Garriga, G.: Consumer journey analytics in the context of data privacy and ethics. Digital Marketplaces Unleashed, pp. 663–674 (2017)
6. Williams, B.P., Hickman, R.M.: Privacy filtering of area description file prior to upload (2016)
7. Luo, C., Fylakis, A., Partala, J., Klakegg, S., Goncalves, J., Liang, K., Sppanen, T., Kostakos, V.: A data hiding approach for sensitive smartphone data. Ubicomp (2016)
8. Owusu, E., Han, J., Das, S., Perrig, A., Zhang, J.: Accessory: Password inference using accelerometers on smartphone (2012)
9. Pin Calmon, F.D.U., Fawaz, N.: Privacy against statistical inference (2012)
10. Han, J., Owusu, E., Nguyen, L.T., Perrig, A., Zhang, J.: Accomplice: Location inference using accelerometers on smartphones (2012)
11. Lu, H., Huang, J., Saha, T., Nachman, L.: Unobtrusive gait verification for mobile phones. In: Proceedings of the 2014 ACM International Symposium on Wearable Computers (ISWC 14), pp. 91–98 (2014)
12. Krumm, J.: A survey of computational location privacy. Pers. Ubiquit. Comput. **13**(6), 391–399 (2009)
13. Darakhshan, J.M.: Information-Theoretic Foundations of Differentail Privacy, Foundations and practice of security, pp. 374–381 (2012)
14. Lockhart, J.W., Weiss, G.M.: The benefits of personalized smartphone-based activity recognition models, In: Proc. SIAM international conference on data mining, society for industrial and applied mathematics, pp. 614–622 (2014)
15. Sankar, L., Rajagopalan, S., Poor, H.V.: A theory of privacy and utility in databases. arXiv e-prints, [Online]. Available: arXiv:1102.3751 (2011)
16. Leoni, D.: Non-interactive differential privacy: a Survey (2014)

17. Enev, M., Jung, J., Bo, L., Ren, X., Kohno, T.: SensorSift: Balancing sensor data privacy and utility in automated face understanding (2012)
18. Lu, M., Guo, Y., Meng, D., Li, C., Zhao, Y.: An information-aware privacy-preserving accelerometer data sharing. International conference of pioneering computer scientists, engineers and educators, pp. 425–432 (2017)
19. Fawaz, N., Salamatian, S., Pin Calmon, F.D., Bhanidipati, S.S., Oliveira, P.C., Taft, N.A., Kveton, B.: Privacy against inference attacks under mismatched prior (2016)
20. Tishby, N., Pereira, F.C., Bialek, W.: The information bottleneck method. In: Proceedings of the 37th annual allerton conference on communication, control and computing, pp. 368–377 (1999)
21. Slonim, N., Tishby, N.: Agglomerative information bottleneck. In: Proceedings of advances in neural information processing systems, pp. 617–623 (1999)
22. Ciaran, O'D.: Privacy in context: Privacy issues in ubiquitous computing applications (2008)
23. Williams, O., McSherry, F.: Probabilistic inference and differential privacy (2014)
24. Wang, K., Wang, P., Fu, W., Wong, C.-W.: Inferential or differential: Privacy laws dictate[J]. Eprint Arxiv, 2012, arXiv:1202.3686
25. Asoodeh, S., Alajaji, F., Linder, T.: Notes on information-theoretic privacy. In: Proceedings of 52nd annual allerton conference on communication, control, and computing, Monticello, IL, USA, pp. 1272–1278 (2014)
26. Shuang, S., Kamalika, C.: Composition properties of inferential privacy for time-series data, CoRR, arXiv:1707.02702 (2017)
27. Chakraborty, S., Shen, C., Raghavan, K.R., Shoukry, Y., Millar, M., Srivastava, M.: ipShield: A framework for enforcing context-aware privacy. In: 11th USENIX symposium on networked systems design and implementation (NSDI 14), pp. 143–156 (2014)
28. Bernecker, T., Graf, F., Kriegel, H.-P., Moennig, C., Dill, D., Tuermer, C.: Activity recognition on 3d accelerometer data (Technical Report) (2014)
29. He, Z., Cai, Z., Jiguo, Y.U.: Latent-data privacy preserving with customized data utility for social network data. IEEE Trans. Veh. Technol. Browse J. Mag. **67**(1), 665–673 (2018)
30. Zheng, Y., Wong, W.-K., Guan, X., Trost, S.: Physical activity recognition from accelerometer data using a multi-scale ensemble method. In: Proceedings of the 25th innovative application of artificial intelligence conference (2014)
31. Liang, Y., Zhou, X., Yu, Z., Guo, B., Yang, Y.: Energy efficient activity recognition based on low resolution accelerometer in smart phones. International conference on grid and pervasive computing: advances in grid and pervasive computing, pp. 122–136 (2012)