scientific reports

Check for updates

OPEN

A study on Pufferfish privacy algorithm based on Gaussian mixture models

Weisan Wu

In real-world scenarios, mixture models are frequently employed to fit complex data, demonstrating remarkable flexibility and efficacy. This paper introduces an innovative Pufferfish privacy algorithm based on Gaussian priors, specifically designed for Gaussian mixture models. By leveraging a sophisticated masking mechanism, the algorithm effectively safeguards data privacy. We derive the asymptotic expressions for the Kullback–Leibler (KL) divergence and mutual information between the original and noise-added private data, thereby providing a solid theoretical foundation for the privacy guarantees of the algorithm. Furthermore, we conduct a detailed analysis of the algorithm's computational complexity, ensuring its efficiency in practical applications. This research not only enriches the privacy protection strategies for mixture models but also offers new insights into the secure handling of complex data.

Keywords Pufferfish privacy, Gaussian mixture models, Taylor series, Differential privacy

The concept of mixture models originated in statistics, with its earliest application tracing back to Karl Pearson's study of biological data in 1894[1]. Pearson employed a mixture of Gaussian distributions to describe a complex biological dataset that could not be adequately modeled by a single normal distribution. The core idea behind mixture models is the assumption that data is generated from multiple distinct statistical distributions, and the combination of these distributions can better capture the complexity of the data.

Mixture models have seen widespread application in statistics and have gradually evolved into more sophisticated forms. The theory behind mixture models has further developed, encompassing more complex structures such as Bayesian mixture models and Hidden Markov Models. In recent years, mixture models have become a key tool in machine learning and artificial intelligence, with wide applications in clustering analysis, pattern recognition, natural language processing, anomaly detection, and more.

Using mixture models in differential privacy protection algorithms offers several advantages:

1. *Flexibility and accuracy*: In the processing of multimodal data, using mixture models, particularly Gaussian Mixture Models (GMM), can effectively handle multimodal data, where the data distribution may consist of multiple different sub-distributions. The combination of differential privacy techniques and mixture models can accurately describe the diversity and complexity of the data while maintaining data privacy[2].
2. *Improved privacy protection*: For example, in the effectiveness of noise addition, a common method in differential privacy is to add noise to data or query results. The application of mixture models can make noise addition more intelligent and targeted, thereby reducing its impact on data analysis results. For instance, adding noise independently to each sub-distribution can effectively prevent excessive noise from interfering with the useful information in the data[3].
3. *Enhanced data utility*: Mixture models can maximize the retention of statistical properties of the data while maintaining privacy. This capability allows algorithms that use differential privacy to protect privacy while still providing highly practical and accurate analysis results[4].
4. *Adaptation to different data structures*: Mixture models can flexibly adapt to the complex structures of datasets. This characteristic allows mixture models to be combined with differential privacy techniques to provide effective privacy protection in complex data scenarios without significantly compromising data utility[5].
5. *Reduced overfitting risk*: By introducing multiple sub-distributions to model the data, mixture models can avoid the overfitting issues that may arise from using a single model. When combined with differential privacy techniques, this approach can further reduce the risk of overfitting when dealing with sensitive data, ensuring the robustness of the generated model or analysis results.Overall, using mixture models to address

IRC-ISS, King Fahd University of Petroleum and Minerals, Dhahran 34463, Saudi Arabia. email: weisan.wu@kfupm.edu.sa

data privacy protection issues can maintain high accuracy and utility in data analysis while protecting data privacy. This combination is particularly effective in handling complex, multimodal, and high-dimensional data and provides an effective and flexible solution for protecting sensitive data.

While differential privacy is often used for its mathematical rigor, it has several limitations in practice, primarily reflected in the following aspects:

Differential privacy mainly focuses on a single type of secret, i.e., protecting the information of individual data points. In differential privacy, all data points are protected by the same mechanism, making it difficult to differentiate which information needs stronger protection or to apply different protection strategies to various types of information. When faced with complex data structures, such as multiple correlated attributes or multi-level data distributions, the application of differential privacy can become less flexible as it relies on a global privacy protection mechanism. differential privacy's protection mechanism usually does not consider domain knowledge, treating all data points equally in terms of privacy protection. This could lead to over-protection in some cases, thus affecting the utility of the data. differential privacy is typically designed as a general privacy protection mechanism suitable for a wide range of scenarios, which means it may not perform optimally in certain specific scenarios.

To address these challenges, we adopt a more flexible Pufferfish privacy approach. Compared to differential privacy, it offers the following advantages:

1. *Flexibility and customizability*: Pufferfish privacy allows users to customize protection strategies based on specific application needs. It can specify which information is considered secret and which assumptions should be applied to protect those secrets. This makes Pufferfish privacy capable of handling more diverse threat models, especially in situations where multiple types of secret information need protection.
2. *Handling complex data structures*: Pufferfish privacy can flexibly protect specific secrets within complex data structures without needing to homogenize the entire data structure. It can assign different protection measures to different secrets, thereby increasing the effectiveness and efficiency of protection.
3. *Encoding domain knowledge*: Pufferfish privacy allows domain knowledge to be encoded into the privacy protection strategy. Users can customize protection strategies based on the characteristics of the data and the specific application context, achieving more precise privacy protection. This capability is particularly useful in applications where a balance between privacy protection and data utility is necessary.
4. *Adaptation to specific application scenarios*: Pufferfish privacy can be adjusted according to specific application scenarios and protection needs, making it more adaptable in certain situations. For instance, in applications involving multiple data sources or multi-level data, Pufferfish privacy can offer more flexible privacy protection.
5. *Better privacy-utility trade-off*: Since Pufferfish privacy can employ different protection strategies for different secrets, it can maintain high levels of privacy protection while maximizing data utility. This allows Pufferfish privacy to provide a better privacy-utility trade-off in some cases.In this paper, we address the challenges of differential privacy by designing a Pufferfish privacy algorithm based on mixture models. Within the masking mechanism of the GMM, we provide a polynomial approximation algorithm to measure the distance between the original and the noise-added data, and we prove that it satisfies Pufferfish privacy guarantees.

The structure of this paper is as follows: first, Sect. 2 presents the theory of mixture models and related privacy concepts. Next, Sect. 3 provides the asymptotic expression for the information entropy of the mixture model masking algorithm. Section 4 then presents the asymptotic formula for the mutual information. Finally, we conclude with a discussion and outlook on future research directions.

Preliminaries

Finite Gaussian mixture models

Finite mixture models can achieve high accuracy when modeling complex data. Researchers can construct mixture models with arbitrary component distributions based on the structure of the data. However, Gaussian distributions are a focal point of research as components in mixture models due to their symmetry, rotational invariance, and other elegant mathematical properties. Below, we present the definition of a Gaussian mixture model.

Definition 1 [6] We called x is a M order Gaussian mixture models if the probability density function of x follows that

$$p(x) = \sum_{i=1}^{M} w_i N(x|\mu_i, \Sigma_i), \tag{1}$$

where w_i is mixing weight satisfying $w_i \geq 0, \sum_{i=1}^{M} w_i = 1$; $\mu_i \in \mathbb{R}^d$ and $\Sigma_i \in \mathbb{R}^{d\times d}$ are the mean and covariance parameter of the i-th component, i.e. $N(x|\mu_i, \Sigma_i)$. For conveniently, we write the parameters of i-th component as $\theta_i = (w_i, \mu_i, \Sigma_i)$, the parameter vector as $\theta = (\theta_1, \cdots, \theta_M)$.
We adopt a more general notation,

$$\mathscr{G}(d, D_i) := \left\{ \sum_{i=1}^{D_i} w_i N(x|\mu_i, \Sigma_i), w_i \geq 0, \sum_{i=1}^{D_i} w_i = 1 \right\}. \tag{2}$$

where d is the dimension of data set, D_i is the number of the components. We employ the following distance measure to quantify the difference between two distinct mixture distributions[7].

$$\mathrm{dist}_{\mathrm{GMM}}\left(\mathscr{G}(d,D_i),\mathscr{G}(d,D_j)\right)=\min_{\pi}\max_{i\in[D_i]}\left\{|w_i-w'_{\pi(i)}|,d_{\mathrm{TV}}\left(N(\mu_i,\Sigma_i),N(\mu'_{\pi(i)},\Sigma'_{\pi(i)})\right)\right\}. \tag{3}$$

The foundations of differential privacy

Definition 2 ($(\epsilon,\delta)-DP$[8]) A mechanism $\mathscr{M}$ satisfied (ϵ,δ) differential privacy for some $\epsilon,\delta>0$, if for any neighboring data set x,x' i.e. only different 1 element between x and x', for any event $\mathscr{A}\subset\mathscr{Y}$ we have

$$\mathbb{P}(\mathscr{M}(x)\in\mathscr{A})\leq e^{\epsilon}\mathbb{P}(\mathscr{M}(x')\in\mathscr{A})+\delta. \tag{4}$$

Specially, if $\delta=0$, we called the $(\epsilon,0)$-DP as a pure DP.

Definition 3 ((ϵ,δ)-indistinguishbale[8]) A mechanism $\mathscr{M}$ satisfied (ϵ,δ)-indistinguishbale for some $\epsilon,\delta>0$, if for any neighboring data set x,x'

$$\mathbb{P}(\mathscr{M}(x)\in\mathscr{A})\leq e^{\epsilon}\mathbb{P}(\mathscr{M}(x')\in\mathscr{A})+\delta, \tag{5}$$

and

$$\mathbb{P}(\mathscr{M}(x')\in\mathscr{A})\leq e^{\epsilon}\mathbb{P}(\mathscr{M}(x)\in\mathscr{A})+\delta. \tag{6}$$

Definition 4 (ϵ-MIDP[8]) A mechanism $\mathscr{M}$ satisfied ϵ-Mutual information differential privacy (MIDP), if

$$\sup_{P_x\in\Theta,g\in\mathscr{G},w\in\mathscr{W}:g\sim w} I\left(g(x);\mathscr{M}(x)\mid w(x)\right)\leq\epsilon. \tag{7}$$

Definition 5 (Pufferfish privacy[8]) Fix $\epsilon,\delta>0$. A random mechanism $\mathscr{M}:\mathscr{X}^{n\times k}\to\mathscr{Y}$ is (ϵ,δ)-private in the pufferfish framework $(\mathscr{S},\mathscr{Q},\Theta)$ if for all $f(x)\in\Theta$, $(\mathscr{R},\mathscr{T})\in\mathscr{Q}$ with $f(x|\mathscr{R})$, $f(x|\mathscr{T})>0$ and $\mathscr{A}\subset\mathscr{Y}$ measurable, we have

$$\mathbb{P}(\mathscr{M}(x)\in\mathscr{A}\mid\mathscr{R})\leq e^{\epsilon}\mathbb{P}(\mathscr{M}(x)\in\mathscr{A}\mid\mathscr{T})+\delta. \tag{8}$$

Definition 6 (Rényi Pufferfish privacy, RPP[9]). Let $\alpha>1$ and $\epsilon\geq0$. A random mechanism $\mathscr{M}$ is said to be (α,ϵ)-Rényi Pufferfish private in a framework $(\mathscr{S},\mathscr{Q},\Theta)$ if for all $\mathscr{R},\mathscr{T}\in\Theta$, we have

$$D_{\alpha}\left(P(\mathscr{M}(X)|\mathscr{R}),P(\mathscr{M}(X)|\mathscr{T})\right)\leq\epsilon, \tag{9}$$

where $D_{\alpha}(\mu,\nu)=\frac{1}{\alpha-1}\log\mathbb{E}_{x\sim\nu}\left[(\frac{\mu(x)}{\nu(x)})^{\alpha}\right]$ is the Renyi divergence of order α.

Pufferfish privacy algorithm

First, we apply the Nuradha's algorithm (Algorithm 1) to perform privacy masking on each mixture component of the original data[8].

Next, we approximate the Kullback–Leibler (KL) divergence $D_{KL}(P\parallel Q)$ between the original data and the two masked mixture models $P(x)=\sum_{i=1}^{D_i}w_iN(x|\mu_i,\Sigma_i)$, and $Q(x)=\sum_{i=1}^{D_i}\hat{w}_iN(x|\hat{\mu}_i,\hat{\Sigma}_i)$.

Finally, we provide an asymptotic upper bound on the mutual information between the two mixture models using the asymptotic KL divergence.

Taylor and Legendre entropy approximations

Because of the fact

$$I(X_i;Y|X^{-i})=\mathbb{E}\left[D_{KL}(P\parallel Q)\right]=\mathbb{E}_P[\log P]-\mathbb{E}_P[\log Q], \tag{10}$$

we consider to use Taylor seriese to approximate KL divergence.

Input: GMM given by $\{(w_i, \mu_i, \Sigma_i)\}_{i\in[k]}$ and parameters $\eta_w^{(0)}, \eta_\mu^{(0)}, \eta_\Sigma^{(0)} > 0$
FOR: $t = 1$ to T
FOR: $i = 1$ to D_i
1. **Computing**: $\max(0, w_i^{(t)} + \eta_w^{(t)} g), g \sim \mathcal{N}(0,1)$
2. **Computing**: $\mu_i^{(t)} + \eta_\mu^{(t)} g, g \sim N(0, \Sigma)$
3. **Computing**: $\Sigma_i^{1/2^{(t)}}(I_d + \eta_\Sigma^{(t)} G)(I_d + \eta_\Sigma^{(t)} G)^\top \Sigma_i^{1/2^{(t)}}$, where $G \in \mathbb{R}^{d\times d}$ be a matrix with independent $N(0,1)$ entries.
4. **Computing**: $\{(\hat{w}_i^{(t)}, \hat{\mu}_i^{(t)}, \hat{\Sigma}_i^{(t)})\} \leftarrow \{\mathscr{B}_{\text{COMP}}(w_{\sigma(i)}, \mu_{\sigma(i)}, \Sigma_{\sigma(i)})\}$
5. **Computing**: $\hat{w}_i^{(t+1)} \leftarrow \hat{w}_i^{(t)} / \sum_{i\in[D_i]} \hat{w}_i^{(t)}$
6. **Return**: $\{(\hat{w}_i^{(t+1)}, \hat{\mu}_i^{(t+1)}, \hat{\Sigma}_i^{(t+1)})\}$
ENDFOR
6. **Return**: $\{(\hat{w}^{(T)}, \hat{\mu}^{(T)}, \hat{\Sigma}_i^{(T)})\}$
ENDFOR

Algorithm 1. Perturbed Gradient EM Algorithm

Definition 7 For a function *f*(*x*), its *n*-th order Taylor polynomial at the point x_0 is

$$T_{f,n}(x_0) = \sum_{i=0}^{n} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n, \tag{11}$$

where $f^{(n)}(x_0)$ is the *n*-th derivative of *f* at the point x_0.
Huber et al. use Taylor series expansion to get approximation of GMM differential entropy[10] as

$$H(p(x)) = -\sum_{i=1}^{m} w_i \sum_{n=0}^{\infty} \frac{h^{(n)}(\mu_i)}{n!} \mathbb{E}_{x\sim N(x|\mu_i,\Sigma_i)}\left[(x - \mu_i)^n\right].$$

We now provide the Taylor series expansion for the expectation of the logarithmic function.

Lemma 3.1 *Let* $P(x) = \sum_{i=1}^{D_i} w_i N(x|\mu_i, \Sigma_i)$, *and let masked data distribution as* $Q(x) = \sum_{i=1}^{D_i} \hat{w}_i N(x|\hat{\mu}_i, \hat{\Sigma}_i)$ *are different two Gaussian mixture models, then the k-th order moment of P*(*x*) *can be write as follows*:

$$\mathbb{E}_P[P(x)^k] = \sum_{j_1+\cdots+j_{D_i}=k} \binom{k}{j_1, \ldots, j_{D_i}} \sum_{i=1}^{D_i} w_i \frac{N(0|\mu_i, \Sigma_i)}{N(0|\mu, \Sigma)} \prod_{t=1}^{D_i} (w_t N(0|\mu_t, \Sigma_t))^{j_t}, \tag{12}$$

where $\Sigma = \left(\sum_{i=1}^{D_i} \Sigma_i^{-1} + \sum_{t=1}^{D_i} \frac{1}{j_t} \Sigma_t^{-1}\right)^{-1}$ *and* $\mu = \Sigma\left(\sum_{i=1}^{D_i} \Sigma_i^{-1}\mu_i + \sum_{t=1}^{D_i} j_t \Sigma_t^{-1}\mu_t\right)$. *The k-th order moment of Q*(*x*) *can be write as follows*:

$$\mathbb{E}_P[Q(x)^k] = \sum_{j_1+\cdots+j_{D_i}=k} \binom{k}{j_1, \ldots, j_{D_i}} \sum_{i=1}^{D_i} \hat{w}_i \frac{N(0|\hat{\mu}_i, \hat{\Sigma}_i)}{N(0|\mu, \Sigma)} \prod_{t=1}^{D_j} (w_t N(0|\mu_t, \Sigma_t))^{l_t}, \tag{13}$$

where $\Sigma = \left(\sum_{i=1}^{D_i} \hat{\Sigma}_i^{-1} + \sum_{t=1}^{D_i} \frac{1}{j_t} \hat{\Sigma}_t^{-1}\right)^{-1}$ *and* $\mu = \Sigma\left(\sum_{i=1}^{D_i} \hat{\Sigma}_i^{-1}\hat{\mu}_i + \sum_{t=1}^{D_i} j_t \hat{\Sigma}_t^{-1}\hat{\mu}_t\right)$.
Dahlke and Pacheco[11] *respectively obtained the Taylor series approximation,*

$$\hat{H}_{N,a}^T(P(x)) = -\log(a) - \sum_{n=1}^{N} \frac{(-1)^{n-1}}{na^n} \sum_{k=0}^{n} \binom{n}{k} (-a)^{n-k} \mathbb{E}_P[P(x)^k], \tag{14}$$

and

$$\hat{H}_{N,a}^T(Q(x)) = -\log(a) - \sum_{n=1}^{N} \frac{(-1)^{n-1}}{na^n} \sum_{k=0}^{n} \binom{n}{k} (-a)^{n-k} \mathbb{E}_P[Q(x)^k], \tag{15}$$

and the Legendre series approximation,

$$\hat{H}^{L}_{N,a}(P(x)) = -\log(a) - \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[P(x)^k]\right), \quad (16)$$

and

$$\hat{H}^{L}_{N,a}(Q(x)) = -\log(a) - \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[Q(x)^k]\right), \quad (17)$$

for the Gaussian mixture distributions P and Q.

Gaussian mixture models Pufferfish privacy

Next, we present the improved formula for calculating mutual information, which in turn provides the guarantees for our privacy algorithm.

First, we present the following lemma[11], which demonstrates that the limits of the entropy expressions obtained through two series approximations exist.

Lemma 4.1 *Let* $P(x) = \sum_{i=1}^{D_i} w_i N(x|\mu_i, \Sigma_i)$, *and let the masked data distribution be denoted as* $Q(x) = \sum_{i=1}^{D_i} \hat{w}_i N(x|\hat{\mu}_i, \hat{\Sigma}_i)$ *are different two Gaussian mixture models. When* $a > 1/2\max\{P(x), Q(x)\}$, *we have*

$$\lim_{N\to\infty} \hat{H}^{T}_{N,a}(P(x)) = H(P(x)), \quad (18)$$

and

$$\lim_{N\to\infty} \hat{H}^{T}_{N,a}(Q(x)) = H(Q(x)). \quad (19)$$

When $a > \max\{P(x), Q(x)\}$, *we have*

$$\lim_{N\to\infty} \hat{H}^{L}_{N,a}(P(x)) = H(P(x)), \quad (20)$$

and

$$\lim_{N\to\infty} \hat{H}^{L}_{N,a}(Q(x)) = H(Q(x)), \quad (21)$$

Theorem 4.2 *Let* $P(x) = \sum_{i=1}^{D_i} w_i N(x|\mu_i, \Sigma_i)$, *and let the masked data distribution be denoted as* $Q(x) = \sum_{i=1}^{D_i} \hat{w}_i N(x|\hat{\mu}_i, \hat{\Sigma}_i)$ *are different two Gaussian mixture models, then we get*:

$$\begin{aligned} & I(x_i; y|x^{-i}) \\ &= \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[Q(x)^k]\right) \\ &\quad - \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[P(x)^k]\right), \end{aligned}$$

Proof When $a > 1/2\max\{P(x), Q(x)\}$, we have

$$\begin{aligned} I(x_i; y|x^{-i}) &= \mathbb{E}\left[D_{KL}\left(P \parallel Q\right)\right] \\ &= \mathbb{E}_P[\log P] - \mathbb{E}_P[\log Q] \\ &= \hat{H}^{T}_{N,a}(P(x)) - \hat{H}^{T}_{N,a}(Q(x)) \\ &= \sum_{n=1}^{N}\frac{(-1)^{n-1}}{na^n}\sum_{k=0}^{n}\binom{n}{k}(-a)^{n-k}\mathbb{E}_P[Q(x)^k] - \sum_{n=1}^{N}\frac{(-1)^{n-1}}{na^n}\sum_{k=0}^{n}\binom{n}{k}(-a)^{n-k}\mathbb{E}_P[P(x)^k], \end{aligned}$$

Furthermore, when $a > 1/2\max\{P(x), Q(x)\}$, we have

$$
\begin{aligned}
I(x_i; y|x^{-i}) &= \mathbb{E}\left[D_{KL}\left(P \parallel Q\right)\right] \\
&= \mathbb{E}_P[\log P] - \mathbb{E}_P[\log Q] \\
&= \hat{H}^L_{N,a}(P(x)) - \hat{H}^L_{N,a}(Q(x)) \\
&= \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[Q(x)^k]\right) \\
&- \sum_{n=0}^{N}(2n+1)\sum_{j=0}^{n}\frac{(-1)^{n+j}(n+j)!((j+1)\log(a)-1)}{(n-j)!(j+1)!^2}L_{[0,a],n}\left(\mathbb{E}_P[P(x)^k]\right).
\end{aligned}
$$

□

In summary, we derive the Pufferfish privacy theorem for mixture models with a Gaussian prior[12].

Theorem 4.3 *Fix $\epsilon > 0$, let $f : \mathscr{X}^{n\times k} \to \mathbb{R}^d$ and consider a random mechanism $Q(x) := P(x) + Z_G$, where $Z_G \sim N(0, \sigma^2 I_d), \sigma > 0$. If*

$$
\sigma^2 \geq \sup_{P_x\in\mathscr{P}, Q_x\in\mathscr{T}} \frac{\sum_{m,l} w^*_{m,l}\left[\sum_{v=1}^{d}(|\mu_{i,m}(v) - \mu_{j,l}(v)^2| + \tau^*(\delta)^2|\sigma_{i,m}(v) - \sigma_{j,l}(v)|^2)\right]}{d(e^{2\epsilon/d}-1)},
$$

then M_G is ϵ-MIPP, where $\tau^*(\delta) = \min\{\tau : P(Z_G > \tau) \leq \delta/2\}$.

Empirical testing framework for Pufferfish privacy with Gaussian mixture models

Given the inherent complexity of the proposed model, we have not conducted direct experiments within the scope of this study. However, it is important to note that while we did not perform empirical experiments, we have outlined several algorithmic frameworks and methodologies that could be effectively used to empirically validate our theoretical results in the future. These frameworks provide a strong foundation for potential empirical exploration of our model and demonstrate its applicability in practical contexts.

- Potential Experimental Approaches and Algorithmic Frameworks. To provide empirical validation of the Pufferfish privacy mechanism based on GMMs, we propose leveraging synthetic data experiments. Simulated datasets offer a controlled environment that allows systematic evaluation of privacy guarantees and utility trade-offs by manipulating parameters like dimensionality, number of components, and mixing weights. This approach has been demonstrated successfully in privacy research, as highlighted by Diao et al.[3] in their study on local differential privacy for GMMs.
- Numerical Approximations and Computational Techniques. Apart from simulated experiments, another promising approach involves using numerical techniques such as Monte Carlo integration or Gaussian quadrature to approximate key performance metrics, including privacy loss (ϵ) and utility metrics. Nuradha and Goldfeld[8] have employed similar techniques in their information-theoretic analysis of Pufferfish privacy, providing a computational means to validate privacy mechanisms without needing a direct experiment. This numerical framework could be adapted to assess the privacy guarantees and utility trade-offs in our algorithm, providing empirical support for our theoretical results.
- Cross-Validation and Comparative Benchmarking. To further strengthen empirical support, we can employ cross-validation and benchmarking against other privacy-preserving algorithms, such as Gaussian differential privacy[5]. Such benchmarking would help illustrate the privacy-utility trade-offs offered by our Pufferfish privacy mechanism relative to well-established differential privacy techniques. The comparative analysis would be valuable for establishing the practical efficiency and relevance of our proposed method.
- Case Study Evaluation. Another potential empirical approach is to apply our privacy mechanism within domain-specific contexts, such as healthcare or finance, where privacy is critical. Kamath et al.[4] used similar case studies to demonstrate privacy guarantees in healthcare data analysis, providing a nuanced perspective on privacy-utility trade-offs in real-world applications. A case study could thus serve as an effective method for validating the practical application of our privacy mechanism, specifically by testing how well it balances privacy preservation with data utility.
- Future Directions for Empirical Work. In conclusion, while experimental work has not been undertaken in the present study due to the complexity of the model, the aforementioned frameworks illustrate feasible approaches for future empirical validation. These methodologies lay the groundwork for validating our theoretical contributions and exploring their practical implications comprehensively. Future work could involve implementing these frameworks to empirically demonstrate the effectiveness of our Pufferfish privacy mechanism in both synthetic and real-world data settings.

Conclusion

In this paper, we investigated the privacy protection problem for mixture models and proposed an effective Pufferfish privacy algorithm. By masking each component in the Gaussian mixture, we protected the privacy of the component distributions. Furthermore, we demonstrated how to calculate the mutual information between the distributions before and after privacy computation using two series approximations. Reducing computational complexity is one of our future research directions, as well as ensuring alignment among components after the masking mechanism, which remains a crucial issue.

Data availability

This study did not involve the generation or analysis of any datasets.

Received: 26 September 2024; Accepted: 19 December 2024
Published online: 06 January 2025

References

1. Pearson, K. Contributions to the mathematical theory of evolution. *Phil. Trans. R. Soc. A* **186**, 343–414 (1894).
2. Wu, Y. *et al.* Differentially private density estimation via gaussian mixtures model. *2016 IEEE/ACM 24th International Symposium on Quality of Service (IWQoS)* 1–6 (2016).
3. Diao, X., Yang, W., Wang, S., Huang, L. & Xu, Y. Privgmm: Probability density estimation with local differential privacy. In *International Conference on Database Systems for Advanced Applications* (2020).
4. Kamath, G., Sheffet, O., Singhal, V. & Ullman, J. Differentially private algorithms for learning mixtures of separated gaussians. *2020 Information Theory and Applications Workshop (ITA)* 1–62 (2019).
5. Dong, J., Roth, A. & Su, W. J. Gaussian di—erential privacy (2019).
6. Chen, J. & Li, P. Hypothesis test for normal mixture models: The em approach. arXiv: Statistics Theory (2009).
7. Arbas, J., Ashtiani, H. & Liaw, C. Polynomial time and private learning of unbounded gaussian mixture models. *ArXiv***abs/2303.04288** (2023).
8. Nuradha, T. & Goldfeld, Z. Pufferfish privacy: An information-theoretic study. *IEEE Trans. Inf. Theory* **69**, 7336–7356 (2022).
9. Pierquin, C., Bellet, A., Tommasi, M. & Boussard, M. Rényi pufferfish privacy: General additive noise mechanisms and privacy amplification by iteration. *ArXiv***abs/2312.13985** (2023).
10. Huber, M. F., Bailey, T. & H. Durrant-Whyte, e. a. On entropy approximation for gaussian mixture random vectors. In *2008 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems* (2008).
11. Dahlke, C. & Pacheco, J. On convergence of polynomial approximations to the gaussian mixture entropy. In *Neural Information Processing Systems* (2023).
12. Ding, N. Approximation of pufferfish privacy for Gaussian priors. *IEEE Trans. Inf. Forensics Sec.* **19**, 5630–5640 (2024).

Acknowledgements

This research was funded by the Natural Science Foundation of Jilin Province (Grant Number YDZ-J202401390ZYTS), the Education Department of Jilin Province (Grant Number JJKH20230021KJ), the Education Department of Jilin Province (Grant Number JJKH20230020CY), Ministry of Education Chunhui plan project China (Grant Number HZKY20220376).

Author contributions

Weisan Wu is the only author and wrote the manuscript text.

Declarations

Competing interests

The authors declare no conflict of interest.

Additional information

Correspondence and requests for materials should be addressed to W.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.