Efficient Non-parametric Neural Density Estimation and Its Application to Outlier and Anomaly Detection

Joseph A. Gallego-Mejia
Universidad Nacional de Colombia
jagallegom@unal.edu.co

Abstract

The main goal of this thesis is to develop efficient non-parametric density estimation methods that can be integrated with deep learning architectures, for instance, convolutional neural networks and transformers. Density estimation methods can be applied to different problems in statistics and machine learning. They may be used to solve tasks such as anomaly detection, generative models, semi-supervised learning, compression, text-to-speech, among others. The present work will mainly focus on the application of the method in anomaly and outlier detection tasks such as medical anomaly detection, fraud detection, video surveillance, time series anomaly detection, industrial damage detection, among others. A recent approach to non-parametric density estimation is neural density estimation. One advantage of these methods is that they can be integrated with deep learning architectures and trained using gradient descent. Most of these methods are based on neural network implementations of normalizing flows which transform an original simpler distribution to a more complex one. The approach of this thesis is based on a different idea that combines random Fourier features with density matrices to estimate the underlying distribution function. The method can be seen as an approximation of the popular kernel density estimation method but without the inherent computational cost.

To achieve all objectives, the methodology is selected according to work packages (WP) depending on each objective. Each work package has tasks and deliverables that must be submitted on time to ensure the success of the research.

Research Questions:

- How to design a nonparametric density estimation method that is efficient in terms of time and space?
- How to design a method capable of integrating with other deep learning methods for nonparametric density estimation?
- Is the approximation given by the proposed method better in terms of efficiency than the state-of-the-art approximation methods for kernel density estimation?
- How to design methods for anomaly and outlier detection using nonparametric density estimation?

Expected Thesis Contributions

- A main algorithm for density estimation using density matrices and Fourier features published in top journals.
- Four novel anomaly detection algorithms based on density matrices and Fourier features presented at leading machine learning conferences and in Q1 research journals.
- Already submitted 4 conference papers and 4 journal articles. I will submit 5 new conference papers and 4 new journal articles.

Work Done

In (Gallego, González, and Nasraoui 2021), we show that least-square estimation (mean calculation) in a reproducing kernel Hilbert space (RKHS) $\mathcal{F}$ corresponds to different M-estimators in the original space depending on the kernel function associated with $\mathcal{F}$. In particular, we present a proof of the correspondence of mean estimation in an RKHS for the Gaussian kernel with robust estimation in the original space performed with the Welsch M-estimator. This result is generalized to other types of M-estimators. This generalization facilitates the definition of new robust kernels associated to Huber, Tukey, Cauchy and Andrews M-estimators. The new kernels were empirically evaluated in different clustering tasks where state-of-the-art robust clustering methods are compared to kernel-based clustering using robust kernels. The results show that some robust kernels perform on a par with the best state-of-the-art robust clustering methods.

In (González et al. 2022), we proposed a new algorithm based on density matrix and random fourier features for neural density estimation called density matrix - kernel density estimation (DMKDE). A density matrix describes the statistical state of a quantum system. It is a powerful formalism to represent both the quantum and classical uncertainty of quantum systems and to express different statistical operations such as measurement, system combination and expectations as linear algebra operations. In this paper, we explore how density matrices can be used as a building block for machine learning models exploiting their ability to straightforwardly combine linear algebra and probability. One of the main results of the paper is to show that density matrices coupled with random Fourier features could approximate arbitrary probability distributions over Rn. Based on this finding the paper builds different models for density estimation, classification and regression. These models are differentiable, so it is possible to integrate them with other differentiable components, such as deep learning architectures and to learn their parameters using gradient-based optimization. In addition, the paper presents optimization-less training strategies based on estimation and model averaging. The models are evaluated in benchmark tasks and the results are reported and discussed.

In (Gallego, Osorio, and Gonzalez 2022), we systematically evaluate the novel DMKDE algorithm and compare it with other state-of-the-art fast procedures for approximating the kernel density estimation method on different synthetic data sets. Our experimental results show that DMKDE is on par with its competitors for computing density estimates and advantages are shown when performed on high-dimensional data. We have made all the code available as an open source software repository.

In (Gallego M and González 2022), we propose a method for neural density estimation that can be seen as a type of kernel density estimation, but without the high prediction computational complexity. The method is based on density matrices, a formalism used in quantum mechanics, and adaptive Fourier features. The method can be trained without optimization, but it could be also integrated with deep learning architectures and trained using gradient descent. Thus, it could be seen as a form of neural density estimation method.

In (Bustos-Brinez, Gallego-Mejia, and González 2022), we present a novel density estimation method for anomaly detection using density matrices (a powerful mathematical formalism from quantum mechanics) and Fourier features. The method can be seen as an efficient approximation of Kernel Density Estimation (KDE). A systematic comparison of the proposed method with eleven state-of-the-art anomaly detection methods on various data sets is presented, showing competitive performance on different benchmark datasets. The method is trained efficiently and does not require optimization, but it can be optimized using gradient descent. The prediction phase complexity of the proposed algorithm is constant relative to the training data size, and it performs well in high dimensional data.

In (Gallego-Mejia, Bustos-Brinez, and González 2022), we present an anomaly detection model that combines the strong statistical foundation of density-estimation-based anomaly detection methods with the representation-learning ability of deep-learning models. The method combines an autoencoder, for learning a low-dimensional representation of the data, with a density-estimation model based on random Fourier features and density matrices in an end-to-end architecture that can be trained using gradient-based optimization techniques. The method predicts a degree of normality for new samples based on the estimated density. A systematic experimental evaluation was performed on different benchmark datasets.

Streaming anomaly detection has gained attention with modern electronic devices and online transactions. In this context, data flows as a stream and it is impractical to store every data point. Classical and deep anomaly detection methods are not designed to cope with conceptual drift and continuous learning. State-of-the-art stream anomaly detection methods rely on fixed memory that possibly forgets old patterns. In (Gallego-Mejia, Bustos-Brinez, and Gonzalez 2022), we present a new incremental quantum measurement anomaly detection method that relies on Fourier features and density matrices. It can process potentially endless data and its update complexity is constant $O(1)$. A systematic evaluation against 12 state-of-the-art streaming anomaly detection algorithms and using 12 streaming data sets is presented.

Date	Milestone
Feb, 2020	Robust kernels for robust location estimation
Dec, 2020	Learning with density matrices and random features
Sep, 2021	Thesis proposal acceptance
Jul, 2022	Anomaly detection paper submissions
Feb, 2023	AAAI conference
Jul, 2023	Variational Anomaly Detection Algorithms
Dec, 2023	PhD Thesis Dissertation

Table 1: Brief Timeline

In (Useche et al. 2022), we present a novel classical-quantum anomaly detection model based on density estimation and the expected values of density matrices. The core subroutine of the proposed method is a qubit-based quantum protocol whose task is to compute the expected value of a density matrix, being flexible enough to support the use of pure and mixed states approaches. The anomaly detection model is tested with pure and mixed states on a synthetic data set for density estimation and on a widely used real-life data set for anomaly detection.

References

Bustos-Brinez, O.; Gallego-Mejia, J.; and González, F. A. 2022. AD-DMKDE: Anomaly Detection through Density Matrices and Fourier Features. *arXiv preprint arXiv:2210.14796*.

Gallego, J. A.; González, F. A.; and Nasraoui, O. 2021. Robust kernels for robust location estimation. *Neurocomputing*, 429: 174–186.

Gallego, J. A.; Osorio, J. F.; and Gonzalez, F. A. 2022. Fast Kernel Density Estimation with Density Matrices and Random Fourier Features. In *Ibero-American Conference on Artificial Intelligence*, 160–172. Springer.

Gallego M, J. A.; and González, F. A. 2022. Quantum Adaptive Fourier Features for Neural Density Estimation. *arXiv e-prints*, arXiv–2208.

Gallego-Mejia, J.; Bustos-Brinez, O.; and Gonzalez, F. 2022. InQMAD: Incremental Quantum Measurement Anomaly Detection. *arXiv preprint arXiv:2210.05061*.

Gallego-Mejia, J.; Bustos-Brinez, O.; and González, F. A. 2022. LEAN-DMKDE: Quantum Latent Density Estimation for Anomaly Detection. *arXiv preprint arXiv:2211.08525*.

González, F. A.; Gallego, A.; Toledo-Cortés, S.; and Vargas-Calderón, V. 2022. Learning with density matrices and random features. *Quantum Machine Intelligence*, 4(2): 1–17.

Useche, D. H.; Bustos-Brinez, O. A.; Gallego, J. A.; and González, F. A. 2022. Computing expectation values of adaptive Fourier density matrices for quantum anomaly detection in NISQ devices. In *arXiv: 2201.10006*.