

A New Anomaly Detection Algorithm based on Quantum Mechanics

Hao Huang*, Hong Qin*, Shinjae Yoo[†] and Dantong Yu[†]

*Department of Computer Science, Stony Brook University (SUNY)

Email: haohuang@cs.stonybrook.edu, qin@cs.stonybrook.edu

[†]Computational Science Center, Brookhaven National Laboratory

Email: sjyoo@bnl.gov, dtyu@bnl.gov

Abstract—The primary originality of this paper lies at the fact that we have made the first attempt to apply quantum mechanics theory to anomaly (outlier) detection in high-dimensional datasets for data mining. We propose Fermi Density Descriptor (FDD) which represents the probability of measuring a fermion at a specific location for anomaly detection. We also quantify and examine different Laplacian normalization effects and choose the best one for anomaly detection. Both theoretical proof and quantitative experiments demonstrate that our proposed FDD is substantially more discriminative and robust than the commonly-used algorithms.

Keywords—Anomaly Detection; Quantum Mechanics;

I. INTRODUCTION AND FUNDAMENTAL IDEA

Our novel unsupervised method in this paper, called Fermi Density Descriptor (FDD), computes an accurate measurement of anomalousness based on quantum mechanics theory [8]. Thus, it has a sound physics theory background. Compared with the existing algorithms, this new algorithm is more reliable than direct statistics based on Euclidean distances or attribute distribution in the original space, and hence more informative in a locally adaptive neighborhood.

A. Related Work

In anomaly detection, the local density of each instance is always the main concern. However, it is far from trivial to measure such value. Many algorithms try to seek a better approximation by means of statistics based on Euclidean distance [15] [3] [6] or attribute distribution in the original data space [12] [11] [18]. Yet, the density of local neighborhood is not as straightforward as pair-wise distance or attribute distribution in the original space. A few techniques [1] adopt spectral anomaly detection techniques with manifold reconstruction in which the anomalous instances can be easily identified [4]. However, the existing algorithms are based on techniques such as isometric feature mapping (ISM) or locally linear embeddings (LLE) [1] which are highly sensitive to local data perturbation [19]. Therefore the subsequent anomaly detection algorithms may fail miserably.

B. Motivation and Key Intuition

Different from common techniques that aim to preserve local embedding for clustering, here manifold reconstruction concentrates on magnifying density difference between

anomalies and normal instances. Consequently, anomalous instances will be more singular and distinctive. The dense neighborhood will become even denser. Although this type of mapping is non-isometric and the original distribution is changed, it is of central interest in anomaly detection, as it becomes more sensitive to local density distribution.

After projecting from the original to the manifold space, we apply quantum mechanics for anomaly detection. For unconstrained movement of free particles, a condensed point set is equivalent to a singularity. A steep density gradient in a data region can render the effect of trapping energy and preventing particles escaping from such a singularity. If the size of a point set (number of instances) is large, the average probability a particle stays within a certain point is low. On the other hand, if this point set represents anomalies, the probability for a particle to stay at each instance is high.

C. Contribution

This paper articulates a novel unsupervised anomaly detection algorithm with the following salient contributions:

- 1) It is the first to quantitatively characterize local density information based on quantum mechanics theory (Section II) which supplies rigorous probabilistic explanation from the perspective of modern physics theory.
- 2) It firstly analyzes different Laplacian normalization effects (Section III) with the goal of anomaly detection. It demonstrates that NN normalization is the best choice for anomaly detection.
- 3) We evaluate FDD with several closely-related baseline algorithms on a number of benchmark datasets (Section V). Our algorithm shows both better average performance and more stable results.

II. FERMI DENSITY DESCRIPTOR

A. Schrödinger Equation and Wave Function

Quantum mechanics [8] is a mathematical machinery for predicting the behavior of microscopic particles. Anomalous instances can be treated as regions of low density that correspond to the aggregation area of maximal free energy, and such area is easier to trap particles. The Schrödinger equation is the key equation in quantum mechanics, which describes how the quantum state of a physical system

changes with time. If we ignore the potential energy, it is directly associated with Laplace operator L as follows:

$$i \frac{\partial \psi}{\partial t}(x, t) = L\psi(x, t), \quad (1)$$

where ψ is the space-time wave function of the quantum system, i is the imaginary unit, x is the position and t is time. The mod square $|\psi(x, t)|^2$ depicts the probability density of a particle at position x at time t . Assume the Laplace spectrum has no repeated eigenvalues, and $L = \phi' \lambda \phi$ (ϕ and λ are the eigenvectors and eigenvalues of L), the wave function can be expressed in the spectral domain as

$$\psi(x, t) = \sum_{k=1}^{\infty} e^{i\lambda_k t} \phi_k(x) f(\lambda_k), \quad (2)$$

where $f(\lambda)$ is the energy distribution since in spectral domain, eigenvalue λ is equivalent to energy level E [8]. So $f(\lambda)$ can be also rewritten as $f(E)$.

Integrating the mod square of wave function $|\psi(x, t)|^2$ over all times, we can get

$$p(x) = \lim_{t' \rightarrow \infty} \frac{1}{t'} \int_0^{t'} |\psi(x, t)|^2 dt = \sum_{k=1}^{\infty} f(\lambda_k)^2 \phi_k(x)^2. \quad (3)$$

The physical meaning of $p(x)$ is the possibility for a particle with an energy distribution $f(\lambda)$ found at position x . The property of quantum mechanics states that due to the fast decaying nature of the evanescent wave, a particle tends to be trapped within the vicinity of region where the strong field enhancement occurs. In high-dimensional dataset, the ‘‘tip’’ regions house those data points with sparse neighborhood. In other words, the particle tends to stay at instances with more sparse neighborhood and rarely shows up at instances with denser neighborhood. Therefore in theory $p(x)$ can intuitively represent the local density of each instance. In practice, however, the key challenge is how to choose a good energy distribution $f(\lambda)$ for $p(x)$.

B. Energy Distribution Function and Definition of Fermi Density Descriptor

$f(\lambda)$ in Equation 3 is the probability that a particle is in energy state λ . In quantum mechanics there are three main distribution functions [8]: Maxwell-Boltzmann distribution (MB), Fermi-Dirac distribution (FD), and Bose-Einstein distribution (BE). Besides quantum mechanics, existing research also explores distributions based on other theoretical assumptions. Sun [17] et al. used heat dissipation (HD) to describe the heat diffusion (distribution) given time t . In 2011, Aubry [2] chose a Gaussian distribution (GD) in the logarithmic energy to define wave kernel signature. Here we will analyze these five distribution functions and their respective performance on anomaly detection.

Maxwell-Boltzmann Distribution (MB)

$$f_{MB}(\lambda) = \frac{1}{e^{\lambda/T}}. \quad (4)$$

Fermi-Dirac Distribution (FD)

$$f_{FD}(\lambda) = \frac{1}{e^{(\lambda-\mu)/T} + 1}, \quad (5)$$

where μ can be obtained from

$$\sum_{\lambda} \frac{1}{e^{(\lambda-\mu)/T} + 1} = n/2. \quad (6)$$

Bose-Einstein Distribution (BE)

$$f_{BE}(\lambda) = \frac{1}{e^{(\lambda-\mu)/T} - 1}, \quad (7)$$

where μ can be obtained from

$$\sum_{\lambda} \frac{1}{e^{(\lambda-\mu)/T} - 1} = n/2. \quad (8)$$

These three functions depend on the absolute temperature T . FD and BE distributions also depend on a chemical potential μ , and n is the number of particles in the whole systems.

Heat Diffusion (HD)

$$f_{HD}(\lambda) = e^{-\lambda t}, \quad (9)$$

where t is the time for heat dissipation. HD describes how the amount of heat dissipates from a heat source to its neighborhood at time t .

Gaussian Distribution (GD)

$$f_{GD}(\lambda) = e^{-\frac{(e-\log(\lambda))^2}{2\sigma^2}}. \quad (10)$$

GD is derived in [2] from a perturbation-theoretic analysis. Under the assumption that the eigenvalues (eigenenergies) of an articulated dataset are log-normally distributed random variables, the author claimed that it is robust to small data perturbations while being as informative as possible.

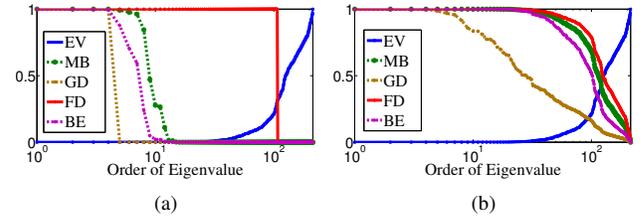


Figure 1. Different energy distribution comparison on Glass dataset. Besides the eigenvalue (EV) ordered by increasing value (decreasing importance) in blue curve, (1(a)) shows four distributions when $T = 0.001$, and (1(b)) shows four distributions when $T = 50$. Green, red, purple and brown curves are the corresponding MB, FD, BE and GD distribution. We can see that FD has the most stable performance as T changes.

HD and MB distributions have similar mathematical properties if we simply replace t in Equation 9 with $\frac{1}{T}$ in Equation 4. So we will ignore HD and only compare the other four distribution functions. For the sake of convenience, we assign σ in Equation 10 with the same value as T in every comparison. Among these distributions, FD is the most

practical one for anomaly detection, especially in terms of stability under different T . In Equation 5, the bias term and balancing term can stabilize FD distribution function performance: the constant smoothing term “plus one” and the balancing term μ in the denominator part. The role of the “plus one” is to damp the contribution of the exponential part from being too small, which results from either extremely small λ or large T . The balancing term μ is a parameter controlling the trade-off between small and large λ . Besides, it helps to tune a sweet range for E according to T , since it has a positive side-effect that it can accelerate the attenuation of contribution from those trivial eigenvalues in Equation 5. Figure 1 shows the value of different distribution functions across different eigenvalues of Glass dataset (Section V-A). In general, FD distribution tends to assign stable weights for the nontrivial eigenvalues compared with the other three distributions regardless how temperature T changes.

Therefore we integrate the FD distribution function (Equation 5) into Equation 3, and define **Fermi Density Descriptor (FDD)** at a point x as:

$$FDD(x) = \frac{1}{C} \sum_{k=1}^{\infty} \left(\frac{1}{e^{(\lambda_k - \mu)/T} + 1} \right)^2 \phi_k(x)^2, \quad (11)$$

where $C = \sum_{k=1}^{\infty} \left(\frac{1}{e^{(\lambda_k - \mu)/T} + 1} \right)^2$, and μ can be derived from Equation 6.

III. NORMALIZATION METHODS

In practice, the discrete Laplace operator in Equation 1 has different normalization ways. Although their effects on clustering has been thoroughly analyzed in [10] and [5], it is still unclear what is the best choice for anomaly detection. This section analyzes the effect of five different normalizations for anomaly detection. Denote the unit matrix as I , the affinity matrix of dataset as W and the corresponding degree matrix as D , there are:

No-normalization (NN)

$$L_{NN} = D - W. \quad (12)$$

Random walk normalization (RW)

$$L_{RW} = I - D^{-1}W. \quad (13)$$

Symmetric normalization (SM)

$$L_{SM} = I - D^{-1/2}WD^{-1/2}. \quad (14)$$

Fokker-Planck normalization (FP)

$$L_{FP} = I - D^{-1}W'_1, \quad (15)$$

where $W'_1 = D^{-1/2}WD^{-1/2}$.

Laplace-Beltrami normalization (LB)

$$L_{LB} = I - D^{-1}W'_2, \quad (16)$$

where $W'_2 = D^{-1}WD^{-1}$.

For clustering purpose, we focus on normal instances and want to recover manifold that is insensitive to the existing anomalies (usually being treated as noise in such applications) [10] [5]. However, from the anomaly detection’s point of view, the recovered manifold should be aware of local density variation, therefore in the manifold space the density differences between anomalies and normal instances should be preserved or even magnified.

Theorem 1: The density impact factors for NN, RW, SM, FP and LB normalization are 2, 1, 1, 0.5, and 0 respectively.

Proof: Define $q(x)$ as the true density function of x , according to [5] the infinitesimal operator can be given by

$$\Delta\phi - \frac{\Delta(q^{1-\alpha})}{q^{1-\alpha}}\phi, \quad (17)$$

where $\phi = fq^{1-\alpha}$. For RW, $\alpha_{RW} = 0$ [5] so the density impact factor is $1 - \alpha_{RW} = 1$. For FP, $\alpha_{FP} = 0.5$ [5] so the density impact factor $1 - \alpha_{FP} = 0.5$. For LB, $\alpha_{LB} = 1$ [5] so $1 - \alpha_{LB} = 0$. SM can be transformed from RW by $L_{SM} = D^{1/2}L_{RW}D^{-1/2}$. From the analysis in [5] we know that D is proportional to the density function q , therefore its $\lim_{\sigma \rightarrow 0} L_{SM, \sigma} f$ depends on density function $q^{1/2}q^{1-\alpha_{RW}}q^{-1/2} = q^1$ where $\alpha_{RW} = 0$. So its density impact factors is also 1. For NN, since $L_{NN} = DL_{RW}$, $\lim_{\sigma \rightarrow 0} L_{NN, \sigma} f$ depends on density function $q \times q^{1-\alpha_{RW}} = q^2$ where $\alpha_{RW} = 0$, therefore its impact factors is 2. \square

As an illustration, Figure 2 shows the effects of different normalizations on Ecoli dataset (Section V-A). We only plot the first three non-trivial eigenvectors derived from the normalized affinity matrix. The red circles are anomalous instances while crosses with other colors represent different clusters of normal instances respectively. We also show AUC score (Section V-A) of anomaly detection results, and NMI score [9] of clustering results from different Laplacian normalizations. With NN normalization, the influence of density is mostly maximized compared with the other four normalizations. It results in that the normal instances with higher density shrink to a condensed area while anomalous instances are far away from the collapsed center. So NN normalization has the strongest ability (with AUC 0.9042) to separate anomaly from normal instances though it is not the sweet choice for clustering (with NMI 0.5432).

IV. ALGORITHMIC FRAMEWORK

Let X be a matrix of size $n \times m$, where n is the number of instances and m is the number of dimensions, our algorithm is detailed in Algorithm 1. It undergoes a kind of data wrapping in the first two steps. Then we perform the eigen-decomposition and compute FDD for each instance. FDD value is used as the final measurement of anomalousness. Anisotropic Gaussian Kernel (AGK) [16] has been used in our experiment to build the affinity matrix in Step 1, because it can better capture the manifold structure of data.

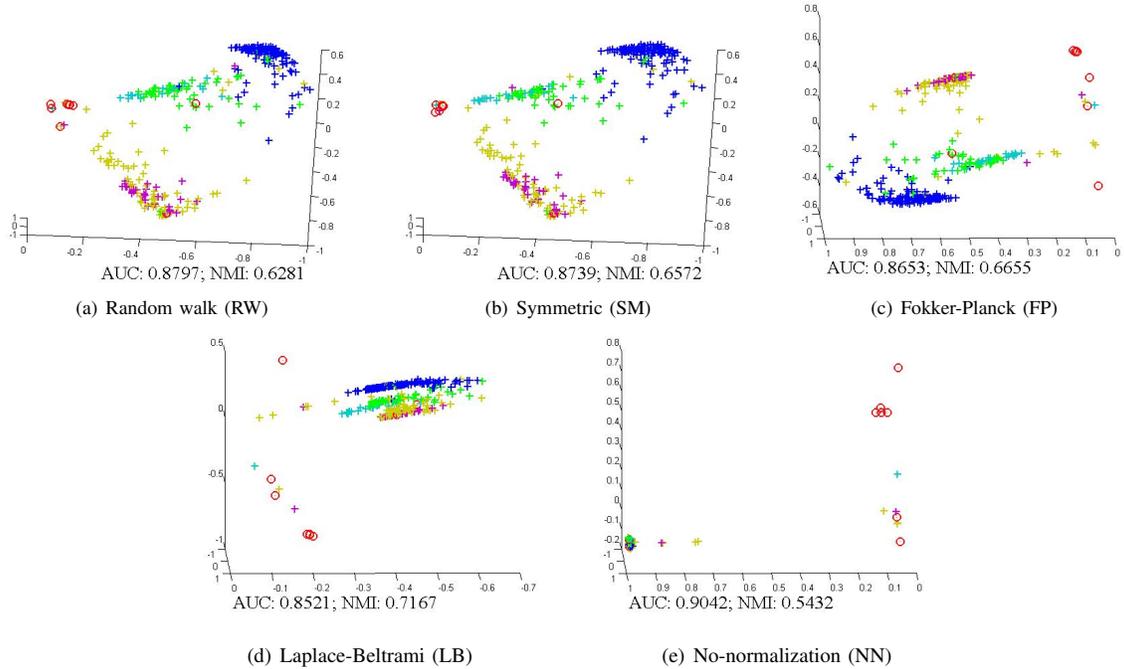


Figure 2. Different normalization effects for anomaly detection (in AUC) and clustering (in NMI).

Table I
STATISTICS OF OUR EVALUATION DATASETS.

	Dataset	# Instance	# Attribute	% Anomalies(classes)	References
1	BRCAN	683	9	35.0% (malignant)	[7]
2	WDBC	569	29	37.3% (malignant)	[14]
3	Pima	768	8	34.9% (positives)	[12] [14]
4	Arrhythmia	452	279	45.0% (abnormal)	[6] [12] [14]
5	Hayes-Roth	132	5	22.7% (class 3)	[14]
6	Ecoli	336	7	2.7% (omL,imL and imS)	[14]
7	Yeast	1484	8	3.7% (vac, pox and erl)	[14]
8	Abalone	4177	7	8.0% ($age < 5$ or > 15)	[14]
9	Pageblocks	5473	10	4.2% (graphic, vertline and picture)	[14]
10	Glass	214	9	4.2% (tableware)	[14]
11	Ionosphere	351	34	35.9% (bad)	[12] [14]
12	Magic	19020	10	35.2% (hadron)	[14]

Algorithm 1: FermiDensityDescriptor(X, T)

Input: Input data $X \in R^{n \times m}$, T is the environmental temperature

Output: FDD score for each instance

- 1 Build affinity matrix W on X ;
 - 2 Construct NN Laplacian normalization, which is $L = D - W$, where D is the degree matrix of W ;
 - 3 Compute generalized eigenvectors $\psi(i)$ and corresponding eigenvalues λ_i , $i = 1, 2, \dots, n$ of L ;
 - 4 Construct Fermi Density Descriptor (FDD) with temperature T using Equation 11 ;
-

V. EXPERIMENTAL ANALYSIS

A. Experimental Setup

Dataset and Baselines. To demonstrate the performance of our new method, we evaluate it on twelve benchmark

datasets listed in Table I. We choose seven state-of-the-art competitors in three categories to show the performance of our new FDD. For kNN-based algorithms, we choose Local Outlier Detection (LOF) [3] and Local Correlation Integral (LOCI) [15]. For attribute-based methods, we include IForest [12] and Mass [18]. For manifold-based methods, we refer readers to three nonlinear techniques including heat kernel signature (HKS) based on random walk (RW) normalization [17], locally linear embeddings (LLE), and isometric feature mapping (ISM) [1] followed by LOF to obtain anomalousness measurement.

Evaluation Metrics. Due to space limitation, AUC (Area under Receiver Operating Characteristics Curve) [13] is used as our evaluation metric here because it is commonly used to evaluate anomaly detectors and it is cut-off independent.

Table II

COMPARISON OF FDD AND OTHER FOUR METHODS USING AUC METRICS. THE BOLD-FACE NUMBERS INDICATE THE BEST METHOD IN A PARTICULAR DATASET; THE NUMBERS IN PARENTHESES INDICATE THE RANKS OF METHODS ACCORDINGLY. AVERAGE AND THE RANK TOTAL IS THE AVERAGE OF PERFORMANCE AND THE SUM OF ALL RANKS FOR EACH METHOD ACROSS DATA SETS RESPECTIVELY. A * INDICATES A P-VALUE OF 5% OR LOWER AND ** INDICATES A P-VALUE OF 1% OR LOWER IN THE STATISTICAL SIGNIFICANCE TEST FOR PERFORMANCE COMPARISON W.R.T. FDD.

Dataset	LOF(k=10)	LOF(k=25)	LOF(k=50)	LOCI(k=10)	LOCI(k=50)	Mass	IForest	FDD
BRCAN	0.2819 (7)	0.3870 (6)	0.7354 (4)	0.7209 (5)	0.7455 (3)	0.2515 (8)	0.9775 (2)	0.9835 (1)
WDBC	0.5874 (8)	0.6192 (7)	0.7784 (4)	0.7573 (6)	0.8119 (2)	0.7974 (3)	0.7760 (5)	0.8989 (1)
Pima	0.4847 (8)	0.5270 (7)	0.6003 (4)	0.5946 (5)	0.6089 (3)	0.5812 (6)	0.6630 (2)	0.7121 (1)
Arrhythmia	0.7419 (7)	0.7547 (1)	0.7482 (4)	0.7482 (4)	0.7483 (3)	0.6363 (8)	0.7456 (6)	0.7484 (2)
Hayes-Roth	0.7033 (4)	0.5948 (7)	0.5464 (8)	0.8176 (3)	0.6686 (5)	0.6581 (6)	0.9894 (2)	0.9918 (1)
Ecoli	0.8260 (7)	0.8614 (4)	0.8454 (6)	0.8641 (3)	0.8549 (5)	0.7699 (8)	0.8754 (2)	0.9042 (1)
Yeast	0.4159 (8)	0.6253 (1)	0.5986 (6)	0.6076 (4)	0.6035 (5)	0.5712 (7)	0.6159 (3)	0.6168 (2)
Abalone	0.5724 (8)	0.6056 (7)	0.6525 (6)	0.6932 (4)	0.7058 (3)	0.6923 (5)	0.7466 (1)	0.7332 (2)
Pageblocks	0.6819 (8)	0.7953 (5)	0.8256 (4)	0.7823 (6)	0.7823 (6)	0.8661 (2)	0.8625 (3)	0.8939 (1)
Glass	0.7474 (6)	0.7008 (7)	0.7528 (4)	0.7480 (5)	0.7593 (3)	0.8922 (1)	0.6933 (8)	0.8737 (2)
Ionosphere	0.9064 (2)	0.8709 (3)	0.7982 (7)	0.8512 (4)	0.7923 (8)	0.8269 (6)	0.8467 (5)	0.9302 (1)
Magic	0.6420 (7)	0.6804 (6)	0.6970 (4)	0.5672 (8)	0.6825 (5)	0.6984 (3)	0.7506 (2)	0.7520 (1)
Average	0.6326** (8)	0.6685** (7)	0.7149** (5)	0.7294** (4)	0.7303** (3)	0.6868* (6)	0.7952* (2)	0.8365 (1)

In our paper we also show that our FDD has the most robust and stable performance for all the datasets by using macro paired t-tests [20] against each competitor respectively.

Parameters. In experiments documented in Table II we fix $T = 1000$ for FDD. In Figure 3(f) and 3(e) we test T of FDD and t of HKS in $[10^{-4}, 10^4]$ in order to show their sensitivity to scaling parameter. For LOF we try size of neighborhood $k = 10, 25, 50$. As for LOCI, we only test Radius coefficient $\alpha = 0.5$, and $k = 10$ and 50 in that LOCI is more robust to k [15]. As for IForest, to conduct safe and fair comparison we set $\rho = 4000$ and the number of trees $nt = 100$ as in [11]. For similar reason, in Mass we set the number of mass estimation $ne = 100$ and the sub-sampling size as the dataset size. For each dataset we run 30 times for both IForest and Mass and use the average AUC in the final comparison. For LLE and ISM experiments, we first fix $d = 3$ and test k in $[10, 50]$ (Figure 3(a) and 3(c)), then fix the best k for each dataset and test d in $[2, 30]$ (Figure 3(b) and 3(d)).

B. Algorithm Performance Comparison

Table II documents the anomaly detection results (in AUC and p-value) of LOF, LOCI, Mass, IForest and FDD, while Figure 3 shows the performance and robustness (as parameters vary) of LLE, ISM, HKS and FDD. In Table II, our FDD shows the best average performance (with AUC 0.8365). Across each dataset our FDD has the best performance for eight out of the twelve selected datasets, and the second best for the other four datasets. For Arrhythmia and Yeast our FDD scores are extremely close (less than 0.01) to the best score by kNN-based algorithms LOF and LOCI. As for Abalone and Glass dataset the AUC score of our FDD are still comparable to the best one (less than 0.02) by attribute-based methods Mass and IForest.

As for the macro paired t-tests, compared with LOF and LOCI respectively, FDD has extremely small p-value (less

than 1%). Compared with Mass with IForest, FDD has p-value less than 5%. This, once again, proves that our FDD has the best and most stable average performance. From Table II, we can see that FDD outperforms the selected k NN-based algorithms and attribute-based algorithms in terms of performance and stability.

To systematically demonstrate the performance and parameter-tuning sensitivity of manifold-based algorithms, we test LLE and ISM with different k and d , and HKS and FDD with changing t and T respectively. Figure 3 shows that our new FDD has the most robustness on parameter tuning. This stable property of FDD (Figure 3(f)) is derived from Fermi-Dirac distribution (Section II-B) and NN normalization (Section III), which is extremely important for reliable data analysis, and for those domain experts who do not have strong machine learning background as they become much more comfortable in utilizing robust anomaly detection algorithms such as our new FDD.

VI. CONCLUSION

We have devised a new unsupervised anomaly detection algorithm with good performance and strong robustness to parameter tuning. It is originated from quantum mechanics and Fermi-Dirac distribution. To further enhance the functionality of our algorithm, we first explored the best choice among different Laplacian normalizations for mining anomalous instances. Extensive experiments and evaluations have demonstrated the sustained superb performance of FDD in comparison with other popular anomaly detection algorithms. Immediate future work will be concentrated on seeking connection between local and global patterns with an emphasis on learning the intrinsic structure of data.

ACKNOWLEDGMENTS

We thank valuable help from Dr. Yan Li in BNL, and all the anonymous reviewers for constructive suggestions. This research is supported in part by NSF grants IIS-0949467,

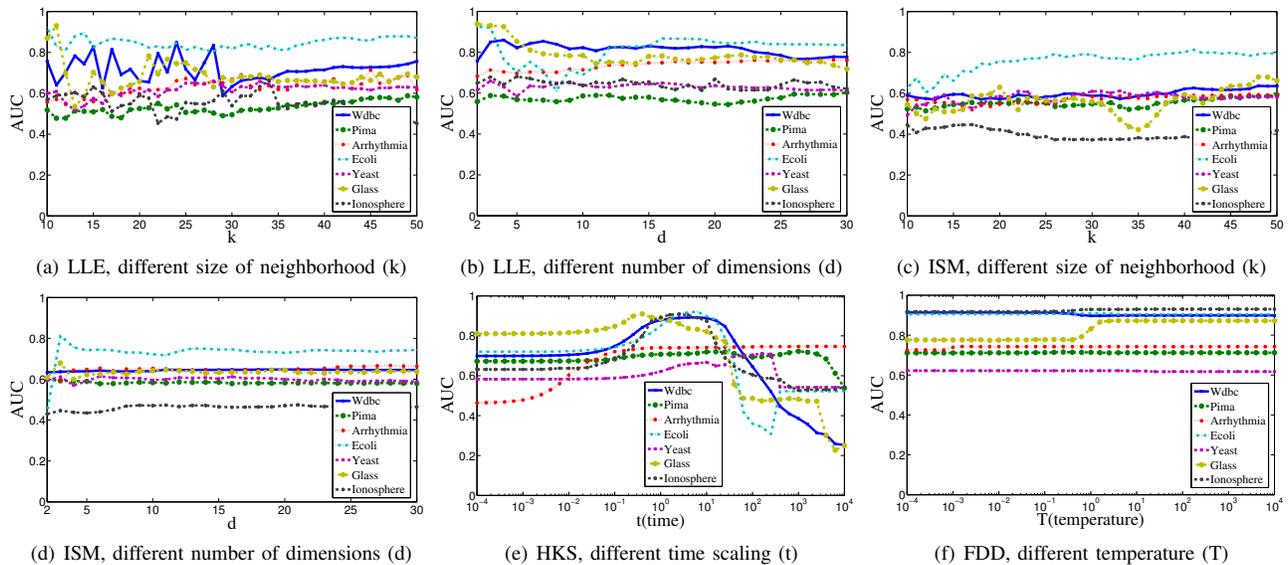


Figure 3. Algorithmic performance on different parameters.

IIS-1047715, IIS-1049448, DOE grants DE-SC0003361 and DE-AC02-98CH10886.

REFERENCES

- [1] A. Agovic, A. Banerjee, A. R. Ganguly, and V. Protopopescu. Anomaly detection in transportation corridors using manifold embedding. *the 1st International Workshop on Knowledge Discovery from Sensor Data*, 2007.
- [2] M. Aubry, U. Schlickewei, and D. Cremers. The wave kernel signature - a quantum mechanical approach to shape analysis. *IEEE CVPR*, pages 1626–1633, 2011.
- [3] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander. Lof: identifying density-based local outliers. *ACM SIGMOD*, pages 93–104, 2000.
- [4] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: a survey. *ACM Computing Surveys*, 41(3):1–72, 2009.
- [5] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1):5–30, 2006.
- [6] T. de Vries, S. Chawla, and M. Houle. Finding local anomalies in very highdimensional space. *IEEE ICDM*, pages 128–137, 2010.
- [7] F. A. Gonzalez and D. Dasgupta. Anomaly detection using real-valued negative selection. *Genetic Programming and Evolvable Machines*, pages 383–403, 2003.
- [8] G. Greenstein and A. Zajonc. *The quantum challenge: Modern research on the foundations of quantum mechanics*. Jones and Bartlett Publishers, 2006.
- [9] J. A. Hartigan and M. A. Wong. Algorithm as 136: a k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1978.
- [10] H. Huang, S. Yoo, H. Qin, and D. Yu. A robust clustering algorithm based on aggregated heat kernel mapping. *IEEE ICDM*, pages 270–279, 2011.
- [11] F. T. Liu and K. M. Ting. Can isolation-based anomaly detectors handle arbitrary multi-modal patterns in data? *Technical Report*, 2010.
- [12] F. T. Liu, K. M. Ting, and Z. H. Zhou. Isolation forest. *IEEE ICDM*, pages 413–422, 2008.
- [13] C. Marzban. A comment on the roc curve and the area under it as performance measures. *Technical Report*, 2004.
- [14] K. Noto, C. E. Brodley, and D. Slonim. Anomaly detection using an ensemble of feature models. *IEEE ICDM*, pages 953–958, 2010.
- [15] S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos. Loci: Fast outlier detection using the local correlation integral. *IEEE ICDE*, pages 315–326, 2003.
- [16] A. Singer and R. R. Coifman. Non-linear independent component analysis with diffusion maps. *Applied and Computational Harmonic Analysis*, 25(2):226–239, 2008.
- [17] J. Sun, M. Ovsjanikov, and L. Guibas. A concise and provably informative multi-scale signature based on heat diffusion. *SGP*, pages 1383–1392, 2009.
- [18] K. M. Ting, G. T. Zhou, F. T. Liu, and J. S. Tan. Mass estimation and its applications. *ACM KDD*, pages 989–998, 2010.
- [19] L. van der Maaten, E. Postma, and J. van der Herik. Dimensionality reduction: A comparative review. *Technical report*, 2009.
- [20] D. W. Zimmerman. A note on interpretation of the paired-samples t test. *Journal of Educational and Behavioral Statistics*, 22(3):349–360, 1997.