EFFICIENT ANOMALY DETECTION USING BIPARTITE k-NN GRAPHS*

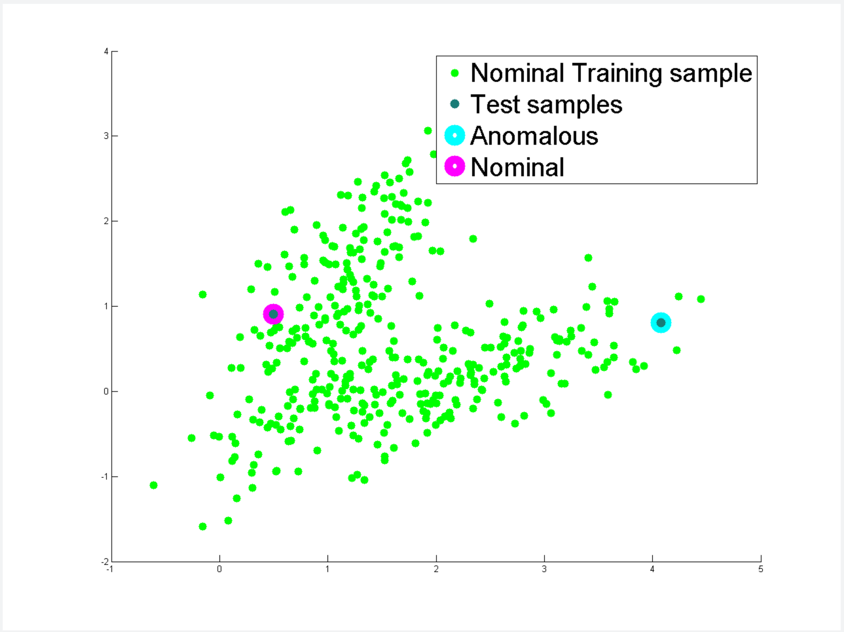
KUMAR SRICHARAN, ALFRED O. HERO III

Department of EECS, University of Michigan, Ann Arbor, MI, USA

*This research is partially supported by the Air Force Office of Scientific Research, grant number FA9550-09-1-0471.

PROBLEM STATEMENT

- Identify unknown, anomalous events that deviate from a training set of normal events.



- Examples: fault detection in mission-critical systems, quality control in manufacturing, medical diagnosis
- Statistical framework:
  - Training sample $\mathcal{X}_T = \{X_1, \ldots, X_T\}$
  - Nominal density $X_i \sim f_0$
  - Test sample $X \sim f$
  - Anomaly detection problem: Test

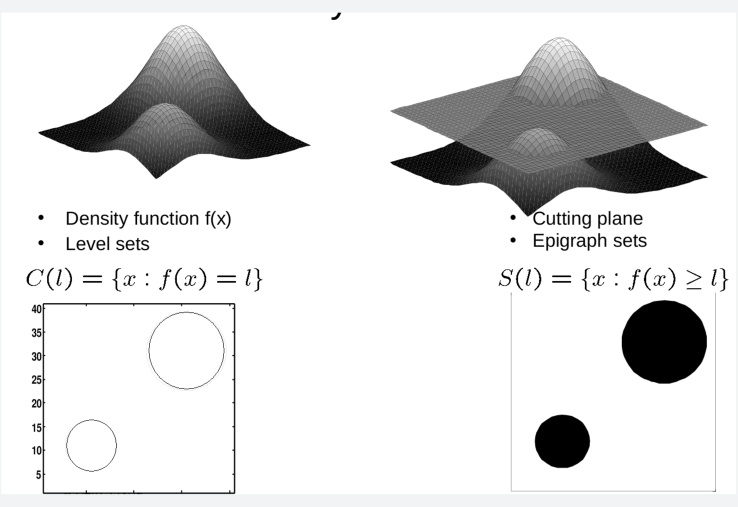
$H_0 : f = f_0$ versus $H_1 : (1-\epsilon)f_0 + \epsilon f$

- Previous approaches based on:
  - Density [5] : Mass, iForest
  - Distance: ORCA [5]
  - Minimum volume sets(MV): Hero [1], K-LPE [2], Park *etal* [4], Scott *etal* [3]

MV SET APPROACH

- Fix false alarm $\alpha \in (0, 1)$
- Seek acceptance region A with minimum volume (MV) or equivalently minimum entropy (ME) that satisfies

$$Pr(X \in A|H_0) \geq 1 - \alpha$$



- MV anomaly detection: If test event falls in MV set, classify as normal; otherwise as anomalous.
- Uniformly most powerful test for $f \sim$ uniform.
- MV set estimation involves approximation of high dimensional quantities [3, 4]: level sets of

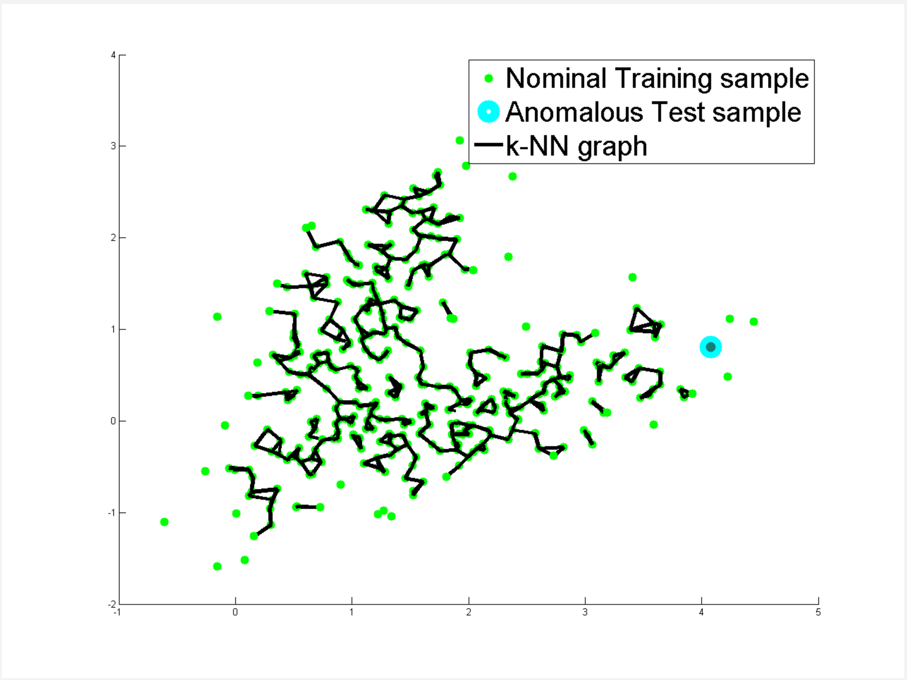
$$\hat{f}_0(x) = \sum K_h(x - X_i).$$

GEM PRINCIPLE

- Geometric entropy minimization (GEM) principle [1] circumvents MV set estimation problem
- Asymptotically consistent in recovering the p-value (1) of the test point
- Let $\mathcal{X}_{K,T+1}$ denote one of the $\binom{T+1}{K}$ K point subsets of $\mathcal{X}_T \cup \{X\}$.
- K-kNNG acceptance region

$$A = \underset{\mathcal{X}_{K,T+1} \in \mathcal{X}}{\operatorname{argmin}} L_{kNN}(\mathcal{X}_{K,T+1})$$

- Declare X to be an anomaly if $X \notin A$. False alarm rate converges to $\alpha = 1 - K/(T+1)$ and runtime is $O(\binom{T}{K}) K \log K$.



- L10-kNNG Hero [1] set $K = T - 1$. Runtime is $O(T^2 \log(T))$. Fixed false alarm rate $1/(T+1)$.
- K-LPE [2] improves on GEM by directly estimating p-values (1) of test samples using k-NN graphs. Runtime is $O(T^2)$.

CENTRAL IDEA

Decompose minimum entropy set estimation problem into two parts by partitioning data.
Use first part for entropy estimation, second part to determine set with minimum entropy.
Advantage: Significant computational savings.

PROPOSED ALGORITHM

1. Input:
   Training samples $\mathcal{X}_T$
   Desired False alarm rate α
   Test samples $\mathcal{Y}_L = \{Y_1, .., Y_L\}$
2. Training phase

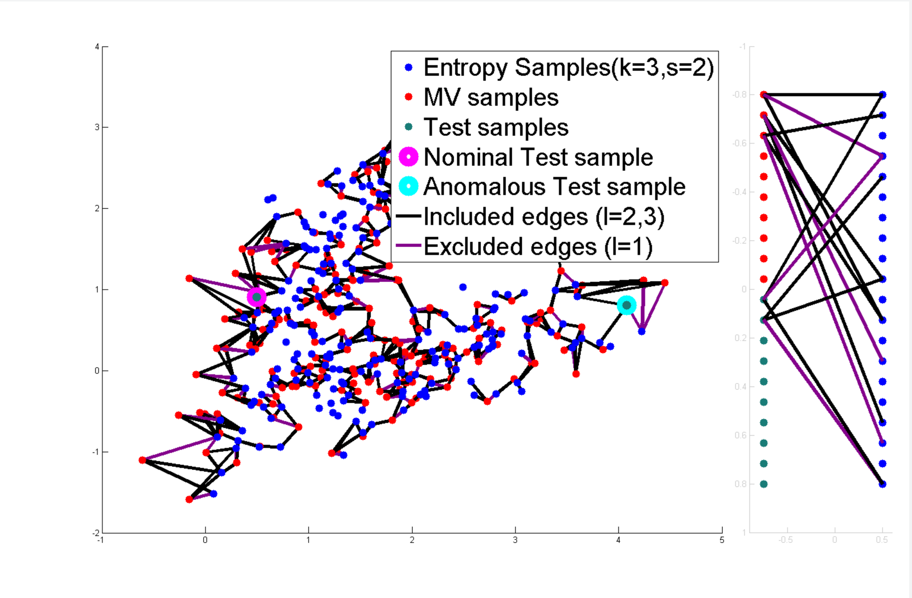
a. Create partition $\{\mathcal{X}_N, \mathcal{X}_M\}$
b. Construct k-NN bipartite graph on partition
   $e_{i(l)}$: l-NN distance of sample X_i wrt $\mathcal{X}_M$
c. Compute k-NN lengths $d_{s,k}(X_i)$ for $X_i \in \mathcal{X}_N$:

$$d_{s,k}(X_i) = \sum_{l=k-s+1}^{k} |e_{i(l)}|^\gamma$$

3. Test phase: detect anomalous points
for each input test sample $X \in \mathcal{Y}_L$ **do**
Compute k-NN length

$$d_{s,k}(X) = \sum_{l=k-s+1}^{k} |e_{X(l)}|^\gamma$$

if $\sum_{X_i \in \mathcal{X}_N} 1(d_{s,k}(X_i) < d_{s,k}(X)) \geq N(1-\alpha)$ **then**
Declare X to be anomalous
else
Declare X to be non-anomalous
end if
end for



RELATION TO GEM

- $\{N, M\}$ partition of $\mathcal{X}_T$: $card\{\mathcal{X}_N\} = N$ and $card\{\mathcal{X}_M\} = M = T - N$.
- Let $\mathcal{X}_{K,N+1}$ denote one of the $\binom{N+1}{K}$ K point subsets of $\mathcal{X}_N \cup \{X\}$.
- Construct bipartite k-NN graph from $\mathcal{X}_{K,N+1}$ to $\mathcal{X}_M$.
- BP-kNNG acceptance region:

$$A = \underset{\mathcal{X}_{K,N+1} \in \mathcal{X}}{\operatorname{argmin}} L_{kNN}(\mathcal{X}_{K,N+1}, \mathcal{X}_M)$$

- Declare X to be an anomaly if $X \notin A$. By GEM principle, false alarm rate converges to $\alpha = 1 - K/(N+1)$.
- Equivalence We can equivalently determine A by ranking. Rank order $\mathcal{X}_N \cup \{X\}$ according to:

$$d_{s,k}(X_{(1)}) \leq \ldots \leq d_{s,k}(X_{(K)}) \ldots \leq d_{s,k}(X_{(N+1)})$$

and set $A = \{X_{(1)}, \ldots, X_{(K)}\}$.

- Computational savings:
  1. Partitioning approach breaks down combinatorial problem into ranking problem
  2. Suffices to construct bipartite graph once on the entire set of nominal data and queries.
- For L queries, runtime per test instance is $O(T(T/L + 1))$. When $T = O(L)$, the runtime is $O(T)$; linear rather than quadratic.

CONSISTENCY

- True p-value

$$p(X) = \int_{\{z: f_0(z) \leq f_0(X)\}} f_0(z)dz \quad (1)$$

- Empirical p-value:

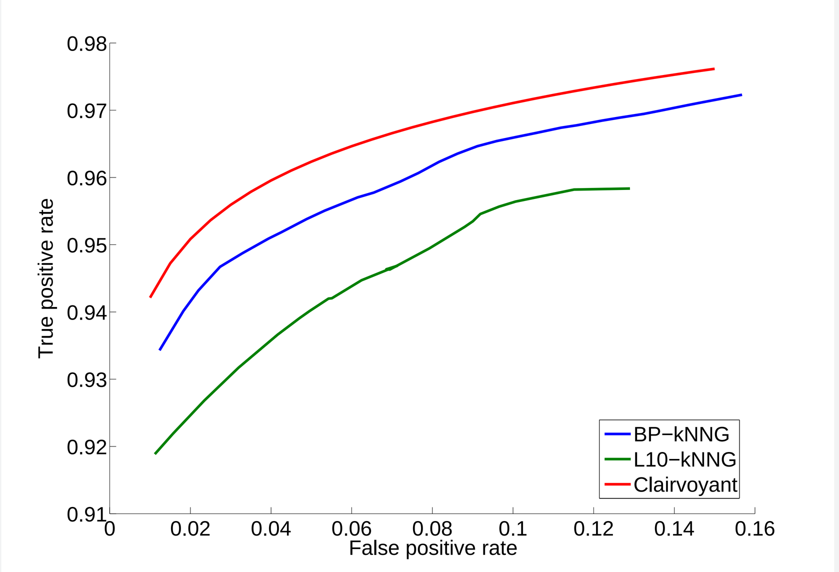
$$p_{bp}(X) = \frac{\sum_{X_i \in \mathcal{X}_N} 1(d_{s,k}(X_i) \geq d_{s,k}(X))}{N}$$

- Asymptotic Consistency: $k \to \infty$, $N \to \infty$, $k/M \to 0$, $s = \theta(1)$

$$\mathbb{E}[(p_{bp}(X) - p(X))^2] = O(1/N + (k/M)^{2/d} + 1/k)$$

- Optimal parameter choice wrt MSE for fixed T:
  $N = O(T^{\frac{4+d}{4+2d}})$, $M = T - N = O(T)$
  $k = O(M^{\frac{2}{2+d}})$
- Faster convergence rate of estimated p-value ($O(T^{-2/(2+d)})$) in comparison to K-LPE ($O(T^{-2/5} + T^{-6/5d})$)

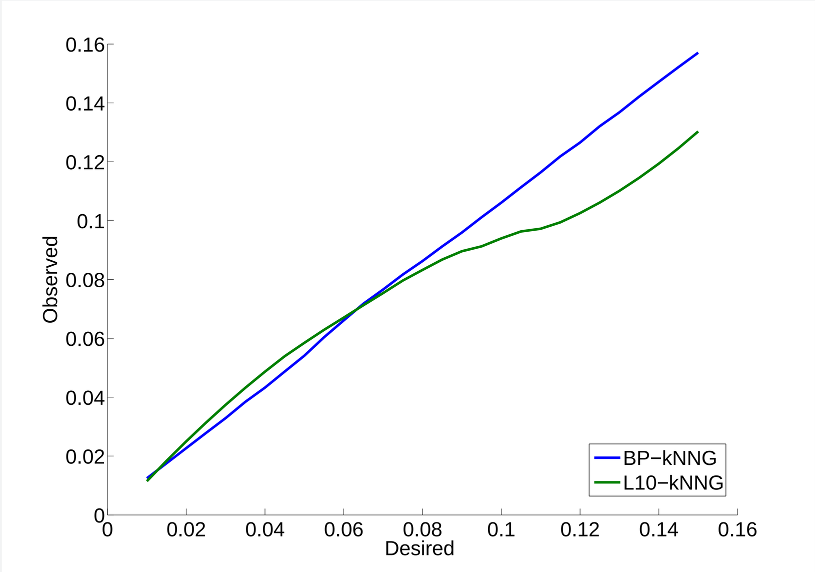
ROC COMPARISON



ROC comparison

Data sets	BP	L10	K-LPE	Mass	iF	ORCA
HTTP	0.99	NA	NA	1.00	1.00	0.36
Forest	0.86	NA	NA	0.91	0.87	0.83
Mulcross	1.00	NA	NA	0.99	0.96	0.33
SMTP	0.90	NA	NA	0.86	0.88	0.87
Shuttle	0.99	NA	NA	0.99	1.00	0.60

FALSE ALARM RATE COMPARISON



False alarm rate comparison

Data sets	Desired false alarm				
	0.01	0.02	0.05	0.1	0.2
HTTP	0.007	0.015	0.063	0.136	0.216
Forest	0.009	0.015	0.035	0.071	0.150
Mulcross	0.008	0.014	0.040	0.096	0.186
SMTP	0.006	0.017	0.046	0.099	0.204
Shuttle	0.026	0.030	0.045	0.079	0.179

RUNTIME COMPARISON

Data sets	BP	L10	K-LPE	Mass	iF	ORCA
HTTP	3.81	.10/i	.19/i	34	147	9487
Forest	7.54	.18/i	.18/i	18	79	6995
Mulcross	4.68	.26/i	.17/i	17	75	2512
SMTP	0.74	.11/i	.17/i	7	26	267
Shuttle	1.54	.45/i	.16/i	4	15	157

CONCLUSIONS

- BP-kNNG is based on GEM principle [1] for MV-set based anomaly detection
  - BP-kNNG inherits theoretical optimality properties of GEM, including asymptotic consistency, unlike L10-kNNG
  - Bipartite construction reduces combinatorial problem to ranking problem
  - Consequence: Runtime of BP-kNNG is significantly better in comparison to K-kNNG, L10-kNNG and K-LPE
- Comparison to state of the art anomaly detection methods:
  - Compares favorably in terms of both ROC performance and run time
  - Consistency in recovering p-value of test sample
  - Detect anomalies at desired false alarm rates.

REFERENCES

[1] A. O. Hero. Geometric entropy minimization (gem) for anomaly detection and localization. In *Proc. Advances in Neural Information Processing Systems (NIPS*, pages 585–592. MIT Press, 2006.

[2] M. Zhao and V. Saligrama. Anomaly detection with score functions based on nearest neighbor graphs. In *Proc. Advances in Neural Information Processing Systems (NIPS*, pages 2250–2258. MIT Press, 2009.

[3] C. Scott and R. Nowak. Learning minimum volume sets. *J. Machine Learning Res*, 7:665–704, 2006.

[4] C. Park, J. Z. Huang, and Y. Ding. A computable plug-in estimator of minimum volume sets for novelty detection. *Operations Research*, 58(5):1469–1480, 2010.

[5] K. M. Ting, G. Zhou, T. F. Liu, and J. S. C. Tan. Mass estimation and its applications. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '10, pages 989–998, New York, NY, USA, 2010. ACM.