Spatio-temporal Anomaly Detection

Adway Mitra

Machine Learning for Earth System Sciences (AI60002)

IIT Kharagpur

Definition of Anomaly

- Earth Science definition: deviation from past behavior

$$Y(s,t) = X(s,t) - \mu(s,t)$$

- Data Science definition: deviation from other values

- Spatio-temporal Anomaly: deviation of values from spatio-temporal neighbors
- Easy problem: isolated anomalies
- Hard problem: bulk anomalies (anomaly events)

Spatio-temporal Dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

- Spatial neighbors: (S1,S2,S3) and (S4,S5,S6)

Spatio-temporal Dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

- $|X(S3,T3) - X(S3,T2)| > thres$, $|X(S3,T3) - X(S2,T3)| > thres$

Spatio-temporal Dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

- $|X(S6,T2) - X(S6,T1)| > thres$, $|X(S6,T2) - X(S5,T2)| > thres$

Spatio-temporal Dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

- $|X(S1,T5) - X(S3,T5)| > thres$, $|X(S2,T6) - X(S3,T6)| > thres$

Spatio-temporal Dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

- $|X(S1,T5) - X(S3,T5)| > thres$, $|X(S2,T6) - X(S3,T6)| > thres$

Anomaly Detection

- Simplest approach: compare with spatio-temporal neighbors or climatology using threshold
- Problems: i) Results totally sensitive on threshold

  ii) Comparison with neighbors can't catch "bulk anomalies"

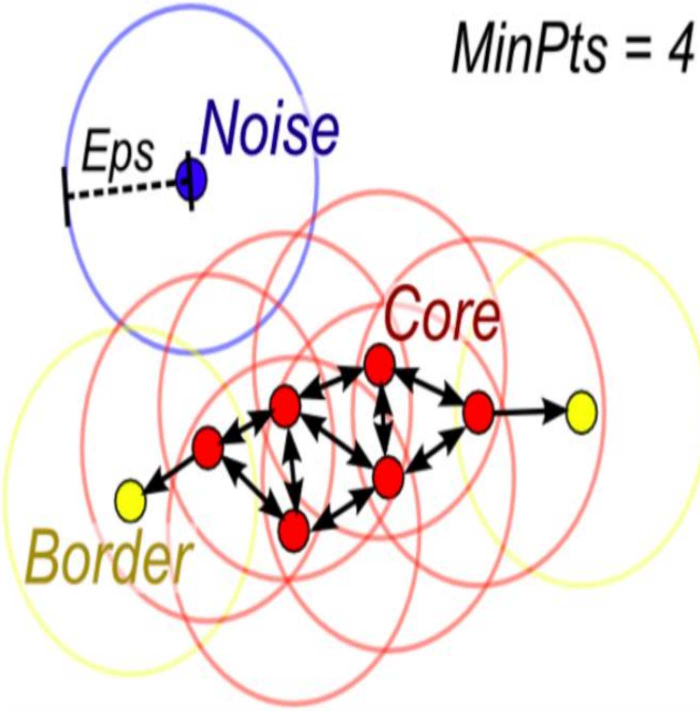
  iii) Climatology may not be available

- Alternatives: i) Clustering

  ii) Latent variable models

Clustering: DB-SCAN

- Density-based Spatial Clustering of Applications with Noise
- Idea: for each point, identify “neighbors” in feature space, and add them in cluster.
- Those points which could not be added to any cluster outlier!
- Need to specify distance threshold



Red: Core Points

Yellow: Border points. Still part of the cluster because it’s within epsilon of a core point, but not does not meet the min_points criteria

Blue: Noise point. Not assigned to a cluster

Spatio-temporal clustering: DBSCAN

- Each data-point has spatial and temporal neighbors
- Each data-point can join only the clusters of its spatial or temporal neighbors
- Joining cluster on the basis of values
- If it cannot join any such cluster, then it is an outlier/anomaly!

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

Spatio-temporal clustering: DBSCAN

- Each data-point has spatial and temporal neighbors
- Each data-point can join only the clusters of its spatial or temporal neighbors
- Joining cluster on the basis of values
- If it cannot join any such cluster, then it is an outlier/anomaly!

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

Spatio-temporal clustering: DBSCAN

- Each data-point has spatial and temporal neighbors
- Each data-point can join only the clusters of its spatial or temporal neighbors
- Joining cluster on the basis of values
- If it cannot join any such cluster, then it is an outlier/anomaly!

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

Can’t handle extended/bulk anomalies!

Alternative-1: Multiple scales

- Difficult to identify bulk anomalies at a single spatial/temporal scale
- Smoothen/coarsen the data at several levels
- Merge locations and consider their mean values
- Merge time-points and consider their mean values
- Repeat same process on coarsened dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

	T1	T2	T3	T4	T5	T6
S1-S2	26	27	26	28	32	28
S2-S3	26	28	25	26	29	28
S4-S5	34	32	29	32	29	29
S5-S6	33	34	29	33	29	28

Alternative-1: Multiple scales

- Difficult to identify bulk anomalies at a single spatial/temporal scale
- Smoothen/coarsen the data at several levels
- Merge locations and consider their mean values
- Merge time-points and consider their mean values
- Repeat same process on coarsened dataset

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

	T1	T2	T3	T4	T5	T6
S1-S2	26	27	26	28	32	28
S2-S3	26	28	25	26	29	28
S4-S5	34	32	29	32	29	29
S5-S6	33	34	29	33	29	28

Anomalies may get smoothed out!

Alternative-2: latent variable models

- At each (s,t) define discrete latent variable Z(s,t)
- Z(s,t) can take two values (anomaly/ no anomaly) or three values (no anomaly/positive anomaly/negative anomaly)
- X(s,t) is a random variable with value known
- X(s,t) ~ f($p_{s,t,k}$) where k = Z(s,t)
- Z(s,t) may also depend on Z(s,t-1) or Z(s’,t) where (s,s’) are neighbors
- Values of Z estimated by Gibbs Sampling

Alternative-2: latent variable models

Example:

Observation of X(s,t) = 34.3

f: Gaussian distribution

Ps,t,1 = [30,10], ps,t,2 = [8, 20]

prob(X(s,t)=34.3 | Z(s,t)=1) = N(34.3; [30,10])

prob(X(s,t)=34.3 | Z(s,t)=2) = N(34.3; [8,20])

prob(Z(s,t) = 1 | X(s,t)=34.3) = ???? (Bayes Theorem)

Z(s,t) should also depend on Z(s,t-1), Z(s',t) etc for bulk anomalies

Alternative-2: latent variable models

- Can handle bulk anomalies
- Used to identify “anomaly events” like heat waves or droughts

	T1	T2	T3	T4	T5	T6
S1	24	26	22	30	31	23
S2	28	27	29	26	32	33
S3	25	29	20	26	25	22
S4	35	33	29	33	34	31
S5	33	31	28	31	24	26
S6	32	37	29	34	33	29

Z	T1	T2	T3	T4	T5	T6
S1	1	1	1	2	2	1
S2	1	1	1	1	2	2
S3	1	1	3	1	1	1
S4	1	1	1	1	1	1
S5	1	1	1	1	2	2
S6	1	2	1	1	1	1