



TERRORISM

빅데이터 분석 및 시각화 개론 Term-project

201411160 허건

201411000 이덕상

Problem Definition

- 테러가 대체로 언제 또는 어느 지역에서 많이 일어나는가?
- 주로 어느 나라가 어느 나라에게 테러를 많이 일으키는가?

Data Sets – Global Terrorism Database



Time period : 1970-2016

Quantity : 170,000+ Terrorist Attacks

Variables : >100 variables

(location, tactics, perpetrators, targets, and outcomes)

Sources : Unclassified media articles

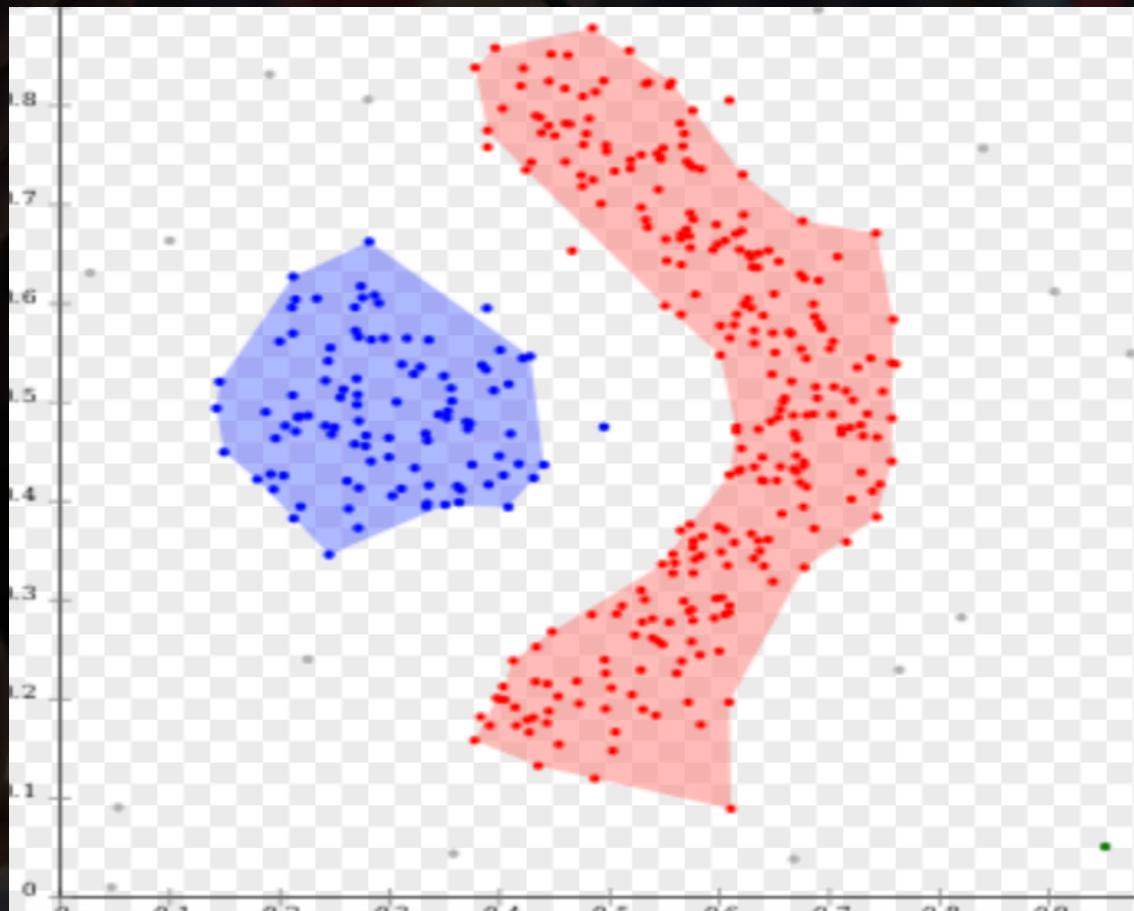
Data Sets – Global Terrorism Database

iyear,imonth,iday,approxdate,extended,resolution,country,country_txt,region,region_txt,provstate,city,latitude,longitude,specificity,vicinity,location,summary,crit1,crit2,crit3,doubtterr,alternative,alternative_txt,multiple,suicide,attacktype1,attacktype1_txt,attacktype2,attacktype2_txt,attacktype3,attacktype3_txt,targtype1,targtype1_txt,targsubtype1,targsubtype1_txt,corp1,target1,natlty1,natlty1_txt,targtype2,targtype2_txt,targsubtype2,targsubtype2_txt,corp2,target2,natlty2,natlty2_txt,targtype3,targtype3_txt,targsubtype3,targsubtype3_txt,corp3,target3,natlty3,natlty3_txt,gname,gsubname,gname2,gsubname2,gname3,gsubname3,motive,guncertain1,guncertain2,guncertain3,individual,nperps,nperpcap,claimed,claimmode,claimmode_txt,claim2,claimmode2,claimmode2_txt,claim3,claimmode3,claimmode3_txt,compclaim,weaptype1,weaptype1_txt,weapsubtype1,weapsubtype1_txt,weaptype2,weaptype2_txt,weapsubtype2,weapsubtype2_txt,weaptype3,weaptype3_txt,weapsubtype3,weapsubtype3_txt,weaptype4,weaptype4_txt,weapsubtype4,weapsubtype4_txt,weapdetail,nkill,nkillus,nkillter,nwound,nwoundus,nwoundte,property,propextent,propextent_txt,propvalue,propcomment,ishostkid,nhostkid,nostkidus,nhours,ndays,divert,kidhijcountry,ransom,ransomamt,ransomamtus,ransompaid,ransompaidus,ransomnote,hostkidoutcome,hostkidoutcome_txt,nreleased,addnotes,scite1,scite2,scite3,dbsource,INT_LOG,INT_IDEO,INT_MISC,INT_ANY,related

Example - 1970,7,2,,0,,58,Dominican Republic,2,Central America & Caribbean,,Santo Domingo,18.456792,-69.951164,1,0,,,1,1,1,0,,,0,1,0,1,Assassination,,,14,Private Citizens & Property,68,Named Civilian,,Julio Guzman,58,Dominican Republic,,MANO-D,,,0,,,0,,13,Unknown,,1,,0,,0,,,0,,0,,PGIS,0,0,0,

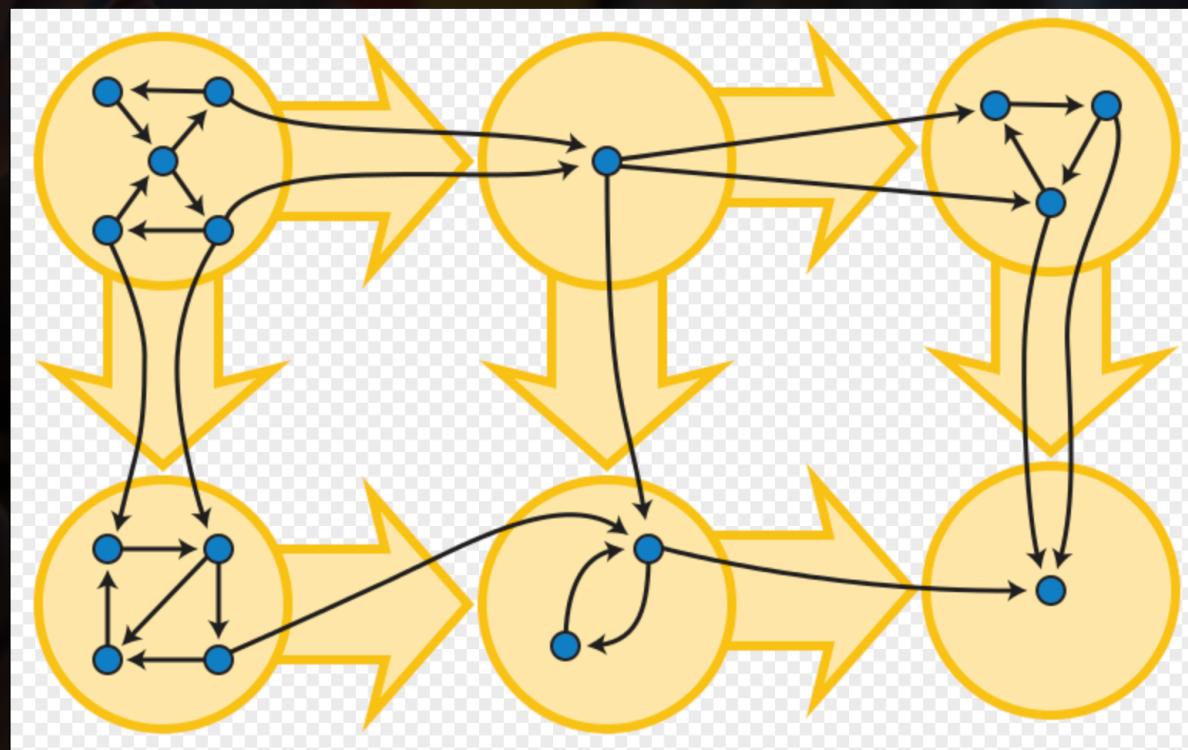
Proposed approach - Theoretical

Density based clustering algorithm(DBSCAN)



Proposed approach - Theoretical

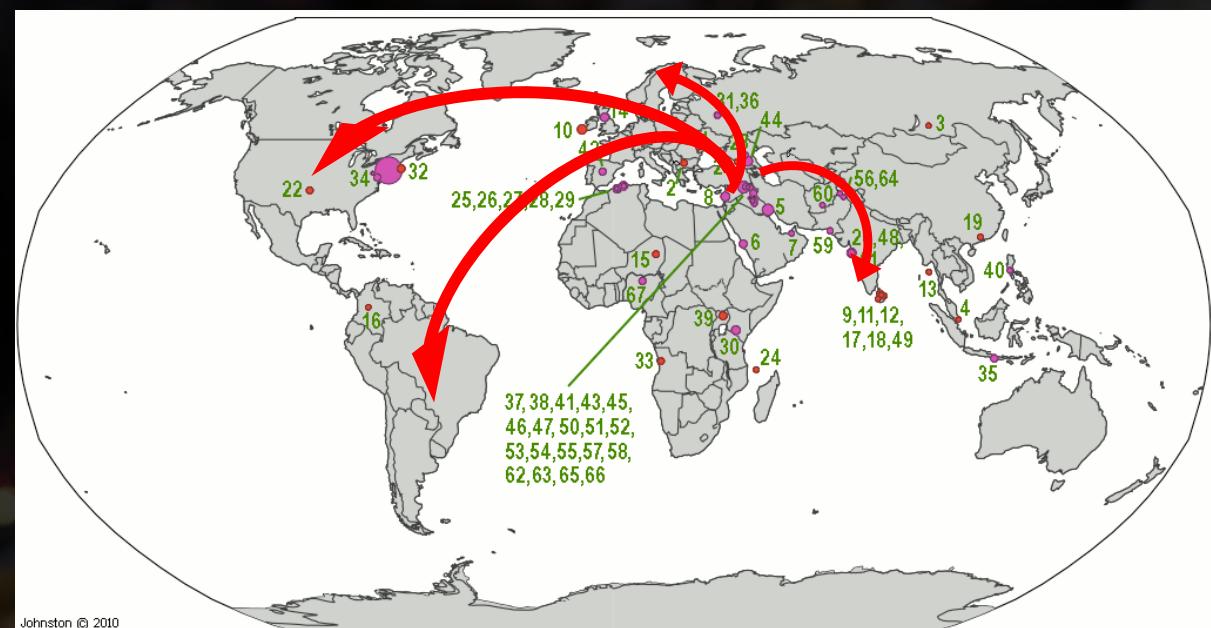
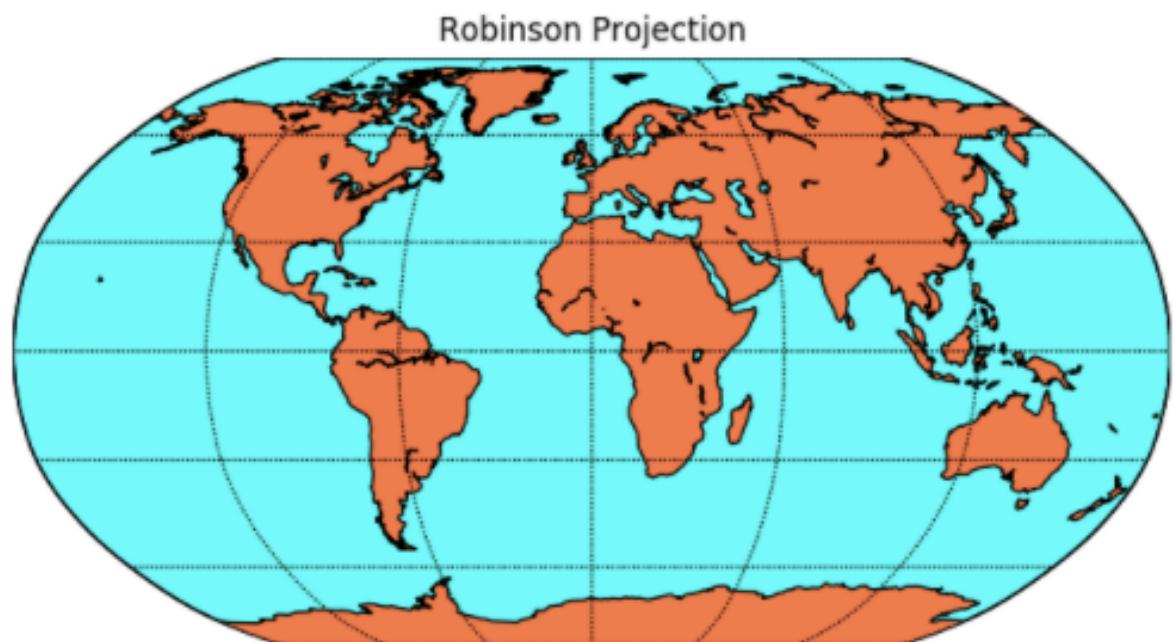
Graph Algorithm



Proposed approach - Practical

matplotlib

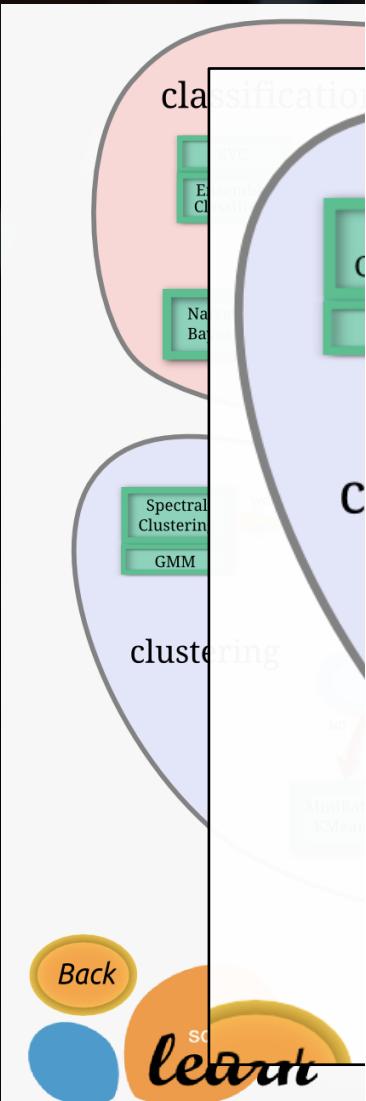
Basemap Toolkit



Proposed approach - Practical



Proposed a



scikit learn

Home Installation Documentation Examples

Google Custom Search Search x

sklearn.cluster.DBSCAN

```
class sklearn.cluster.DBSCAN (eps=0.5, min_samples=5, metric='euclidean', metric_params=None, algorithm='auto', leaf_size=30, p=None, n_jobs=1)
```

[source]

Perform DBSCAN clustering from vector array or distance matrix.

DBSCAN - Density-Based Spatial Clustering of Applications with Noise. Finds core samples of high density and expands clusters from them. Good for data which contains clusters of similar density.

Read more in the [User Guide](#).

Parameters:

eps : float, optional

The maximum distance between two samples for them to be considered as in the same neighborhood.

min_samples : int, optional

The number of samples (or total weight) in a neighborhood for a point to be considered as a core point. This includes the point itself.

metric : string, or callable

The metric to use when calculating distance between instances in a feature array. If metric is a string or callable, it must be one of the options allowed by `metrics.pairwise.calculate_distance` for its metric parameter. If metric is "precomputed", X is assumed to be a distance matrix and must be square. X may be a sparse matrix, in which case only "nonzero" elements may be considered neighbors for DBSCAN.

New in version 0.17: metric precomputed to accept precomputed sparse matrix.

metric_params : dict, optional

Additional keyword arguments for the metric function.

New in version 0.19

Expected Results

- 특정 기간동안 어떤 나라의 어떤 단체가 어떤 나라에 테러를 많이 일으켰는지 분석 및 시각화
- 테러가 자주 일어나는 지역에 대한 경향
 - **다음 테러가 일어날 가능성이 높은 지역 예측
 - **테러의 대략적인 주기 경향성과 앞으로 일어날 확률이 높은 시기 예측

Reference

- Density-Based Method
https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=0ahUKEwiny_bjkq_XAhUIXbwKHUGpD4gQFgg3MAE&url=https%3A%2F%2Fwww.cse.buffalo.edu%2F~jing%2Fcse601%2Ffa12%2Fmaterials%2Fclustering_density.pdf&usg=AOvVaw1pxrRhW3p676BNu0q6QZGS
 - https://www.google.co.kr/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwi3qpWTk6_XAhWMVrwKHbXCAQIQFgg0MAA&url=https%3A%2F%2Fen.wikipedia.org%2Fwiki%2FStrongly_connected_component&usg=AOvVaw0wgK4rqKwKKPls2axOjrs
- Scikit learn – clustering
- <http://scikit-learn.org/stable/modules/clustering.html>