NZRS
an InternetNZ company

Understanding traffic sources seen at .nz

Jing Qiao - NZRS

OARC 27

Heterogenous Sources

- DNS operators assume all sources sending queries are DNS resolvers
- Different types of sources
  - Big ISPs
  - Small ISPs
  - Monitoring hosts
  - Domainers
  - And many others
- Previous work: “In the search of resolvers”, OARC 25

Data Overview

- Total size
  - ns1 ~ ns4: 500K~600K unique source IPs each day
  - Now ns5 and ns6 also available thanks to CIRA
- Known samples (< 1K)
  - Resolvers
    - Local ISP IPs: 52
    - Google DNS IPs: 766
    - OpenDNS IPs: 90
  - Non-resolvers:
    - ICANN monitoring hosts: 60

Feature Engineering

- May ~ July 2017 (92 days)
- 30 features per source IP
  - fraction of active days, weekdays fraction
  - unique query types per day
  - unique domains queried per day, registered domains queried per day
  - fraction of common query types (A, AAAA, NS, MX, etc.)
  - fraction of RCODE (NXDOMAIN, NOERROR, REFUSED)
  - average number of queries per domain per day
  - number of queries per day
  - **means, stddev, percentiles of above measures**

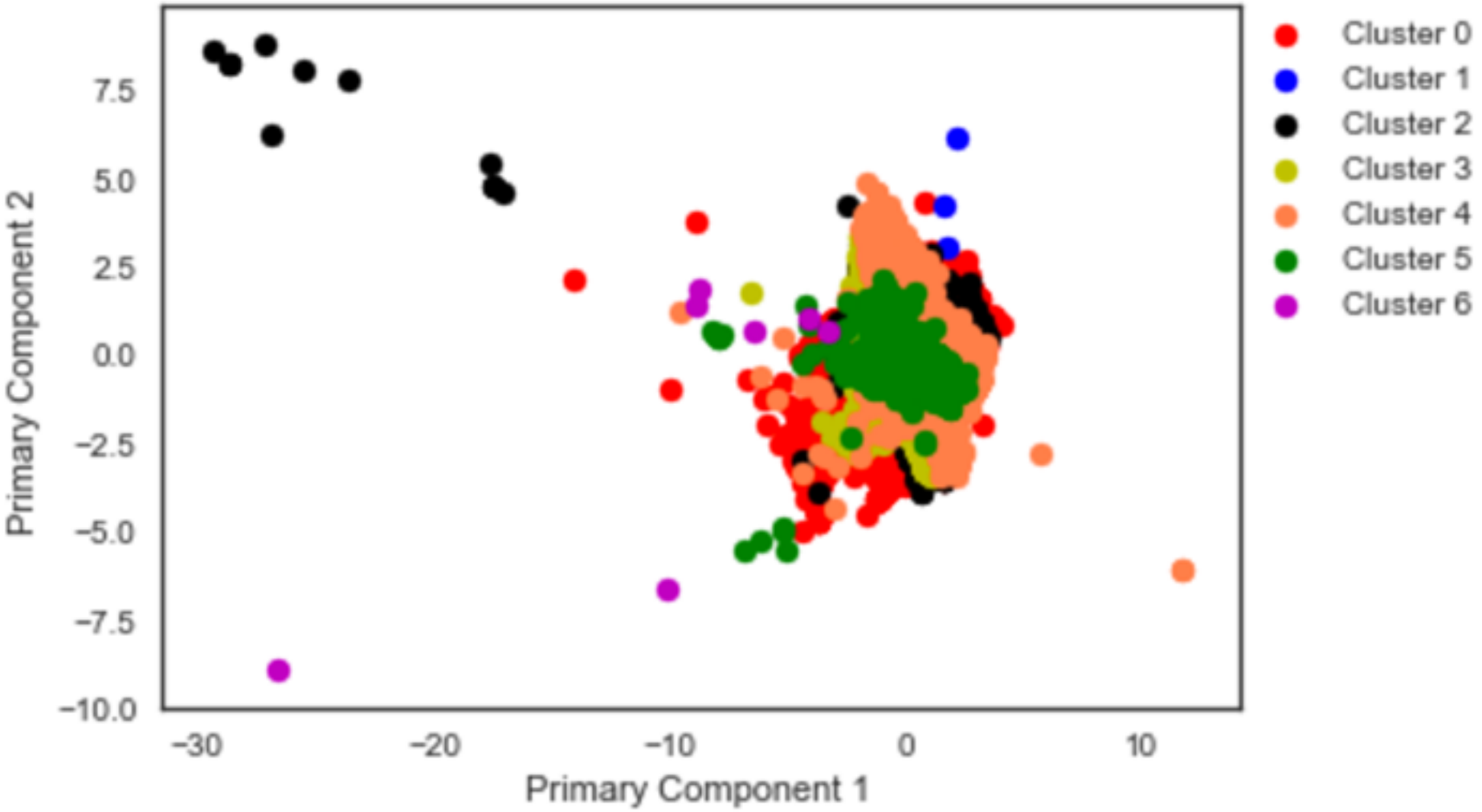
Machine Learning

- Data scaling: StandardScaler, MaxMinScaler, etc.
- Model selection:
  - KMeans (large sample size)
  - DBSCAN (memory expensive)
  - Hierarchical model (small sample size)
- Evaluate the clustering
  - homogeneity & completeness
  - silhouette score
  - verify with known samples
- Overfitting: PCA

Preliminary Results

- 7 clusters
  - All ICANN monitors fall into one cluster
  - Known resolvers splitted into 5 other clusters
    - ISPs of different scale
    - Validating resolvers
    - Highly distributed Google DNS
  - One cluster has no known data

Visualising high dimensions is challenging



Future Work

- Explore the pattern in each cluster
- Iterations of training with adjusted features and model parameters to find more ground truth
- Use the ground truth to improve the source classifier to recognise the resolvers
- Research into any interesting certain pattern revealed by the clustering

Thank You

Contact: Jing Qiao / jing@nzrs.net.nz
www.nzrs.net.nz

NZRS
an InternetNZ company