

۹۸/۹/۱۰

Anomaly Detection

کشف نابهنه های غیرمعمول

Some of the contents were taken from the authors of "Anomaly Detection : A Survey ACM Computing Surveys, Vol. 41(3), Article 15, July 2009 and Anupam Das CS 568MCC Spring 2013

December 1, 2014

Introduction

- ♦ We are drowning in the deluge of data that are being collected world-wide, while starving for knowledge at the same time
- ♦ Anomalous events occur relatively infrequently
- ♦ However, when they do occur, their consequences can be quite dramatic and quite often in a negative sense



What are Anomalies?

(اقلیتی از مشاهدات در مجموعه داده که الگوی متفاوتی با اکثریت داده ها دارند)

- Anomaly is a pattern in the data that does not conform to the expected behaviour
- Also referred to as outliers, exceptions, leverage points, surprise, etc.
- Anomalies translate to significant (often critical) real life entities

Department of Computer Science, UJUC

December 1, 2014

Real World Anomalies

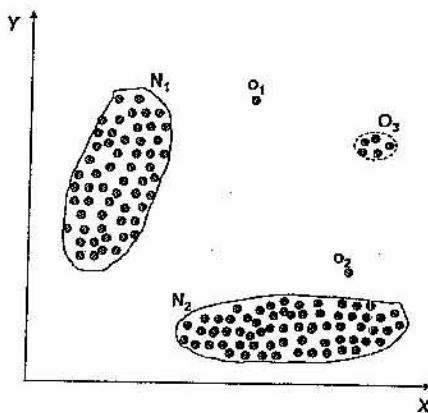
- Credit Card Fraud
- Cyber Intrusions
- Healthcare Informatics / Medical diagnostics
- Industrial Damage Detection
- Image Processing / Video surveillance
- Novel Topic Detection in Text Mining



December 1, 2014

Simple Example

- N_1 and N_2 are regions of normal behavior
- Points o_1 and o_2 are anomalies
- Points in region O_3 are anomalies



Department of Computer Science, UIUC

Key Challenges

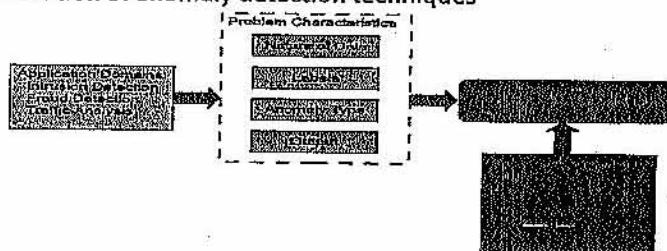
- Defining a representative normal region is challenging
- The boundary between normal and outlying behavior is often not precise
- The exact notion of an outlier is different for different application domains
- Availability of labeled data for training/validation
- Malicious adversaries
- Data might contain noise
- Normal behavior keeps evolving

Department of Computer Science, UIUC

December 1, 2014

Aspects of Anomaly Detection Problem

- Nature of input data
- Availability of supervision
- Type of anomaly: point, contextual, structural
- Output of anomaly detection
- Evaluation of anomaly detection techniques



Department of Computer Science, UIUC

December 1, 2014

Input Data

- Input Data could be
 - Univariate : single variable
 - Multivariate: multiple variable
- Nature of attributes
 - Binary
 - Categorical
 - Continuous
 - Hybrid

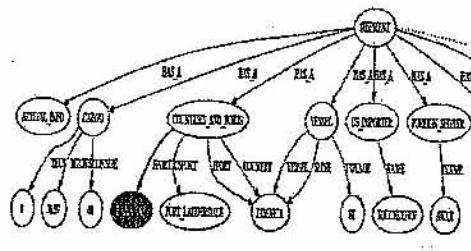
	String	Double	Double	Number of bytes	Interpreted
1	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
2	206.163.37.89	160.94.179.243	160.94.179.243	16	IP
3	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
4	160.94.123.45	160.94.179.243	160.94.179.243	16	IP
5	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
6	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
7	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
8	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
9	206.163.37.11	160.94.179.243	160.94.179.243	16	IP
10	206.163.37.11	160.94.179.243	160.94.179.243	16	IP

Department of Computer Science, UIUC

December 1, 2014

Input Data – Complex Data Types

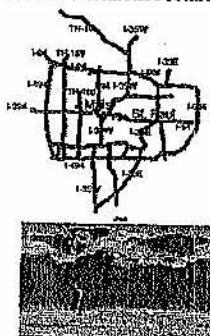
- Relationship among data instances
 - Sequential
 - Temporal
 - Spatial
 - Spatio-temporal
 - Graph



Department of Computer Science, UIUC

December 1, 2014

GCTTCCCGCTTGACGCCCCCG
 CGCAAGGCCCGCCCGCGCGCG
 GAGAAGGGCGCGCGCGCGCG
 GGCGGGAGGGGGGGCGCGCGCG
 GCAACCGAGTCGCACCCAGCTGCC
 GCGCTCGCTCGCGCTTGACCGCG
 GCGCATTAAGGGCGAGCGCGACG
 GCGCACTGAGACACCCGCGCGCG



Department of Computer Science, UIUC

December 1, 2014

Data Labels

- Supervised Anomaly Detection
 - Labels available for both normal data and anomalies
 - Similar to rare class mining
- Semi-supervised Anomaly Detection
 - Labels available only for normal data
- Unsupervised Anomaly Detection
 - No labels assumed
 - Based on the assumption that anomalies are very rare compared to normal data

Department of Computer Science, UIUC

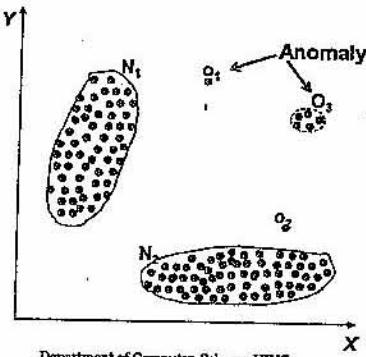
December 1, 2014

Type of Anomaly

- Point Anomalies
- Contextual Anomalies
- Collective Anomalies

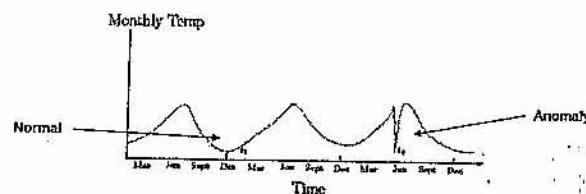
Point Anomalies

- An individual data instance is anomalous if it deviates significantly from the rest of the data set.



Contextual Anomalies

- An individual data instance is anomalous within a context
- Requires a notion of context
- Also referred to as conditional anomalies*

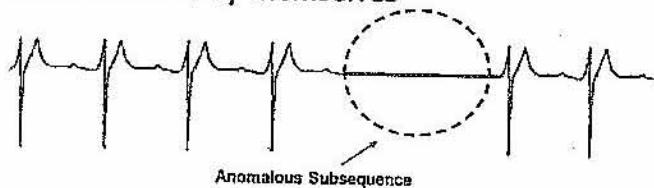


Department of Computer Science, UIUC

December 1, 2014

Collective Anomalies

- A collection of related data instances is anomalous
- Requires a relationship among data instances
 - Sequential Data
 - Spatial Data
 - Graph Data
- The individual instances within a collective anomaly are not anomalous by themselves



Department of Computer Science, UIUC

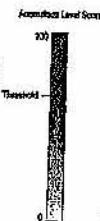
December 1, 2014

طبعات اسفلتیاتی ریاضیات فیبر لاین پر
 تابعیت فیبر لاین اسفلتیاتی کوی
 (طباطبیت اسفلتیاتی فیبر لاین) طبقه بندی: FP

12/1/2014

Output of Anomaly Detection

- Label
 - Each test instance is given a *normal* or *anomaly* label
 - This is especially true of classification-based approaches
- Score
 - Each test instance is assigned an anomaly score
 - Allows the output to be ranked
 - Requires an additional threshold parameter



Department of Computer Science, UIUC

December 1, 2014

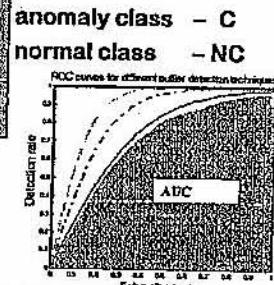
P = Precision

FP = False Precision

Evaluation of Anomaly Detection

- ♦ Accuracy is not sufficient metric for evaluation
 - Example: network traffic data set with 99.9% of normal data and 0.1% of intrusions
 - Trivial classifier that labels everything with the normal class can achieve 99.9% accuracy !!!!!

CONFUSION MATRIX		DEPICTED BY	
		ANOMALY CLASS	NORMAL CLASS
ANOMALY CLASS	NORMAL CLASS	TP	FN
ANOMALY CLASS	NORMAL CLASS	TP	FN

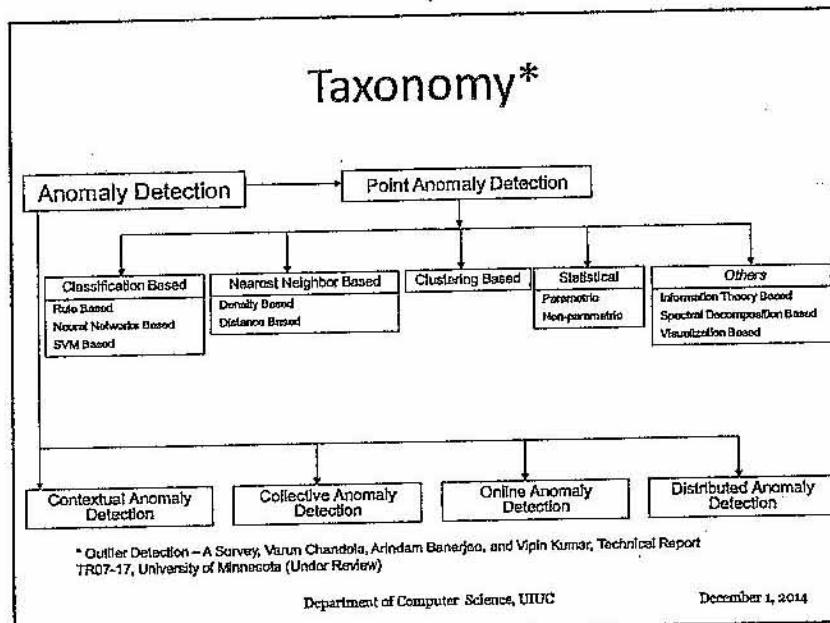


Department of Computer Science, UIUC

December 1, 2014

- Focus on both recall and precision
 - Recall /Detection (R)= $TP/(TP + FN)$
 - Precision (P)= $TP/(TP + FP)$
 - False rate (F)= $FP/(TN+FP)$

طبعات اسفلتیاتی ریاضیات فیبر لاین پر
 تابعیت فیبر لاین اسفلتیاتی کوی
 (طباطبیت اسفلتیاتی فیبر لاین) طبقه بندی: FP



Classification Based Techniques

- Main idea: build a classification model for normal (and anomalous) events based on labelled training data, and use it to classify each new unseen event
- Classification models must be able to handle skewed (imbalanced) class distributions
- Categories:
 - *Supervised classification techniques*
 - Require knowledge of both **normal** and **anomalous** classes
 - Build classifier to distinguish between normal and known anomalies
 - *Semi-supervised classification techniques*
 - Require knowledge of **normal** class only!
 - Use modified classification model to learn the normal behavior and then detect any deviations from normal behavior as anomalous

Pros and Cons

Department of Computer Science, UIUC December 1, 2014

Classification Based Techniques

- Some techniques
 - Neural network based approaches
 - Support Vector machines (SVM) based approaches
 - Bayesian networks based approaches
 - Rule based techniques
 - Fuzzy Logic
 - Genetic Algorithms
 - Principle Component Analysis

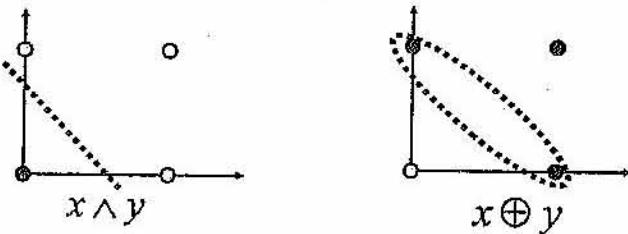
Department of Computer Science, UIUC

December 1, 2014

Using Neural Networks

Multi-layer NNs

- Creating hyper-planes for separating between various classes
- Good when dealing with huge data sets and handles noisy data well
- Bad because learning takes a long time

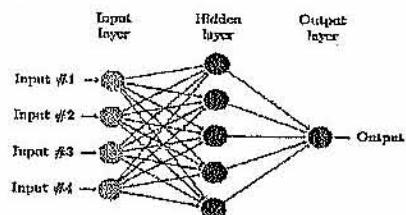


Department of Computer Science, UIUC

December 1, 2014

Neural Networks

- **INPUT LAYER-** $X = \{x_1, x_2, \dots, x_n\}$, where n is the number of attributes. There are as many nodes as no. of inputs.
- **HIDDEN LAYER** – the number of nodes in the hidden layer and the number of hidden layers depends on implementation.
- **OUTPUT LAYER** – corresponds to the class attribute. There are as many nodes as classes.
- **Back Propagation** learns by iteratively processing a set of training data (samples).



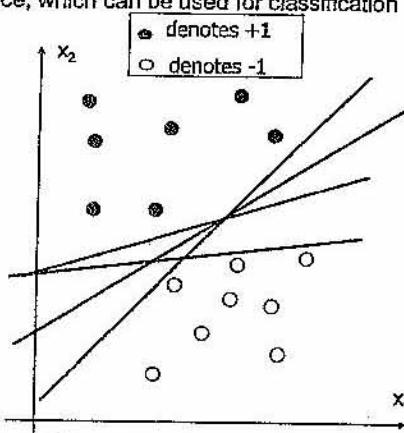
Department of Computer Science, UIUC

December 1, 2014

SVM

Support vector machine constructs a hyperplane or set of hyperplanes in a high or infinite-dimensional space, which can be used for classification

- How would you classify these points using a linear discriminant function in order to minimize the error rate?
- Infinite number of answers!
- Which one is the best?



Department of Computer Science, UIUC

December 1, 2014

Linear SVM

- denotes +1
- denotes -1

Support Vectors are those data points that the margin pushes up against

- Why it is the best?
 - Robust to outliers and thus strong generalization ability

Department of Computer Science, UIUC

Linear SVM

- Formulation:

$$\mathbf{w}^T \mathbf{x}^+ + b = 1$$

$$\mathbf{w}^T \mathbf{x}^- + b = -1$$
- The margin width is:

$$M = (\mathbf{x}^+ - \mathbf{x}^-) \cdot \mathbf{n}$$

$$= (\mathbf{x}^+ - \mathbf{x}^-) \cdot \frac{\mathbf{w}}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}$$
- Goal:

$$\text{maximize } \frac{2}{\|\mathbf{w}\|}$$

$$\text{such that } \forall i [(\mathbf{w}^T \mathbf{x}_i + b) \geq 1]$$

● denotes +1
 ○ denotes -1

Quadratic programming with linear constraints

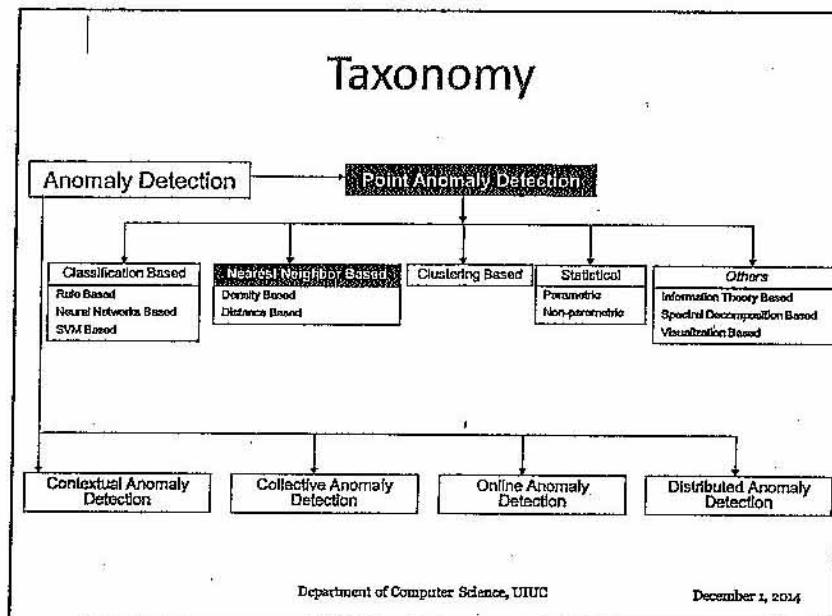
x₂

x₁

Margin

Department of Computer Science, UIUC

December 1, 2014



Nearest Neighbor Based Techniques

- **Key assumption:** normal points have close neighbors while anomalies are located far from other points
 - General two-step approach
 1. Compute neighborhood for each data record
 2. Analyze the neighborhood to determine whether data record is anomaly or not
 - Categories:
 - Distance based methods
 - Anomalies are data points most distant from other points
 - Density based methods
 - Anomalies are data points in low density regions



Distance Based Anomaly Detection

For each object o , examine the # of other objects in the r -neighborhood of o , where r is a user-specified **distance threshold**

An object o is an outlier if most (taking π as a **fraction threshold**) of the objects in D are far away from o , i.e., not in the r -neighborhood of o

An object o is a $DB(r, \pi)$ outlier if

$$\frac{\|\{o' | dist(o, o') \leq r\}\|}{\|D\|} \leq \pi,$$



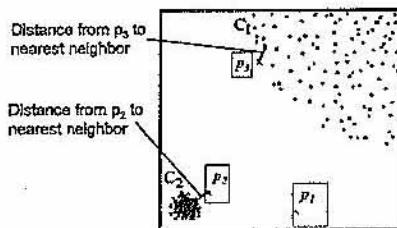
Department of Computer Science, UIUC

December 1, 2014

Density Based Anomaly Detection

Compute local densities of particular regions and declare instances in low density regions as potential anomalies

Approach: **Local Outlier Factor (LOF)**



In the **NN** approach, p_2 is not considered as outlier, while the **LOF** approach find both p_1 and p_2 as outliers

NN approach may consider p_3 as outlier, but LOF approach does not

Department of Computer Science, UIUC

December 1, 2014

Local Outlier Factor (LOF)

For each data point o compute the # of points in k -distance:

$$N_k(o) = \{o' | o' \in D, \text{dist}(o, o') \leq \text{dist}_k(o)\}.$$

Compute *reachability distance (reachdist)* for each data example o with respect to data example o' as:

$$\text{reachdist}_k(o \leftarrow o') = \max\{\text{dist}(o), \text{dist}(o, o')\},$$

Compute *local reachability density (lrd)*:

$$\text{lrd}_k(o) = \frac{\|N_k(o)\|}{\sum_{o' \in N_k(o)} \text{reachdist}_k(o' \leftarrow o)}.$$

LOF is the average of the ratio of local reachability density of o 's k -nearest neighbors and local reachability density of the data record o

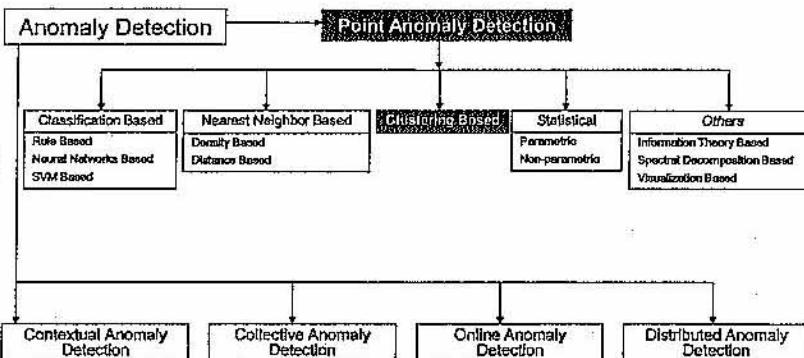
$$\text{LOF}_k(o) = \frac{\sum_{o' \in N_k(o)} \frac{\text{lrd}_k(o')}{\text{lrd}_k(o)}}{\|N_k(o)\|}$$

Higher the LOF the more likely its an outlier

Department of Computer Science, UIUC

December 1, 2014

Taxonomy



Department of Computer Science, UIUC

December 1, 2014

Clustering Based Techniques

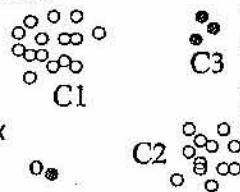
- **Key assumption:** normal data records belong to large and dense clusters, while anomalies belong do not belong to any of the clusters or form very small clusters
- **Categorization according to labels**
 - Semi-supervised – cluster normal data to create modes of normal behavior. If a new instance does not belong to any of the clusters or it is not close to any cluster, is anomaly
 - Unsupervised – post-processing is needed after a clustering step to determine the size of the clusters and the distance from the clusters is required for the point to be anomaly
- **Anomalies detected using clustering based methods can be:**
 - Does not belong to any cluster,
 - Large distance between the object and its closest cluster
 - Belongs to a small or sparse cluster

Department of Computer Science, UIUC

December 1, 2014

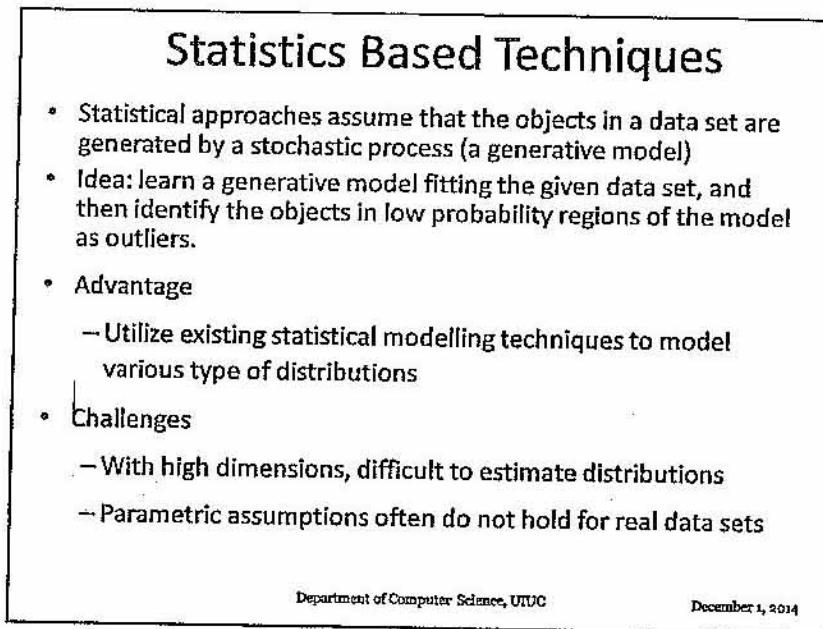
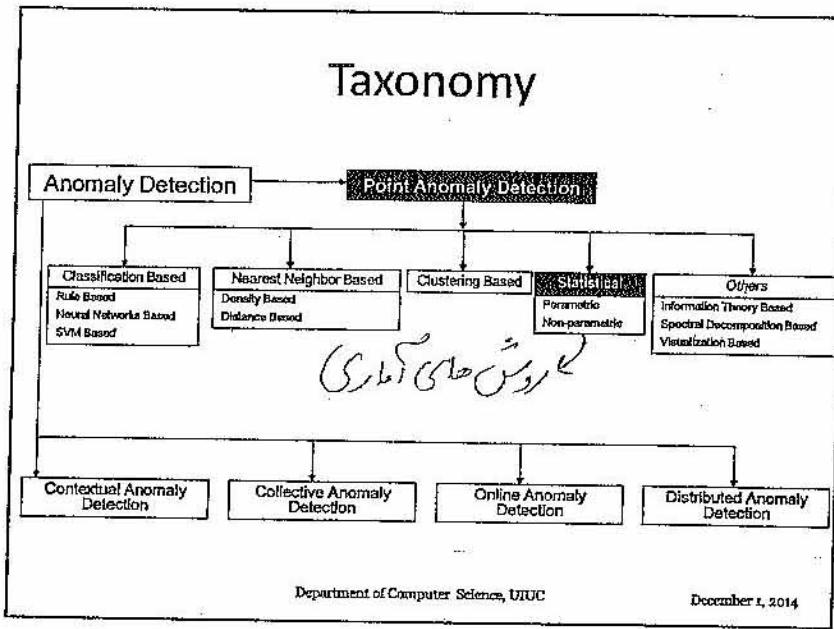
Cluster Based Local Outlier Factor

- **FindCBLOF:** Detect outliers in small clusters
 - Find clusters, and sort them in decreasing size
 - To each data point, assign a *cluster-based local outlier factor* (CBLOF):
 - If obj p belongs to a large cluster, CBLOF = cluster_size X similarity between p and cluster
 - If p belongs to a small one, CBLOF = cluster size X similarity betw. p and the closest large cluster
- Ex. In the figure, o is outlier since its closest large cluster is C₁, but the similarity between o and C₁ is small. For any point in C₃, its closest large cluster is C₂ but its similarity from C₂ is low, plus |C₃| = 3 is small



Department of Computer Science, UIUC

December 1, 2014



Types of Statistical Techniques

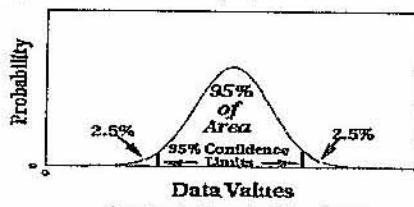
- Parametric Techniques
 - Assume that the normal data is generated from an underlying parametric distribution
 - Learn the parameters from the normal sample
 - Determine the likelihood of a test instance to be generated from this distribution to detect anomalies
- Non-parametric Techniques
 - Do not assume any knowledge of parameters
 - Not completely parameter free but consider the number and nature of the parameters are flexible and not fixed in advance
 - Examples: histogram and kernel density estimation

Department of Computer Science, UIUC

December 1, 2014

Parametric Techniques

- Univariate data: A data set involving only one attribute or variable
- Often assume that data are generated from a normal distribution, learn the parameters from the input data, and identify the points with low probability as outliers
- Ex: Avg. temp.: {24.0, 28.9, 29.0, 29.1, 29.1, 29.2, 29.2, 29.3, 29.4}
 - Compute μ and σ from the samples



Department of Computer Science, UIUC

December 1, 2014

Parametric Techniques

- Univariate outlier detection: The Grubb's test (another statistical method under normal distribution)
- For each object x in a data set, compute its z-score:

$$z = \frac{|x - \bar{x}|}{s}$$

- Now x is an outlier if

$$z \geq \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\alpha/(2N), N-2}^2}{N-2 + t_{\alpha/(2N), N-2}^2}}$$

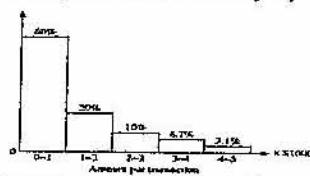
where $t_{\alpha/(2N), N-2}^2$ is the value taken by a t-distribution at a significance level of $\alpha/(2N)$, and N is the # of objects in the data set

Department of Computer Science, UIUC

December 1, 2014

Non-Parametric Techniques

The model of normal data is learned from the input data without any *a priori* structure.



Outlier detection using histogram:

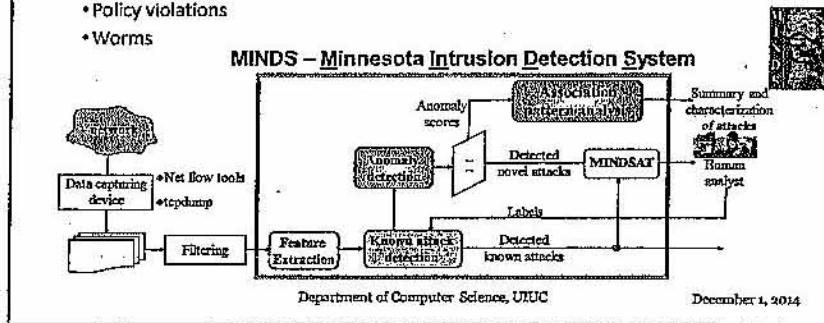
- Figure shows the histogram of purchase amounts in transactions
- A transaction in the amount of \$7,500 is an outlier, since only 0.2% transactions have an amount higher than \$5,000
- Problem: Hard to choose an appropriate bin size for histogram
- Solution: Adopt kernel density estimation to estimate the probability density distribution of the data.

Department of Computer Science, UIUC

December 1, 2014

Anomaly Detection on Real Network Data

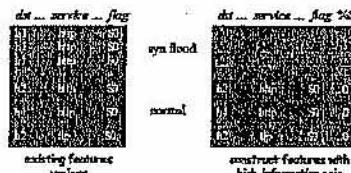
- Anomaly detection was used at U of Minnesota and Army Research Lab to detect various intrusive/suspicious activities
- Many of these could not be detected using widely used intrusion detection tools like SNORT
- Anomalies/attacks picked by MINDS
 - Scanning activities
 - Non-standard behavior
 - Policy violations
 - Worms



Feature Extraction

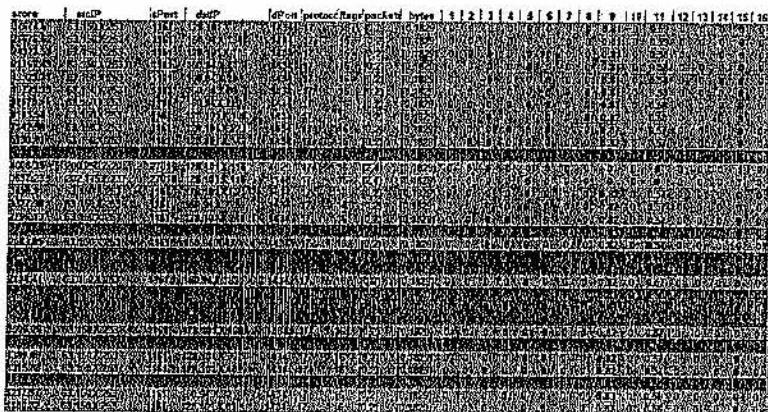
- Three groups of features
 - Basic features of individual TCP connections

• source & destination IP	Features 1 & 2
• source & destination port	Features 3 & 4
• Protocol	Feature 5
• Duration	Feature 6
• Bytes per packets	Feature 7
• number of bytes	Feature 8
 - Time based features
 - For the same source (destination) IP address, number of unique destination (source) IP addresses inside the network in last T seconds – Features 9 (13)
 - Number of connections from source (destination) IP to the same destination (source) port in last T seconds – Features 11 (15)
 - Connection based features
 - For the same source (destination) IP address, number of unique destination (source) IP addresses inside the network in last N connections - Features 10 (14)
 - Number of connections from source (destination) IP to the same destination (source) port in last N connections - Features 12 (16)



Typical Anomaly Detection Output

- 48 hours after the "slammer" worm



Department of Computer Science, UIUC

December 1, 2014

Conclusions

- Anomaly detection can detect critical information in data
- Highly applicable in various application domains
- Nature of anomaly detection problem is dependent on the application domain
- Need different approaches to solve a particular problem formulation

Department of Computer Science, UIUC

December 1, 2014

Related problems

- Rare Class Mining
- Chance discovery
- Novelty Detection
- Exception Mining
- Noise Removal
- Black Swan*

Department of Computer Science, UIUC

December 1, 2014

Subject:

Year: Month: Date: ١٢

بيانات Data warehousing

بيانات متعددة من مصادر مختلفة مثل قاعدة بيانات MySQL و PostgreSQL.

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

Data warehouse

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

Data warehouse

بيانات متغيرة في كل يوم، مثل حركة المرور على موقع ويب.

Data warehouse

Data warehouse هي بيئة موجهة نحو الموضوع.

Data warehouse هي بيئة موجهة نحو الموضوع.

Subject:

Year. Month. Date. ()

ایجاد (Create) داده ایل (Data) warehouse

ویرایش (Update) داده ایل (Data) warehouse

حذف (Delete) داده ایل (Data) warehouse

اندیکس (Index) کردن داده ایل (Data) warehouse

محدود (Restrict) کردن داده ایل (Data) warehouse

DBMS db14105, b1

ایجاد (Create) داده ایل (Data) warehouse time-Variant

ویرایش (Update) داده ایل (Data) warehouse

حذف (Delete) داده ایل (Data) warehouse

DBMS 1 Datawarehouse قریب

PAPCO

درینه نیازمند بـ جای افسـه نابـع خـطاـسـه طـارـمـه رـا دـاشـتـه باـشـم اـز هـسـرـنـه
بلـدـه بـرـگـردـه اـنـبـاع خـطاـسـه اـسـنـه وـسـمـه.

بروش حاکم آنها به دوستی قائم می‌شوند:

بروش حاکم آنها پارامتری ۲- برش حاکم آنها نایاب امری

بروش علی شبیه علی علیم را درین مرتبه جزو شرکت نایاب امری می‌شوند

بروش علی شبیه علی علیم را درین مرتبه جزو شرکت نایاب امری می‌شوند
بروش علی شبیه علی علیم را درین مرتبه جزو شرکت نایاب امری می‌شوند Free from Distribution

بروش علی پارامتری جواب آن درست است لایه کن منتهی خواهد بود و

حصه از نفع علی او نیز درست باشد بجز این

بروش علی شبیه علیم را درین مرتبه جزو شرکت نایاب امری می‌شوند

آن likelihood را که در این مرتبه جزو شرکت نایاب امری می‌شوند

تفصیل در محدوده اینجا اینجا نیست

بروش نایاب امری طبق ۱۸ بدل این شبیه علیم را درین مرتبه مائمه

منه این در این است که نایاب امری دستگاه کلست این شبیه علیم را می‌شوند

با این برش صورت

خوبی رشتهای کمالی ایضاً است از جمله دستم از اینها خود مطلع نیز نیز

ساده ساخته می‌شوند
اگر راهی در میان آنها که نسبت پلند چنین رشتهای کمالی می‌شوند می‌توانند
دینامیکی نباشند و شرایط نیاز ایشان
درینهای خود را در طور معمول برای اینها می‌توانند

در صورت استفاده از روش های برگبندی ملکه بار چنین فرض اویسی

حالت نهم:

✓ طبقه ای از مواردی به همینام از کتاب ترمهاسن مطابق رسادا که حضرت

است.

✓ در اینجا خواسته می شود این داشته باشند

✓ در رساد Category لازمه طبقه باشند بتوانند رساد کتاب ترمهاسن

رویاظ نیم. درینکجا انتفاف نیز برایمده

در صورتی که این بخواسته باشد رساد روش ملک خوشبختی

رسام را بخواهد رساد خوشبختی که این دلیل اینکه در عین رساد رساد

ز طوری ریاضت این روش دلیلی خواهد بود

درانی برای سه ماه مسکونی در چهار کشور طبق دارم که پیش داشت
که فرستاد سه ماه مسکونی در این کشور آن تا ماه ابرil است که
میری بزمی با خدا آشنا شویم و آن کار را بگام بردند.

کشناختی از راه و مسیر طاری در سال ۱۷۰۰ درست شد که
در آن با خود رفته اند و مراقب می شوند جواب خواهند
گرفتند و این را می بینند

حداد و امداد احمد فرشتگان
در سرکاری نهاد، طاوله‌ای پرست با عشیره سرمه مادرسته. در امارت ناصبی
بنجیم سیم کان طاوله این درجه نداش. اداره طاره طاری با طاوله کش ناصبی کار
کش اداره کار این احترم بود اداره طاره طاری به این درجه کش محل بدل لرستان

باید حضری کن رفته سوون
حال بسیار سخن درباره طاوه، محل اینبار طاوه است رسه در سرعن
باید حضری در کل سخنی و دسته همچو این سرود.
در کتابرو نادری مائش و مادر طاوه نامهایی مهدیان

داده ۴۰ ناچیار زیاده غریب شد و درین ریخت آنها را به شد که انداد
نمایم که برخلاف آنست و درین میل مادر لزمه بدل شوند و اینها میزد

مختصر است. مفهومی نایر جنگ، بادا و میکروپلی نشان دارد در رسم و دست نمایش
برای ماده (نایر جنگ از درودره) Leverage Point، outlayer (نقطه احتراز) است
استفاده می شوند. در ماده (نایر جنگ) مفهومی نایر جنگ در خط رسم صندوق و در خط
Leverage Point، outlayer مفهومی نایر جنگ باشند خواهند داشت. missing Data

لارش نهاد بارگ دیگر نمی خواهد باشد همچو که شد. بارگ از مردم.

چهارمین مادل برای دستگاه اسکان و سیم ۸
از داشت که نایاب است و از این داشت که این دستگاه
Model based detection

نایاب است و داشت
پیشنهادی SVM، شبکه عصبی، متریک داده، متریک برای کار خالص
کاربری، PCA،
برای کار خالص بازگشت

برای کار خالص

خود قبیض دهد و در شش خوب است. برای مانع بگیرید که این زنگنه را
که این کوشش را تقویت نمایند کنید / ۳ برجسته باشم و سه چهل
در پروری چهل کوشش امده است هست (اگر ماده شدند پس نهایت پیشین
بلطفت ماه علی نایب صغار ایش طام)

تاخنومی و زیست دلگزینی تفصیل نایب صغار ایش

زیست محیطی برای نیمه سینه پرور کردن بیش برای نسبت نایب صغار ایش
کل میان عده موادی که در تغذیه کاروی میباشد و
بنابراین با این خوب برای ماده ایش که از ماده علی کن کمتر
نمیتواند را ایام داشد. در اینجا غیر و دشوار داشته باشند که این
ایجاد کرد. چنان که خوش نیم کار این اتفاقی آن خوب نمیباشد.
برای این امر دو مفترض را در میانمایم خوب یا نهایت کم.
در درون دسته های ملی بین دیگرین تراک لذت گشتن میباشد بعنوان خطاوی
کام نیز خوب استه کرد علی این ۸۹ بام ایشانه کرد

حدرویه Anomaly Detection می‌بیند که ابتدا است و در نتیجه بدلان باشند همان‌ست کرد
ویسیه دارد و در هر عرضه این را بینم. علاوه بر رسمیت عدد در فحصیه (نحوه)
یافته ارسال شده است که آن بعدم دلیل داشت با اینکه فحص در مردم درست نباشد

برای داده فحصیه بگیر

برای رعایت این خواسته روش مانع: روش مدلی در طبقه‌بندی روش ساده‌بیسیه جعلی
کارآمد است که براساس آن تغییر در خود را که در آن طبقه‌بندی می‌تواند محدود نباشد
آنرا انتخاب نماییم؟ معلم این است که مدلی داده تابع خوارجیم

که در آن طبقه‌بندی مشبه استفاده نماییم از درجه‌بندی همچو ۹۹٪ درجه‌بندی
می‌شود. اگر روش مدلی باشی که درین اسماه کم مدل‌بندی می‌نماید
آن‌ها در درجه‌بندی مدل‌بندی که انتخاب داده بیسیه همچو
آن‌ها در درجه‌بندی مدل‌بندی که انتخاب داده بیسیه همچو

(نحوه) طبعی که از مدل نزدیک (نحوه) طبعی که از مدل نزدیک (نحوه) طبعی

می‌نماید (NC). این نسبت داده است اما این مدل می‌بیند منی بینم که در
می‌نماید NC و دست است. FP نیز را آنچه آنچه فرامل می‌نماید اما این می‌نماید
FP می‌نماید همچو در مدل می‌نماید شو است. آنچه که فرامل می‌نماید در مدل
که از مدل می‌نماید شو است این را نیز در مدل می‌نماید FP می‌نماید. در آنچه که از مدل
می‌نماید FP می‌نماید زنگ ایجاد کریم FN می‌نماید. آنچه که داده زنگ ایجاد کریم
ما فرامل می‌نماید که درین می‌نماید که از مدل می‌نماید بسیار درین می‌نماید که می‌نماید

اگر بایکسیکل نباشد که خود را می‌تواند پنهان نماید سپس بدینها
Anomaly Detection / Change Point Detection
همچو خود را درست نماید می‌تواند
outlier ، leverage Point یا درست نماید
از هر چیز نفع نماید . می‌تواند اینجا همچو خود را
باشد (مثل سرمه ایست) نماید یا درست نماید می‌تواند
و منظمه باشند و این بصیرت جماعتی از داده های
که داشته باشند و یک جایی نسبت خاصیتی داشته باشند

بجود طبقات صخريات مع

نقطة انomalie point Anomaly
هي نقطة طبيعية يزيد حجمها عن المعدل الاجمالي

در خود کارهای باتوجهی سه سوکارهایم. مثل DNA که ۴۰٪ از این اسید را تشکیل می‌دهد آنها (مثل اسید عروزانی) اما هشت سه کم آنقدر در میان مشتقات هم ندارد بلکه در دینامیک این اسید را تشکیل می‌نمایند.

داده های مکانی، داده های نسبتی، داده های ماتریسی و داده های طبقه بندی شده (Spatial Data) هستند.

(اده) عزیز امیر متفاوت از راه دیگر زمانی است. علت است خادمه عجایبی بیش از ۱۰۰۰ مایل باشد
مثل آندریه دشتربر (Dr. Andrei Shchastlivyy) در مطالعات خود اندیشی در زمان پاره خلیق
خود را که احوال طرد و زمان هم دارد تفسیر نمود. مثل طاده عزیز شیوه ایجاد این کوچک
کوچک را می توان از زمانی ساخت (ست در زمانی).

کن نایم هزاری بان اخراج دسته عمل بند است که وارد بول شوئی از دهداره
آن کا گاراگر عمل شوئی کرد. لایک بحسب بول شوئی برگان هنوز. عمل ما
(رسان) (طبیعت) بول شوئی از دهن

بررسی خودکار (Self-supervised Learning) برای پنهان کردن این خطا است. در این روش، مدل برای این خطا را با آنچه معمول است در داده‌ها نمایش نمی‌دهد و مدل این خطا را با آنچه معمول نیست نمایش می‌دهد. این مدل معمول را می‌توان از مجموعه داده‌های معمول (Normal Data) یا مجموعه داده‌های غیرمعمول (Anomaly Data) بدست آوردن. این مدل معمول می‌تواند این خطا را پنهان کند.

دائرہ ادارہ unsupervised ہے جو جس سطح پر داد و صورت کا درجہ اعلیٰ ہے اس کے لئے سارے داد و صورت کا نمونہ کو کامیاب کرنے کے لئے برداشت کرنے اور اس کو ایک classification کا نمونہ کر کے پسندیدہ کرنے کا کام کرتا ہے۔

بنای سیاست خود را نایاب همچنان باید جست داده و این بسیم و اعمال تقدیرت برخواهد
که از این شرط باشیم، فرع داده علی نایاب همچنان بجهود کنندگان میتوانیم این مطلب
را مشخص کنیم. در فرع خوبیم طلب این شرط کنیم
داده اول میتواند را صفت است که میتوان با boxPlot داده علی نایاب همچنان را مشخص کرد
که چهار چیز هست که خط باقیمانده چهار رسم، خط درست چهار کوچک داریم اما این خطوط باقیمانده
چهار اول است در ناخالیم هست که خط باقیمانده برابر است. هر داده ای که مشخص از این
دو نعمت از این شرط است.

داده عزیز خان میرزا محمد و میرزا علی (است) واحد اعلیه ریگ اکنام حکم داشت باشد اسرار
پرنس، بنت، نشسته، میتوانند در کاربری اکنام مشغول باشند.

اگر راههاییم را خوب نمایند بالا ساخته و نزدیک ساده طبقه ایم را خود را بخواهیم باشیم خیلی خوب بر
حضره دانیش. اگر سعادت یعنی خوشبود شده باشیم ماضی داشته باشند مارا دچار مشکل
نمایند.

داده‌های نابهنجار را به راهی عنوان کنید. در اینجا داده‌ها در دیدگاه راهی معرف کنیم.
در داده‌های آماری از دسته دیگر داده‌های خودکار است که در آنها داده‌ها از یکدیگر جدا شوند. (clustering)
در داده‌های بایوپلیکاماریستیم داده‌های visualizatoin برای اینها باید را کمتر همراه باشند
تا مفهوم در تفسیر سه‌بعدی داده‌های نابهنجار را بسیار بخوبی تفسیر کنند. تفسیر داده‌های خودکار را می‌توان در دو روش داشت:
برای داده‌های آبجکتی برای سیستم کنترل اکثر روش‌های تفسیر داده‌های خودکار را می‌توان در چهار روش مختلف تفسیر کرد:
نابهنجاری در کسریم مدلیم. و در داخل خود داده‌های اولیه که در پایه ترسیم شده‌اند.
برای داده‌های نابهنجار کار خوبی داشته است Project