What an “Outlier” Possibly Means

Outlier Detection Methods

The Big Picture

Back to the Future

Applications

Conclusion

References

There and Back Again

Outlier Detection between Statistical Reasoning and Efficient Database Methods

Arthur Zimek

University of Alberta
Edmonton, AB, Canada

Talk at University of Waterloo, Nov. 28, 2012

UNIVERSITY OF ALBERTA
FACULTY OF SCIENCE

Outline

What an “Outlier” Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

Back to the Future: Probability Estimates for Potential Outliers

Applications of Outlier Probability Estimates

Conclusion

What is an Outlier?

The intuitive definition of an outlier would be "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism".

[Hawkins, 1980]

An outlying observation, or "outlier," is one that appears to deviate markedly from other members of the sample in which it occurs.

[Grubbs, 1969]

An observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data

[Barnett and Lewis, 1994]

What is an Outlier?

The intuitive definition of an outlier would be "an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism".

[Hawkins, 1980]

An outlying observation, or "outlier," is one that appears to deviate markedly from other members of the sample in which it occurs.

[Grubbs, 1969]

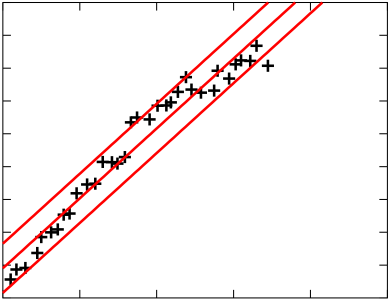
An observation (or subset of observations) which appears to be inconsistent with the remainder of that set of data

[Barnett and Lewis, 1994]

Where Can This Happen?

- measurement errors
- unusually extreme deviations
- data input, processing, transmission errors
- attacks, manipulation, fraud

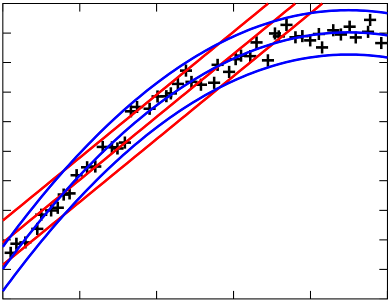
What's the Conclusion from Having an Outlier?



outliers should be treated generally as an indication that either the model or the cases may be in error, and they often provide useful diagnostic information

[Beckman and Cook, 1983]

What's the Conclusion from Having an Outlier?

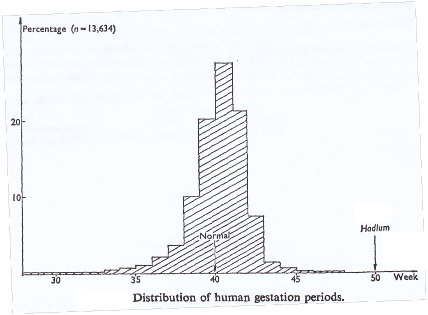


outliers should be treated generally as an indication that either the model or the cases may be in error, and they often provide useful diagnostic information

[Beckman and Cook, 1983]

Example [Barnett, 1978]: the Legal Case of Hadlum vs. Hadlum (1949)

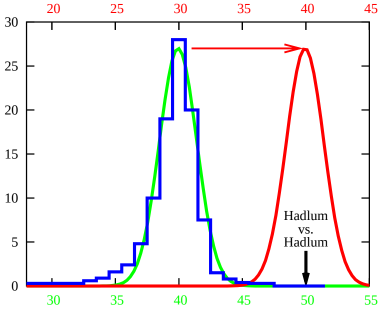
- The birth of a child to Mrs. Hadlum happened 349 days after Mr. Hadlum left for military service.
- Average human gestation period is 280 days (40 weeks).
- Statistically, 349 days is an outlier.



(Figure from [Barnett, 1978].)

Example (contd.): the Legal Case of Hadlum vs. Hadlum (1949)

- blue: statistical basis (13,634 observations of gestation periods)



- green: assumed underlying Gaussian process
  - very low probability for the birth of Mrs. Hadlums child for being generated by this process
- red: assumption of Mr. Hadlum
  - another Gaussian process responsible for the observed birth, where the gestation period starts later
  - Under this assumption the gestation period has an average duration and the specific birthday has highest-possible probability.

So What Does an “Outlier” Mean?

- An “outlier” is “suspicious” – when designing a meaningful evaluation scenario the researcher should keep this vagueness in mind.
- Whether or not the “outlier” should be removed (actually *is* a contaminant, fraud, measurement error,…) is a delicate question for the domain expert.
- In scientific data, there are even more subtle questions from a point of view of philosophy of science: remove the evidence from your data that your theory is wrong?

Outline

What an “Outlier” Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

Back to the Future: Probability Estimates for Potential Outliers

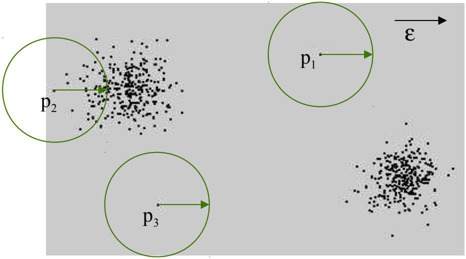
Applications of Outlier Probability Estimates

Conclusion

Distance-based Outliers

DB(ε, π)-outlier [Knorr and Ng, 1997]

- given ε, π
- A point p is considered an outlier if at most π percent of all other points have a distance to p less than ε

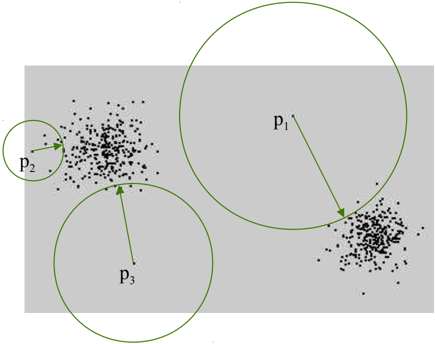


$$OutlierSet(\varepsilon, \pi) = \left\{ p \middle| \frac{Cardinality(q \in \mathcal{DB} | dist(q,p) < \varepsilon)}{Cardinality(\mathcal{DB})} \leq \pi \right\}$$

Distance-based Outliers

Outlier scoring based on kNN distances:

- Take the kNN distance of a point as its outlier score [Ramaswamy et al., 2000]
- Aggregate the distances for the 1-NN, 2-NN, …, kNN (sum, average) [Angiulli and Pizzuti, 2002]



Density-based Local Outliers

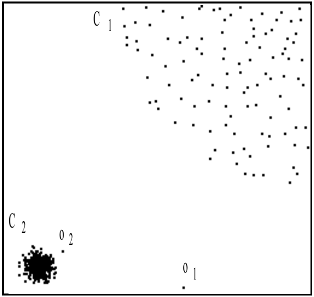


Figure from Breunig et al. [2000].

- DB-outlier model: no parameters ε, π such that o_2 is an outlier but none of the points of C_1 is an outlier
- kNN-outlier model: kNN-distances of points in C_1 are larger than kNN-distances of o_2

Density-based Local Outliers

Local Outlier Factor (LOF) [Breunig et al., 2000]:

- reachability distance (smoothing factor):
  $reachdist_k(p,o) = \max\{kdist(o), dist(p,o)\}$
- local reachability distance (lrd)
  $lrd_k(p) = 1 / \frac{\sum_{o \in kNN(p)} reachdist_k(p,o)}{Cardinality(kNN(p))}$
- Local outlier factor (LOF) of point p:
  average ratio of lrds of neighbors of p and lrd of p

$$LOF_k(p) = \frac{\sum_{o \in kNN(p)} \frac{lrd_k(o)}{lrd_k(p)}}{Cardinality(kNN(p))}$$

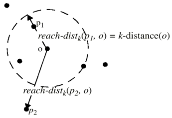


Figure from [Breunig et al., 2000]

- LOF ≈ 1: homogeneous density
- LOF $\gg 1$: point is an outlier (meaning of "$\gg$" ?)

Variants of Outlier Models

- connectivity-based (COF) [Tang et al., 2002]
- reverse neighborhood (INFLO) [Jin et al., 2006]
- local outlier integral (LOCI) [Papadimitriou et al., 2003]
- local distance-based outlier (LDOF) [Zhang et al., 2009]
- angle-spectrum variance (ABOD) [Kriegel et al., 2008]
- subspace distances/densities [Kriegel et al., 2009, Müller et al., 2010, Keller et al., 2012, Kriegel et al., 2012] (survey: [Zimek et al., 2012])
- ...

Efficiency Variants

- for DB-outlier (index-based, nested-loop-based, grid-based) [Knorr and Ng, 1998]
- for kNN
  - nested-loop [Ramaswamy et al., 2000]
  - linearization [Angiulli and Pizzuti, 2002]
  - nested-loop with randomization and pruning [Bay and Schwabacher, 2003]
  - approximate solution (reference-points) [Pei et al., 2006]
  - ...
  - overview and framework: [Orair et al., 2010]
- for LOF:
  - top-n [Jin et al., 2001]
  - random projections [de Vries et al., 2010]

Outline

What an “Outlier” Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

Back to the Future: Probability Estimates for Potential Outliers

Applications of Outlier Probability Estimates

Conclusion

Current Outlier Detection Research

…has plenty of:

- Faster variations (approximate, top-k)
- “New” outlier detection methods

…common shortcomings:

- Little or no statistical reasoning
- Just outlier rankings, no “outlierness measures”
- Evaluation using precision@k and ROC curves

No evaluation of *result usability*!

Outlier Score Usability

Outlier scores are defined using:

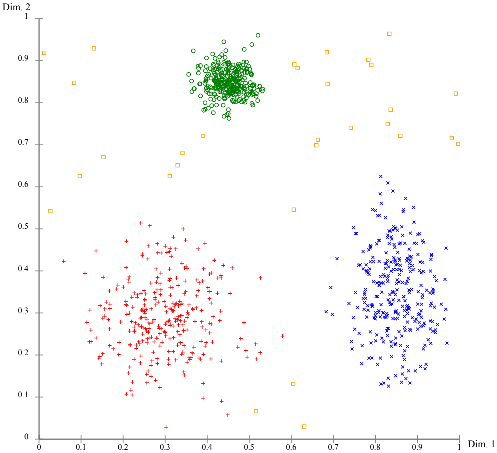
- Distances [Knorr and Ng, 1998, Ramaswamy et al., 2000, Angiulli and Pizzuti, 2002, Pei et al., 2006]
- Density quotient [Breunig et al., 2000, Papadimitriou et al., 2003]
- Distance quotient [Zhang et al., 2009]
- Angle spectrum variance [Kriegel et al., 2008]
- ...

So which points *are* outliers?

The scores convey little information!

Score Visualization

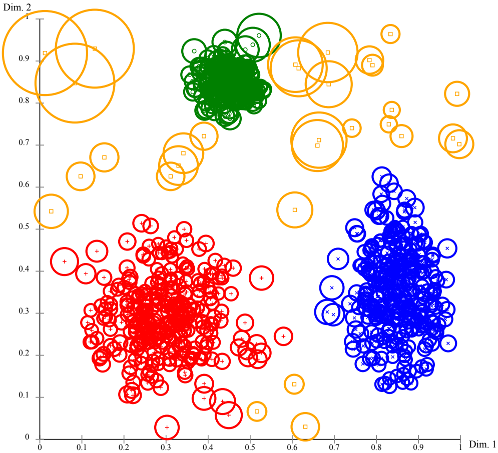
Simple data set with Gaussians (colored by label)



Visualized using the ELKI framework [Achtert et al., 2010].

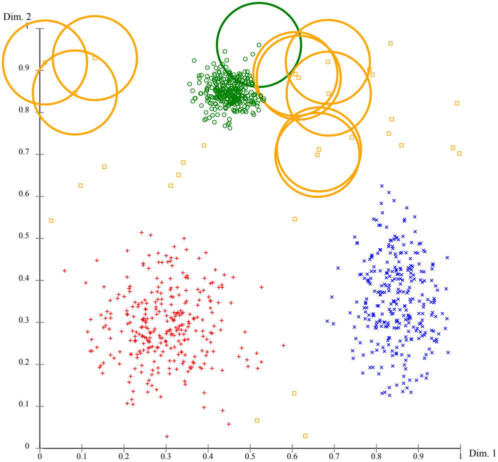
Score Visualization

LOF [Breunig et al., 2000] – naïvely scaled (linear)



Visualized using the ELKI framework [Achtert et al., 2010].

LOF [Breunig et al., 2000] – top-k



Visualized using the ELKI framework [Achtert et al., 2010].

Please Mind the Gap

I see no way of drawing a dividing line between those [observations] that are to be utterly rejected and those that are to be wholly retained

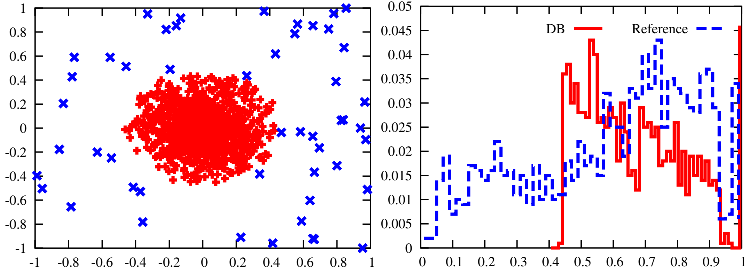
[Bernoulli, 1777]

a sample containing outliers would show up such characteristics as large gaps between ‘outlying’ and ‘inlying’ observations and the deviation between outliers and the group of inliers, as measured on some suitably standardized scale

[Hawkins, 1980]

Outlier Score Histograms

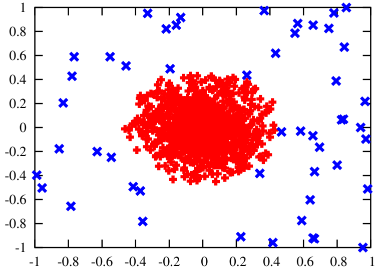
DB-outlier [Knorr and Ng, 1998],
Reference-based [Pei et al., 2006]

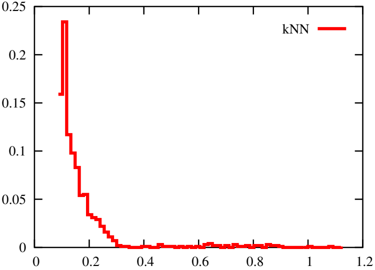


So what do the scores mean?

Outlier Score Histograms

kNN [Ramaswamy et al., 2000]

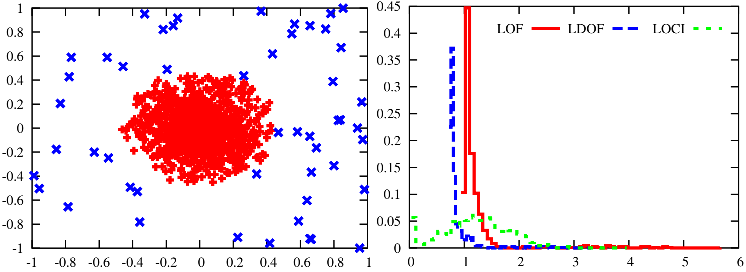




So what do the scores mean?

Outlier Score Histograms

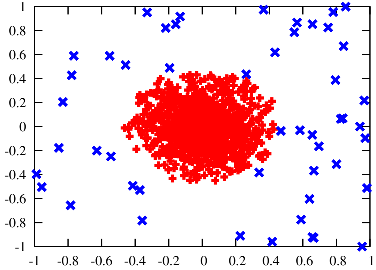
LOF [Breunig et al., 2000], LDOF [Zhang et al., 2009], and LOCI [Papadimitriou et al., 2003]

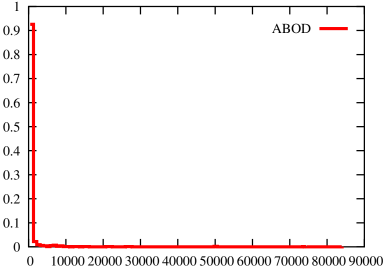


So what do the scores mean?

Outlier Score Histograms

ABOD [Kriegel et al., 2008]





So what do the scores mean?

Outline

What an "Outlier" Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

Back to the Future: Probability Estimates for Potential Outliers

Applications of Outlier Probability Estimates

Conclusion

Unified Scores

We [Kriegel, Kröger, Schubert, and Zimek, 2011] call a score S "unified" when it is:

- regularized
  ($Reg_S(o) \approx 0$ for inliers, $Reg_S(o) \gg 0$ for outliers)
- normalized
  - in the range of $[0 \ldots 1]$
  - (clear) inliers at 0, (clear) outliers at 1
- no decision at 0.5
- same ranking as original score
- intuitively the "outlier probability"

Goal: improve *interpretability*
of the scores of existing methods!

Score Unification

Unification would be possible using various transformations:

- Naïve: linear scaling
- Naïve: fractional rank
- Range clipping (e.g. LOF to $[1 \ldots 3]$)
  loses ranking information for inliers and extreme outliers
- Specialized: $-\log$ inversion e.g. for ABOD
- Statistical, using:
  - Gaussian distribution
  - Gamma distribution (including χ^2, exponential)
  - Half-normal distribution
- Combinations

Good news: depends mostly on algorithm, not the data set!

Score Unification

Statistical unification:

1. Regularize (e.g. $-\log$ for ABOD)
2. Assume a score distribution (e.g. Gaussian)
3. Fit distribution parameters (mean, stddev, ...)
4. Compute error function to get probabilities

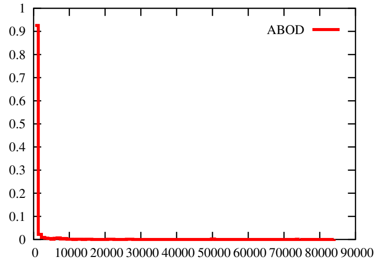
Properties:

- Monotone: no ranking changes (depending on the baseline, no *strict* monotony: ties in the ranking of inliers are possibly introduced)
- Precision and ROC AUC unchanged
- Brings back the statistics into outlier detection!

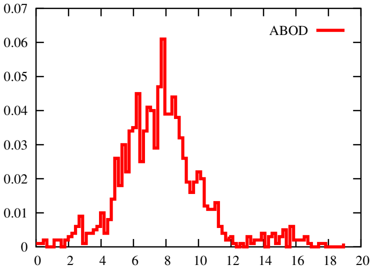
Score Unification - Example

Effect of regularization on ABOD scores – regularization by:

$$Reg_S^{loginv}(o) := -\log\left({}^{S(o)}/S_{\max}\right)$$



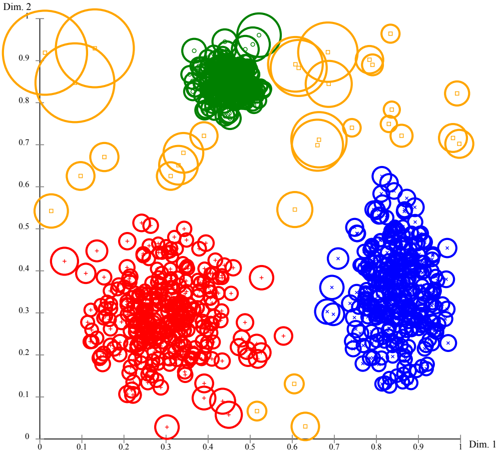
original



regularized

Unified Score Visualization

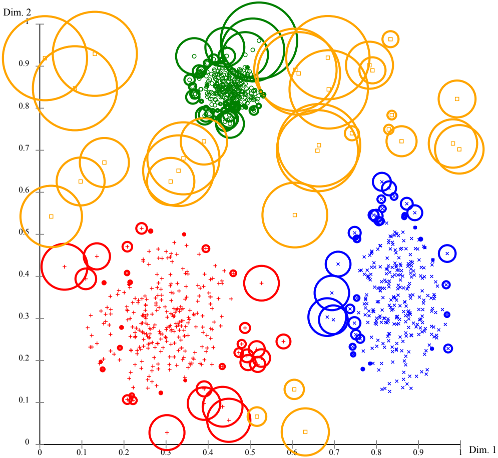
Local Outlier Factor – naïvely scaled



Visualized using the ELKI framework [Achtert et al., 2010].

Unified Score Visualization

Local Outlier Factor – Gaussian unification



Visualized using the ELKI framework [Achtert et al., 2010].

Outline

What an “Outlier” Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

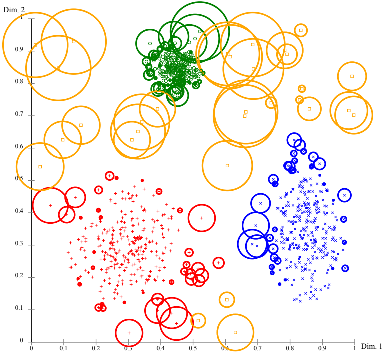
Back to the Future: Probability Estimates for Potential Outliers

Applications of Outlier Probability Estimates

Conclusion

Applications

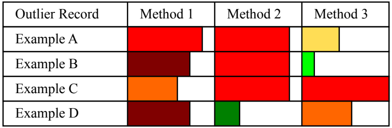
- Visualization
- Reporting
- Evaluation
- Comparison of scores
- Combination of scores: outlier ensembles



Essentially, anything that uses the *numbers* and not just the ranking!

Applications

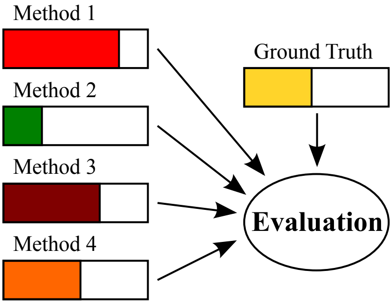
- Visualization
- Reporting
- Evaluation
- Comparison of scores
- Combination of scores: outlier ensembles



Essentially, anything that uses the *numbers* and not just the ranking!

Applications

- Visualization
- Reporting
- Evaluation
- Comparison of scores
- Combination of scores: outlier ensembles



Essentially, anything that uses the *numbers* and not just the ranking!

Applications

What an "Outlier" Possibly Means

Outlier Detection Methods

The Big Picture

Back to the Future

Applications

Overview

Ensemble Experiment

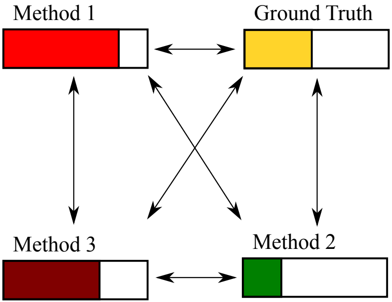
Comparison of Scores

Another Ensemble Experiment

Conclusion

References

- Visualization
- Reporting
- Evaluation
- Comparison of scores
- Combination of scores: outlier ensembles

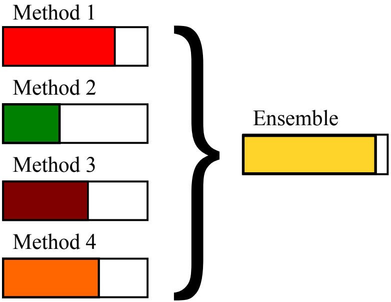


Essentially, anything that uses the *numbers* and not just the ranking!

What an "Outlier" Possibly Means
Outlier Detection Methods
The Big Picture
Back to the Future
Applications
Overview
Ensemble Experiment
Comparison of Scores
Another Ensemble Experiment
Conclusion
References

Applications

- Visualization
- Reporting
- Evaluation
- Comparison of scores
- Combination of scores: outlier ensembles



Essentially, anything that uses the *numbers* and not just the ranking!

What an "Outlier" Possibly Means
Outlier Detection Methods
The Big Picture
Back to the Future
Applications
Overview
Ensemble Experiment
Comparison of Scores
Another Ensemble Experiment
Conclusion
References

Ensemble Experiment

Competing methods:

- *Naive ensemble*: mean unified score (Gaussian)
- Feature bagging [Lazarevic and Kumar, 2005]
- Outlier probability estimates [Gao and Tan, 2006]
- HeDES [Nguyen et al., 2010]

Scenario:

- Data sets: 1. KDDCup1999, 2. ALOI images [Geusebroek et al., 2005] subset
- Ensemble 1: 10-fold feature bagging
- Ensemble 2: LOF with different parameters k
- Ensemble 3: LOF, LDOF, kNN, agg. kNN
- Evaluation: traditional ROC AUC score

What an "Outlier" Possibly Means
Outlier Detection Methods
The Big Picture
Back to the Future
Applications
Overview
Ensemble Experiment
Comparison of Scores
Another Ensemble Experiment
Conclusion
References

Ensemble Results – KDDCup1999

unified score [Kriegel et al., 2011]:

Ensemble construction	ROC AUC	Combination method
Feature Bagging LOF	0.7201	unscaled mean [Lazarevic and Kumar, 2005]
10 rounds,	0.7257	sigmoid mean [Gao and Tan, 2006]
dim $\in [d/2 : d-1]$,	0.7300	mixture model mean [Gao and Tan, 2006]
$k = 45$	0.7312	HeDES scaled mean [Nguyen et al., 2010]
	0.7327	maximum rank [Lazarevic and Kumar, 2005]
	0.7447	mean **unified score**
LOF [Breunig et al., 2000]	0.6693	mixture model mean [Gao and Tan, 2006]
$k = 20, 40, 80, 120, 160$	0.7078	unscaled mean [Lazarevic and Kumar, 2005]
	0.7369	sigmoid mean [Gao and Tan, 2006]
	0.7391	HeDES scaled mean [Nguyen et al., 2010]
	0.7483	maximum rank [Lazarevic and Kumar, 2005]
	0.7484	mean **unified score**
Combination of	0.5180	mixture model mean [Gao and Tan, 2006]
different methods:	0.9046	maximum rank [Lazarevic and Kumar, 2005]
LOF [Breunig et al., 2000],	0.9104	unscaled mean [Lazarevic and Kumar, 2005]
LDOF [Zhang et al., 2009],	0.9236	sigmoid mean [Gao and Tan, 2006]
kNN [Ramaswamy et al., 2000],	0.9386	HeDES scaled mean [Nguyen et al., 2010]
agg.kNN [Angiulli and Pizzuti, 2002]	0.9413	mean **unified score**

Ensemble Results – ALOI Images Subset

unified score [Kriegel et al., 2011]:

Ensemble construction	ROC AUC	Combination method
Feature Bagging LOF	0.7812	mixture model mean [Gao and Tan, 2006]
10 rounds,	0.7832	sigmoid mean [Gao and Tan, 2006]
dim $\in [d/2 : d-1]$,	0.7867	maximum rank [Lazarevic and Kumar, 2005]
$k = 45$	0.7990	unscaled mean [Lazarevic and Kumar, 2005]
	0.7996	HeDES scaled mean [Nguyen et al., 2010]
	0.8000	mean **unified score**
LOF [Breunig et al., 2000]	0.7364	mixture model mean [Gao and Tan, 2006]
$k = 10, 20, 40$	0.7793	maximum rank [Lazarevic and Kumar, 2005]
	0.7805	sigmoid mean [Gao and Tan, 2006]
	0.7895	HeDES scaled mean [Nguyen et al., 2010]
	0.7898	unscaled mean [Lazarevic and Kumar, 2005]
	0.7902	mean **unified score**
Combination of	0.7541	mixture model mean [Gao and Tan, 2006]
different methods:	0.7546	maximum rank [Lazarevic and Kumar, 2005]
LOF [Breunig et al., 2000],	0.7709	unscaled mean [Lazarevic and Kumar, 2005]
LDOF [Zhang et al., 2009],	0.7821	sigmoid mean [Gao and Tan, 2006]
kNN [Ramaswamy et al., 2000],	0.7997	mean **unified score**
agg.kNN [Angiulli and Pizzuti, 2002]	0.8054	HeDES scaled mean [Nguyen et al., 2010]

Diversity for Better Ensembles

We [Schubert, Wojdanowski, Zimek, and Kriegel, 2012] propose to measure and use diversity of individual outlier detectors to build improved ensembles:

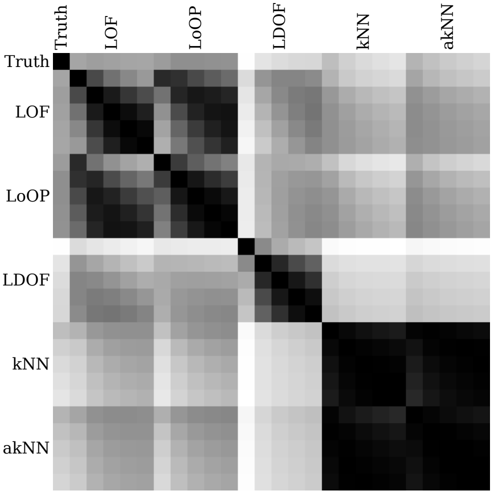
- similarity between rankings: does not use all information available from outlier scorings
- outlier scores as vector fields:
  - each data object is an axis (continuum of outlier scores)
  - each outlier scoring result is a point in this vector field
- similarity-measure: weighted Pearson correlation

$$\rho_\omega(X,Y) := \frac{\mathrm{Cov}_\omega(X,Y)}{\sigma_\omega(X)\sigma_\omega(Y)}$$

- use weights in order to balance between outliers and inliers

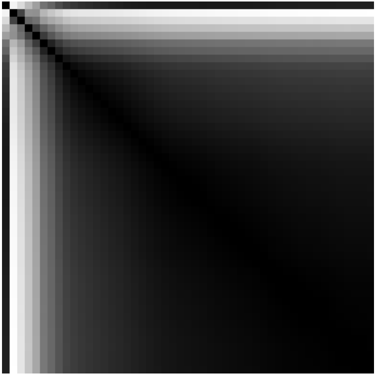
Similarity of Methods

ALOI data, $k = \{5, 10, 15, 20, 25\}$, Euclidean distance

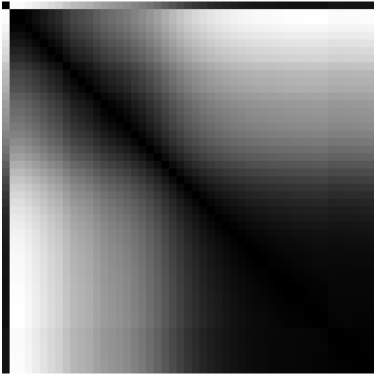


Parameter Stability

Wisconsin Breast Cancer (WBC) data, $k = 3, \ldots, 50$, Manhattan distance



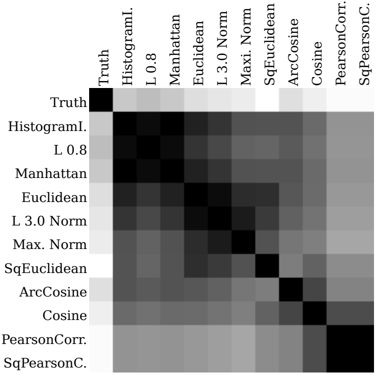
LOF



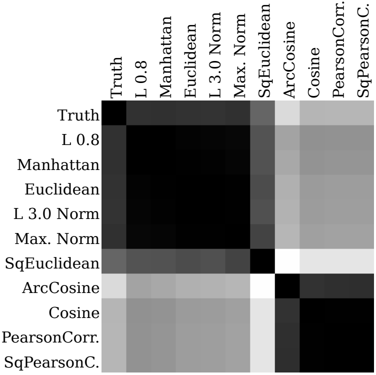
LDOF

Distance Measures

LOF, $k = 20$



ALOI



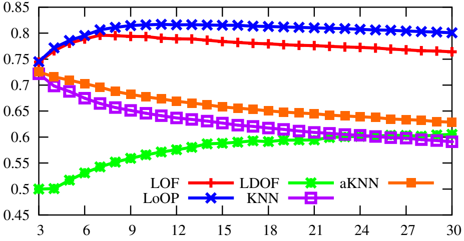
WBC

Diversity vs. Accuracy for Combinations

gain by combination of outlier detectors as compared to their individual performance: the relative improvement towards the target AUC score of 1 over the best of the combined detectors

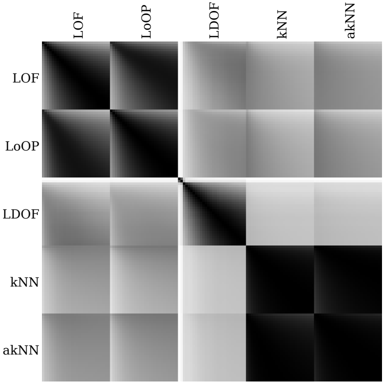
$$\text{gain}(M_1, M_2) := 1 - \frac{1 - AUC(M_1 + M_2)}{1 - \max\left(AUC(M_1), AUC(M_2)\right)}$$

accuracy of the algorithms (on ALOI) over choice of k:

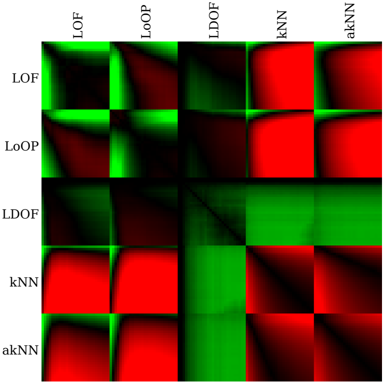


Similarity and Gain Combining Different Methods and Parametrization

combining pairs (ranked average scores):



Similarity



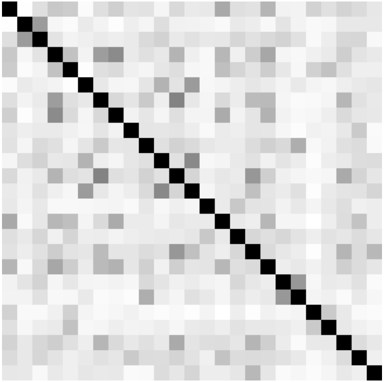
Gain (green: improved, red: deteriorated)

Combination of Diverse Pairs vs. Ensemble Methods

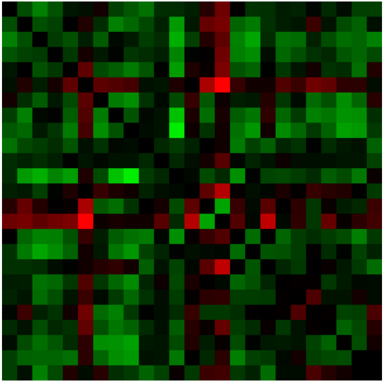
ROC	gain	combined methods	correl.
0.7218	-	kNN $k = 3$	-
0.7663	-	LOF $k = 4$	-
0.7716	-	LoOP $k = 4$	-
0.7767	-	LOF $k = 20$	-
0.8007	-	LoOP $k = 30$	-
0.8253	0.2176	LOF $k = 20$ + LoOP $k = 4$	0.4006
0.7952	0.1237	LOF $k = 4$ + kNN $k = 3$	0.4226
0.7938	0.0769	LOF $k = 20$ + kNN $k = 3$	0.5014
0.8275	0.1344	LOF $k = 4$ + LoOP $k = 30$	0.5373
0.7814	0.0427	LOF $k = 4$ + LoOP $k = 4$	0.8458
0.7932	-0.0375	LOF $k = 20$ + LoOP $k = 30$	0.9311
reference: existing ensemble methods			
0.7541	mixture model mean[Gao and Tan, 2006]		
0.7546	maximum rank[Lazarevic and Kumar, 2005]		
0.7709	unscaled mean[Lazarevic and Kumar, 2005]		
0.7821	sigmoid mean [Gao and Tan, 2006]		
0.7997	unified score [Kriegel et al., 2011]		
0.8054	HeDES scaled mean [Nguyen et al., 2010]		

Similarity and Gain Combining Feature Bags

combining pairs of feature bags (ALOI)



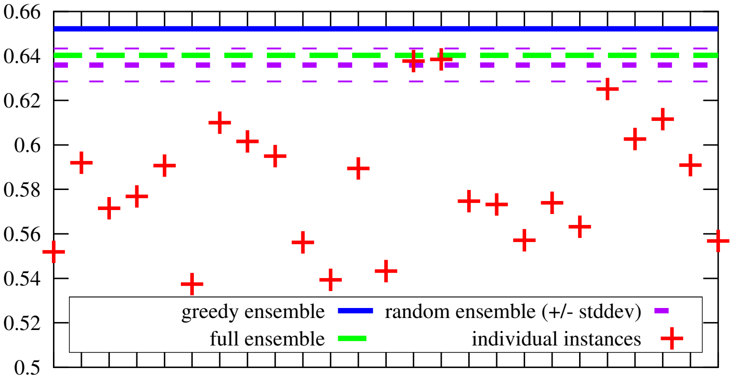
Similarity



Gain (green: improved, red: deteriorated)

Greedy Ensemble

Combining the most diverse individuals
(feature bags on ALOI)



Greedy Ensemble

Combining different methods/parameterizations

Method	AUC	significance	gain compared to full	gain compared to random
Metabolic dataset ($5 \cdot 13 = 65$ instances, $k = 100, 125, \ldots, 400$)				
Full ensemble	0.9201	n/a	$:= 0$	$+56.6\%$
Random ensemble	0.8159	± 0.1221	-130%	$:= 0$
Greedy ensemble	0.9530	$= \mu + 1.12\sigma$	$+41.2\%$	$+74.5\%$
Pen digits dataset ($6 \cdot 98 = 588$ instances, $k = 3 \ldots 100$)				
Full ensemble	0.9656	n/a	$:= 0$	$+74.6\%$
Random ensemble	0.8648	± 0.1669	-293%	$:= 0$
Greedy ensemble	0.9697	$= \mu + 0.63\sigma$	$+11.8\%$	$+77.6\%$
ALOI images dataset ($5 \cdot 28 = 140$ instances, $k = 3 \ldots 30$)				
Full ensemble	0.7903	n/a	$:= 0$	$+2.36\%$
Random ensemble	0.7853	± 0.0222	-2.42%	$:= 0$
Greedy ensemble	0.8380	$= \mu + 2.37\sigma$	$+22.7\%$	$+24.6\%$
KDDCup 1999 dataset ($5 \cdot 10 = 50$ instances, $k = 5 \ldots 50$)				
Full ensemble	0.8861	n/a	$:= 0$	$+15.3\%$
Random ensemble	0.8655	± 0.0414	-18.1%	$:= 0$
Greedy ensemble	0.9472	$= \mu + 1.97\sigma$	$+53.6\%$	$+60.7\%$

Outline

What an "Outlier" Possibly Means

A Short History of Outlier Detection Methods

The Big Picture: Rise and Decline of Outlier Detection Models

Back to the Future: Probability Estimates for Potential Outliers

Applications of Outlier Probability Estimates

Conclusion

Conclusion

status quo

- statistical reasoning about outliers: rich literature, results accumulated over centuries
- database/data mining research: $\approx$ 15 years, some models, many variants for efficiency
- efficiency variants aim at approximating the basic models, not the statistical intuition

  They are approximating approximations!
- even if the ranking is good, outlier scores are often quite meaningless

Conclusion

our focus: reconciliation of statistical reasoning and efficient, database-oriented solutions

- unification of outlier scores:
  - regularization, normalization
  - interpretability ("outlier probability")
  - comparability of different methods, parameterizations
  - comparability between different samples (subspace methods – see also Zimek et al. [2012])
  - combination of different methods (ensembles)
- open questions:
  - unification of more methods
  - calibration of outlier probabilities
  - optimizing contrast between outliers and inliers
  - improved evaluation procedures
  - outlier detection on multi-represented data
  - ensembles for outlier detection as better approximations of "true" outlierness

Thank you for your attention!

References I

E. Achtert, H.-P. Kriegel, L. Reichert, E. Schubert, R. Wojdanowski, and A. Zimek. Visual evaluation of outlier detection models. In *Proceedings of the 15th International Conference on Database Systems for Advanced Applications (DASFAA), Tsukuba, Japan*, pages 396–399, 2010. doi: 10.1007/978-3-642-12098-5_34.

F. Angiulli and C. Pizzuti. Fast outlier detection in high dimensional spaces. In *Proceedings of the 6th European Conference on Principles of Data Mining and Knowledge Discoverys (PKDD), Helsinki, Finland*, pages 15–26, 2002. doi: 10.1007/3-540-45681-3_2.

V. Barnett. The study of outliers: Purpose and model. *Applied Statistics*, 27(3): 242–250, 1978.

V. Barnett and T. Lewis. *Outliers in Statistical Data*. John Wiley&Sons, 3rd edition, 1994.

S. D. Bay and M. Schwabacher. Mining distance-based outliers in near linear time with randomization and a simple pruning rule. In *Proceedings of the 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Washington, DC*, pages 29–38, 2003. doi: 10.1145/956750.956758.

R. J. Beckman and R. D. Cook. Outlier..........s. *Technometrics*, 25(2):119–149, 1983.

References II

D. Bernoulli. Diiudicatio maxime probabilis plurium observationum discrepantium atque verisimillima inductio inde formanda. *Acta Academiae Scientiarum Imperialis Petropolitanae*, pages 3–23, 1777.

M. M. Breunig, H.-P. Kriegel, R.T. Ng, and J. Sander. LOF: Identifying density-based local outliers. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Dallas, TX*, pages 93–104, 2000.

T. de Vries, S. Chawla, and M. E. Houle. Finding local anomalies in very high dimensional space. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM), Sydney, Australia*, pages 128–137, 2010. doi: 10.1109/ICDM.2010.151.

J. Gao and P.-N. Tan. Converting output scores from outlier detection algorithms into probability estimates. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), Hong Kong, China*, pages 212–221, 2006. doi: 10.1109/ICDM.2006.43.

J. M. Geusebroek, G. J. Burghouts, and A.W.M. Smeulders. The Amsterdam Library of Object Images. *International Journal of Computer Vision*, 61(1): 103–112, 2005. doi: 10.1023/B:VISI.0000042993.50813.60.

F. E. Grubbs. Procedures for detecting outlying observations in samples. *Technometrics*, 11(1):1–21, 1969.

References III

D. Hawkins. *Identification of Outliers*. Chapman and Hall, 1980.

W. Jin, A.K. Tung, and J. Han. Mining top-n local outliers in large databases. In *Proceedings of the 7th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), San Francisco, CA*, pages 293–298, 2001. doi: 10.1145/502512.502554.

W. Jin, A. K. H. Tung, J. Han, and W. Wang. Ranking outliers using symmetric neighborhood relationship. In *Proceedings of the 10th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Singapore*, pages 577–593, 2006. doi: 10.1007/11731139_68.

F. Keller, E. Müller, and K. Böhm. HiCS: high contrast subspaces for density-based outlier ranking. In *Proceedings of the 28th International Conference on Data Engineering (ICDE), Washington, DC*, 2012.

E. M. Knorr and R. T. Ng. A unified notion of outliers: Properties and computation. In *Proceedings of the 3rd ACM International Conference on Knowledge Discovery and Data Mining (KDD), Newport Beach, CA*, pages 219–222, 1997.

E. M. Knorr and R. T. Ng. Algorithms for mining distance-based outliers in large datasets. In *Proceedings of the 24th International Conference on Very Large Data Bases (VLDB), New York City, NY*, pages 392–403, 1998.

References IV

H.-P. Kriegel, M. Schubert, and A. Zimek. Angle-based outlier detection in high-dimensional data. In *Proceedings of the 14th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Las Vegas, NV*, pages 444–452, 2008. doi: 10.1145/1401890.1401946.

H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. Outlier detection in axis-parallel subspaces of high dimensional data. In *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Bangkok, Thailand*, pages 831–838, 2009. doi: 10.1007/978-3-642-01307-2_86.

H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. Interpreting and unifying outlier scores. In *Proceedings of the 11th SIAM International Conference on Data Mining (SDM), Mesa, AZ*, pages 13–24, 2011.

H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. Outlier detection in arbitrarily oriented subspaces. In *Proceedings of the 12th IEEE International Conference on Data Mining (ICDM), Brussels, Belgium*, 2012.

A. Lazarevic and V. Kumar. Feature bagging for outlier detection. In *Proceedings of the 11th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Chicago, IL*, pages 157–166, 2005. doi: 10.1145/1081870.1081891.

References V

E. Müller, M. Schiffer, and T. Seidl. Adaptive outlierness for subspace outlier ranking. In *Proceedings of the 19th ACM Conference on Information and Knowledge Management (CIKM), Toronto, ON, Canada*, pages 1629–1632, 2010. doi: 10.1145/1871437.1871690.

H. V. Nguyen, H. H. Ang, and V. Gopalkrishnan. Mining outliers with ensemble of heterogeneous detectors on random subspaces. In *Proceedings of the 15th International Conference on Database Systems for Advanced Applications (DASFAA), Tsukuba, Japan*, pages 368–383, 2010. doi: 10.1007/978-3-642-12026-8_29.

G. H. Orair, C. Teixeira, Y. Wang, W. Meira Jr., and S. Parthasarathy. Distance-based outlier detection: Consolidation and renewed bearing. *Proceedings of the VLDB Endowment*, 3(2):1469–1480, 2010.

S. Papadimitriou, H. Kitagawa, P.B. Gibbons, and C. Faloutsos. LOCI: Fast outlier detection using the local correlation integral. In *Proceedings of the 19th International Conference on Data Engineering (ICDE), Bangalore, India*, pages 315–326, 2003. doi: 10.1109/ICDE.2003.1260802.

Y. Pei, O. Zaïane, and Y. Gao. An efficient reference-based approach to outlier detection in large datasets. In *Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), Hong Kong, China*, pages 478–487, 2006. doi: 10.1109/ICDM.2006.17.

References VI

S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. In *Proceedings of the ACM International Conference on Management of Data (SIGMOD), Dallas, TX*, pages 427–438, 2000.

E. Schubert, R. Wojdanowski, A. Zimek, and H.-P. Kriegel. On evaluation of outlier rankings and outlier scores. In *Proceedings of the 12th SIAM International Conference on Data Mining (SDM), Anaheim, CA*, pages 1047–1058, 2012.

J. Tang, Z. Chen, A. W.-C. Fu, and D. W. Cheung. Enhancing effectiveness of outlier detections for low density patterns. In *Proceedings of the 6th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Taipei, Taiwan*, pages 535–548, 2002. doi: 10.1007/3-540-47887-6_53.

K. Zhang, M. Hutter, and H. Jin. A new local distance-based outlier detection approach for scattered real-world data. In *Proceedings of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), Bangkok, Thailand*, pages 813–822, 2009. doi: 10.1007/978-3-642-01307-2_84.

A. Zimek, E. Schubert, and H.-P. Kriegel. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining*, 5(5):363–387, 2012. doi: 10.1002/sam.11161.