

Supplementary Material for

On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study

by G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent and M. E. Houle

Data Mining and Knowledge Discovery 30(4): 891-927, 2016, DOI: 10.1007/s10618-015-0444-8

PenDigits (version#01)

The 10 classes contained in this data set correspond to the digits from 0 to 9, with examples created by different hand writings. Class 4, defined here as outlier, was downsampled to only 20 objects. After the preprocessing, this database has 16 numeric attributes and 9868 instances, divided into 20 outliers (0.2%) and 9848 inliers (99.8%). This dataset is already normalized, i.e., all 16 attributes (spatial coordinates) have the same range [0,100]. It has been used in this form in [1,2].

References:

[1] H.-P. Kriegel, P. Kroeger, E. Schubert, and A. Zimek. Interpreting and unifying outlier scores. In Proc. SDM, pages 13-24, 2011.

[2] E. Schubert, R. Wojdanowski, A. Zimek, and H.-P. Kriegel. On evaluation of outlier rankings and outlier scores. In Proc. SDM, pages 1047-1058, 2012.

[Download all data set variants used \(2.1 MB\)](#). You can also access the [original data](#). (merge train and test [pendigits.tes and pendigits.tra])

Normalized, without duplicates

This version contains 16 attributes, 9868 objects, 20 outliers (0.20%)

[Download raw algorithm results \(83.1 MB\)](#) [Download raw algorithm evaluation table \(57.4 kB\)](#)

Best Parameters

The following table contains the best (overall and per-method) results for each method and evaluation measure (when the same score was achieved twice, only the smallest k is given). The Maximum F1-Measure is complimentary in addition to the measures in the original publication.

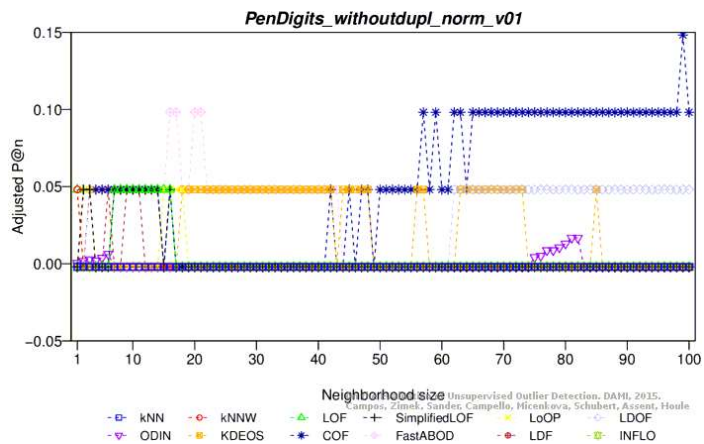
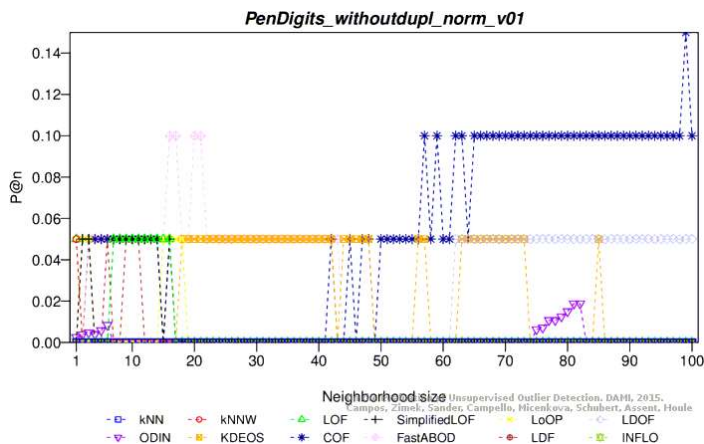
Algorithm	k	P@n	Adj. P@n	AP	Adj. AP	Max-F1	Adj. MF1	ROC AUC
KNN	1	0.00000	-0.00203	0.03741	0.03546	0.10345	0.10163	0.93188
KNN	12	0.00000	-0.00203	0.07243	0.07055	0.15603	0.15431	0.98681
KNN	19	0.00000	-0.00203	0.06984	0.06795	0.17143	0.16975	0.98505
KNNW	1	0.05000	0.04807	0.04703	0.04509	0.13333	0.13157	0.92636

KNNW	24	0.00000	-0.00203	0.06517	0.06327	0.14978	0.14805	0.98542
KNNW	26	0.00000	-0.00203	0.06568	0.06378	0.14346	0.14172	0.98548
LOF	1	0.05000	0.04807	0.00896	0.00695	0.06452	0.06262	0.62443
LOF	5	0.00000	-0.00203	0.00815	0.00614	0.08000	0.07813	0.48821
LOF	60	0.00000	-0.00203	0.01569	0.01369	0.04016	0.03821	0.91466
LOF	69	0.00000	-0.00203	0.01568	0.01368	0.04255	0.04061	0.91680
SimplifiedLOF	2	0.05000	0.04807	0.01485	0.01285	0.06452	0.06262	0.63034
SimplifiedLOF	3	0.05000	0.04807	0.01573	0.01373	0.11538	0.11359	0.62107
SimplifiedLOF	98	0.00000	-0.00203	0.01486	0.01286	0.04082	0.03887	0.91070
LoOP	2	0.05000	0.04807	0.01595	0.01396	0.07143	0.06954	0.63279
LoOP	7	0.00000	-0.00203	0.01007	0.00806	0.08571	0.08386	0.55128
LoOP	99	0.00000	-0.00203	0.01471	0.01271	0.03974	0.03778	0.90398
LDOF	3	0.05000	0.04807	0.03148	0.02951	0.09091	0.08906	0.69827
LDOF	42	0.00000	-0.00203	0.00593	0.00391	0.03175	0.02978	0.71811
ODIN	81	0.01875	0.01676	0.01603	0.01403	0.04444	0.04250	0.91463
ODIN	96	0.00000	-0.00203	0.01729	0.01530	0.04103	0.03908	0.92255
ODIN	97	0.00000	-0.00203	0.01763	0.01563	0.04188	0.03994	0.92250
FastABOD	5	0.05000	0.04807	0.03720	0.03525	0.12000	0.11821	0.92775
FastABOD	6	0.05000	0.04807	0.03320	0.03123	0.12245	0.12067	0.92428
FastABOD	16	0.10000	0.09817	0.03261	0.03064	0.10000	0.09817	0.94058
FastABOD	94	0.00000	-0.00203	0.03329	0.03132	0.08333	0.08147	0.96105
KDEOS	18	0.05000	0.04807	0.00555	0.00353	0.05556	0.05364	0.53232
KDEOS	33	0.05000	0.04807	0.05261	0.05069	0.09524	0.09340	0.63018
KDEOS	37	0.05000	0.04807	0.05277	0.05085	0.09524	0.09340	0.65215
KDEOS	100	0.00000	-0.00203	0.00691	0.00490	0.04348	0.04154	0.81132
LDF	3	0.05000	0.04807	0.00632	0.00430	0.05556	0.05364	0.53295
LDF	10	0.05000	0.04807	0.02266	0.02067	0.10714	0.10533	0.89785
LDF	60	0.00000	-0.00203	0.02438	0.02240	0.06642	0.06452	0.95337
LDF	72	0.00000	-0.00203	0.02404	0.02205	0.06289	0.06099	0.95450
INFLO	1	0.00000	-0.00203	0.00712	0.00510	0.03659	0.03463	0.60988
INFLO	2	0.00000	-0.00203	0.01458	0.01258	0.07018	0.06829	0.67489
INFLO	99	0.00000	-0.00203	0.01274	0.01073	0.03704	0.03508	0.89086
COF	81	0.10000	0.09817	0.05143	0.04950	0.15789	0.15618	0.95506
COF	90	0.10000	0.09817	0.05711	0.05519	0.14458	0.14284	0.95792
COF	99	0.15000	0.14827	0.05026	0.04833	0.15000	0.14827	0.96097

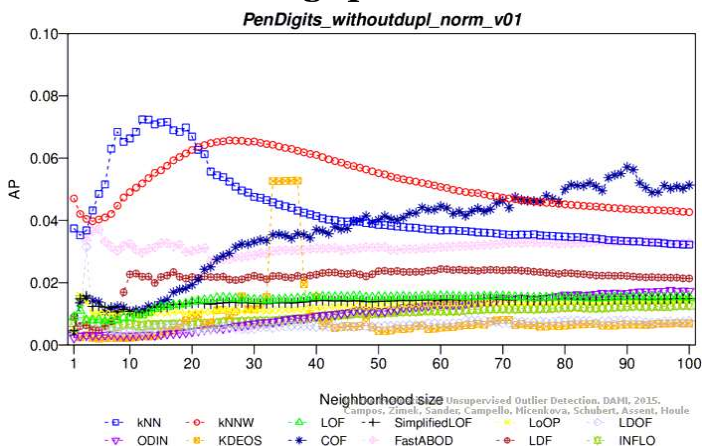
Plots

Precision at n

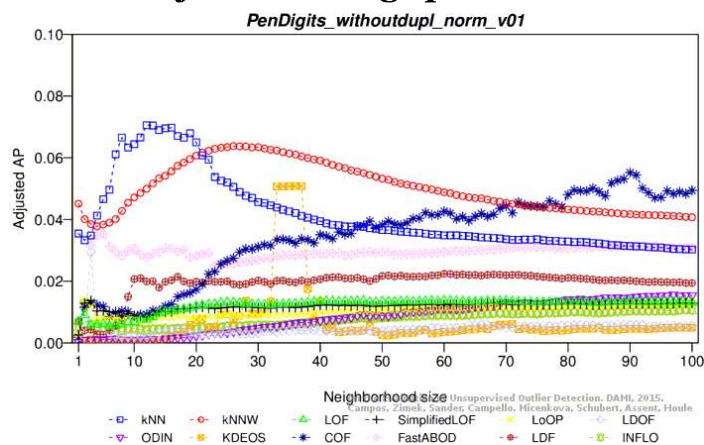
Adjusted precision at n



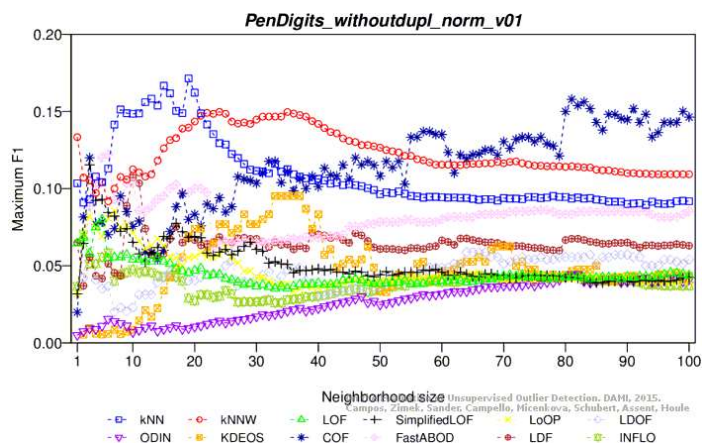
Average precision



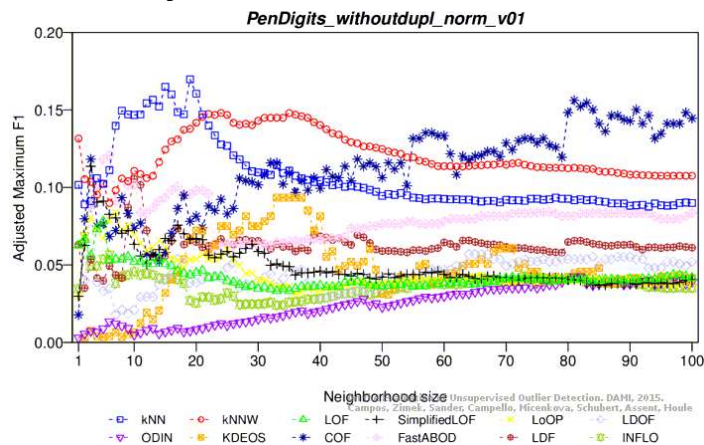
Adjusted average precision



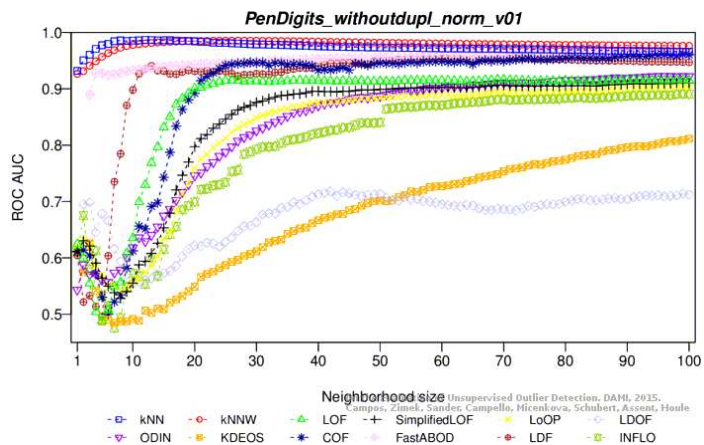
Maximum F1 score



Adjusted maximum F1 score



ROC AUC



Diversity

