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

Guilherme O. Campos[1], Arthur Zimek[2], Jörg Sander[3],
Ricardo J. G. B. Campello[1], Barbora Micenková[4], Erich Schubert[5],
Ira Assent[4], and Michael E. Houle[6]

[1] University of São Paulo, {gocampos,campello}@icmc.usp.br
[2] University of Southern Denmark, zimek@imada.sdu.dk
[3] University of Alberta, jsander@ualberta.ca
[4] Aarhus University, {barbora,ira}@cs.au.dk
[5] Ludwig-Maximilians-Universität München, schube@dbs.ifi.lmu.de
[6] National Institute of Informatics meh@nii.ac.jp

The evaluation of unsupervised outlier detection algorithms is a constant challenge in data mining research. Little is known regarding the strengths and weaknesses of different standard outlier detection models, and the impact of parameter choices for these algorithms. The scarcity of appropriate benchmark datasets with ground truth annotation is a significant impediment to the evaluation of outlier methods. Even when labeled datasets are available, their suitability for the outlier detection task is typically unknown. Furthermore, the biases of commonly-used evaluation measures are not fully understood. It is thus difficult to ascertain the extent to which newly-proposed outlier detection methods improve over established methods. We performed an extensive experimental study [1] on the performance of a representative set of standard k nearest neighborhood-based methods for unsupervised outlier detection, across a wide variety of datasets prepared for this purpose. Based on the overall performance of the outlier detection methods, we provide a characterization of the datasets themselves, and discuss their suitability as outlier detection benchmark sets. We also examine the most commonly-used measures for comparing the performance of different methods, and suggest adaptations that are more suitable for the evaluation of outlier detection results.

We present the results from our previous publication [1] as well as additional observations and measures. All results are available online in the repository at: http://www.dbs.ifi.lmu.de/research/outlier-evaluation/

References

[1] G. O. Campos, A. Zimek, J. Sander, R. J. G. B. Campello, B. Micenková, E. Schubert, I. Assent, and M. E. Houle. "On the Evaluation of Unsupervised Outlier Detection: Measures, Datasets, and an Empirical Study". In: *Data Mining and Knowledge Discovery* 30 (4 2016), pp. 891–927. DOI: 10.1007/s10618-015-0444-8.