SDU
UNIVERSITY OF
SOUTHERN DENMARK

Dr. Arthur Zimek

Professor, Head of Section Data Science and Statistics
Dept. of Mathematics and Computer Science
University of Southern Denmark
Campusvej 55, 5230 Odense M, Denmark

Review report on the dissertation thesis of Ing. Vít Škvára, Neural network-based generative models for anomaly detection

March 24, 2024

The submitted dissertation thesis "Neural network-based generative models for anomaly detection" by Ing. Vít Škvára, dated 2023, states as objectives (Section 1.3):

1. providing an overview on state-of-the-art classical (shallow) and (deep) generative models for anomaly detection and "deeper insights into the behaviour of generative models in anomaly detection";
2. an "extensive experimental comparison of selected methods under different operating conditions";
3. a novel anomaly detection method based on deep generative models.

The thesis is a coherent text, addressing these objectives in chapters 2-5. Each of these chapters relates to a dedicated publication, but is apparently thoroughly rewritten or reorganized to contribute in a harmonic way to the text of the thesis. Of the related 4 publications, one is unpublished but "considered for publication", 1 is an archival preprint, 2 are published in peer reviewed journals.

1. The topic is up-to-date
   Anomaly detection is a very active and competitive research topic in fundamental methods research (in areas such as machine learning, data mining, databases, AI, data science, statistics) as well as in various applications (of which the candidate explored physics and image data). The literature is enormous and difficult to survey, not the least as different terminology is used in different areas but also in the same community when problems are characterized in different ways (e.g., anomaly, outlier, novelty, out-of-distribution). The thesis has a focus on deep generative models, but keeps classic methods for comparison, which adds considerable value and interest to the contribution. There is a tendency in the research community that different characterizations of the problem are studied without relating findings to other characterizations. This thesis presents an attempt to compare supervised (deep) models and classic (unsupervised) models and as such adds clearly value to the research corpus in the area.
   A relevant contemporary study has been considered in the thesis (ref. [35]) and apparently was inspirational and influential on the candidate. Other studies also comparing supervised and unspervised methods have been overlooked, such as [SMS$^+$16] and some work referenced therein – the recent more extensive journal version of which [MSS$^+$23] has been published perhaps too late to find consideration in the thesis). This related work compares other supervised and unsupervised methods, but does not consider deep generative models in turn. However, completeness is not achievable, especially for a very up-to-date topic with rich ongoing research in several communities.

2. Formal structure and organization
   The thesis comprises 6 chapters on 84 pages and additional appendices plus bibliography.
   - Chapter 1 provides an introduction and names the objectives.
   - Chapter 2 discusses quality measures used for evaluation in anomaly detection and gives an overview of some classic ("shallow") methods.
   - Chapter 3 describes deep anomaly detectors with a special focus on generative models.
   - Chapter 4 surveys deep generative models for anomaly detection theoretically and comprises an experimental comparison of several such models and some representative classic (unsupervised) anomaly detection methods.
   - Chapter 5 introduces a novel method aiming in particular at addressing some shortcomings of existing methods identified in Chapter 4.
   - Chapter 6 summarizes the contributions.

3. Methods applied in the thesis
   The thesis comprises both theoretical and practical comparative analysis of existing and new models for anomaly detection. A concise and coherent survey of existing methods identifies shortcomings in the state of the art and subsequently addresses these with a novel method designed from scratch.
   The study relates to relevant concurrent publications as well as classic methods. The construction of the presented novel method is theoretically well-founded and thoroughly evaluated on a range of relevant datasets.

4. Completion of the dissertation objectives
   The first objective is achieved in Chapters 2 and 3. The second objective is achieved in Chapter 4. The third objective is achieved in Chapter 5.
   While of course a good study always raises more questions to explore, the achieved results are satisfactory with no obvious shortcomings or defects. However, one interesting addition to the experimental comparison would be the inclusion of computational requirements, time complexity, and time and space scalability of the compared methods.

5. Evaluation of the scientific value of the results and the contributions
   The four chapters each relate to a publication, of which only two are actually published in peer reviewed journals. However, the other two, in my opinion, also constitute interesting publishable material and the related publications (one preprint, one "considered for publication", which supposedly means submitted work currently under review) should have a good chance of being published in peer reviewed journals or conferences eventually.

6. Remarks for discussion or minor corrections
   - The attempt to define anomalies as being below a given density threshold (Section 1.1) gives a first intuition that can be used to understand what very classic methods try to achieve, such as the kNN anomaly detection or DB-outlier [KN97], but already falls short to capture the different notions of outlierness in LOF and its numerous variants, ABOD, and many others.
     I think it is actually not possible to give a general mathematical definition of outlierness. What is an outlier can perhaps better be seen as operationally defined by any applied method, be it statistical tests, algorithms applying various forms of density estimators or other measurements of properties, or generative models trying to capture the behaviour of normal and anomalous distributions [ZF18].
     Should the definition presented in Section 1.1 actually be satisfying, we can identify outliers perfectly as the "noise" object of DBSCAN [EKSX96]. No alternative method would ever be needed. The thesis itself is proof that this cannot be the intention behind this definition. It would have been good to discuss the preliminary nature of this definition clearly.

- The discussion of evaluation methodology overlooks the (relatively recent) existence of internal evaluation measures, i.e., quality measures that do not require label information [MCSZ20, MZCS22].
- The illustrative case study with the two bananas dataset (Fig. 2.3 and subsequent figures) can be a bit confusing. Later in the thesis it seems that anomalies are not present in the training set. It is not clearly discussed what this means for unsupervised methods in the test bed (but see [MSS^{+}23] for a discussion). In these figures the reader gets the impression that there are as many anomalies as normal data points. How are the methods applied in that context? For most classic methods, the assumption of "rareness" or "unusualness" is crucial, which would relate to an extremely imbalanced classification scenario.
- The discussion of distance-based methods (Section 2.2.2) remains a bit vague in some places.
  What is meant by "a very local behaviour"? Clearly this is not based on the same notion of "locality" that comes with LOF a few lines later. See [SZK14] for a deeper discussion of the notion of locality in outlier detection.
  What means "does not scale well" precisely? Since this issue is raised for LOF, how do all the other methods scale in comparison?
- The discussion of the "curse of dimensionality" and its consequences (Section 3.4.3) is vague and unclear. See [ZSK12] to clear up some (common) misunderstandings.
- The phrase "which is disappointing" (p. 52) could be omitted.
- The method introduced in Chapter 5 is very interesting. I tend to disagree on the use of the terminology "semantic". Supposedly, this terminology was taken from the literature [52], but still it is more catchy than telling and worth a critical discussion, albeit at a more philosophical level. Effectively, the method can be guided to work in specific (latent) subspaces. It is the guidance (by the researcher) where semantics is placed, not the outliers as such.
  It would perhaps be beneficial (not the least for the ongoing publication) to place this method in the context of explainable or interpretable anomaly detection [LZvL24].
- A corrupted sentence on p. 74 should be corrected: "For additional These models include".
- Figures 5.5 and 5.6, when comparing measurements I would recommend to normalize the scale of the plots.

7. The overall evaluation
   The formulated objectives are meaningful, relevant, and can be considered adequate for a PhD thesis, and they have been completed in the thesis.

8. Recommendation
   The author of the dissertation proved the ability to conduct research and to achieve scientific results. I do recommend the thesis for the presentation and defense with the aim of receiving the Ph.D. degree.

Odense, March 24, 2024

Dr. Arthur Zimek

References

[EKSX96] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the 2nd ACM International Conference on Knowledge Discovery and Data Mining (KDD), Portland, OR*, pages 226–231, 1996.

[KN97] 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.

[LZvL24] Z. Li, Y. Zhu, and M. v. Leeuwen. A survey on explainable anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 18(1):23:1–23:54, 2024.

[MCSZ20] H. O. Marques, R. J. G. B. Campello, J. Sander, and A. Zimek. Internal evaluation of unsupervised outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 14(4):47:1–47:42, 2020.

[MSS+23] H. O. Marques, L. Swersky, J. Sander, R. J. G. B. Campello, and Arthur Zimek. On the evaluation of outlier detection and one-class classification: a comparative study of algorithms, model selection, and ensembles. *Data Mining and Knowledge Discovery*, 37(4):1473–1517, 2023.

[MZCS22] H. O. Marques, A. Zimek, R. J. G. B. Campello, and J. Sander. Similarity-based unsupervised evaluation of outlier detection. In *Proceedings of the 15th International Conference on Similarity Search and Applications (SISAP), Bologna, Italy*, 2022.

[SMS+16] L. Swersky, H. O. Marques, J. Sander, R. J. G. B. Campello, and A. Zimek. On the evaluation of outlier detection and one-class classification methods. In *Proceedings of the IEEE International Conference on Data Science and Advanced Analytics (DSAA), Montreal, QC, Canada*, pages 1–10, 2016.

[SZK14] E. Schubert, A. Zimek, and H.-P. Kriegel. Local outlier detection reconsidered: a generalized view on locality with applications to spatial, video, and network outlier detection. *Data Mining and Knowledge Discovery*, 28(1):190–237, 2014.

[ZF18] A. Zimek and P. Filzmoser. There and back again: Outlier detection between statistical reasoning and data mining algorithms. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, page e1280, 2018.

[ZSK12] 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.

Phone: +45 6550 9071 `http://www.imada.sdu.dk/~zimek/` `zimek@imada.sdu.dk`