Homework 5

David Östling
dostl@kth.se
DA2210

November 20, 2021

Main (a and b)

a) Chosen Article: Kevin S. Killourhy R. Maxion (2009). Comparing anomaly-detection algorithms for keystroke dynamics [1].

b) Judging from the citation *"International Conference on Dependable Systems Networks"* we can distinguish that the article [1] has appeared as the proceedings of a conference called DSN; hence the article was published in a *conference* [2]. The medium is a "*print*" and the publication is (as mentioned above) a conference. The scientific field of the conference is *Computer Science, Algorithms and Heuristics* [1]. The article provides a deep insight into anomaly detection algorithms for keystroke dynamics and it can be used for further studies in the respective field (when cited).

Template for Taking Notes on Research Articles c)

Complete citation: Kevin S. Killourhy R. Maxion (2009). Comparing anomaly-detection algorithms for keystroke dynamics," 2009 IEEE/IFIP International Conference on Dependable Systems & Networks, 2009, pp. 125-134, doi: 10.1109/DSN.2009.5270346.
Web Access: https://www.cs.cmu.edu/ maxion/pubs/KillourhyMaxion09.pdf
Key Words: Heuristic algorithms, Rhythm, Detectors, Laboratories, Error analysis, Computer science, Algorithm design and analysis, Benchmark testing, Security, Biometrics [3]
General Subject: Computer Science
Specific Subject: Anomaly detection-algorithms
Hypothesis: The authors did not really have a hypothesis as it was purely a research study designed to test what detectors were most efficient, not to propose what detectors might be

more or less efficient.

Methodology: To evaluate all detectors using the same password-timing data procedure. All detectors were trained and tested using this exact method. This then returned the margin of error for each detector which then was used in a comparison between all detectors. [1]

Results: See figure 1

	Detector	equal-error rate			Detector	zero-miss false-alarm rate	
1	**Manhattan (scaled)**	**0.096**	**(0.069)**	**1**	**Nearest Neighbor (Mahalanobis)**	**0.468**	**(0.272)**
2	**Nearest Neighbor (Mahalanobis)**	**0.100**	**(0.064)**	**2**	**Mahalanobis**	**0.482**	**(0.273)**
3	**Outlier Count (z-score)**	**0.102**	**(0.077)**	**3**	**Mahalanobis (normed)**	**0.482**	**(0.273)**
4	SVM (one-class)	0.102	(0.065)	**4**	**SVM (one-class)**	**0.504**	**(0.316)**
5	Mahalanobis	0.110	(0.065)	5	Manhattan (scaled)	0.601	(0.337)
6	Mahalanobis (normed)	0.110	(0.065)	6	Manhattan (filter)	0.757	(0.282)
7	Manhattan (filter)	0.136	(0.083)	7	Outlier Count (z-score)	0.782	(0.306)
8	Manhattan	0.153	(0.092)	8	Manhattan	0.843	(0.242)
9	Neural Network (auto-assoc)	0.161	(0.080)	9	Neural Network (auto-assoc)	0.859	(0.220)
10	Euclidean	0.171	(0.095)	10	Euclidean	0.875	(0.200)
11	Euclidean (normed)	0.215	(0.119)	11	Euclidean (normed)	0.911	(0.148)
12	Fuzzy Logic	0.221	(0.105)	12	Fuzzy Logic	0.935	(0.108)
13	k Means	0.372	(0.139)	13	k Means	0.989	(0.040)
14	Neural Network (standard)	0.828	(0.148)	14	Neural Network (standard)	1.000	(0.000)

Figure 1: The results are presented within this table, ranked from best to worst

Summary of key points: *"The results-along with the shared data and evaluation methodology-constitute a benchmark for comparing detectors and measuring progress."* [1]

Context: This article relates to anything that includes anomaly-detection as it provides a steady guideline for what detectors are most efficient. This can also be used for future studies in the field and potentially even my own work in school (if I for instance pick the machine learning track).

Significance: If I ever need to use an anomaly-detector I now know what detector to side with and why.

Important Figures and/or Tables: Table 1 *("Seven different studies investigate 11 different anomaly detectors and report evaluation results. The diversity of evaluation conditions makes a direct comparison of the anomaly-detector performance results impossible").* (p.3), Figure 1 *(Nearest Neighbor (Mahalanobis) Subject 19)* (p.7) and Table 2 *("The average equal-error rates (left side) and average zero-miss false-alarm rates (right side) from the evaluation of the 14 detectors are ranked from best to worst (with standard deviations in parentheses). The set of top-performing detectors is indicated in bold-face (i.e., those that are not signifi-*

cantly worse than the best-performing detector)") (p.8).
Cited References to follow up on: E.g. S. Bleha, C. Slivinsky, and B. Hussien. Computeraccess security systems using keystroke dynamics. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(12):1217–1222, 1990 as it is heavily related to the topic and might provide some further insight. The same goes for: CENELEC. European Standard EN 50133-1: Alarm systems. Access control systems for use in security applications. Part 1: System requirements, 2002. Standard Number EN 50133-1:1996/A1:2002, Technical Body CLC/TC 79, European Committee for Electrotechnical Standardization (CENELEC). and S. Cho, C. Han, D. H. Han, and H. Kim. Web-based keystroke dynamics identity verification using neural networks. Journal of Organizational Computing and Electronic Commerce, 10(4):295–307, 2000.

Main (d, e, f and g)

d) As above (one article among the references); e.g. "S. Bleha, C. Slivinsky, and B. Hussien. Computeraccess security systems using keystroke dynamics. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(12):1217–1222, 1990" as it is heavily related to the topic and might provide some further insight.

An article that has cited my chosen article [1] which I found very interesting was the following: "Aythami Morales, Julian Fierrez, Ruben Tolosana, Javier Ortega-Garcia, Javier Galbally, Marta Gomez-Barrero, André Anjos, Sébastien Marcel, "Keystroke Biometrics Ongoing Competition", Access IEEE, vol. 4, pp. 7736-7746, 2016". I chose this article as it provides a deeper dive into the topic while using my chosen article as a reference for their claims.

e) As mentioned above, my chosen article [1] clearly follows the IMRD format and it made the text easier to read following the great structuring. However if we look at another article [4] it does not really follow the same structure. In my eyes though this does not mean that the article was harder to read, it was just a bit different from some of the other articles and I believe it would not really benefit from following the IMRD format (when compared to what was used instead). I believe that the IMRD format is more substantial when writing e.g. reports or research studies in the same sense as my chosen article [1] it was not really needed as much here as I could still understand what the authors intended with their paper.

f) The article [1] has been cited a total 224 times [3]

g) The article has provided a great guideline for finding the most efficient types of anomaly-detection algorithms and poses a good starting point for future research in the corresponding field [1].

References

[1] K. S. Killourhy and R. A. Maxion, "Comparing anomaly-detection algorithms for keystroke dynamics," 2009 IEEE/IFIP International Conference on Dependable Systems Networks, 2009, pp. 125-134, doi: 10.1109/DSN.2009.5270346.

[2] Purugganan, M., & Hewitt, J. (2004). How to read a scientific article. Rice University.

[3] IEEEXplore (2021) "Comparing anomaly-detection algorithms for keystroke dynamics" Available: https://ieeexplore.ieee.org/document/5270346/citations?tabFilter=patents

[4] Arute, Frank, et al. "Quantum supremacy using a programmable superconducting processor." Nature 574.7779 (2019): 505-510.