Anomalieerkennung mit unüberwachten Autoencodern durch Beschränkung der Trainingsdatenauswahl

Exposé Masterarbeit

Jonas Helm

Anomalieerkennung ist von entscheidender Bedeutung in verschiedenen Bereichen, da sie dazu beiträgt, Abweichungen von normalen Mustern oder Verhalten in Daten frühzeitig zu identifizieren. In der Finanzindustrie beispielsweise hilft die Anomalieerkennung dabei, betrügerische Transaktionen aufzudecken.[1] Im Gesundheitswesen kann sie dazu beitragen, Krankheiten frühzeitig zu diagnostizieren.[2]

Eine Möglichkeit um Anomalien zu erkennen besteht in der Verwendung von Autoencodern. Diese sind neuronale Netze, die aus zwei Hauptbestandteilen bestehen. Einem Encoderteil um eine komprimierte Kodierung der Eingabe zu erhalten und einem Decoderteil, der aus den kodierten Daten die Eingabe rekonstruiert. Das Encoding zwingt das Netzwerk die wesentlichen Eigenschaften der Daten zu extrahieren, um eine möglichst gute Rekonstruktion zu ermöglichen.[3]

Im Zusammenhang mit der Erkennung von Anomalien wird der Autoencoder in der Regel durch Minimierung des Rekonstruktionsfehlers unter Verwendung von normalen Daten trainiert. Dieser Rekonstruktionsfehler wird anschließend als Indikator für Anomalien verwendet. Im Allgemeinen wird davon ausgegangen, dass der Rekonstruktionsfehler bei normalen Eingaben geringer ist, da sie nahe an den Trainingsdaten liegen, während der Rekonstruktionsfehler bei abnormalen Eingaben höher wird. Jedoch muss für diesen Fall zu Beginn ein Datensatz mit normalen Daten vorlegen.[3]

Im Fall, dass nur ein Datensatz vorliegt, der auch Anomalien beinhaltet und für den der Aufwand des Labelns und Aussortierens zu groß ist, spricht man von der sogenannten unüberwachten Anomalieerkennung. Da beim Trainieren auch Anomalien enthalten sind, kann es im Modell passieren, dass der Rekonstruktionsfehler auch bei Anomalien gering ist und diese somit nicht erkannt werden.[4]

In der Forschung gibt es viele verschiedene Ansätze um dem Problem dieser gemischten Datensätze zu begegnen: „Memory-augmented deep Autoencoder“[5] verwenden einen Speicher in dem typische Elemente während des Trainings gespeichert werden. Für die Rekonstruktion wird nicht das Encoding der Eingabe verwendet. Stattdessen wird mithilfe des Encoding das relevanteste Element des Speichers zur Konstruktion verwendet. In der Testphase können anomale Daten nicht gut mit den gespeicherten normalen Daten konstruiert werden, und können daher von den normalen Daten getrennt werden.

Ein weiteres Beispiel ist das „Deep autoencoding gaussian mixture model“[6]. Bei diesem gibt es neben dem eigentlichen Autoencoder ein zusätzliches Bewertungsnetzwerk, dass neben den rekonstruierten Daten auch die encodierten Daten als Eingabe erhält. Der „RSR-based AutoEncoder“ (RSRAE)[7] verwendet eine „robust subspace recovery“-Schicht. Diese Schicht zielt darauf ab, den zugrunde liegenden Subraum aus einer „normalen“ Darstellung der gegebenen Daten zu extrahieren und Ausreißer zu entfernen, die sich von diesem Subraum entfernt befinden. „Robust Deep AutoEncoder“ (RDAE)[8] ist von der robusten Hauptkomponentenanalyse inspiriert.Die Eingangsdaten werden in zwei Teile aufgeteilt, wobei ein Teil effektiv durch einen tiefen Autoencoder rekonstruiert werden kann und der andere Teil die Anomalien in den ursprünglichen Daten enthält. Beim „Deep Structured Energy Based Model“ (DSEBM)[9] wird auf sogenannte EMBs zurückgegeriffen, welche die Dichteverteilung der Daten abschätzen. Dazu wird eine Energiefunktion verwendet und diese durch ein tiefes neuronales Netz dargestellt. Für den „Improved AutoEncoder for unsupervised Anomaly Detection“ (IAEAD)[4] wird der Featurespace mithilfe des Anomalieerkennungsloss manipuliert, sodass normale Datenpunkte näher beieinander liegen.

Die Aufgabe der Masterarbeit soll sein, ein neues Autoencoder Modell zu bauen und zu untersuchen, dass lediglich mit den Trainingsdaten trainiert wurde, die einen Loss aufweisen, der kleiner ist, als ein festgelegter Schwellwert. Der Gedanke hierbei ist, dass nach einigen Iterationen der Loss für anomale Daten größer wird, als für normale Daten, da diese deutlich häufiger im Trainingsdatensatz vorkommen als anomale Daten. Zusätzlich soll der Einfluss von Regularisierung auf dieses Modell untersucht werden, um den Einfluss anomaler Daten weiter zu reduzieren. Der Einfluss verschiedener Anteile von anomalen Daten am gesamten Datensatz soll hinsichtlich der Anomalieerkennung bei der Untersuchung des Modells berücksichtigt werden. Außerdem soll die Effektivität dieses Modell im Vergleich mit den bisher vorgestellten Autoencodern untersucht werden.

Die Forschungsfrage wird wie folgt bearbeitet: Zuerst müssen passende Datensätze gefunden werden, die für das Training der Modelle verwendet werden können. Hierfür könnten Bilddatensätze wie zum Beispiel der MNIST oder der CIFAR Datensatz verwendet werden, da diese auch in [5] und [4] verwendet werden. Durch die Verwendung von vorverarbeitenden Datensätzen entfällt eine aufwendige Aufbereitung. Die Hyperparameter der Modelle müssen festgelegt werden. Dazu zählen die Anzahl der Schichten und die Anzahl der Neuronen pro Schicht, die Parameter für die Regularisierung, die Aktivierungsfunktionen und die Anzahl an Trainingsepochen. Orientiert wird sich hierbei an der aktuellen Literatur [10, 11]. Außerdem muss herausgefunden werden, ob von Anfang an die Daten anhand des Losses zum Trainieren verwendet werden, oder ob am Anfang alle Daten verwendet werden.

Zuletzt werden die erstellten Modelle gegen die in der Einleitung genannten Modelle verglichen. Diese müssen entweder ebenfalls implementiert werden oder es muss eine geeignete Implementation gefunden werden, die für einen Vergleich verwendet werden kann. Dabei muss auf ähnliche Hyperparameter geachtet werden, um die Vergleichbarkeit zu gewährleisten. Als Vergleichsmetrik bieten sich Precison, Recall, der F1-Score und AUROC an.

References

[1] L. Sabetti, R. Heijmans, Shallow or deep? training an autoencoder to detect anomalous flows in a retail payment system, Latin American Journal of Central Banking 2 (2) (2021) 100031. doi:https://doi.org/10.1016/j.latcb.2021.100031.
URL https://www.sciencedirect.com/science/article/pii/S2666143821000119

[2] M. E. Tschuchnig, M. Gadermayr, Anomaly detection in medical imaging - a mini review, in: P. Haber, T. J. Lampoltshammer, H. Leopold, M. Mayr (Eds.), Data Science – Analytics and Applications, Springer Fachmedien Wiesbaden, Wiesbaden, 2022, pp. 33–38.

[3] U. Michelucci, An introduction to autoencoders, CoRR abs/2201.03898 (2022). arXiv:2201.03898.
URL https://arxiv.org/abs/2201.03898

[4] Z. Cheng, S. Wang, P. Zhang, S. Wang, X. Liu, E. Zhu, Improved autoencoder for unsupervised anomaly detection, International Journal of Intelligent Systems 36 (12) (2021) 7103–7125. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/int.22582, doi:https://doi.org/10.1002/int.22582.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22582

[5] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, A. van den Hengel, Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection, CoRR abs/1904.02639 (2019). arXiv:1904.02639.
URL http://arxiv.org/abs/1904.02639

[6] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, H. Chen, Deep autoencoding gaussian mixture model for unsupervised anomaly detection, in: International Conference on Learning Representations, 2018.
URL https://openreview.net/forum?id=BJJLHbb0-

[7] C.-H. Lai, D. Zou, G. Lerman, Robust subspace recovery layer for unsupervised anomaly detection (2019). arXiv:1904.00152.

[8] C. Zhou, R. C. Paffenroth, Anomaly detection with robust deep autoencoders, in: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17, Association for Computing Machinery, New York, NY, USA, 2017, p. 665–674. doi:10.1145/3097983.3098052.
URL https://doi.org/10.1145/3097983.3098052

[9] S. Zhai, Y. Cheng, W. Lu, Z. Zhang, Deep structured energy based models for anomaly detection (2016). arXiv:1605.07717.

[10] E. Ordway-West, P. Parveen, A. Henslee, Autoencoder evaluation and hyper-parameter tuning in an unsupervised setting, in: 2018 IEEE International Congress on Big Data (BigData Congress), 2018, pp. 205–209. doi:10.1109/BigDataCongress.2018.00034.

[11] F. Dernoncourt, S. Nemati, E. B. Kassis, M. M. Ghassemi, Hyperparameter Selection, Springer International Publishing, Cham, 2016, pp. 419–427. doi:10.1007/978-3-319-43742-2_29.
URL https://doi.org/10.1007/978-3-319-43742-2_29