UNIKASSEL VERSITÄT

mrt
Mess- und Regelungstechnik
Univ.-Prof. Dr.-Ing. Andreas Kroll

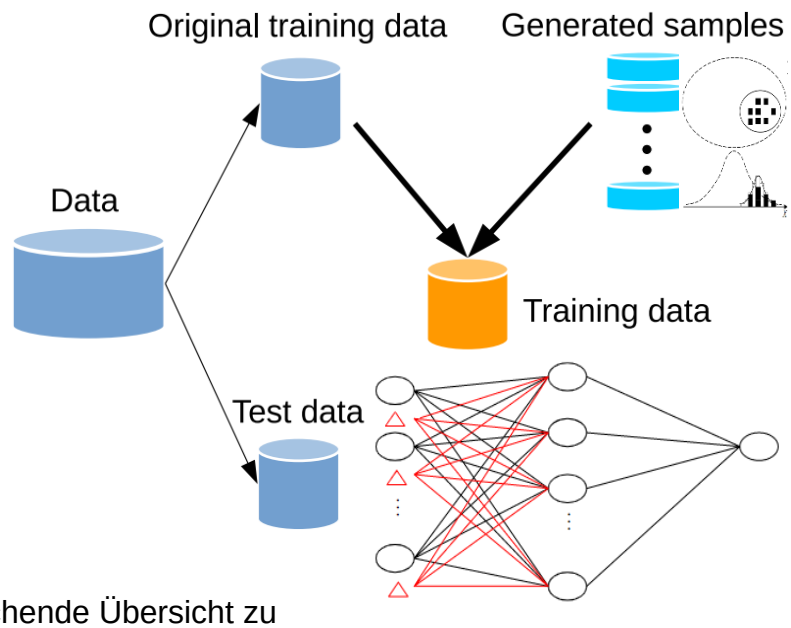
Seminar-, Semester-, Projekt- oder Bachelorarbeit
Data-Augmentation zur Erhöhung der Genauigkeit maschineller Lernalgorithmen

NN

Im Forschungsprojekt EEp-Beton wird ein datengetriebenes Modell für die Echtzeitvorhersage von Betoneigenschaften in der Produktion und für die Qualitätssicherung entwickelt. Das Hauptproblem bei der Modellierung für Betonqualitätsvorhersagen ist die geringe verfügbare Datenmenge.

Maschinelle Lernalgorithmen (ML) sind interessant, weil mittel ML aus gesammelten Daten automatisch die in diesen Daten bereits vorhandene Struktur analysiert und gelernt werden kann. Aber die Anwendung von ML erfordert große Datenmengen, die bei technischen und wissenschaftlichen Problemen nicht immer verfügbar sind.

Um mit kleinen Datenmengen auszukommen, sollte die Anzahl der Trainingsdaten erhöht werden. Ein Ansatz zur Lösung dieses Problems besteht darin, den gemessenen Daten künstliche Daten hinzuzufügen. Die Verwendung solcher Data-Augmentation (DA) Techniken kann zu einer Verbesserung der Genauigkeit des Lernalgorithmus führen, ohne zusätzliche Kosten für die Datenerzeugung wie bei Experimenten zu verursachen.



In dieser Arbeit soll eine vergleichende Übersicht zu verschiedenen entwickelten Data-Augmentation-Techniken erstellt und ausgearbeitet werden.

Folgende Teilaufgaben sind vorgesehen:

- Recherchieren und Einarbeitung in die verschiedenen Konzepte, Typen und Darstellungen von DA-Algorithmen.
- Methodischer Vergleich von DA-Algorithmen, dabei Untersuchung ihrer Grenzen, Herausforderungen, Anwendungsbereiche und Entwurfsverfahren sowie von Vor- und Nachteilen.
- Implementierung und Test einer exemplarischen Methode auf der Basis des DIAB-Datensatzes in MATLAB oder Python und Auswertung der Ergebnisse.
- Dokumentation der Untersuchungen und Kolloquiumsvortrag

Betreuer: F. Rezazadeh M.Sc., Univ.-Prof. Dr.-Ing. A. Kroll

Beginn: Ab sofort möglich
Geplantes Ende: Nach Vereinbarung

Literaturhinweise:

[1] Lateh, M. A., Muda, A. K., Yusof, Z. I. M., Muda, N. A., & Azmi, M. S. “Handling a small dataset problem in prediction model by employ artificial data generation approach: A review”. In *Journal of Physics: Conference Series*, 2017.

[2] Tsai, T. I., Yeh, C. W., Lin, L. S., & Wen, I. H. “Improving the learning accuracy of the mega trend diffusion technique for small data sets”. *International Congress on Advanced Applied Informatics IEEE,* 2019.

[3] Li, D. C., Wu, C. S., Tsai, T. I., & Lina, Y. S. “Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge”. *Computers & Operations Research*, 2007.

[4] Ohno, H. “Auto-encoder-based generative models for data augmentation on regression problems”. *Soft Computing*, 2020.

[5] Lu, L., Dao, M., Kumar, P., Ramamurty, U., Karniadakis, G. E., & Suresh, S. “Extraction of mechanical properties of materials through deep learning from instrumented indentation”. *Proceedings of the National Academy of Sciences*, 2020.

[6] Li, D. C., Hsu, H. C., Tsai, T. I., Lu, T. J., & Hu, S. C. “A new method to help diagnose cancers for small sample size”. *Expert Systems with Applications*, 2007.

[7] Goodfellow, I. J., Shlens, J., & Szegedy, C. “Explaining and harnessing adversarial examples”. *arXiv preprint arXiv:1412.6572*, 2014.

[8] Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems* 30, 2017.