

## Finding the signal regions in unprocessed 1H NMR spectra with deep learning

Das Finden von Signalregionen in den unverarbeiteten 1H-Nuclear-Magnetic-Resonance-Spektren (1H-MNR-Spektren) ist selbst für geschulte Experten auf diesem Gebiet eine zeitaufwändige und eintönige Aufgabe. Messabweichungen wie Phasenverschiebungen, Baseline-Verformungen oder Verunreinigungen deformieren das ansonsten klar definierte Spektrum. Berechnungsmethoden zur Erkennung von Signalbereichen existieren bereits, sind jedoch nicht stabil genug, um eine hohe Genauigkeit zu erzielen.

In dieser Bachelorarbeit werden die drei Deep-Learning-Architekturen Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) und Deep Q-Learning Networks (DQNs) zur Erkennung von Signalregionen verglichen. Die Modelle sind darauf trainiert, vielversprechende Signalbereiche in unverarbeiteten 1H-NMR-Spektren automatisch zu erkennen, wobei Phasenverschiebungen, Baseline-Verformungen und Verunreinigungen berücksichtigt werden. Die CNNs, RNNs und DQNs werden gegen synthetisch erzeugte Spektren getestet, durch Ändern von Hyperparametern, Hinzufügen von Schichten oder Erhöhen der Trainingsdaten verbessert und sowohl miteinander als auch mit dem Baseline Model verglichen. Die Genauigkeit der besten Modelle wird anhand experimenteller Spektren, welche von Experten gelabelt wurden, weiter analysiert.

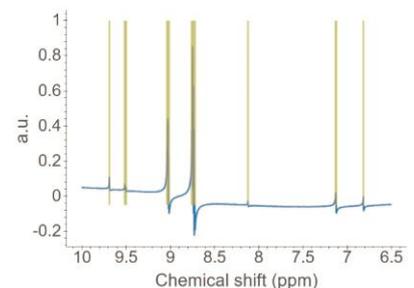
Die Ergebnisse zeigen, dass diese Algorithmen leistungsstarke Werkzeuge zum Auffinden relevanter Peaks in synthetisch erzeugten 1H-NMR-Spektren sind, da sie das Baseline Model um einen signifikanten Betrag übertreffen. Im Vergleich zum Baseline Model mit einem *mean Intersection over Union (mean IoU)* von 0,767, erreichten das CNN, RNN und DQN einen *mean IoU* von 0,933, 0,960 bzw. 0,950 für Spektren ohne signifikante Messabweichungen. Für Spektren mit zufällig generierten Phasenverschiebungen, Baseline-Verformungen und Verunreinigungen erreichte das Baseline Model einen *mean IoU* von 0,088, während das CNN, RNN und DQN weiterhin hohe *mean IoUs* von 0,929, 0,959 und 0,879 aufwiesen. Das Testen von experimentellen Spektren war jedoch eine herausfordernde Aufgabe, da die Experten die Signalbereiche zu inkonsistent gelabelt hatten und der Normalisierungsprozess nicht optimal war.

Weitere Studien könnten versuchen, den Normalisierungsprozess zu optimieren oder die Modelle auf die Signal-to-Noise-Ratio anstelle der Signalbereiche zu trainieren. Dies würde es einem Benutzer ermöglichen, einen eigenen Threshold zu definieren, ab welchem ein Peak zur Signalregion gezählt wird.

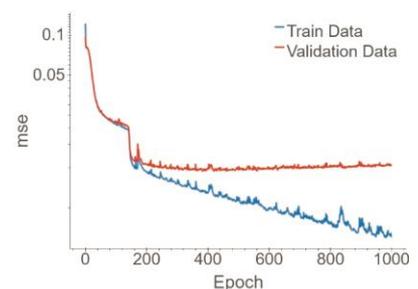


Diplomierende  
Cornel Toni Kleindl  
Benjamin Ricchiuto

Dozierende  
Samuel Beer  
Thomas Oskar Weinmann  
Heidi Gebauer



Markierte Signalregion in unverarbeitetem 1H-NMR-Spektrum, vorhergesagt von einem neuronalen Netz



Plot des Loss der Trainings- und Validierungs-Daten pro Epoche eines Recurrent Neural Network