

Zusammenfassung

Eine Verteilungsverschiebung zwischen Trainings- und Testdaten beeinträchtigt die Leistung tiefer neuronaler Netze. Sowohl beim Transfer-Lernen als auch beim *Few-shot learning* kann mit Zugang zu wenigen Daten das Modell an die Verteilungsverschiebung angepasst werden. Leider ist es unklar, wie das feinabgestimmte Modell das Wissen aus vortrainierten Modellen nutzt.

Daher analysieren wir die Ähnlichkeit der Repräsentationen der vortrainierten und der feinabgestimmten Netzwerke. Wir stellen fest, dass eine große Repräsentationsänderung in den Schichten vor der letzten Schicht für die domänenübergreifende Anpassung entscheidend ist. Umgekehrt findet in den frühen Schichten keine Veränderung der Repräsentation statt, wenn das Modell nur mit einer begrenzten Datenmenge oder sogar nur mit ganz wenigen Daten pro Klasse gelernt wird, wie in *Few-shot learning*.

Diese Beobachtung veranlasst uns, eine Methode zu entwickeln, die eine Repräsentationsänderung in den früheren Schichten bei *Few-shot learning*-Problemen bewirkt. Wir schlagen eine nicht-parametrische Erweiterung der *Model-agnostic Meta-Learning*-Methode vor, die eine effektive Initialisierung eines Netzwerks über eine Verteilung von Problemen mit Gradientenabstieg löst. Wir ersetzen die letzte Schicht des Netzes durch einen nicht-parametrischen Kopf, um Gradientenfluss weiterhin zu ermöglichen. Dies führt zu einer größeren Änderung der Repräsentation in früheren Schichten, sodass das Modell sich besser an neue Domänen anpasst, was zu einer verbesserten Leistung führt. Darüber hinaus zeigt der Ansatz die Fähigkeit, flexibel auf Veränderungen in der Dimensionalität der Probleme zu reagieren, insbesondere auf Veränderungen bei der Anzahl der Klassen.

Die Möglichkeit, sich an wenigen Daten anzupassen, ist bei der Domänengeneralisierung nicht gegeben. Daher muss das Modell alternatives Wissen nutzen, wie z. B. die Aufteilung der Daten in Domänen. Wir schlagen vor, domänenspezifische Informationen einzubeziehen, um die Daten zu erweitern und dadurch klassenrelevante Merkmale hervorzuheben. Darüber hinaus untersuchen wir die Durchführbarkeit des Lernens domäneninvarianter Repräsentationen mit *Auto-encodern*. Trotz der geringfügig besseren Leistung der domänenspezifischen

Erweiterung behindert die Nutzung domänenspezifischen Wissens die Übertragbarkeit auf andere Probleme. Der domäneninvarianter *Auto-encoder* hat das Potenzial, die Domänengeneralisierung zu verbessern. Allerdings stellt das Ausbalancieren unterschiedlicher Optimierungsfunktionen eine Herausforderung dar, sodass der Erfolg stark von einer gelungenen Modellselektion abhängt.

Insgesamt untersucht diese Dissertation die negativen Effekte von Verteilungsschiebungen und entwickelt darauf aufbauend Methoden zu deren Minderung.

Abstract

Distribution shift between the training and test data diminishes the performance of deep neural networks. In transfer learning and few-shot learning tasks, access to limited data enables fine-tuning to adapt to the shift. However, it remains unclear how a fine-tuned model leverages knowledge from a pre-trained task.

Thus, we analyze the similarity of the representations of the pre-trained and fine-tuned networks. We elucidate that the representation change in the layers preceding the final layer is crucial for cross-domain adaptation. Conversely, representation change does not occur in the early layers if we update the model on limited data or even only a few samples per class, as in few-shot learning.

This observation leads us to develop a method that raises representation change in few-shot learning. We propose a non-parametric extension of the model-agnostic meta-learning method, which meta-learns an effective initialization across multiple tasks with gradient descent. We replace the last layer of the network with a non-parametric head while still enabling gradients to flow through. This induces representation change in earlier layers such that the approach can better adapt to novel domains, resulting in an improved performance. Simultaneously, the approach demonstrates the ability to respond flexibly to changes in task dimensionality, specifically variations in the number of classes.

The opportunity to adapt to a few samples is absent in domain generalization. Thus, the model has to leverage alternative priors, such as the organization of data in domains. We incorporate task-specific domain information to augment the data, emphasizing class-relevant features. Furthermore, we investigate the feasibility of learning domain-invariant representations using an auto-encoder, which aims for sparse reconstruction. Although task-specific augmentation marginally improves performance, utilization of task-specific knowledge of one problem impedes transferability to other problems. Our domain-invariant auto-encoder exhibits the potential to enhance the task. However, balancing loss functions presents a challenge, such that success depends on fortuitous model selection.

Overall, this thesis examines the adverse effects of distribution shifts and, based on this analysis, proposes methodologies to mitigate them.