h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Ensemble-basiertes Handling von Concept Drifts durch Messung der Modell-Unsicherheit

Paula Möller
Hochschule Darmstadt - Fachbereiche Informatik und Mathe und Naturwissenschaften

Ziel der Arbeit

Oft wird in Einsatzbereichen von Machine Learning bzw. Deep Learning auf regelmäßiges Re-Training der Modelle gesetzt, um deren Aktualität zu erhalten. Gerade wenn dies aber nur an zuvor festgelegten Zeitpunkten oder nach dem Eintreffen einer bestimmten Mengen an Daten in Streams ausgelöst wird, kann ein Modell vor einer erneuten Anpassung schon nicht mehr aktuell sein. Ursächlich liegt dem häufig eine Veränderung der Verteilungen zugrunde. Diese Veränderung wird auch als Concept Drift bezeichnet.
Wenn sich die Veränderung nur in der Verteilung der Features zeigt, die aber keinen Einfluss auf die Zielvariable hat, wird von einem virtuellen Drift gesprochen. Um reale Concept Drifts als eine Veränderung der gemeinsamen Wahrscheinlichkeitsverteilung $P^t(X, y) \neq P^{t+\Delta}(X, y)$ von den Features und der Zielvariablen (Label) zu erkennen, werden häufig überwachte Lernansätze verwendet. Diese führen allerdings zu dem Nachteil, von dem Vorhandensein von den Labels abhängig zu sein. Gerade in Streams von Daten kann also ein Concept Drift somit erst erkannt werden, wenn auch die Abweichung von der Vorhersage zu dem wahren Wert ermittelt werden kann. Im schlechtesten Fall liegen große Zeitabschnitte dazwischen. Abhilfe kann durch unüberwachte Ansätze geschaffen werden, die bspw. auf der prädiktiven Modell-Unsicherheit (Model Uncertainty) beruhen.

Model Uncertainty

Neben der Data Uncertainty, die das konstante oder veränderliche Rauschen in den Daten beschreibt, gibt die Model Uncertainty die Unsicherheit bezogen auf das datengenerierende Modell an, die durch die Modell-Parameter wiedergegeben wird. Damit kann Model Uncertainty auch als Abweichung der Vorhersage beschrieben werden. Diese Abweichung ließe sich durch das Vorliegen mehrerer Vorhersagen berechnen. Besonders häufig verwendet werden Sampling-Methoden, zu denen Monte-Carlo Dropouts oder Ensembles gezählt werden.

Uncertainty Drift Detection (UDD)

In der UDD [2] wird die Model Uncertainty durch ein Neuronales Netz unter Monte-Carlo Dropouts quantifiziert. Um die Abweichung in Form der Varianz zu berechnen, werden die Daten mehrfach durch das Netz propagiert, während das Neuronale Netz unter dem Einfluss von Monte-Carlo Dropouts veränderliche Strukturen aufweist. Dafür werden wir in Abbildung 1 randomisiert einzelne Knoten durch Dropouts zu Trainings- und Test-Zeit ausgeblendet.

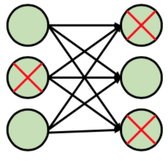

Fig. 1: Monte-Carlo Dropouts[1]

Laut den Veröffentlichenden konnte durch die Nutzung von einem Neuronalen Netz unter Monte-Carlo Dropouts das beste Ergebnis im Vergleich zu anderen Modellen erzeugt werden.

Algorithmus der EUDD

Da sich in der Quantifizierung von Model Uncertainty in zahlreichen Studien Ensembles als die konservativeren Schätzer gezeigt haben und flexibler in den Anpassungsstrategien nach einem Concept Drift sind, wird die UDD um ein Ensemble erweitert. Der neue Algorithmus wird als **Ensemble Uncertainty Drift Detection (EUDD)** bezeichnet. Verwendete Datensätze werden gestreamt. Die ersten 5% der Daten dienen dem initialen Training der Modelle des Ensembles. In Klassifikations-Modellen wird die Model Uncertainty anhand der Softmax-Ausgaben geschätzt. Regressions-Modelle schätzen als Ausgabe sowohl den wahren Wert als auch die Varianz als Maß für die Model Uncertainty. Die Model Uncertainty wird anschließend durch einen ADWIN Detektor ausgewertet. Dieser wird mit den nachfolgenden 10% der Daten parametrisiert. Auf den restlichen 85% der Daten finden Vorhersagen des wahren Wertes und der Model Uncertainty statt, um Concept Drifts zu detektieren. Auf eine Detektion folgt ein Re-Training aller Modelle des Ensembles.

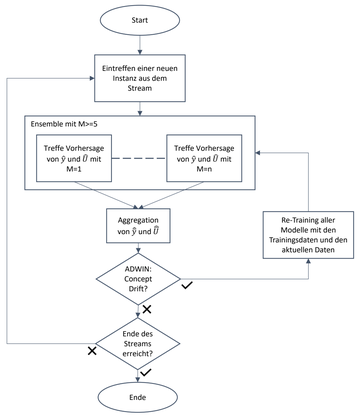

Fig. 2: Ablaufdiagramm des Algorithmus

Für den Nachweis der Wirksamkeit der EUDD wurden synthetische Datensätze verwendet. Um einen direkten Vergleich zu liefern, wurden weitere Strategien der Modellanpassung verwendet.
Außerdem fand ein Vergleich mit der UDD statt, indem die gleichen Real-World Datensätze und Metriken verwendet wurden.

Ergebnisse: synthetische Datensätze

In Klassifikations-Daten konnten alle realen Concept Drifts detektiert werden. Die Ergebnisse einer reaktiven Anpassung auf Concept Drifts waren besser als die der Vergleichsstrategien.

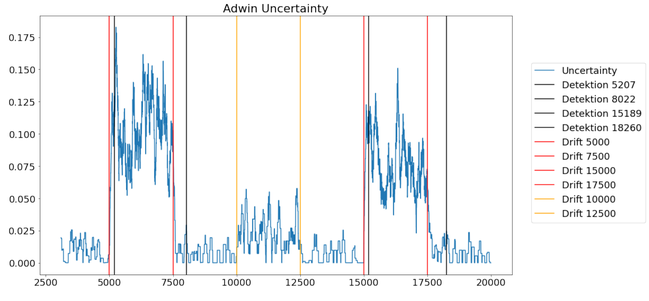

Fig. 3: Darstellung der Concept Drifts und Detektionen auf einem synthetischen Datensatz

Auf Regressions-Daten zeigte das Ensemble eine Reaktion auf virtuelle Drifts.

Ergebnisse: Real-World Daten

Für Regressions-Daten wird der Vergleich anhand des RMSE vorgenommen. Hier gilt: je kleiner, desto besser.
Bei den Klassifikations-Datensätzen findet der Vergleich anhand des MCC statt. Dieser liegt im Bereich von $[-1, 1]$ und gibt die Korrelation der Vorhersagen zu den tatsächlichen Werten an, wobei genau gegensätzliches gilt: je größer, desto besser.
Von den neun Datensätzen konnte die UDD nur in einem Fall ein besseres Ergebnis erreichen. Zu erwähnen ist hier, dass die Menge der Air Quality Daten verhältnismäßig klein ist. Für alle anderen Datensätze konnte mit der Ermittlung der Model Uncertainty durch ein Ensemble in der EUDD ein besseres Ergebnis erzielt werden.

Datensatz	Algorithmus	RMSE
Air Quality	*UDD*	$\mathbf{1.15(14)}$
	EUDD	$13.92(30)$
Bike Sharing	*UDD*	$129.93(5)$
	EUDD	$\mathbf{0.85(13)}$

Tab. 1: Vergleich auf Regressions-Daten

Datensatz	Algorithmus	MCC
Insects Abrupt	*UDD*	$0.52(9)$
	EUDD	$\mathbf{0.54(5)}$
Insects Inc	*UDD*	$0.24(4)$
	EUDD	$\mathbf{0.48(3)}$
Insects IncAbr	*UDD*	$0.52(22)$
	EUDD	$\mathbf{0.53(20)}$
Insects IncReo	*UDD*	$0.21(10)$
	EUDD	$\mathbf{0.61(21)}$
KDDCUP99	*UDD*	$0.96(20)$
	EUDD	$\mathbf{0.99(0)}$
Gassensor	*UDD*	$0.48(39)$
	EUDD	$\mathbf{0.66(44)}$
Electricity	*UDD*	$0.44(13)$
	EUDD	$\mathbf{0.48(9)}$

Tab. 2: Vergleich auf Klassifikations-Daten

Diskussion und Ausblick

Das Ensemble der EUDD konnte die Verwendung eines Neuronalen Netzes unter Monte-Carlo Dropouts in der UDD übertreffen. Verstärkt wird dies durch eine optimierte Konfiguration des Ensembles. Für die Klassifikations-Daten zeigten sich bessere Ergebnisse bei einer Verringerung der Hidden Layer der einzelnen Modelle und einer Erhöhung der Anzahl der Modelle im Ensemble. Für die Regressions-Daten konnte nur letzteres eine Verbesserung erreichen.
Zudem war es am erfolgreichsten, alle Modelle neu zu trainieren. Einzelne oder die Hälfte der Modelle neu zu trainieren zeigte nicht den erwarteten Erfolg.
In zukünftiger Forschung könnte die EUDD um verschiedenen Modelltypen erweitert werden, die auch eine Schätzung der Varianz im Output-Layer ermöglichen. Somit könnten bspw. LSTM-Netze verwendet werden. Zudem könnte zusätzlich die Data Uncertainty quantifiziert werden, um eine gegenüber virtuellen Drifts robustere Detektion durchzuführen.

Literatur

[1] M. Abdar et al. *A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges*. 2021. arXiv: 2011.06225 [cs.LG].

[2] L. Baier et al. *Detecting Concept Drift With Neural Network Model Uncertainty*. 2021. arXiv: 2107.01873 [cs.LG].