

Kurzfassung

Basierend auf einem speziellen maschinellen Lernverfahren, bestehend aus einem künstlich neuronalen Netz (ANN) als Approximationsfunktion und einem erweiterten Kalman Filter (EKF) als Optimierungsmethode, wird eine Methode zur Identifikation von geodätischen Systemen in dieser Dissertation abgeleitet. Dieser ANN-EKF-Lernansatz wird durch die Berücksichtigung von daten-, modell- oder daten- und modellbezogenen Unsicherheiten erweitert. Je nach geodätischer Anwendung wird jeweils ein Aspekt der drei Unsicherheitsgruppen berücksichtigt. Bei den zu identifizierenden Systemen handelt es sich um einen Roboterarm, für den die Positionsunsicherheit durch Kalibrierung verbessert werden soll, und um Immobilien, deren Wert bestimmt werden soll. Während der Roboterarm ein steuerbares, datenproduzierendes System mit einer realisierbaren Referenz und geringer lokaler Heterogenität ist, handelt es sich bei den Immobilien um ein System, das in den Einflussgrößen, in den Datenstichproben und in der Realisierung einer Referenz begrenzt ist und wenig empirische Evidenz und starke lokale Heterogenität aufweist.

Die Unterschiede in den geodätischen Systemen führen zu unterschiedlichen Forschungszielen. Das Roboterarmsystem bietet eine kontrollierte Umgebung (Daten können erzeugt werden, eine Referenz kann realisiert werden) und Wissen über geometrische/physikalische Gesetze ist verfügbar. Daher ist es möglich, zwei Weiterentwicklungen des ANN-EKF-Lernansatzes für die Roboterarm-Anwendung zu verfolgen. Diese basieren auf daten- und modellbezogenen Unsicherheiten und werden durch den EKF unterstützt. Die inhärente Varianzfortpflanzung im EKF ermöglicht die Berücksichtigung der zufälligen Variabilität der Messungen. Es werden realistische Zufallsvariabilitäten der durch eine Referenz gemessenen Roboterarm-Positionen abgeleitet. Aufgrund der vorhandenen Systembeschreibung im EKF ist es möglich, auch ein geometrisches Modell in das ANN-EKF-Lernverfahren zu integrieren. Ziel ist es, einen methodischen Rahmen zu schaffen, mit dem die Unvollständigkeit des Modells reduziert werden kann.

Immobilien sind ein begrenztes und komplexes System. Der größte limitierende Faktor der Immobilienbewertung auf Basis der deutschen Kaufpreissammlung sind die kleinen Datenstichproben. Um ein vorteilhaftes ANN-EKF-Lernen zu erreichen, wird die Datenstichprobe durch Aggregation von Teilmärkten vergrößert. Dies vergrößert zwar die Stichprobe, erhöht aber auch die Komplexität des Systems. Die unvollständigen Daten verursachen eine modellbedingte Unsicherheit. Daher wird die Interaktion zwischen daten- und modellbezogener Unsicherheit analysiert.

Die erste Erkenntnis befasst sich mit den Auswirkungen der Berücksichtigung von datenbezogenen Unsicherheiten beim ANN-EKF-Lernen. Das beste Generalisierungsergebnis (minimaler Testfehler) wird durch die Berücksichtigung realistischer datenbezogener Unsicherheiten beim ANN-EKF-Lernen der Positionskorrekturen des Roboterarms erreicht. Das Lernverfahren wird durch eine sinnvolle Adaption der Lernrate aussagekräftiger. Darüber hinaus werden unbedeutende ANN-Parameter eliminiert und kleinere Modellstrukturen können erreicht werden. Die zweite Erkenntnis behandelt die Entwicklung eines Rahmens für die Integration eines parametrischen Modells in das ANN-EKF-Lernverfahren, um eine Reduktion der modellbedingten Unsicherheit zu ermöglichen. Der integrierte ANN-EKF-Ansatz wird als Residualmodell aufgebaut. Die Prüfung der Funktionalität auf der Basis simulierter Daten zeigt, dass die Einführung einer zusätzlichen ANN-Iteration notwendig ist, um die Adaptivität von ML-Ansätzen zu erreichen. Die Analyse der Modellkomplexität in Abhängigkeit von reduzierten Stichproben ermöglicht die Ableitung von minimal erforderlichen Stichprobengrößen für das ANN-EKF-Lernen und entspricht der dritten Erkenntnis dieser Dissertation. Die Methoden der Kreuzvalidierung (CV) und der struk-

turellen Risikominimierung (SRM) werden zur Beschreibung der Modellkomplexität verwendet und für die Zwecke der geodätischen Systeme angepasst. Während die SRM eine kritische Anzahl von benötigten Stichproben bei 5000 ermittelt, liefert die CV eine stabile Region ab 9000 Stichproben für einen räumlich aggregierten Datensatz. Die Aggregation von Teilmärkten ist ein geeigneter Ansatz zur Vergrößerung des Stichprobenumfangs. Die Immobilienbewertung auf Basis des teilmarktübergreifenden ANN-EKF-Lernverfahrens ist mit lokalen Bewertungsverfahren vergleichbar.

Die ersten beiden Erkenntnisse beziehen sich auf zwei aktuelle Themen im wissenschaftlichen maschinellen Lernen (sciML). Dies sind die Berücksichtigung von Unsicherheiten im ML und physikalisch informierte maschinelle Lernverfahren (PIML), die für viele Anwendungsbereiche wie z.B. die Erdbeobachtung von Interesse sind. Der entwickelte Baukasten der ANN-EKF-Lernmethode ermöglicht die Berücksichtigung dieser beiden Aspekte. Die dritte Erkenntnis aus dieser Dissertation ergänzt diesen Baukasten, um diese mit Vorinformation integrierten Ansätze auf kleinere Modellkomplexitäten oder kleinere minimale Stichprobengrößen zu testen.