

Kurzfassung der Dissertation

Wissen in evolutionären und lernenden Systemen

von Yaochu Jin, Future Technology Research, Honda R&D Europe, Offenbach/Main

Wissen über Zusammenhänge und Systeme kann in unterschiedlichen Repräsentationen vorliegen. Die Nutzbarkeit (z.B. in Algorithmen) und die Verständlichkeit (z.B. für Menschen) des Wissens hängen von der Art der Repräsentation ab. In dieser Arbeit werden Ansätze zur Übertragbarkeit von Wissen zwischen unterschiedlichen Repräsentationen vorgeschlagen und untersucht.

Evolutionäre Algorithmen und Lernverfahren sind wirkungsvolle Ansätze zur Modellierung und Optimierung nichtlinearer Systeme. Im Allgemeinen benötigen beide Verfahren jedoch eine große Anzahl von Daten, die durch die Berechnung einer Qualitätsfunktion oder in Form von Trainingsdaten zur Verfügung gestellt werden müssen, was in vielen Fällen sehr aufwändig und nur eingeschränkt möglich ist. Gleichzeitig ist jedoch häufig Vorwissen, z.B. in Form von Heuristiken, über das System vorhanden. Ein Ziel dieser Arbeit ist es, solches Vorwissen für evolutionäre Algorithmen und Lernverfahren nutzbar zu machen. Dazu werden neue Algorithmen auf der Basis von Fuzzy-Logik und neuronalen Netzen entwickelt. Die grundlegende Idee ist es, Wissen mittels Fuzzy-Logik und neuronaler Netze zu erwerben und in einer Form zu repräsentieren, die seine weitere Nutzung in evolutionären Algorithmen und Lernverfahren ermöglicht. Dazu werden drei Ansätze entwickelt. Die erste Methode stellt ein Verfahren für die Übertragung menschlicher Heuristiken auf neuronale Netze dar. Das Einbeziehen von Expertenpräferenz in mehrkriterielle Optimierung bildet den zweiten in dieser Arbeit verfolgten Ansatz. Die letzte Methode befasst sich mit der Wissensübermittlung zwischen unterschiedlichen evolutionären und lernenden Systemen.

Die Ergebnisse der vorliegenden Arbeit zeigen zum einen, dass es von großer Bedeutung ist, Wissen in geeigneter Form in evolutionäre Algorithmen und Lernverfahren einzubringen. Auf der anderen Seite stellt der umgekehrte Prozeß, durch Evolution bzw. Lernen gefundenes Wissen zur Verfügung zu stellen, eine nicht weniger wichtige Aufgabe dar. Im Allgemeinen liegt das in trainierten neuronalen Netzen gespeicherte Wissen in einer für den Menschen nicht zugänglichen Repräsentation vor. Obwohl Wissen in Form von Fuzzy-Regeln im Prinzip zwar verständlich ist, kann die Interpretierbarkeit dieser Regeln jedoch verloren gehen, insbesondere wenn sie aus Daten erzeugt worden sind. Daher ist das zweite Ziel der Dissertation, Wissen in Form von *interpretierbaren* Fuzzy-Regeln zu extrahieren. Dazu wird eine Definition von Interpretierbarkeit vorgeschlagen, die erfüllt werden muß, um die Verständlichkeit des von den Regeln dargestellten Wissens zu gewährleisten. Zwei Methoden werden untersucht. In dem ersten Ansatz werden interpretierbare Regeln direkt aus Daten mit Hilfe evolutionärer Algorithmen gewonnen. Bei der zweiten Methode wird die Ähnlichkeit zwischen Fuzzy-Systemen und trainierten radialen Basisfunktionsnetzwerken genutzt, um Wissen zu extrahieren.

Anhand verschiedener Testprobleme und industrieller Anwendungen wird gezeigt, dass das Einbringen von Vorwissen in evolutionäre Algorithmen und Lernverfahren nicht nur sehr hilfreich, sondern oftmals zwingend notwendig ist. Gleichzeitig ist die Wissensextraktion, d.h. die Darstellung von erworbenem Wissen in einer verständlichen Repräsentation, entscheidend, wenn die Überprüfbarkeit dieses Wissens erforderlich ist, z.B. bei sicherheitsrelevanten Applikationen.