

Mehrebenenvalidierung von kognitiven Benutzermodellen

Jeronimo Dzaack

Leon Urbas

Abstract: In diesem Beitrag werden verschiedene Methoden für die Mehrebenenvalidierung kognitiver Modelle im Kontext der Entwicklung und Gestaltung von Mensch-Maschine-Systemen vorgestellt und kurz diskutiert. Das Ziel liegt darin die Adäquatheit der kognitiven Benutzermodelle zu bewerten. Es wird gezeigt, dass dieses Methodenrepertoire eine wichtige Ergänzung heutiger Werkzeuge der kognitiven Modellbildung darstellt und dass die einzelnen Methoden für die Validierung einzelnen kognitiven Modelle sowie von Teams kognitiver Agenten eingesetzt werden können.

1 Einleitung

Ein wichtiger Aspekt bei der Bedienermodellierung und der Konstruktion von kognitiven Modellen stellt die Prüfung der Adäquatheit der kognitiven Modellierung dar. Dabei werden die Ergebnisse der Modellsimulation im Bezug zu einer empirischen Datenbasis validiert und strukturell verglichen, um deren Gültigkeit im Vergleich zum Menschen zeigen zu können. Das übergeordnete Ziel liegt in der Herstellung von Glaubwürdigkeit und Vertrauen in die Simulationsmodelle [BL92]. Besteht bei mathematischen und numerischen Modellen die Möglichkeit statistische Verfahren für die Adäquatheitsprüfung zu nutzen, können diese Verfahren für die Prüfung kognitiver Modelle, die hypothetische Konstrukte der menschlichen Kognition abbilden, kaum genutzt werden [Wal98]. Eine letztendliche Bewertung eines korrekten Ansatzes ist daher nicht möglich. Ein Simulationsmodell kann aber gegenüber konkurrierenden Ansätzen plausibilisiert werden [Sch93]. Dabei werden Experimente mit natürlichen und künstlichen kognitiven Systemen betrachtet (kognitive Empirie; [Str95]). Beispielsweise können empirische und simulierte Ergebnisse miteinander in Bezug gesetzt werden oder zwei kognitive Modelle, die auf unterschiedlichen Annahmen beruhen, verglichen werden. Zu diesem Zweck müssen die theoretische und die empirische Adäquatheit eines Modells gezeigt werden [Wal98]. Die empirische Adäquatheit beinhaltet den Nachweis über eine genügende Korrespondenz zwischen dem Verhalten des Benutzermodells und des Menschen. Die theoretische Adäquatheit kann anhand der Modelle und der zugrundeliegenden kognitiven Architektur geprüft werden. Das Ziel liegt darin zu zeigen, dass die gewählte Instanz zu theoretischen Annahmen korrekt ist.

2 Ebenenbetrachtung kognitiver Modelle

Für die Validierung kognitiver Benutzermodelle können drei Kontrollebenen der Modelle betrachtet werden [Gra07]. Die erste Kontrollebene stellt die Kommunikation zwischen funktionalen Modulen des kognitiven Apparates dar (z. B. visuelles Modul, deklaratives Modul). Die internen Prozesse eines kognitiven Moduls stellen die zweite Kontrollebene dar (z. B. die implementierten modulspezifischen Theorien). Die dritte Kontrollebene kombiniert aufgabenspezifisches Wissen mit den Merkmalen der kognitiven Architekturen, um Strategien oder Methoden für die Aufgabenbearbeitung zu erzeugen. Adressieren die ersten zwei Kontrollebenen die theoretische Adäquatheit einer kognitiven Simulation, fokussiert die dritte Ebene auf die empirische Adäquatheit. Werden die drei Ebenen betrachtet und die Modelle jeweils hinreichend validiert, kann von einem kognitiv adäquaten Modell gesprochen werden. Im Zusammenhang mit der Betrachtung von Teams kognitiver Agenten kann diese Ebenenbetrachtung übertragen werden. Die erste Ebene bezeichnet somit die grundlegende Analyse der kognitiven Agenten innerhalb der von Gray vorgeschlagenen Kontrollstrukturen. Die zweite Ebene beinhaltet die Interaktion der Agenten untereinander. Die dritte Ebene bezeichnet die Integration der Agenten und des Systems in einen speziellen Kontext.

3 Validierungsmethoden

Für die Validierung künstlicher kognitiver Systeme und deren Interaktion können fünf Arten von Korrespondenzen der zu bewertenden quantitativen Daten betrachtet werden: (1) Die Produktkorrespondenz, (2) die Korrespondenz von Zwischenschritten, (3) die temporale Korrespondenz, (4) die Lernkorrespondenz und (5) die Fehlerkorrespondenz (weiterführende Informationen bei [Wal98]). Aus der Auflistung ist ersichtlich, dass von (1) bis (5) die Granularität der Analyse ansteigt und eine tiefere Analyse der kognitiven Prozesse und Strukturen unterstützt. Je nach Betrachtungsgrad der Analyse können diese fünf Parameter variiert und kombiniert werden. Mit Verfahren aus der Methode *Goodness-of-fit* können Daten mit unterschiedlicher Herkunft für einen Untersuchungsgegenstand im Bereich der Modellbildung verglichen werden (für einen Überblick siehe [SW05]). Dabei werden visuelle und numerische Aggregationen der Daten genutzt, um zum einen das Auffinden von Unterschieden und Ähnlichkeiten zu unterstützen und zum andern die Genauigkeit der Modellierung bestimmen zu können. Eine weitere Methode stellt die modellvergleichende Argumentation dar [VBG84]. Dabei werden zentrale Annahmen, Strukturen, Mechanismen und Interaktionen einer Modellierung zu alternativen Annahmen verglichen, um die theoretische und strukturelle Stütze der Modellierung vergleichend validieren zu können. Zunehmend werden Werkzeuge entwickelt, die eine unterstützte Analyse der quantitativen Simulationsdaten ermöglichen, mit dem Ziel den Aufwand für die Einzelanalyse und die Gruppenanalyse (z. B. empirische und simulierte Daten) zu minimieren. Das Werkzeug ProtoMatch [SM05] wurde entwickelt, um die Einzelanalyse von Daten hinsichtlich der Gleichheit von Sequenzen, zu unterstützen (z. B. Blickdaten, Maus- und Tastatureingaben). Das Werkzeug SimTrA (Simulation Trace Analyzer; [Dza08]) ermög-

licht die Analyse von Daten unter Berücksichtigung deren inhärenten Komplexität (z. B. komplexe Analysemethoden für Blickdaten) und die Analyse und den Vergleich empirischer und simulierter Daten. Beide Werkzeuge unterstützen durch Importalgorithmen die Analyse von Simulationsdaten unterschiedlicher kognitiver Architekturen. Im Bereich der Modellierung mit speziellen kognitiven Architekturen (z. B. ACT-R – Atomic components of thought – rational; [ABB⁺04]) werden zunehmend spezifische Methoden entwickelt, die eine numerische Analyse der Simulationsdaten erlauben, um beispielsweise die kognitive Performanz der Modelle untereinander vergleichen und bewerten zu können (siehe [BCK03]). Zu diesem Zweck werden die Modelldaten extrahiert und algorithmisch mit Daten, die aus der formalisierten Theorie extrahiert wurden, verglichen. Das Ziel liegt in der Bestimmung der Qualität und Genauigkeit deren Übereinstimmung zueinander. Im Bereich der Teaminteraktion kognitiver Agenten können diese Methoden angewandt werden, um die kognitiven Agenten zu validieren. Für die übergreifende Analyse eignen sich Verfahren, die auf der Methode Goodness-of-fit beruhen. Zusätzlich können die klassischen heuristischen Maße (wie z. B. Dauern, Fehler) genutzt werden, die sich aus den fünf Graden der Modellvalidierung ableiten lassen.

4 Diskussion und Ausblick

Es konnte gezeigt werden, dass für die Analyse und Validierung kognitiver Modelle verschiedene Verfahren bekannt sind, die den Anwender unterstützen und eine wertvolle Ergänzung heutiger Werkzeuge in der kognitiven Modellbildung darstellen. Eine Validierung der Simulationsdaten einzelner kognitiver Agenten kann daher unterstützt und zum Teil automatisiert durchgeführt werden. Die Validierung von Teaminteraktion stellt eine neue Herausforderung dar, soweit diese nicht auf die Analyse der einzelnen Teile beschränkt bleiben soll. Für die übergreifende Analyse können Methoden übertragen und modifiziert werden. Bei der Analyse kognitiver Modelle müssen zwei Aspekte beachtet werden, um die Ergebnisse wissenschaftlich betrachten und nutzen zu können. Kognitive Architekturen und Modelle beruhen zumeist auf verbal kommunizierten Modellen. In der Modellierung kann es daher zu Unterschieden zwischen Theorie und tatsächlicher Implementierung kommen, die beispielsweise auf die geeignete Abstraktion des Gegenstandsbereiches zurückzuführen sind (Theory implementation gap; [CS95]). Zusätzlich kommt es durch das Hinzufügen von Zusatzannahmen während der Implementierung für die korrekte Ablaufsteuerung eines kognitiven Modells oder dessen Einbindung in einen größeren Verbund zu einer überfrachteten Implementierung, die oft keine eindeutige Trennung von essentiellen Teilen der Theorie und den Details der Implementierung zulässt (Irrelevant specification; [New90]). Die beispielhaft aufgeführten Methoden zeigen, dass erst spezifische Aspekte des Simulationsverhaltens abgedeckt und berücksichtigt werden. Ein Grund dafür liegt in dem Aufwand für die Analyse der Simulationsdaten aufgrund der Datenfülle und der oft nicht gesicherten Kenntnisse über die kognitiven und sozialen Prozesse (z. B. im Rahmen der Teaminteraktion). Es ist zu erwarten, dass in zukünftigen Forschungs- und Anwendungskontexten der Fokus mehr auf die Validierung der simulierten Daten verschoben wird, da diese eine grundlegende Voraussetzung für den Einsatz kognitiver Modelle im

Zuge der Evaluation von Mensch-Maschine-Systemen darstellt. Somit bleibt zu hoffen, dass diese ersten vielversprechenden Ansätze den Grundstein legen für eine weiterführende Erforschung und Erprobung von anwendbaren Verfahren.

Literatur

- [ABB⁺04] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere und Y. Qin. An Integrated Theory of the Mind. *Psychological Review*, 111(4):1036–1060, 2004.
- [BCK03] R. S. Baker, A. T. Corbett und K. R. Koedinger. Statistical techniques for comparing act-r models of cognitive performance. In *Proceedings of the tenth annual ACT-R workshop*, Seiten 129–134. Carnegie Mellon University, 2003.
- [BL92] W. Bub und P. Lugner. Systematik in der Modellbildung – Teil 2: Verifikation und Validierung. In *Modellbildung für Regelung und Simulation: Methoden, Werkzeuge, Fallstudien*, Seiten 19–43. VDI Verlag, 1992.
- [CS95] R. P. Cooper und T. Shallice. Soar and the case for unified theories of cognition. *Cognition*, 55:115–149, 1995.
- [Dza08] J. Dzaack. *Analyse kognitiver Benutzermodelle für die Evaluation von Mensch-Maschine-Systemen*. Dissertation, Technische Universität Berlin, 2008.
- [Gra07] W. D. Gray. Composition and control of integrated cognitive systems. In W. D. Gray, Hrsg., *Integrated models of cognitive systems*, Seiten 3–12. Oxford University Press, 2007.
- [New90] A. Newell. *Unified theories of cognition*. Harvard University Press, 1990.
- [Sch93] H. Schaub. *Modellierung der Handlungsorganisation*. Huber, 1993.
- [SM05] M. J. Schoelles und C. W. Myers. ProtoMatch: A tool for analyzing high-density, sequential eye gaze and cursor protocols, 2005.
- [Str95] H. Stroher. *Kognitive Systeme – Eine Einführung in die Kognitionswissenschaft*. West-deutscher Verlag, 1995.
- [SW05] C. D. Schunn und D. Wallach. Evaluating goodness-of-fit in comparison of models to data. Seiten 115–154. University of Saarland Press, 2005.
- [VBG84] K. VanLehn, J. S. Brown und J. Greeno. Competitive Argumentation in Computational Theories of Cognition. In W. Kintsch, J. R. Miller und P. G. Polson, Hrsg., *Method and Tactics in Cognitive Science*, Seiten 235–262. Laurence Erlbaum Associates, 1984.
- [Wal98] D. Wallach. *Komplexe Regelungsprozesse – Eine kognitionswissenschaftlicher Analyse*. Deutscher Universitäts Verlag, 1998.