



Computational Intelligence und Data Mining

Moderne statistische Verfahren zur experimentellen Versuchsplanung

Prof. Dr. Thomas Bartz-Beielstein
thomas.bartz-beielstein@fh-koeln.de, Tel. 02261/8196-6391

Prof. Dr. Wolfgang Konen
wolfgang.konen@fh-koeln.de, Tel. 02261/8196-6275

Prof. Dr. Hartmut Westenberger
hartmut.westenberger@fh-koeln.de, Tel. 02261/8196-6385

Zielsetzung

Methoden der statistischen Versuchsplanung kommen immer dann zum Einsatz, wenn mit möglichst geringem Aufwand (Kosten, Zeit) eine verlässliche Aussage über das Verhalten (Leistung) eines Systems zu treffen ist. Unter einem System kann in diesem Zusammenhang ein Objekt mit klar spezifizierten Eingabegrößen und messbaren Ausgabegrößen verstanden werden. Durch geschickte Variation der Belegungen der Eingabegrößen (sog. experimentellen Designs, siehe Abbildung 1) kann der Anwender wichtige Systemeigenschaften und ihre Abhängigkeit von den Eingabegrößen bestimmen.

So kann im Rahmen der industriellen Optimierung der Füllstand Y eines Klärbeckens in Abhängigkeit von der Regenmenge R , der Sonnenscheindauer S und der Bodenbeschaffenheit B experimentell untersucht werden. Gesucht ist ein Modell (z. B. ein neuronales Netz), das möglichst gut das Verhalten des Systems $Y \sim R + S + B$ widerspiegelt, um computergestützte Prognosen zur Vermeidung von Überflutungen zu generieren. In diesem Zusammenhang sprechen wir von problemspezifischen Parametern.

Auch Computerprogramme selbst können Gegenstand dieser experimentellen Untersuchungen sein. Die meisten Optimier- und Simulationsprogramme erfordern die Festlegung bestimmter Parameter, die die Güte der Verfahren maßgeblich beeinflussen. So ist die Güte Y eines neuronalen Netzes von der Anzahl der Neuronen N oder der Wahl der Schwellenwertfunktion S abhängig. Gesucht ist ein neuronales Netz, so dass $Y \sim N + S$ optimiert wird. Die in Frage kommenden Parameter heißen in diesem Zusammenhang algorithmenspezifische Parameter. Einen guten Einstieg in diese Thematik stellt [1] dar.

Durch die in den letzten Jahren enorm gewachsene Rechnerleistung stehen umfangreiche Möglichkeiten zur statistischen Analyse von Experimenten zur Verfügung, die die klassischen Regressionstechniken und Varianzanalysemethoden ideal ergänzen und erweitern. So werden Regressionsbäume und stochastische Prozessmodelle von uns erfolgreich zur Analyse komplexer Problemstellungen eingesetzt. Für die Datenvorverarbeitung haben sich Data Mining Methoden als sinnvoll herausgestellt. Somit kann durch eine sinnvolle Kombination unterschiedlicher klassischer mathematischer und moderner CI Methoden ein auf die spezielle Problemstellung zugeschnittenes Lösungsverfahren mit geringen Kosten entwickelt werden.

Forschungsarbeiten

Einladungen zu Vorträgen, Seminaren und Tutorien auf den wichtigsten Konferenzen im Bereich der Computational Intelligence wie z. B. der Genetic and Evolutionary Computation Conference (GECCO) in London, der Parallel Problem Solving From Nature (PPSN) Konferenz in Reykjavik und dem renommierten Dagstuhl-Seminar belegen die Aktualität und den Einfluss unserer Forschungsarbeiten [9, 13, 14].

Im Bereich der Grundlagenforschung besteht ein Austausch mit Vertretern des Neuen Experimentalismus. Ziel dieser einflussreichen philosophischen Strömung, zu der führende Forscher wie Sir David Cox oder Deborah Mayo zählen, ist die wissenschaftliche Fundierung experimentell gewonnener Erkenntnisse als Erweiterung des von Sir Karl Raimund Popper vorgeschlagenen Wissenschaftsbegriffs [8].



Abbildung 1. Glasfenster, das ein experimentelles Design darstellt. In jeder Zeile und jeder Spalte tritt jede Farbe nur einmal auf. Jede Farbe kann als die Einstellung eines Parameters interpretiert werden. Dieses Fenster wurde zu Ehren von R.A. Fisher (Statistiker und Genetiker) im Gonville and Caius College in Cambridge (UK) errichtet.

Neben dieser internationalen Sichtbarkeit gibt es eine enge Zusammenarbeit auf nationaler Ebene, die sich durch Mitgliedschaft und aktive Teilnahme an VDI/VDE Workshops und der Mitarbeit als Program Chair während des Workshops Hybrid Metaheuristics in Dortmund zeigt [15]. Hinzu kommen noch Gutachtertätigkeiten für diverse Fachzeitschriften und die Mitarbeit in zahlreichen Programmkomitees.

Kooperationen mit Industriepartnern im Bereich der experimentellen Versuchsplanung und dem Einsatz von CI Methoden entstanden ebenfalls im Laufe unserer Arbeiten. Exemplarisch ist die Kooperation mit Fujitec Japan im Bereich der Fahrstuhloptimierung zu nennen. Diese Zusammenarbeit wird durch eine Publikation in der Reihe „Advances in Industrial Control“ bei Springer [2] dokumentiert.

Intensive Kontakte auf Hochschulebene existieren zur Universität Dortmund [5, 7, 11], Universität Karlsruhe [6] und weiteren nationalen und internationalen Hochschulen. Unsere Forschungen werden auch innerhalb der Fachhochschule Köln vernetzt. Momentan existieren mehrere Wahlpflichtveranstaltungen an der FH Köln, die aktuelle CI Forschungsvorhaben zum Inhalt haben. Diplomarbeiten, die in zusammen mit Industriepartnern wie der Firma Protagen (Bioinformatik) oder dem Dortmund Intelligence Project (Portfoliooptimierung) durchgeführt werden, belegen den Austausch zwischen Hochschule und Industrie. Innerhalb der FH Köln gibt es enge Kooperationen in Forschung und Lehre im Umfeld von Computational Intelligence [3, 4, 12].

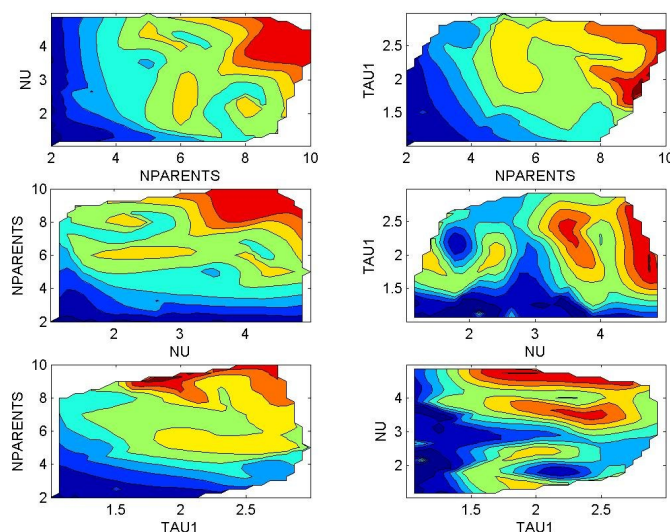
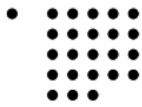


Abbildung 2. Performance eines Computerprogramms in Abhängigkeit von den algorithmenspezifischen Parametern. Blaue Bereiche kennzeichnen gute Einstellungen, rote Bereiche deuten auf eine schlechte Performance des Algorithmus hin.



Literatur

- [1] T. Bartz-Beielstein, *Experimental Research in Evolutionary Computation*. Springer. 2006.
- [2] S. Markon, H. Kita, H. Kise, T. Bartz-Beielstein. *Control of Traffic Systems in Buildings*. Springer. 2006.
- [3] T. Bartz-Beielstein, W. Konen, H. Westenberger. *Datenanalyse und Prozessoptimierung am Beispiel Kläranlagen*. Forschungsbericht FH Köln, 2007.
- [4] H. Westenberger, W. Konen, T. Bartz-Beielstein, *Business Intelligence an Hochschulen*, Forschungsbericht FH Köln, 2007.
- [5] B. Naujoks, D. Quagliarella, T. Bartz-Beielstein. *Sequential parameter optimisation of evolutionary algorithms for airfoil design*. ERCOFTAC. *Design and Optimization: Methods & Applications*. Las Palmas de Gran Canaria, Spain. 2006.
- [6] T. Bartz-Beielstein, D. Blum and J. Branke. *Particle Swarm Optimization and Sequential Sampling in Noisy Environments*. Springer Operations Research / Computer Science Interfaces Book Series. Karl F. Doerner et al., Eds. Berlin, Heidelberg, New York: Springer, 2006.
- [7] M. Preuß and T. Bartz-Beielstein. "Self-adaptation in evolution strategies—an experimental analysis based on sequential parameter optimization," in *Parameter Setting in Evolutionary Algorithms*, ser. *Studies in Computational Intelligence*, F. Lobo, C. Lima, and Z. Michalewicz, Eds. Berlin, Heidelberg, New York: Springer, 2006.
- [8] T. Bartz-Beielstein *Neyman–Pearson Theory of Testing and Mayo’s Extensions Applied to Evolutionary Computation* (Invited talk). Presented at the First Symposium on Philosophy, History, and Methodology of E.R.R.O.R (Experimental Reasoning, Reliability, Objectivity & Rationality: Induction, Statistics, Modelling). Virginia Tech, Blacksburg, VA. 2006.
- [9] T. Bartz-Beielstein und M. Preuß. *GECCO Tutorial on Experimental Research in EC*, GECCO, Seattle, 2006.
- [10] T. Bartz-Beielstein und M. Preuß. *GECCO Tutorial on Experimental Research in EC*, GECCO, London, 2007.
- [11] T. Bartz-Beielstein, A. Chmielewski, M. Janas, B. Naujoks, and R. Scheffermann. *Optimizing door assignment in LTL-terminals by evolutionary multi objective algorithms*, in *Proceedings 2006 Congress on Evolutionary Computation (CEC’06) Within Fourth IEEE World Congress on Computational Intelligence (WCCI’06)*, Seattle WA, D. B. Fogel et al., Eds. Piscataway NJ: IEEE, 2006.
- [12] T. Bartz-Beielstein, W. Konen, H. Westenberger. *Methoden zur Portfoliooptimierung unter Nebenbedingungen*. Forschungsbericht FH Köln, 2007.
- [13] M. Preuss and T. Bartz-Beielstein. *Sequential parameter optimization (SPO) and the role of tuning in experimental analysis*. (Invited talk). EMEA - Workshop on Empirical Methods for the Analysis of Algorithms. Reykjavik, Iceland, 2007.
- [14] M. Preuss and T. Bartz-Beielstein. *Considerations of Budget Allocation for Sequential Parameter Optimization (SPO)*. EMEA - Workshop on Empirical Methods for the Analysis of Algorithms. Reykjavik, Iceland, 2007.
- [15] T. Bartz-Beielstein, M. Preuss. *Moderne Methoden zur experimentellen Analyse evolutionärer Verfahren*. In R. Mikut, M. Reischl, Hrsg., *Proc. 16. Workshop Computational Intelligence*. Universitätsverlag Karlsruhe, 2006.