

Farbbildbasierte Flammendiagnose zur Führung von Kohlenstaubfeuerungsprozessen¹

V. Stephan, K. Debes und H.-M. Groß

Technische Universität Ilmenau, Fachgebiet Neuroinformatik
D-98684 Ilmenau (Thür.), PF 100565, Tel. 03677-692858, Fax 691665
Klaus.Debes@informatik.tu-ilmenau.de

H. Wintrich, F. Wintrich

ORFEUS Combustion Engineering GmbH
Kleiststr. 10, 45128 Essen, Tel. 0201-8207230, Fax 0201-8207241
info@orfeus.de

Zusammenfassung

Mit dieser Arbeit soll gezeigt werden, dass die Nutzung von direkt am Verbrennungsprozess gewonnenen visuellen Informationen kombiniert mit einem adaptiven, modellfreien und selbstoptimierenden Reinforcement-Lernsystem eine effektive Prozessführung gestattet.

1 Problemstellung

Primäres Ziel der fossilen Energieerzeugung ist neben der maximalen Energieausbeute unter Einhaltung definierter Randbedingungen heute und in Zukunft der ökonomische Einsatz unterschiedlicher Brennstoffqualitäten. Deshalb ist es wünschenswert, den Prozess möglichst schnell in eine solche Situation zu führen, in der obige Forderungen optimal erfüllt werden. Gegenwärtige Steuer- und Regelstrategien beruhen auf totzeitbehafteten und nur punktförmig erfassten Messgrößen. Durch den Einsatz von Videotechnik unmittelbar am Verbrennungsprozess können zwar direkte Informationen gewonnen werden - jedoch fällt es schwer, eine Führungsgröße für einen Regler im klassischen Sinne zu definieren (man kennt kein „optimales“ Flammenbild). Deshalb wird ein Prozessführungssystem gesucht, welches durch Exploration selbständig einen optimalen Arbeitspunkt anfährt.

Aufgrund dieser Anforderungen bietet sich das aus zahlreichen Veröffentlichungen der Robotik bekannte Reinforcement-Learning (RL) an. RL-Verfahren sind in ihrer Vorgehensweise ähnlich dem in der Natur vorkommenden Lust- und Schmerz-Prinzip. Durch trial-and-error können mit einfachen, skalaren und evtl. zeitlich verzögerten Bewertungen ohne Lehrer komplexe Zusammenhänge gelernt werden. Als Grundlage dient dabei eine Menge von Zuständen, in denen sich der Agent befinden kann. Zu jedem dieser Zustände existieren mehrere Aktionsmöglichkeiten, die während des Lernvorganges ausprobiert und entsprechend bewertet werden, so dass dann durch eine geeignete „Politik“ ein maximales Bewertungssignal (reward) von der Umwelt erreicht werden kann. Verwendet werden beispielsweise der $TD(\lambda)$ -Algorithmus [7] oder das Q-Learning [8] zur Bewertung von Aktionen (*value function*) in einem Prozesszustand. In Systemen mit kontinuierlichen Zustands- und Aktionsräumen muss diese *value function* mit reellwertigen Variablen arbeiten, die die Zustände und Aktionen repräsentieren. Typischerweise wird diese *value function* durch *neuronale Funktionsapproximatoren* realisiert, die endliche Ressourcen nutzen, um diese kontinuierlichen Zustands- Aktionspaare zu kodieren. Funktionsapproximatoren sind sehr nützlich, weil sie durch die Generalisierung einen Erwartungswert auch von bisher im Zustandsraum unbekanntem Zustands-Aktionspaaren vorhersagen können.

¹gefördert durch das BMBF, Projekt-Nr.: 032 6843 B

2 Stand der Forschung

In industriellen Verbrennungsanlagen zur Wärme- bzw. Stromerzeugung übernimmt ein Prozessleitsystem mit standardisierten Steuerelementen u.a. die Visualisierung, Alarmindikation und klassische Prozesssteuerungs- und Regelungsaufgaben. Zur Führung komplexerer Prozesse wurde eine Reihe weiterführender Strategien entwickelt, die aber detaillierte Informationen über den Prozess erfordern, um daraus Prozessmodelle zu entwickeln [5, 9] oder Expertenwissen nutzen, um eine intelligente Prozessführung zu realisieren [3, 1]. Voraussetzung und zugleich Nachteil dieser Ansätze sind detaillierte und anlagenspezifische Informationen, die einerseits die Performanz des Systems definieren und andererseits dessen Portierbarkeit limitieren.

3 Das Kraftwerk

Die Versuche wurden in einem Kraftwerk der „Hamburgischen Elektrizitätswerke“ durchgeführt. Das kontrollierte Subsystem besteht aus einem Ofen mit 6 Brennern (je 2 Brenner auf den 3 Ebenen 10, 20 und 30) und einer Maximalleistung von 252 MW (Abb. 1).

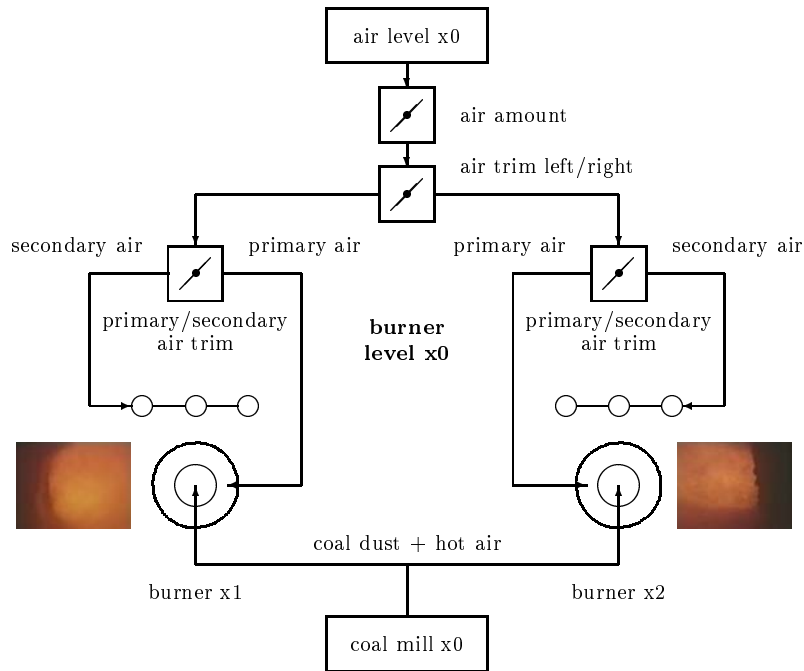


Abbildung 1: Schematische Ansicht der Verbrennungskammer mit Kohle- und Luftversorgung für eine der 3 Brenner-Ebenen.

Beide Brenner einer Ebene werden durch eine Kohlemühle versorgt, die einen nur theoretisch gleichen Kohlestrom für beide Brenner liefert. Die Verschiebung des Gleichgewichtes und der tatsächliche Betrag des Kohlestromes sind nicht messbar und schliessen damit klassische Ansätze wie Regelungen mit Störgrößenaufschaltung aus. Folgende Stellengriffe durch das vorgeschlagene System sind erlaubt:

Steuereingriff	Bedeutung
Primärluftvertrimmung in den Ebenen 10, 20, 30	Luftvertrimmung zwischen dem rechten und linken Brenner einer Ebene
Vertrimmung Primär-, Sekundärluft an den Brennern 11, 12, 21, 22, 31 and 32	Verteilung zwischen Primär- und Sekundärluft eines Brenners
Luftmenge in den Ebenen 10, 20, 30	Gesamtluftmenge der betrachteten Ebene

Es ist zu bedenken, dass diese 12 Steuereingriffe (Abb. 1) nur die Luftmenge und deren Verteilung auf die 6 Brenner berücksichtigen – nicht die Menge und Verteilung der eingebrachten Kohle! Zur Reduktion dieses immensen Aktionsraumes nutzen wir relative statt absolute Stelleingriffe, d.h., es werden pro Steuereingriff 3 Aktionen definiert: +1%, 0, -1%. Selbst unter Verwendung dieser relativen Stelleingriffe bleibt ein Aktionsraum von $3^{3+6+3} = 531.441$).

Die zur Prozessführung verwendeten Prozessdaten werden normalerweise außerhalb der Brennkammer gemessen (z.B. im Abgasteil der Anlage), so dass keine direkten Informationen über den Verbrennungsprozess selbst verfügbar sind. Genau diese Informationen (Kohleverteilung, Flammenform, Temperatur) sind aber erforderlich, um den Prozess *gezielt* zu beeinflussen. Deshalb werden alle 6 Brenner mit Spezial-Farbkameras überwacht, die von der Orfeus Combustion Engineering GmbH speziell für diesen Zweck entwickelt wurden. Aus den Flammenbildern werden mittels der klassischer Bildverarbeitung Merkmale extrahiert, die den Verbrennungsprozess direkt in der Brennkammer beschreiben.

4 Architektur

Wie bereits erwähnt, muss unsere Architektur für jede Prozesssituation eine Aktion auswählen, die den höchsten *reward* erwarten lässt. Das Schlüsselproblem dabei ist der ungeheure Aktionsraum in Kombination mit einem sehr breit gefächerten Eingangssignalraum. Ein monolithischer Architekturansatz ist in diesem Fall nicht sinnvoll, da die Explorationszeit, in der alle Aktionen für alle möglichen Zustände vom System erprobt werden können, für die vorliegende Prozessführung nicht praktikabel ist. Mit der oben genannten Zahl für die möglichen Stelleingriffe und einer (sehr sparsamen) Charakterisierung des Zustandsraumes mit 100 Knoten ergibt sich bei einer Versuchsdauer pro Stelleingriff von 10 min eine Explorationszeit von über 1000 Jahren!

4.1 Problemdekomposition

Konsequenterweise schlagen wir aus diesem Grunde ein System mit mehreren Agenten vor, die jeder für sich einen relevanten Teil des Eingangssignalraumes beobachten. Die Abbildung 2 (links) zeigt die Dekomposition in 4 Agenten mit ihren Eingängen und korrespondierenden Prozesseingriffen. AGENTL10, AGENTL20, und AGENTL30, steuern die Luftverteilung auf der jeweiligen Brennerenebene, während AGENTO2 die Gesamtluftmenge für jede Brennerenebene kontrolliert. Dadurch reduzieren sich die Dimensionen der einzelnen Agenten wie folgt:

	AgentL10	AgentL20	AgentL30	AgentO2
Eingänge	3	3	3	7
NG-Neuronen	20	20	20	50
Aktionen	27	27	27	27

4.2 Neuronaler Funktions-Approximator

Jeder der 4 Agenten enthält einen Neuronalen Funktions-Approximator, für den hier ein erster, einfacher Ansatz vorgestellt werden soll. Wir kombinieren den neuronalen Vektor-Quantifizierer Neural Gas (NG) [6] für eine optimale Clusterung eines hochdimensionalen, kontinuierlichen Eingangsraumes [4] (Gleichung 1) mit einem Assoziativspeicher zur Bewertung der möglichen Aktionen (Abb. 2). Gleichung 1 zeigt die Wichtungsadaption

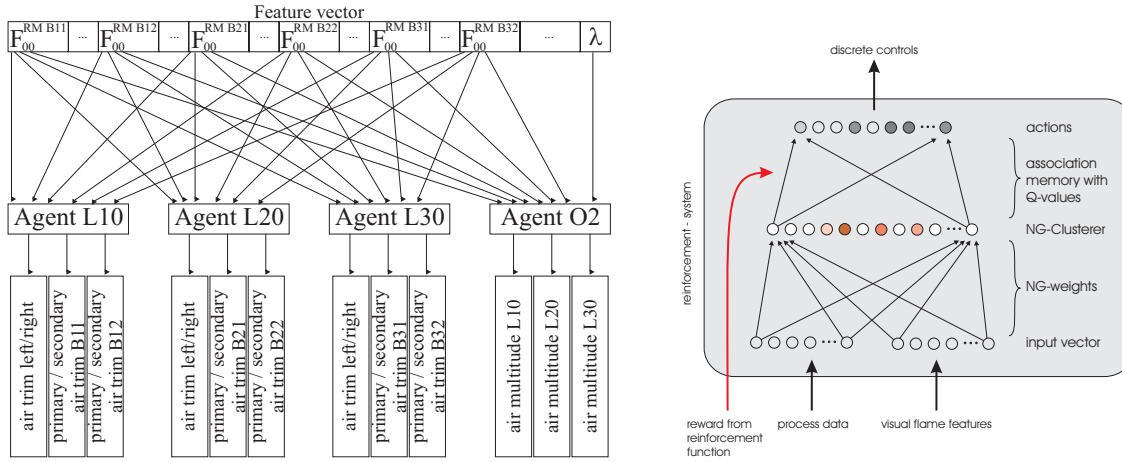


Abbildung 2: Links: Aufteilung der situationsbeschreibenden Größen und der Stellgrößen auf die 4 Agenten. Rechts: Neuronale Steuerarchitektur für das RL-System eines Agenten. Der Input-Vektor besteht aus den Merkmalen, die die Flamme beschreiben; der Neuronale Clusterer bildet den kontinuierlichen und hochdimensionalen Eingangssignalraum auf einen diskreten Zustandsraum ab, mit dem die Q-Werte der ausgewählten Aktionen bestimmt werden.

des NG-Netzwerkes $\Delta \underline{w}_k(t)$ für jedes Neuron k , wobei $\eta^{NG}(t)$ die Lernrate ist, $i(k)$ der Index des Neurons k sortiert nach dem Abstand zum Eingangsvektor $\underline{x}(t)$ und $h(t)$ der Lernradius.

$$\Delta \underline{w}_k(t+1) = \eta^{NG}(t) \cdot e^{-\frac{i(k)}{h(t)}} \cdot [\underline{x}(t) - \underline{w}_k(t)] \quad (1)$$

Für die action-value Approximation Q des Zustandes s^t und der Aktion a^t wird Q-Learning [8] verwendet (Gleichung 2).

$$\Delta Q(s^t, a^t) = \eta \{ r^t + \gamma V(s^{t+1}) - Q(s^t, a^t) \} \quad \text{mit} \quad (2)$$

$$V(s^{t+1}) = \max_a Q(s^{t+1}, a^{t+1}) \quad (3)$$

Für unsere Experimente nutzten wir einen Discount-Faktor von $\gamma = 0.5$ und eine Q-Lernrate von $\eta = 0.2$. Das Reinforcement r ist das Ergebnis einer Agenten-spezifischen Reinforcement-Funktion: Die Agenten AGENTL10, AGENTL20 und AGENTL30 erhalten eine Belohnung, wenn die NO_x oder die O_2 Konzentration sinkt und eine Bestrafung, wenn diese Konzentrationen steigen. Das Reinforcement hängt auch von der O_2 Konzentration ab, weil diese Agenten durch eine bessere *Verteilung* der Luft eine vollständigere Verbrennung der Kohle erzielen können. Agent AGENTO2 erhält eine Belohnung, wenn die NO_x Konzentration oder der totale Betrag der verbrauchten Luft sinkt.

Unabhängig von diesen Aktionsbewertungen wurde aus sicherheitstechnischen Gründen ein Arbeitsbereich für das RL-System durch Schwellwerte definiert. Bei Verlassen des Arbeitsbereiches erfährt das System ein stark negatives Feedback.

5 Ergebnisse

Zur Reduzierung der Explorationszeit wurde der verwendete Multiagenten-Ansatz mit zuvor aufgezeichneten Prozessdaten vortrainiert, wobei eine sinnvolle Vorstrukturierung des RL-Systems an den realen Prozess beobachtet werden konnte.

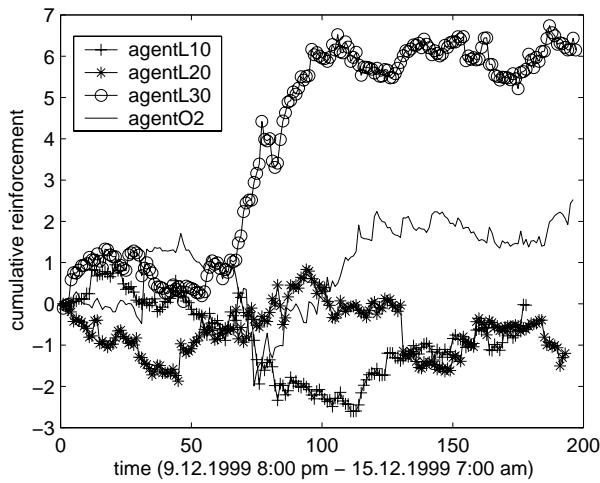


Abbildung 3: Entwicklung des kumulativen Reinforcements aller 4 Agenten im Verlauf von 2 Tagen nach einem Vortraining und relativ hohem Explorationsfaktor. Wie man sieht, erhalten besonders *AgentL30* und *AgentO2* trotz der permanent vorhandenen Exploration meistens positive Reinforcements. Dies ist plausibel, da die oberen Brennebenen als Konsequenz auf die Strömungsverhältnisse im Verbrennungssofen den stärkeren Einfluss auf die Abgaskonzentrationen haben.

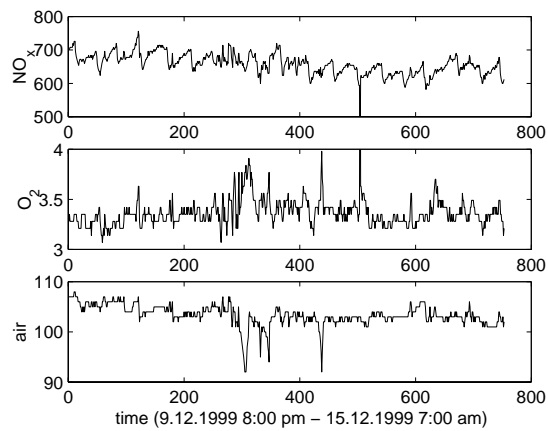
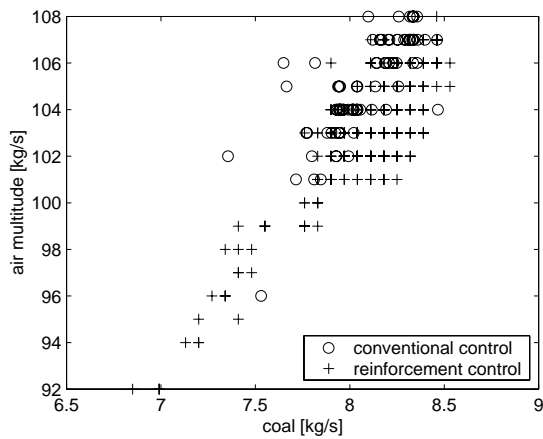


Abbildung 4: Vergleich von konventioneller und RL-basierter Prozessführung bezüglich verbrauchter Luft (links) über eine Testperiode von 6 Tagen. Über den Versuchszeitraum konnte das RL-System sowohl NO_x als auch Luftverbrauch kontinuierlich senken (rechts). Die Anlage lief mit 90% Last.

Nach dem Vortraining wurde das Multiagenten-RL-System im Kraftwerk implementiert. Zur Sicherung des Explorationsverhaltens wurden die Q-Werte mit einem Rauschterm beaufschlagt, der im Verlauf der Zeit zu einem festen Wert größer Null abgesenkt wurde. Somit wird eine zunehmende Anwendung des erlernten Verhaltenswissens realisiert (Exploitation), wobei gleichzeitig ein gewisses Explorationsverhalten die Anpassungsfähigkeit dieses Ansatzes an sich ändernde Prozesseigenschaften (z.B. Verschleiß) garantiert. Abbildung 3 zeigt die kumulativen Reinforcements aller 4 Agenten auf der Basis des vortrainierten Netzwerkes. Aus Abbildung 4 (links) ist ersichtlich, dass der Betrag der eingesetzten Luftmenge mit dem RL-System signifikant geringer wurde (links). Im Gegensatz dazu blieben die NO_x und O_2 Abgaskonzentrationen bei diesen ersten Versuchen etwa gleich, wobei zu beachten ist, dass das Potenzial zur NO_x and O_2 Senkung mit geringer werdendem Lastfaktor der Anlage steigt. Abbildung 4 (rechts) zeigt die relevanten Emissionsdaten mit sinkender NO_x Konzentration und sinkendem Lufteinsatz für die Betriebszeit mit dem RL-System.

6 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde ein auf neuronalen Netzwerken basierendes Multiagenten-Reinforcement-System für die Führung eines industriellen Verbrennungsprozesses vorgestellt. Zur Bewältigung des gewaltigen Aktions- und Zustandsraumes einer solchen Verbrennungsanlage wurde das komplexe System in mehrere Teile zerlegt. Das vorgeschlagene Multiagenten-Reinforcement-System besteht aus 4 Agenten, die durch relativ einfache neuronale Funktionsapproximatoren realisiert werden. Letztere sind sehr nützlich, da sie die erwartete Aktion der Zustands-Aktions-Paare eines Agenten sehr gut im Zustands-Aktions-Raum generalisieren können. Aus diesem Grund kann der Erfolg von Zustands-Aktions-Paaren bewertet werden, die vorher nie erprobt worden sind.

Zukünftige Arbeiten werden der Entwicklung verbesserter Funktionsapproximatoren dienen, da die verwendete Art rein datengetrieben die Wahrscheinlichkeitsdichteverteilung der Eingangsdaten im Merkmalsraum abbildet. Alternativen wären Q-Fehler getriebene inkrementelle neuronale Netzwerke, wie z.B. das Growing Neural Gas nach Fritzke [2]. In diesem Zusammenhang spielt das Stabilitäts-Plastizitäts-Dilemma eine wichtige Rolle, da die Änderung der Kohlequalität, Verschleißerscheinungen an der Anlage usw. den Prozess im Laufe der Zeit maßgeblich beeinflussen. Die ersten Resultate sind sehr vielversprechend – der weitere Einsatz der RL-Methoden ist eine anspruchsvolle Herausforderung für derartig komplexe und schwer exakt erfassbare Industrieprozesse.

Literatur

- [1] P. Eklund and F. Klawonn. Neural Fuzzy Logic Programming. *IEEE Trans. on Neural Networks*, 3(5), 1992.
- [2] B. Fritzke. A Growing Neural Gas Network Learns Topologies. *Advances in Neural Information Processing Systems 7*, MIT-Press, Cambridge MA, 1995.
- [3] S. Gehlen, M. Hormel, and J. Kopecz. Einsatz neuronaler Netze zur Kontrolle komplexer industrieller Prozesse. *Automatisierungstechnik*, 2, 1995.
- [4] H.-M. Gross, V. Stephan, and M. Krabbes. A Neural Field Approach to Topological Reinforcement Learning in Continuous Action Spaces. In *Proc. of WCCI-IJCNN'98, Anchorage*, pages 1992–1997. IEEE Press, 1998.
- [5] H. Maier. *Experimentelle Untersuchungen der Kohlenstaubverbrennung unter Beruecksichtigung der Brennstoffaufbreitung*. PhD thesis, Universitaet Stuttgart, Fakultae Energietechnik, 1997.
- [6] T.M. Martinetz and K. Schulten. A “neural gas” network learns topologies. In T. Kohonen, Mäkisara, K., O. Simula, and J. Kangas, editors, *Artificial Neural Networks*, pages 397–402. Elsevier Amsterdam, 1991.
- [7] R.S. Sutton. Learning to predict by the methods of temporal differences. *Machine Learning*, 3:9–44, 1988.
- [8] Ch. Watkins and P. Dayan. Q-learning. *Machine Learning 8, 1992*, pages 279–292, 1992.
- [9] S. Wirtz. *Mathematische Modellierung der Kohlenstaubverbrennung*. PhD thesis, Ruhr-Universitaet Bochum, Fakultae fuer Maschinenbau, 1989.