

Antrag auf Einrichtung eines neuen
DFG-Schwerpunktprogramms

Autonomes Lernen

(Kurzversion)

Marc Toussaint¹, Helge Ritter²
Jürgen Jost³, Christian Igel⁴

28. Mai 2010

Initiatorengruppe: siehe Abschnitt ??.

Inhaltsverzeichnis

Zusammenfassung	2
1 Stand der Forschung	3
2 Wissenschaftliche Ziele	8
3 Arbeitsprogramm	9

¹Technische Universität Berlin, AG Maschinelles Lernen

²Universität Bielefeld, Institut für Neuroinformatik

³Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig

⁴Ruhr-Universität Bochum, Institut für Neuroinformatik

Zusammenfassung

Die Erforschung der Grundlagen des Lernens hatte in den letzten Jahrzehnten beachtliche Erfolge vorzuweisen. Die Methoden des Maschinellen Lernens und der statistischen Lerntheorie sind heute aus vielen technischen und naturwissenschaftlichen Anwendungen nicht mehr wegzudenken. In der Praxis sind diese Methoden jedoch in erheblichen Maße von einem menschlichen Experten abhängig, der die Trainingsbeispiele auswählt und die Daten in einer vorverarbeiteten Form repräsentiert, der entscheidet, welcher Lernalgorithmus mit welchen Parametern angewandt wird, und der schließlich die Struktur und die intern genutzten Repräsentationen (z. B. Merkmalsvektoren) des Systems festlegt. Dies widerspricht der Intention, dass die Fähigkeit zu Lernen ein System flexibler und autonomer machen sollte, und steht im Kontrast zum Lernen, wie wir es in biologischen Systemen finden. Der Kernpunkt der Autonomie des Lernens wurde im Kontext des Maschinellen Lernens bislang unzureichend verstanden und betrachtet. Autonomie ist hier nicht im Sinne eines vorprogrammierten Systems zu verstehen, das zwar autonom agiert, sich aber nicht weiterentwickelt oder lernt, sondern im Sinne des eigenständigen Sammelns und Vernetzens von Information in einer komplexen Domäne und der eigenständigen Bildung eines strukturierten, generalisierenden Modells als Repräsentation dessen, was gelernt wurde. Der Mangel an Autonomie heutiger Lernalgorithmen verursacht einen erheblichen Entwurfsaufwand für lernende Systeme in der technologischen Anwendung und verdeutlicht die Kluft zwischen dem momentanen Stand des Maschinellen Lernens und dem Lernen in biologischen Systemen.

Um dem entgegenzuwirken brauchen wir neue Forschungsansätze in Richtung eines autonomen Lernens in Systemen, die mit einer variablen, nur partiell modellierbaren Umwelt interagieren und diese explorieren. Einige Kernaspekte des autonomen Lernens sind

- die Unabhängigkeit des lernenden Systems von einem menschlichen Experten, der spezifisch für das Problem die Lernparameter, die Daten, deren Repräsentation und die Struktur des Systems vorgibt;
- die autonome Exploration und die aktive Suche nach Information statt des Lernens aus vorgegebenen Datensätzen;
- die autonome Formation geeigneter Repräsentationen, insbesondere das Lernen hierarchischer Repräsentationen, bei dem inkrementell immer abstraktere Ebenen der Repräsentation von Stimuli, Handlungen und Prozessen aufgebaut werden.

Existierende Methoden des Maschinellen Lernens, vor allem des Reinforcement-Lernens, und der autonomen Robotik bieten Ausgangspunkte und Basismethoden für die Erforschung solcher Lernprozesse. Darauf aufbauend muss nun der nächste Schritt in Richtung vollständiger Autonomie des Lernsystems gemacht werden. Das wollen wir mit diesem Antrag erreichen.

Die Anwendungsszenarien für das autonome Lernen sind vielfältig. Das generelle Problemfeld sind Systeme, die eigenständig ihre Umwelt explorieren und von ihr lernen – etwa in der direkten sensomotorischen Interaktion mit der natürlichen Umwelt (Robotik) oder durch die selbstständige Exploration im Internet oder in Simulationen. Der Schwerpunkt des Programms liegt auf der Entwicklung neuer Theorie und Methoden für das autonome Lernen. Beispielhafte Themenbereiche in denen Projekte gefördert werden sind

- neue Methoden, um geeignete Repräsentationen und computationale Strukturen in eigenständiger Weise zu lernen, z. B. tiefe Repräsentationen von Stimuli, Handlungen und Prozessen, das iterative Verwenden „flacher Lerner“, neue Modellselektions-Kriterien;
- Forschung an autonomen Systemen, z. B. die autonome Exploration in der direkten Interaktion mit der Umwelt, die eigenständige Entwicklung hierarchischer sensomotorischer Repräsentationen, die Entwicklung hierarchischer Handlungsstrukturen im Rahmen des RL;
- neue, autonome Methoden des Maschinellen Lernens, z. B. die automatische Merkmals-Konstruktion, die automatische Suche in einer Menge von Algorithmen, Parametern und Strukturen, die automatische Kombination (parallel oder hierarchisch) von Algorithmen, die automatische Exploration extrem großer Datensätze.

1 Stand der Forschung

Die statistische Lerntheorie und das Maschinelle Lernen (ML) haben in den letzten zwei Jahrzehnten beachtliche Fortschritte und Erfolge vorzuweisen. Beispiele für den erfolgreichen Einsatz maschineller Lernverfahren sind neue Methoden zur Text- und Sprachverarbeitung, wo statistische Modelle (z. B. *hidden Markov models* und *conditional random fields*) heute wesentlich bessere Ergebnisse liefern als frühere Ansätze, die auf grammatikalischem Expertenwissen basierten (Rabiner, 1989; Roark et al., 2007); oder in der Bioinformatik, wo die von statistischen Lernverfahren vorhersagte Gensegmentierung von *C. elegans* so gut war, dass die von Expertenhand erstellte gängige Genannotation revidiert werden musste (Rätsch et al., 2007; Ben-Hur et al., 2008). Ähnliche Erfolge sind bei der Gesichtserkennung (Wiskott et al., 1997; Draper et al., 2003), bei der Erkennung handschriftlicher Adressen in Postsortiermaschinen (Matan et al., 1992; LeCun et al., 1995; Pfister et al., 2000), in Internet-Suchmaschinen und E-Mail-Filtern oder bei der automatischen Strategeanalyse von Fußballspielen (Beetz et al., 2007; Leo et al., 2008) vorzuweisen. Das ML hat sich zu wesentlichen Teilen aus der mathematischen Abstraktion neuronaler informationsverarbeitender Prozesse herausgebildet und ermöglichte umgekehrt auch Erfolge bei der neuronalen Modellierung, z. B. bei der Theorie des Lernens in künstlichen neuronalen Netzwerken (Rojas, 1993; Anthony und Bartlett, 1999; Hammer, 2000) oder beim Verständnis der Organisation visueller Repräsentationen im Cortex (Simoncelli und Olshausen, 2001; Carandini et al., 2005). Im ML weiterentwickelte Algorithmen des Reinforcement-Lernens werden zur Erklärung grundlegender belohnungsgetriebener Lernvorgänge bei Menschen und Tieren herangezogen (Doya et al., 2002; Daw et al., 2006; Doya, 2007; Fiete et al., 2007), und ein beeindruckendes Beispiel der Analyse von Hirnaktivität sind die Erfolge des MLs, EEG-Signale intensionaler Gedanken zu klassifizieren und so ein Brain-Computer Interface zu schaffen (Blankertz et al., 2007).

Bei genauerer Betrachtung der hier skizzierten Erfolge des Maschinellen Lernens zeigt sich, dass diese vor allem bei Problemen der Datenanalyse in einem vom Experten sehr klar vor-spezifizierten Kontext liegen. Ein menschlicher Experte hat die Daten und deren Repräsentation ausgewählt und sie entsprechend vorverarbeitet. Das Problem wurde vom menschlichen Experten als Klassifikations-, Regressions- oder Optimierungsproblem abstrahiert. Oftmals wurden bekannte strukturelle Abhängigkeiten zwischen Ein- und Ausgabe vom Experten vordefiniert: bei der Textverarbeitung durch strukturierte generative Modelle (CRFs), bei denen grammatikalisches Expertenwissen einfließt; bei der Gensegmentierung durch sehr komplexe und für die spezielle Anwendung von Hand entwickelte Ausgabe-Strukturen; bei der Bildverarbeitung durch eine Vielzahl sorgfältig ausgewählter Merkmale (*features*) und komplexe Strukturen zu deren Verarbeitung (beispielsweise hierarchische Graphische Modelle). Der Bezug zu den Neurowissenschaften war besonders dann erfolgreich, wenn neuronale Strukturen dezidiert einem konkreten Informationsverarbeitungsproblem zugeordnet werden können, wie beispielsweise im visuellen Cortex in den obigen Arbeiten.

Die große Abhängigkeit der Methoden von der Vorstrukturierung und dem Eingriff eines Experten ist ein entscheidender Unterschied zum biologischen Lernen.⁵ Ein aktueller Diskurs greift dieses Problem auf und thematisiert die Diskrepanz zwischen dem momentanen Stand des Maschinellen Lernens und dem ursprünglichen Ziel des Verständnisses biologischen Lernens oder der künstlicher Intelligenz (z. B. Bengio und LeCun, 2007; Dietterich et al., 2008). Die fehlende Autonomie des Lernens, die sich darin widerspiegelt, dass ein Experte die Trainingsbeispiele auswählen muss, die Repräsentation der Daten bestimmen, Teilziele setzen und die Struktur bzw. innere Repräsentation des Systems festlegen muss, ist als Kernproblem des MLs identifiziert worden, wie etwa in einem zusammenfassenden Report eines NSF Expertengremiums zum Thema „Future Challenges for the Science and Engineering of Learning“ (Douglas et al., 2008) dargestellt ist. In dessen Rahmen wird auch der Begriff des **autonomen Lernens**⁶ geprägt, im

⁵Wir wollen nicht suggerieren, dass Lernen gänzlich ohne Vorwissen, Vorstrukturierung oder Prior möglich ist – siehe die Diskussion in Abschnitt 3.1(ii) des Arbeitsprogramms und Bousquet et al. (2004).

⁶Der Begriff „Autonomes Lernen“ wird auch in der Sozialpädagogik genutzt und beschreibt z. B. Situationen, in

Sinne eines MLs, das wie biologisches Lernen in Autonomie und unabhängig von menschlicher Vorstrukturierung stattfinden muss.

Eine neue Forschungsrichtung im ML, die diesen jüngsten Diskurs begleitet, ist das *deep learning*, d.h. das Lernen in vielschichtigen Hierarchien von Repräsentationen (Boltzmann-Maschinen oder Auto-Assoziatoren) mit der Kernfrage, wie geeignete Repräsentationen in diesen tiefen Ebenen gelernt (oder initiiert) werden können. Ausgehend von Hinton et al. (2006) haben eine Reihe von Arbeiten die traditionelle Argumentation wieder aufgegriffen, dass es ein wesentliches Merkmal des Gehirns ist, immer tiefere Schichten der Repräsentation zu nutzen (siehe auch Hinton, 2007), um Generalisierungsfähigkeit und Abstraktionsvermögen zu erzielen. Die Arbeiten wollen Ähnliches erreichen, indem Lernverfahren für vielschichtige neuronale Strukturen entwickelt werden. In einigen Anwendungen zeigen sich kompetitive Resultate im Vergleich mit herkömmlichen „flachen“ Klassifikatoren oder Regressionsmethoden (Bengio und LeCun, 2007; Hadsell et al., 2008; Collobert und Weston, 2008; Weston et al., 2008). Obwohl diese Ansätze durch die „Rückbesinnung“ auf ein neuronales Substrat und die Untersuchung tiefer Repräsentationen den Bezug zum biologischen Lernen wieder herstellen wollen, so bleiben sie doch in vielerlei Hinsicht hinter dem biologischen Vorbild zurück. Die neuen Lernalgorithmen erweisen sich als wenig robust (Fischer und Igel, 2009). Da weiterhin konventionelle Regressions- und Klassifikationsprobleme betrachtet werden, anstatt Systeme, die wie biologische Systeme ihre Umwelt explorieren und mit dieser interagieren, kommen hier Kernaspekte des autonomen Lernens zu kurz.

Wir wollen zwei Punkte festhalten. (1) Mit dem Maschinellen Lernen und der statistischen Lerntheorie sind wesentliche mathematische und theoretische Grundbausteine des Lernens gelegt. Wir können auf dieser Ebene sehr genau definieren, was Lernen aus informationstheoretischer Sicht oder aus Sicht eines statistischen Schätzers bedeutet. Daraus ableitbare Methoden sind extrem erfolgreich bei der Datenanalyse. Diese theoretischen Grundlagen bilden ein wertvolles Fundament, auf das wir aufbauen können. (2) Die erwähnte aktuelle Diskussion um die Autonomie des Lernens und die deutliche Begrenzung der Erfolge des ML auf von Experten abstrahierte Regressions- oder Klassifikationsprobleme zeigen jedoch auch, dass diese Grundbausteine nicht hinreichend sind, um Systeme zu verstehen oder zu bauen, die in autonomer Weise durch Interaktion mit der Umwelt lernen. Die Zielstellung des vorliegenden Programms ist genau das Verständnis solchen *autonomen Lernens*. Einige Aspekte unseres Programms, die wir in Abschnitt 2 genauer definieren, wie die Unabhängigkeit der Lernalgorithmen von einem menschlichen Experten, der die Daten selektiert und vorverarbeitet, wurden schon von Douglas et al. (2008) oder Bengio und LeCun (2007) als zukunftsweisende Forschungsrichtungen diskutiert. Andere, wie der Fokus auf mit der Umwelt interagierende Systeme statt auf die Lösung von Klassifikations- und Regressionsproblemen, gehen darüber hinaus.

Methodische Grundlagen im Detail. Im Folgenden wollen wir einige für unser Programm wichtige Grundlagen und Methoden näher betrachten und im Einzelnen deren Bezug zu unserem Forschungsziel klarstellen.

(i) *Strukturierte Ein- und Ausgaberräume.* Die Entwicklung kernbasierter Lernverfahren, wie Support-Vektor-Maschinen (SVMs, Vapnik 1998; Schölkopf und Smola 2002; Steinwart und Christmann 2008) oder Gauß'sche Prozesse (Rasmussen und Williams, 2006), hat maßgeblich zum aktuellen Erfolg des MLs beigetragen. Mit Hilfe einer sogenannten Kernfunktion können Daten effizient in einem ggf. hochdimensionalen metrischen Merkmalsraum verarbeitet werden. Kernbasierte Algorithmen bieten den Vorteil, dass viele Fragestellungen im Bereich des Lernens vergleichsweise einfach mathematisch formuliert und untersucht werden können. Die Qualität der erreichten Lösung hängt allerdings entscheidend von der Wahl der Kernfunktion ab, die die Repräsentation der Daten im Merkmalsraum bestimmt. Zwar gibt es Ansätze, die Kernfunktion ebenfalls datengetrieben zu ermitteln (Chapelle et al., 2002; Friedrichs und Igel, 2005; Glasmachers und Igel, 2005; Igel et al., 2007; Mersch et al., 2007; Igel, 2009), doch auch diese setzen voraus, dass ein Experte zuvor die Klasse der möglichen Kernfunktionen stark eingeschränkt hat.

denen die Lernenden alleine lernen, oder das Recht der Lernenden, die Richtung ihres Lernens selbst zu bestimmen. Wir definieren den Begriff hier neu in einem technischen, lerntheoretischen Kontext.

Während Kernfunktionen die Strukturierung des Eingabe- bzw. Signalraums ermöglichen, konnten ebenfalls bei der Behandlung strukturierter Ausgaben wichtige Fortschritte erzielt werden, indem probabilistische Inferenzmethoden in den Regressionsprozess eingebunden werden. Statt wie bei der herkömmlichen Regression oder Klassifikation einer Eingabe einen einzelnen Wert oder eine Klasse zuzuweisen, können neue Algorithmen in komplexe Räume abbilden, z. B. auf Sequenzen, Graphen oder eine Segmentierung eines Bildes. Ausgehend von den grundlegenden Veröffentlichungen (Lafferty et al., 2001; Altun et al., 2003; Taskar et al., 2004; Hammer, 2002; Hammer et al., 2005a; Tsochantaridis et al., 2005) hat sich dieser Ansatz in vielen Anwendungen durchgesetzt und erhebliche Verbesserungen gebracht. Er ist konzeptionell interessant, weil uns hiermit eine Methode in die Hand gelegt wird, den Zusammenhang zwischen Eingaben, z. B. der momentanen Situation eines Systems, und komplexen Ausgaben, wie Handlungsstrategien, zu lernen (siehe z. B. Zhang et al., 2007). Jedoch muss zur Anwendung der Methode die Struktur der Ausgabe bekannt sein, und damit steht wieder die Frage der geeigneten Repräsentation, hier der zu lernenden Ausgabe, im Mittelpunkt.

(ii) *Strukturierte Informationsverarbeitung und Handlungsplanung mittels probabilistischer Inferenz.* Zu einem Grundprinzip der Informationsverarbeitung haben sich Methoden der probabilistischen Inferenz entwickelt. Diese können auch als informationstheoretische Abstraktion neuronaler Informationsverarbeitung interpretiert werden (Ott und Stoop, 2007; Rao, 2004; Doya et al., 2007). Durch rekurrenten Informationsaustausch (z. B. *belief propagation*) kann effizient Kohärenz in komplexen, hierarchisch tiefen Modellen hergestellt werden. Eine wichtige Erkenntnis in diesem Kontext ist, dass auch Entscheidungs-, Steuerungs- oder Planungsprobleme als probabilistische Inferenzprobleme betrachtet und mit den gleichen Methoden gelöst werden können (Toussaint, 2009a; Toussaint und Storkey, 2006; Hoffman et al., 2008). Dass dieser Ansatz trägt, wurde in verschiedenen Gebieten gezeigt. In der Psychologie dient er als Modell menschlicher Entscheidungsfindung (Botvinick und An, 2009) und in der Robotik zur Bewegungsplanung (Toussaint, 2009b). Der probabilistische Ansatz ist auch auf hierarchische und relationale Planungsprobleme anwendbar (Toussaint et al., 2008; Lang und Toussaint, 2009). Die neue Sichtweise hat wichtige Implikationen: Das gleiche generische Prinzip der Informationsverarbeitung kann für die sensorische Informationsverarbeitung, die motorische Steuerung, die Planung von Aktionen (implizit durch interne Simulation) und die Kopplung all dieser Ebenen verwendet werden. Die Hoffnung ist, mit diesen Prinzipien einen generischen Mechanismus der Informationsverarbeitung in komplexen sensomotorischen Repräsentationen zur Hand zu haben. Diese Methoden können auch auf tiefe Strukturen effizient angewandt werden.

(iii) *Unüberwachtes, halbüberwachtes und aktives Lernen.* Unüberwachtes Lernen bezeichnet das Lernen aus Daten, die keine Sollausgaben enthalten, und ist in diesem Sinne unabhängig von der Bereitstellung solcher Sollausgaben durch einen Experten. Die Idee des unüberwachten Lernens basiert auf der Annahme, dass statistische Regularitäten in den sensorischen Signalen wichtige Informationen über die Beschaffenheit der Welt liefern. Das Ziel ist es, eine geeignete Beschreibung der beobachteten Daten zu finden, die die charakteristischen, relevanten Eigenschaften und Muster der Daten widerspiegelt, um nachfolgend Entscheidungen basierend auf Beobachtungen, die Vorhersage neuer Beobachtungen oder die Kommunikation der Daten zu erleichtern (Ghahramani, 2004). Als Beispiele für neue Algorithmen, die nichtlineare Zusammenhänge aufdecken können, seien *kernel principal component analysis* (Schölkopf et al., 1998), *locally linear embedding* (Roweis und Saul, 2000), *Isomap* (Tenenbaum et al., 2000) und *unsupervised kernel regression* (Meinicke et al., 2005) genannt. Auch auf dem Gebiet der selbstorganisierenden Karten und verwandter Methoden (Ritter et al., 1991; Hammer und Villmann, 2002; Hammer et al., 2005b) und nichtlinearen ICA (*independent component analysis*, Hyvärinen et al., 2001) konnten neue Erkenntnisse erzielt werden. Während die bisherigen Arbeiten zum unüberwachten Lernen eine wichtige Grundlage für das autonome Lernen bilden und auch die Aufklärung von allgemeineren Zusammenhängen zwischen verschiedenen Modellklassen vorangetrieben hat Sinz et al. (2008), so haben sich die bisher vorgeschlagenen Algorithmen in ihrer Leistungsfähigkeit nach wie vor als sehr beschränkt erwiesen. Dies wurde insbesondere durch die vor kurzem

vorangetriebene Entwicklung von verlässlichen Methoden deutlich, die es erlauben, die Leistung von unüberwachten Lernverfahren quantitativ erfassen zu können (Bethge, 2006; Bethge et al., 2007; Sinz und Bethge, 2009; Lyu und Simoncelli, 2008; Bethge und Berens, 2008).

Oft gibt es Situationen, in denen Eingaben bzw. Stimuli für das System ausreichend zur Verfügung stehen, die zugehörigen Sollausgaben aber schwierig oder teuer zu ermitteln sind. Oder die Menge potentieller Eingaben ist so groß, dass sie nicht in Gänze von der Lernmaschine verarbeitet werden kann. In solchen Fällen bietet es sich an, das lernende System selbst – in unserem Sinne autonom – diejenigen Eingaben auswählen zu lassen, aus denen es mutmaßlich am meisten lernt. Dies wird im ML als aktives Lernen bezeichnet (*active learning*, Cohn et al., 1996; Schohn und Cohn, 2000; Glasmachers und Igel, 2008). Das erstgenannte Problem adressiert auch das halbüberwachte Lernen (*semi-supervised learning*), welches überwachtes und unüberwachtes Lernen verbindet (Chapelle et al., 2006). Zum Lernen vorliegende Eingaben, zu denen keine Zielausgabe bekannt ist, können den Lernprozess unterstützen, indem aus ihnen unüberwacht die Struktur des Eingaberaums gelernt wird.

(iv) *Reinforcement-Lernen*. Das Reinforcement-Lernen (RL, Verstärkungslernen) ist ein sehr allgemeines Lernparadigma, bei dem eine Handlungsstrategie durch Interaktion mit der Umwelt aus ggf. spärlichen und verzögerten Bewertungssignalen gelernt werden muss. Ein klassisches Erfolgsbeispiel sind die Ergebnisse beim Backgammonspiel (Tesauro, 1989, 1995), bei dem Spielstrategien gelernt werden konnten, die menschlichen Spielern ebenbürtig sind. Zur gleichen Zeit konnte die Konvergenz grundlegender RL Algorithmen nachgewiesen werden, z. B. von Q-Lernen (Watkins, 1989). Neuere Erfolge sind die Anwendung im Roboterfußball (Müller et al., 2007; Latzke et al., 2007; Riedmiller et al., 2009) und bei der Steuerung von ferngesteuerten Helikoptern (die gelernten Algorithmen können Manöver fliegen, die für Menschen kaum mehr beherrschbar sind, Abbeel et al., 2008). Das algorithmische RL hatte auch entscheidenden Einfluss auf die Neurowissenschaften: Es steht außer Frage, dass die Lösung von belohnungsbasierten Entscheidungs- und Handlungsproblemen eine entscheidende Rolle für autonomes Handeln spielt und somit stellt sich die Frage, wie das Gehirn solche Probleme löst. Daw et al. (2006), Doya (2007) und Fiete et al. (2007) konnten explizite Bezüge zwischen Strukturen des Nervensystems und typischen Größen des RL herstellen.

Analog zu unserer Kritik an den heutigen Methoden des Maschinellen Lernens kann man auch bei diesen Erfolgsbeispielen des RL feststellen, dass die Vorstrukturierung des Problems durch den menschlichen Experten bisher entscheidend für den Erfolg ist: Beim Backgammonspiel hatte Tesauro sehr raffinierte Spielkonfigurationsmerkmale als Zustandsrepräsentation von Hand vorgegeben; bei der Helikoptersteuerung floss viel Expertenwissen in eine geschickte Kombination von modell-basiertem und modell-freiem Reinforcement-Lernen und Imitations-Lernen von menschlichen Flugexperten ein. Vor allem die Wahl der Repräsentation des Zustandsraumes ist kritisch. Die Effizienz und Skalierbarkeit des RL hängt in entscheidender Weise von der Repräsentation und Struktur ab, die genutzt wird, sowohl um das Problem (die Welt) als auch den Agenten (dessen interne Struktur) zu repräsentieren. Ein Beispiel dafür, was hier mit „strukturiert“ gemeint ist, sind die frühen Ideen zur hierarchischen Dekomposition von Verhaltensproblemen mit Optionen und *macro policies* (auch Abstraktionen genannt) (Hauskrecht et al., 1998; Sutton et al., 1999; McGovern und Barto, 2001). Hier werden explizit hierarchisch verschiedene Ebenen der Beschreibung unterschieden. Ein Satz von Elementarverhalten (Optionen, *sub policies*) wird darauf trainiert, gewisse Subziele (Sollzustände einer hierarchisch niedrigeren Variable) zu erreichen, während gleichzeitig eine *macro policy* gelernt wird, die entscheidet, welches Elementarverhalten in welchen Situationen auszulösen ist. Boutilier et al. (1995) entwickelten zunächst unabhängig davon einen grundlegenden Formalismus für strukturierte Repräsentationen. Sie behandeln den Fall, dass schon bei der Problemdefinition der Zustand mit mehreren Variablen beschrieben wird – im allgemeinsten in Form eines Dynamischen Bayes'schen Netzwerkes (DBNs). Seither gab es beträchtliche Fortschritte bei Lösungsmethoden für solche faktorisierten MDPs (Koller und Parr, 1999, 2000; Schuurmans und Patrascu, 2002; Guestrin et al., 2002; Guestrin und Gordon, 2002; Guestrin et al., 2003, 2004; Kveton und Hauskrecht, 2005, 2006; Toussaint et al.,

2006, 2008). Das Problems des autonomen Findens hierarchischer Handlungsstrukturen ist jedoch keineswegs gelöst. Toussaint et al. (2008) zeigen einen ersten Ansatz, wie hierarchische Lösungen in POMDPs zu finden sind. Littman et al. (2002) schlagen *predictive state representations* als alternative (nicht hierarchische) Repräsentation in POMDPs vor. Andere Ansätze nutzen einen beliebigen internen Automaten als Repräsentation der Policy und implizit auch des Zustands und lernen diese durch *direct policy search*-Methoden (Meuleau et al., 1999a,b; Poupart und Boutilier, 2004; Brazianus und Boutilier, 2004), zu denen auch auf evolutionären Algorithmen basierende Verfahren zählen (Pellecchia et al., 2005; Whiteson und Stone, 2006; Gomez et al., 2008; Heidrich-Meisner und Igel, 2009a,b). Diese Ansätze erlauben effiziente Lösungsmethoden für manche POMDPs, sie kommen aber nicht dem Ziele näher, *strukturierte* Repräsentationen des Zustands oder eines Weltmodells zu konstruieren oder solche gegebenen Repräsentationen zu nutzen.

Ein zweites wichtiges Merkmal autonomen Lernens neben der Organisation von Repräsentationen ist die Exploration und Auswahl der Daten, die gelernt werden. Hier bietet die Theorie des Reinforcement-Lernens eine klare Grundlage. In einem spezifizierten Kontext (unbekannter *Markov Decision Process*, endliche Zustands- und Aktionsräume) lässt sich klar definieren, was die optimale Handlungsstrategie ist, die vorausschauend die Umwelt exploriert, um zielführende Information zu sammeln und gleichzeitig diese zu nutzen, um Belohnung zu erhalten. Dies leisten beispielsweise *model-based Bayesian RL* (Poupart et al., 2006) als optimale Bayesianische Lösung oder *explicit explore or exploit* (Kearns und Singh, 2002) als polynomielle Lösung (siehe auch frühere Heuristiken, z. B. von Schmidhuber, 1991b,a). Das früher oft diskutierte Explorations-Exploitations-Dilemma ist theoretisch gelöst. Allerdings sind diese optimalen Lösungen beweisbar ineffizient. Die Herausforderung besteht also in effizienten Approximationen solcher optimaler Strategien.

Wir wollen festhalten: (1) Das RL formalisiert ein allgemeines Handlungsproblem in unbekanntem Umwelten. Bisherige Methoden sind typischerweise aber nur dann effizient, wenn vom menschlichen Experten eine geeignete Repräsentation und Strukturierung des Problems vorgegeben wird. (2) Das RL bietet zudem eine klare Definition, was optimale Exploration im Allgemeinen bedeutet. Die Umsetzung solcher theoretischer Ansätze in großen und hochdimensionalen Zustandsräumen ist jedoch eine enorme Herausforderung.

(v) *Die Evolution von Strukturen.* In biologischen Systemen werden Repräsentationen durch Selbstorganisations- und Entwicklungsprozesse geschaffen, insbesondere der Evolution. Evolutionäre Algorithmen haben sich als Such- und Optimierungsverfahren aufgrund ihrer Leistungsfähigkeit etabliert, und ihre theoretische Grundlage wurde in jüngerer Zeit erheblich gestärkt (siehe z. B. Vose, 1999; Beyer, 2001; Beyer et al., 2002; Droste et al., 2006; Jägersküpper, 2006; Jansen und Wegener, 2006). Evolutionäre Algorithmen eröffnen eine Möglichkeit, geeignete Strukturen und Repräsentationen für effizientes Lernen automatisiert hervorzubringen, und können damit zur Autonomie des Lernprozesses beitragen. Dieses biologisch inspirierte Vorgehen hat sich bereits in ersten technischen Anwendungen bewährt (Yao, 1999; Friedrichs und Igel, 2005; Igel, 2009; Igel und Sendhoff, 2008; Floreano et al., 2008), aber noch keine qualitativen Fortschritte erzielt.

(vi) *Robotik.* Autonomie ist eines der großen Ziele der Roboterforschung, daher ist die Robotik ein wichtiges, prototypisches Anwendungsfeld für autonomes Lernen. Bei der Forschung an intelligenteren, lernenden oder adaptiven Robotern gab es Fortschritte vor allem in den einzelnen Teildisziplinen, wie der Bild- und Sprachverarbeitung, Trajektorienplanung, Motorsteuerung und Lokalisation, Systemidentifikation, Modellierung und Simulation, Verhaltensorganisation, Reinforcement-Lernen und Imitationslernen sowie Mensch-Maschine-Interaktion und -Kommunikation. Die Schwierigkeit, diese existierenden Methoden in autonomen Gesamtsystemen zusammenzuführen wird häufig als Integrationsproblem bezeichnet. Ein Grund für dieses Problem ist die Heterogenität der Paradigmen, Methoden und vor allem Repräsentationen, die in den erwähnten Teilgebieten genutzt werden. Ein möglicher Ausweg wären allgemeine Prinzipien der Repräsentation, Abstraktion und Verarbeitung von Information, die kohärent in jedem Modul, auf jeder Modalität und auf

jeder Ebene der Abstraktion genutzt werden können (siehe z. B. Ritter et al., 2003).

Die Robotik bietet heute exzellente Testplattformen, die uns aufzeigen können, was die tatsächlichen Herausforderungen und notwendigen Funktionen sind, mit denen Systeme, die mit der natürlichen Welt interagieren und autonom von ihr lernen sollen, konfrontiert sind. So vermögen in vielerlei Hinsicht heutige robotische Systeme die Welt auf ähnliche Weise zu erfahren wie Menschen und Tiere – man kann Objekte (be-)greifen, manipulieren und die Konsequenzen von solchen Handlungen spüren, die natürliche Statistik der Welt erfahren, mit einem ferngesteuerten Roboter handwerkliche Aufgaben erledigen. Die Tatsache, dass trotz dieser physischen Fähigkeiten all diese Dinge bis heute kaum in Autonomie erlernt werden können, hilft, die Lücken unseres Verständnisses offenzulegen. Die Interaktion von Theoretikern mit Robotikern und das Testen der theoretischen Ansätze ist nicht zuletzt deshalb notwendig, damit Ansätze erzwungen werden, die die sensomotorische Kopplung an die Umwelt beinhalten, die Integration vieler Sensor- und Motormodalitäten ermöglichen und die das genannte Integrationsproblem vermeiden, indem kompatible Prinzipien der Informationsverarbeitung entwickelt werden. Umgekehrt sind wir der Meinung, dass besonders die Robotik von unserem Forschungsprogramm profitieren wird, weil es Grundprobleme autonomer Systeme adressiert. Hier sei betont, dass die großen Fortschritte im Bereich autonomer Robotersysteme, zum Beispiel bei der Entwicklung autonomer Fahrzeuge im Rahmen der DARPA Challenges (Thrun et al., 2006; Seetharaman et al., 2006), Autonomie nur in dem Sinne erreichen, dass ein vorprogrammiertes, nicht-lernendes System ohne menschlichen Eingriff eine vorgegebene Aufgabe erfüllt – also nicht im Sinne des autonomen Lernens oder der autonomen Exploration einer Umwelt.

2 Wissenschaftliche Ziele

Eine der großen wissenschaftlichen Herausforderungen ist das Verständnis der Effizienz, der Autonomie, der Flexibilität und der Intelligenz biologischer Systeme – auch in Hinblick auf die Anwendung in neuen Technologien. Ein charakteristisches Merkmal ist hierbei ihre Fähigkeit zu *lernen*, d. h. sich aus der Interaktion mit der Umwelt Wissen anzueignen, um dieses später zu nutzen. Das vorliegende Programm widmet sich in diesem Kontext dem Problem der *Autonomie*. Wir meinen damit nicht eine Autonomie, bei der Systeme fest vorprogrammiert sind und dann ohne menschliche Überwachung auskommen. Statt dessen verfolgen wir das Ziel des *autonomen Lernens*, d. h. das eigenständige Lernen, das eigenständige Sammeln und Vernetzen von Information in einer komplexen Umwelt und die eigenständige Bildung eines strukturierten, generalisierenden Modells als Repräsentation dessen, was gelernt wurde. Die Erforschung des Lernens ist zentrales Thema sowohl in den Neurowissenschaften als auch in den entsprechenden Disziplinen der Statistik und der Informatik: der statistischen Lerntheorie und dem Maschinellen Lernen. Der Ansatzpunkt unseres Programms ist es, einerseits auf den theoretischen Fundamenten und Methoden, die in diesen Disziplinen gelegt wurden (siehe Stand der Forschung), aufzubauen, andererseits jedoch aufzuzeigen, dass der Kernpunkt der Autonomie in diesem Kontext bislang unzureichend verstanden und betrachtet wurde. Trotz der großen Erfolge bei Anwendungen der Datenanalyse – oder vielleicht gerade wegen dieser Erfolge – scheint das Maschinelle Lernen wenig Fortschritte in Richtung eines seiner ursprünglichen Ziele, intelligente autonome Systeme in der natürlichen Umwelt zu schaffen, gemacht zu haben. Wenn Lernmethoden bei Anwendungen zum Einsatz kommen, so meist als Lösung eines Teilproblems, z. B. eines dezierten Regressions- oder Klassifikationsproblems, das vom Entwickler als solches identifiziert und genau spezifiziert wurde. In der Praxis muss ein menschlicher Experte

- das konkrete Lernproblem (Datenquellen, Zielfunktion) vordefinieren,
- das Gesamtproblem in Teilprobleme zerlegen,
- die Trainingsbeispiele auswählen,
- die Daten in einer wohlspezifizierten, vorverarbeiteten Form repräsentieren,
- entscheiden, welcher Lernalgorithmus mit welchen Parametern angewandt wird, und
- die Struktur und die intern genutzten Repräsentationen des Systems festlegen.

All dies steht im Widerspruch zu dem Ziel der Autonomie. Die eigentliche Intelligenz in dieser Praxis liegt zu großen Teilen bei dem menschlichen Experten. Das verbleibende Lernen ist ein passiver Prozess. Wenn wir biologische Systeme zum Vorbild nehmen, so erscheint Lernen jedoch als aktiver Prozess. In einer interaktiven Umwelt bestimmen die Aktionen eines Systems die Daten, die es sammelt. In komplexen Szenarien ist es meist unmöglich, alle Daten zu erfassen, und es hängt von dem bereits gesammelten Wissen ab, welche Daten sinnvoll integriert werden können.

Das wissenschaftliche Ziel dieses Programms ist die Erforschung der Grundlagen des autonomen Lernens. Insbesondere müssen neue Konzepte und Methoden entwickelt werden, die den nächsten Schritt machen, vom bisherigen Maschinellen Lernen zu einem autonomen Lernen in Systemen, die mit einer variablen, nur partiell modellierbaren Umwelt hoher Komplexität interagieren und diese explorieren. Kernaspekte des autonomen Lernens sind:

- die autonome Wahl von Parametern, Repräsentationen und genutzten Strukturen beim Lernen, unabhängig von einem menschlichen Experten,
- die autonome Auswahl dessen, was gelernt wird, d. h. die autonome Exploration und die aktive Suche nach Information, statt des Lernens aus vorgegebenen Datensätzen,
- das Finden geeigneter Repräsentationen; insbesondere das Lernen hierarchischer Repräsentationen, das inkrementelle Aufbauen immer abstrakterer Ebenen der Repräsentation von Stimuli, Handlungen und Prozessen aufbauen,
- das eigenständige Vernetzen von Information verschiedener Modalitäten oder Quellen, sowohl im Kontext des Schließens der sensomotorischen Kopplung wie auch die Vernetzung verschiedener Quellen des Wissens und
- generische Prinzipien der Informationsverarbeitung und des Lernens in tiefen hierarchischen Strukturen.

Wie im Stand der Forschung erwähnt, ist dieses Programm damit in Teilen eine Antwort auf die aktuelle Diskussion um wesentliche Merkmale des biologischen Lernens, die bisher unzureichend im Rahmen des Maschinellen Lernens, der statistischen Lerntheorie, oder auch den Neurowissenschaften verstanden sind.

Wir sind an dem Punkt, an dem auf der Ebene der theoretischen Grundbausteine und Basismethoden hervorragend ausgearbeitete und mächtige Verfahren zur Verfügung stehen. Jetzt ist der ideale Zeitpunkt, den nächsten Schritt in Richtung vollständiger Autonomie des Lernsystems zu machen. Das wollen wir mit diesem Antrag erreichen.

3 Arbeitsprogramm

3.1 Fragestellungen und Ansatzpunkte

Die Erforschung des autonomen Lernens verlangt die Entwicklung neuer theoretischer Konzepte auf verschiedenen Ebenen. Aus unserer Sicht besonders dringende theoretische Fragestellungen ergeben sich bei der Frage nach der eigenständigen Organisation tiefer Repräsentationen. Damit verbunden sind Fragen der Generalisierung, der genutzten Struktur und des Priors (d. h. des Vorwissens über die Umwelt) sowie Fragen zur Kopplung verschiedener paralleler Lernprozesse. Im Folgenden werden ein mögliches Arbeitsprogramm und entsprechende methodische Ansätze vorgeschlagen, die sich auf diese Kernaspekte konzentrieren. Das Programm wird besonders Projekte fördern, die auf diesen Gebieten methodischen Fortschritt versprechen, aber auch Projekte, die alternative vielversprechende Herangehensweisen an das autonome Lernen vorschlagen.

(i) Repräsentationen. Natürliche Umwelten sind extrem variabel. Keine Situation wird exakt ein zweites Mal erlebt. Systeme, die in natürlichen Umwelten agieren, sind also gezwungen zu generalisieren und sich flexibel immer neuen Situationen anzupassen. Für das Lernen bedeutet das, dass es nicht hinreichend ist, sich alle erlebten Daten zu merken. Vielmehr muss Erfahrung in ein strukturiertes, abstrahierendes und deshalb generalisierendes Wissen integriert werden.

Natürliche Umwelten sind zudem außerordentlich komplex in dem Sinne, dass zwar viele Gesetzmäßigkeiten herrschen, diese aber sehr tiefschichtige und hierarchische Prozesse und Stimuli hervorbringen. Es ist ein grundlegendes Prinzip der Natur (wie auch des Systemdesigns), zunehmend komplexere Systeme aus modularen Bausteinen zusammenzufügen. Ein lernendes System in einer solchen Welt muss in der Lage sein, Teile dieser Struktur zu erkennen und diese zu nutzen, um immer weitere Strukturen höherer Ordnung darauf aufbauend erkennen zu können. Dies gilt sowohl für Hierarchien verschiedener Zeit- und Raumskalen von Prozessen wie auch für Konzept- und Objekthierarchien.

Die computationale Struktur eines lernenden oder adaptiven Systems ist deshalb von großer Wichtigkeit. Beim Maschinellen Lernen gibt es nur wenige Beispiele für Lernprinzipien, die auf der inkrementellen Organisation zunehmend tiefer Repräsentationen basieren, statt statistische Abhängigkeiten direkt zu modellieren. Im Gegensatz hierzu basieren viele adaptive und lernende Prozesse der Biologie auf der inkrementellen Organisation zunehmend tiefer Repräsentationen. Beim genetischen Code beispielsweise können einzelne genotypische Variationen komplex strukturierte, modulare, hierarchische Variationen des Phänotyps hervorbringen. Der genetische Code ist eine extrem „tiefe“ Repräsentation biologischer Organismen, ähnlich zu tief verschachtelten Grammatiken. Bei Betrachtung der Organisation des Gehirns ist es offensichtlich, dass biologisches Lernen tiefere Strukturen nutzt. Auf höheren Ebenen werden zunehmend abstraktere Konzepte (Merkmale) repräsentiert und aus einfacheren Signalen geformt. Diese Re-Repräsentation ist ein wichtiges Merkmal biologischer Adaptation und ein Spiegel der komplexen hierarchischen Struktur der natürlichen Welt.

In diesem Arbeitspaket werden neue Konzepte und Methoden zur autonomen Organisation hierarchisch tiefer Repräsentationen gefunden. Zu beantwortende Fragen sind:

- Wie können lernende Systeme inkrementell immer abstraktere Ebenen der Repräsentation von Stimuli, Handlungen und Prozessen aufbauen?
- Kann durch solches Lernen eine Integration subsymbolischer und symbolischer Repräsentationen erreicht werden?
- Was sind generische Prinzipien der Informationsverarbeitung und des Lernens, die auf allen Ebenen der Repräsentation operieren?
- Wie kann Wissen im System verteilt werden?
- Wie kann ein System dazu gebracht werden, Probleme nicht nur in Teilprobleme zu zerlegen, sondern auch eine hierarchische Repräsentation zu entwickeln, damit spezifische erworbene Problemlösungsfähigkeiten generalisiert und auf andere, bis zu einer bestimmten Stufe der Hierarchie aber ähnliche Probleme übertragen werden können?
- Wie können Adaptationsprozesse auf verschiedenen Zeitskalen – insbesondere Evolutions-, Selbstorganisations- und Lernprozesse – gekoppelt werden?

Erste methodische Ansätze zu solchen Fragen sind in der aktuellen Forschung zum *deep learning* zu finden (siehe Stand der Forschung, Seite 4). Inspiriert von der Organisation des Gehirns wird dabei das Lernen in tiefen Schichtungen neuronaler Ebenen untersucht. Dies geschieht bisher nur im Kontext von Klassifikations- und Regressionsaufgaben – unzureichend verstanden ist, ob und wie analoge Prinzipien auch zur Organisation von Handlungsrepräsentationen genutzt werden können. Sind sie auf Bewegungslernen und Planung, zum Beispiel durch interne Simulation, anwendbar? Ein alternativer methodischer Ansatz besteht darin, konventionelle Lernsysteme des Maschinellen Lernens iterativ zu koppeln: Können „flache Lerner“ (d. h. konventionelle Lernalgorithmen wie SVMs) iterativ verwendet werden, um Strukturen inkrementell aufzubauen? Was sind geeignete Zwischenrepräsentationen? Die Forschung an diesen Fragestellungen ist ein Schlüssel für das autonome Lernen und zum Verständnis des biologischen Lernens.

(ii) *Generalisierung, Prior & Struktur.* Die Diskussion von vollkommener Autonomie verleitet zu der Annahme, dass Systeme in jeder beliebigen Welt ohne Vorwissen autonom lernen und sich immer weiter Wissen aneignen könnten. Aus theoretischer Sicht verlangt jedoch jede Art der

Generalisierung einen „Prior“ (sei es in Form einer Komplexitätsregularisierung, der Wahl eines Hypothesenraums oder der Wahl einer Kernfunktion), d. h. Annahmen und Vorwissen über den zugrundeliegenden Prozess oder die Welt, die die Daten erzeugt hat. Wenn ein lernendes System sich nicht wie erwartet verhält, so liegt das – aus theoretischer Sicht – an einem falschen Prior (oder an unausweichlichen Approximationen, die gemacht werden mussten). In dieser reduzierten Sicht verbirgt sich hinter dem Wort „Prior“ jedoch geradezu alles, was ein lernendes System ausmacht: dessen Struktur, die Art der Repräsentationen, die es nutzen kann, wie es generalisiert, etc. Unser Ziel kann nur das autonome Lernen in einer konkreten, wenngleich komplexen Umwelt sein, insbesondere das autonome Lernen in unserer realen natürlichen Umwelt oder anderen Domänen, die ebenfalls hinreichend komplex sind. Dieser theoretische Hintergrund ist bei der Forschung am autonomen Lernen essentiell. Grundlegende Fragestellungen in diesem Kontext sind:

- Was ist geeignetes Vorwissen, ein geeigneter Prior, für ein autonomes Lernen in der konkreten Welt?
- Wie übersetzt sich solches Vorwissen in Strukturannahmen über das System?
- Wie viele und welche Strukturelemente müssen vorgegeben werden (System-Design) und was kann sich selbstorganisierend im aktiven Wechselspiel mit einer herausfordernden Umwelt entwickeln (System-Anpassung)?
- Kann Vorwissen in Form einer heterogenen Sammlung von Modellen gegeben werden?

Ein möglicher Ansatzpunkt ist es, die existierende Lerntheorie als Basis zu nehmen und nach alternativen Regularisierungskriterien zu fragen, die unser Vorwissen widerspiegeln. Die heute übliche Komplexitätsregularisierung hat Maße wie die VC-Dimension, die Rademacher-Komplexität oder die Zahl der effektiven Parameter/Freiheitsgrade zur Grundlage. Wären stattdessen alternative Kriterien möglich, die z. B. Hierarchien bevorzugen? Können Strukturen begünstigt werden, in denen Inferenz effizient ist (z. B. indem die maximale Cliquengröße in graphischen Modellen als Komplexitätsmaß betrachtet wird)? Im Idealfall sollte ein Prior oder eine Regularisierung Strukturen begünstigen, von denen wir glauben, dass sie der natürlichen Welt entsprechen, d.h. einen zur natürlichen Welt „passenden“ Prior induzieren.

(iii) Vernetzung, Kopplung und parallele Lernprozesse. Ein typisches Merkmal natürlicher Umgebungen ist es, dass es vielerlei Quellen der Information gibt – einerseits „Orte“, andererseits auch „Modalitäten“. Man kann argumentieren, dass der Ursprung neuronaler Informationsverarbeitung gerade in dieser Vernetzung und Kopplung von Information liegt, z. B. in der Kopplung einer motorischen mit einer sensorischen Variablen. Die Vernetzung von Information ist aber auch in anderen Szenarien grundlegend: Zum Beispiel wenn Information aus verschiedenen Datenbanken (z. B. im Internet) vernetzt werden soll oder wenn Wissen zunächst in verschiedenen Kontexten gelernt wird, sich dann jedoch gegenseitig ergänzen soll. Die Untersuchung solcher Vernetzung verlangt zumindest Szenarien oder Domänen, bei denen die Information an verschiedenen Quellen vorliegt (mehrere Sensoren, mehrere Datenbanken, mehrere zunächst getrennte Repräsentationen des Wissens, etc.).

Im Kontext des Lernens stellt sich ein zusätzlicher entscheidender Aspekt: die *Vernetzung und Kopplung lernender Subsysteme*. In der Biologie finden wir eine Kopplung vieler autarker Lernprozesse, die in ihrer Interaktion das Gesamtproblem lösen. Diese Kopplung birgt einerseits zusätzliche Probleme, wie die Rollenverteilung und die Gesamtstabilität, andererseits ermöglicht gerade dies eine angepasste Strukturierung und Dekomposition des Systems, wie wir sie in den letzten beiden Abschnitten diskutiert haben. Grundlegende Fragen in diesem Kontext sind:

- Wie kann Information verschiedener Quellen und Modalitäten vernetzt und eigenständig eine gemeinsame Repräsentation gefunden werden?
- Wie können parallele Lernprozesse und -algorithmen kombiniert werden?
- Wie können sich diese parallelen Prozesse gegenseitig unterstützen?

Konkrete Ansatzpunkte in diesem Arbeitspaket lassen sich direkt im Rahmen des ML erkennen:

committee machines und *boosting*-Mechanismen sind bekannte Prinzipien, mit denen sich verschiedene Lernalgorithmen („schwache Lerner“) kombinieren lassen, die jedoch alle das gleiche Lernziel verfolgen. Durch ähnliche Kombination existierender Lernalgorithmen ließe sich die Dekomposition des Lernproblems untersuchen. Auch die Kooperation verschiedenartiger Lernalgorithmen, z. B. eines Reinforcement-Lerners, der Klassifikatoren reguliert, sind auf Basis existierender Algorithmen zu untersuchen.

(iv) *Neue mathematische Paradigmen.* Die heute gängigen Lernalgorithmen stützen sich mathematisch vornehmlich auf die Gebiete der Statistik, der Wahrscheinlichkeitstheorie und der Optimierung. Lernen wird üblicherweise als Optimierung einer vorgegebenen, starren Zielfunktion formalisiert. Dies erfasst jedoch nicht alle Aspekte des Lernens, wie ein Blick auf biologische Systeme unmittelbar zeigt. Lernen in der Natur ist zeitlich situiert und muss als eine flexible Anpassung an veränderliche Umweltbedingungen verstanden werden. Dabei ist zu beachten, dass das System in ständiger Interaktion mit seiner Umwelt nicht nur die benötigten Daten, sondern auch die Kriterien, nach denen der Erfolg eigener Aktionen bewertet wird, aktiv generieren kann und muss. Eine formale Beschreibung von autonomem Lernen muss folglich in die Theorie dynamischer Systeme eingebettet sein. Der dynamische Aspekt des Lernens wird jedoch in der (statistischen) Lerntheorie i. d. R. vernachlässigt. Allgemein stellt sich die Frage, ob nicht neue mathematische Methoden und Paradigmen zur Entwicklung der Konzepte des autonomen Lernens herangezogen werden können und müssen. Generelle Fragen im Rahmen unseres Programms sind demnach:

- Wie kann Systemautonomie mathematisch formalisiert werden?
- Wie repräsentiert und verarbeitet ein dynamisches System Information über seine Umwelt und wie kann Lernen als beteiligter dynamischer Prozess beschrieben werden? Wie lässt sich Information aus der Innenperspektive des Systems beschreiben und verstehen? Wie können die statistische Lerntheorie und die Theorie dynamischer Systeme enger gekoppelt werden?
- Wie kann Lernen autonomer Systeme, d. h. unter Berücksichtigung der Systemstruktur und in Interaktion mit der Umwelt mathematisch formalisiert werden? Was sind geeignete Gütekriterien, um autonomes Lernen in Interaktion mit der Umwelt anzutreiben?
- Wie kann die Komplexität einer Umgebung gemessen und bewertet werden, um einen Maßstab für die inkrementelle Entwicklung von Hierarchien und Repräsentationen zu gewinnen?
- Welche Formen der Systemdifferenzierung, die z.B. modular, hierarchisch strukturierte Subsysteme generiert, lassen sich identifizieren? Wie müssen Subsysteme beschaffen sein, damit sie sowohl Informationen austauschen als auch integriert agieren können?

Ein Ansatzpunkt bzgl. der Zielfunktion ist die Erweiterung der üblichen Gütekriterien des aktiven und unüberwachten Lernens auf Szenarien der Exploration einer Umwelt. Zu untersuchen sind beispielsweise informationstheoretische Lernmethoden, die Prädiktion und Postdiktion kombinieren und auch interne Simulationen einschließen. Der Ansatz von Still (2009) beruht auf der Wechselwirkung von informationstheoretischen Prinzipien mit Effizienzkriterien im Rahmen des interaktiven Lernens. Ansatzpunkte zu alternativen mathematischen Paradigmen zur Beschreibung von Autonomie und Lernen finden sich beispielsweise bei Bertschinger et al. (2006, 2008), Ay et al. (2008) oder Jost et al. (2007). Weitere Verbindungen von Informationsverarbeitung und dynamischen Systemen bietet der Ansatz der Liquid State Maschine (Maass et al., 2002). Hierbei wird benutzt, dass hochdimensionale, weitgehend zufällig verbundene neuronale Netzwerke Information über eingehende Signale auf vielfältige Weise verarbeiten. Der Netzwerkzustand dient so als Repräsentation vergangener Eingabesignale, dessen Kodierung im Gegensatz zu anderen Ansätzen des maschinellen Lernens nicht vom Designer vorgegeben ist. Ziel des Projektes ist es diese und weitere Ansätze auszubauen, um zu einer mathematischen Formalisierung der obigen Punkte zu gelangen.

(v) *Autonome Systeme und Robotik.* Die Synthese konkreter autonomer Systeme in der Interaktion mit der natürlichen Umwelt kann nicht nur als Anwendungs- und TestszENARIO des autonomen Lernens verstanden werden, sondern auch als Grundlage für die Entwicklung neuer Methoden. Die direkte sensomotorische Interaktion mit der physischen Umwelt bietet enorm reichhaltige und strukturierte Daten. In den Neurowissenschaften wird diskutiert, dass wesentliche Merkmale des Gehirns originär aus der Notwendigkeit entstanden, mit diesen Daten umzugehen um Bewegung zu steuern. Das heißt, die konkrete sensomotorische Interaktion ist der treibende Faktor für die Entwicklung neuronaler informationsverarbeitender Prozesse und damit auch für das biologische Lernen und die Fähigkeit, hierarchische Sensor-, Motor- und Verhaltensrepräsentationen zu organisieren. Tatsächlich fällt bei der konkreten Forschung an der Steuerung und Verhaltensorganisation robotischer Systeme die Notwendigkeit hierarchischer Organisationsstrukturen besonders drastisch auf, ebenso wie der Mangel an existierenden Methoden, die solche Repräsentationen autonom organisieren oder lernen können. Die Forschung an autonomen Systemen kann also treibende Kraft für die Entwicklung neuer Methoden des autonomen Lernens sein. Speziell im sensomotorischen Kontext können die grundlegenden Fragestellungen, die im Allgemeinen schon in den vorigen Abschnitten gestellt wurden, konkretisiert werden:

- Wie können hierarchische sensomotorische Repräsentationen auf Basis der direkten Interaktion mit der physischen Umwelt gelernt werden?
- Welche Rolle spielt dabei die rekurrente Informationsverarbeitung in Sensor-Motor-Zyklen?
- Wie können hierarchische Handlungsstrukturen im Rahmen des Reinforcement-Lernens gelernt werden?
- Wie können gelernte Teilstrategien (*skills*) in neuen Aufgabenstellungen genutzt werden?
- Wie kann die aktive sensomotorische Exploration analog zur Neugierde bei Kleinkindern organisiert werden?

Methodische Ansatzpunkte für diese Fragestellungen sind einerseits wiederum die allgemeinen Methoden, die in den vorherigen Abschnitten genannt wurden. Im konkreten sensomotorischen Kontext lassen sich jedoch weitere Ansatzpunkte aus der Biologie ableiten. Die Forschung an Motorprimitiven schlägt beispielsweise einige konkrete Modelle für die Hierarchisierung der motorischen Steuerung vor. Die existierenden Ansätze des Reinforcement-Lernens (siehe Stand der Forschung) beinhalten Algorithmen für das Finden hierarchischer Problemdekompositionen, die jedoch bisher nur in relativ kleinen, diskreten Umwelten effizient sind.

3.2 Anwendungsszenarien

Die Liste potenzieller Teilnehmer dieses Programms (Abschnitt ??) zeigt, dass die jeweiligen konkreten Anwendungsgebiete dieser Forscher divers sind (Robotik, Bildverarbeitung, Bioinformatik, *semantic web*, Neurowissenschaften). Dennoch basieren die verwendeten Ansätze auf gemeinsamen theoretischen und methodischen Grundlagen der Informationsverarbeitung, der Lerntheorie, und des Maschinellen Lernens. Die Forschung am Lernen ist also eine Basiswissenschaft und Grundlage für den Fortschritt in den genannten Disziplinen. Aus diesem Grund ist das Ziel dieses Schwerpunktprogramms nicht eine einzelne technologische Anwendung, sondern die Entwicklung neuer Theorie und Methoden des Lernens, die in den aktuellen Stand der technologischen Forschung einfließen und das jeweilige Anwendungsgebiet vorantreiben sollen. Die entscheidenden Kriterien für die Förderung von Projekten innerhalb dieses Programms sind: (1) die Innovation der Lerntheorie und -methoden an sich, wie im vorherigen Abschnitt diskutiert und (2) die Integration dieser Methoden in kompetitive Technologien im jeweiligen Anwendungsgebiet, in dem die Antragsteller Experten sind. Wir skizzieren im Folgenden Anwendungsszenarien, die eine zentrale Rolle einnehmen werden.

(i) *Repräsentationen in der Bildverarbeitung.* In der Bildverarbeitung ist das Problem des autonomen Lernens von Repräsentationen (Punkt 1 des Arbeitsprogramms) sehr direkt greifbar. Beispielsweise sind statistische Verfahren relativ effizient darin, zu klassifizieren, ob ein Objekt einer bestimmten Klasse in einem Bild enthalten ist; andererseits sind sie extrem schlecht

(im Vergleich zum Menschen) darin, dieses Objekt zu lokalisieren und zu segmentieren (siehe z. B. die letztjährigen Ergebnisse der VOC challenge⁷). Vollkommen ungelöst ist das Problem, eigenständig generalisierende und robuste Repräsentationen von Objekten und visuellen Merkmalen zu bilden, die ein wirkliches Szenen-„Verständnis“ ermöglichen. Existierende statistische Bildanalyseverfahren stehen im Kontrast zur sakkadischen Exploration einzelner Bildstrukturen bei Tieren und Menschen, wo das Sehen ein aktiver explorativer Prozess ist. Die Forschung unseres Programms zielt genau auf diese grundlegenden Probleme. Als Ziel der Anwendung in der Bildverarbeitung stellen wir uns Systeme vor, die autonom lernen die visuellen Daten einer natürlichen Umwelt zu explorieren und inkrementell abstrahierende Repräsentationen zu bilden. Die Bildverarbeitung ist eine klassische Testanwendung für die Forschung am unüberwachten und tiefen Lernen (siehe Stand der Forschung), weil sie reichhaltig strukturierte und natürliche Daten betrachtet. Eine wirkliche Szenenanalyse verlangt unweigerlich geeignete Prior und Strukturen der Repräsentationen (Punkt 2 des Arbeitsprogramms). Die Bildverarbeitung ist auch eine Basistechnologie für autonomes Verhalten in natürlichen Umwelten und damit Teil des folgenden Anwendungsszenarien.

(ii) *Autonomes Lernen und Robotik.* Wie in Punkt 5 des Arbeitsprogramms diskutiert bietet die Robotik ein zentrales Anwendungs- und Testgebiet für die Forschung am autonomen Lernen. Ein exemplarisches Testszenario ist folgendes: Ein Roboter mit Mitteln der Fortbewegung und Manipulation (im einfachsten Falle zwei Räder und Manipulatoren) hat das Problem, eine vorher unbekannte Umgebung (ein Zimmer, eine Wohnung) autonom zu explorieren und insbesondere die funktionale Interaktionen mit Objekten (Lichtschalter, Schubladen, Tischeverschieben, Geräte an/ausschalten, etc.) in dieser Umgebung zu lernen. Dieses Szenario ist ähnlich zu existierenden „Challenges“⁸, konzentriert sich aber auf das Problem des Lernens und „manipulativen Verstehens“ einer Umgebung, statt des Lösen einer vordefinierten Aufgabe. Obwohl ein solches System nicht direkt auf eine industrielle Anwendung zielt, würde es doch ein Meilenstein aus Sicht der Forschung am Lernen und dem Verständnis autonomen, explorativen, und lernenden Verhaltens sein. Dabei ist es nicht das Ziel unseres Programms, die Konstruktion neuer Robotik-Plattformen zu fördern. Wie bei der Bildverarbeitung muss es das Ziel sein, die Forschung an den grundlegenden Fragen des autonomen Lernens zu forcieren und sie in moderne Robotik-Systeme einfließen zu lassen. Solche Systeme existieren bereits in den Laboren vieler der potenziellen Teilnehmer. Oben wurde bereits die Wichtigkeit der Bildverarbeitung für die Autonome Robotik erwähnt, jedoch muss hier das Ziel ein Szenenverständnis sein, das die visuellen Daten mit den möglichen Aktionen und möglichen Manipulationen der Objekte in Relation setzt.

(iii) *Autonomes Lernen im Internet.* Punkt 3 des Arbeitsprogramms diskutiert die Vernetzung gelernter Wissens und paralleler lernender Systeme. Die Kopplung und Kombination paralleler Modalitäten und Algorithmen ist auch in den vorher genannten Anwendungen relevant. Ein besonders eindrückliches Anwendungsszenario in diesem Kontext ist jedoch die Kombination verschiedener Informationsquellen und multimodaler Datenbanken, die im Internet zu finden sind. Beispielsweise finden sich zunehmend Daten im Internet, die Worte mit Bildern oder semantischem Wissen assoziieren – neben den unspezifischen *Google images* beispielsweise *LabelMe*⁸, bei der einzelne Objekte in den Bildern markiert und per Hand benannt sind, oder *TextRunner*⁸, das durch relationale Analyse von Wortkoinzidenzen im Internet semantische Fragen beantworten kann. Wenn Systeme fähig werden, diese Information mit reellen Szenen und Objekten der Welt zu koppeln, wäre dies ein großer Schritt hin zum semantischen Verständnis der natürlichen Umwelt. Durch die Nutzung des Internet als Informationsquelle können Systeme Objekte der aktuellen Umgebung mit Bildern, die im Internet zu finden sind, abgleichen und so Objekte erkennen und ihnen semantische Bedeutung zuschreiben, ohne diese zuvor gesehen zu haben. Das Internet kann aktiv exploriert werden, um zu Objekten der Umgebung Informationen zu fin-

⁷PASCAL Visual Object Classification challenge. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/>

⁸RoboCub@Home <http://www.ai.rug.nl/robocupathome/>, Semantic Robot Vision Challenge <http://www.semantic-robot-vision-challenge.org/>, LabelMe <http://labelme.csail.mit.edu/>, TextRunner <http://www.cs.washington.edu/research/textrunner/>

den. Namen (Spracheingaben) können mit Hilfe solcher Datenbanken konkreten Objekten in der Umgebung zugeordnet werden. Zentral ist dabei die Vernetzung der verschiedenen Quellen der Information und des Vorwissens, das in diesem Fall implizit durch die Daten des Internet gegeben ist. Ein mögliches Zielszenario ist das oben skizzierte Szenario der autonomen Robotik, jedoch mit den erweiterten Möglichkeiten der Vernetzung mit mehreren Datenquellen des Internets. Eine vereinfachte Variante besteht darin, solche Quellen in die Bildverarbeitung einzubeziehen und so die Brücke zwischen dem Bild und der semantischen Bedeutung für die möglichen Aktionen und Manipulationen herzustellen.

Zusammenfassend wollen wir festhalten, dass autonomes Lernen, wie es in diesem Programm im Zentrum der wissenschaftlichen Fragestellungen steht, eine Basiswissenschaft darstellt die in vielerlei Disziplinen zur Anwendung kommen wird. Das Ziel dieses Programms ist vornehmlich die Erforschung der Grundlagen des autonomen Lernens und damit auch das Verständnis des biologischen Vorbildes. Die am Programm teilnehmenden Projekte werden nicht auf ein gemeinsames integratives Anwendungsszenario festgelegt. Dies wäre insbesondere für neue theoretische Ansätze hinderlich und würde den Fokus zu stark auf die Anwendung und Integration existierender Methoden in einer Plattform setzen statt auf die Weiterentwicklung von Theorie und Neuentwicklung von Methoden.

Literatur

- P. Abbeel, A. Coates, M. Quigley und A. Y. Ng. An application of reinforcement learning to aerobatic helicopter flight. In: *Advances in Neural Information Processing Systems (NIPS 2007)*, 2008.
- Y. Altun, I. Tsochantaridis und T. Hofmann. Hidden markov support vector machines. In: *Proc. of the 20th Int. Conf. on Machine Learning*, S. 3–10, 2003.
- M. Anthony und P. L. Bartlett. *Neural Network Learning: Theoretical Foundations*. Cambridge University Press, 1999.
- N. Ay, N. Bertschinger, R. Der, F. Güttler und E. Olbrich. Predictive information and explorative behavior of autonomous robots. *The European Physical Journal B-Condensed Matter and Complex Systems*, 63: 329–339, 2008.
- M. Beetz, S. Gedikli, J. Bandouch, B. Kirchlechner, N. von Hoyningen-Huene und A. Perzylo. Visually tracking football games based on TV broadcasts. In: *Proc. of the Twentieth Int. Joint Conf. on Artificial Intelligence (IJCAI)*, 2007.
- A. Ben-Hur, C. Ong, S. Sonnenburg, B. Schölkopf und G. Rätsch. Support vector machines and kernels for computational biology. *PLoS Computational Biology*, 4, 2008.
- Y. Bengio und Y. LeCun. Scaling learning algorithms towards ai. In: L. Bottou, O. Chapelle, D. DeCoste und J. Weston (Hrsg.), *Large-Scale Kernel Machines*. MIT Press, 2007.
- N. Bertschinger, E. Olbrich, N. Ay und J. Jost. Information and closure in systems theory. In: S. Artmann und P. Dittrich (Hrsg.), *Explorations in the Complexity of Possible Life*, Proceedings of the 7th German Workshop on Artificial Life, Amsterdam, 2006. IOS Press.
- N. Bertschinger, E. Olbrich, N. Ay und J. Jost. Autonomy: an information theoretic perspective. *Biosystems*, 91(2):331–345, 2008.
- M. Bethge. Factorial coding of natural images: how effective are linear models in removing higher-order dependencies? *Journal of the Optical Society of America A*, 23(6):1253–1268, 2006.
- M. Bethge und P. Berens. Near-maximum entropy models for binary neural representations of natural images. In: *Advances in Neural Information Processing Systems (NIPS 2007)*, S. 97–104. MIT Press, 2008.
- M. Bethge, T. V. Wiecki und F. A. Wichmann. The independent components of natural images are perceptually dependent. In: *Human Vision and Electronic Imaging XII*, S. 1–12. SPIE, 2007.
- H.-G. Beyer. *The Theory of Evolution Strategies*. Springer, Berlin, 2001.
- H.-G. Beyer, H.-P. Schwefel und I. Wegener. How to analyse evolutionary algorithms. *Theoretical Computer Science*, 287(1):101–130, 2002.
- B. Blankertz, G. Dornhege, M. Krauledat, G. Curio und K.-R. Müller. The non-invasive berlin brain-computer interface: Fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550, 2007.
- M. Botvinick und J. An. Goal-directed decision making in prefrontal cortex: A computational framework. In: *Advances in Neural Information Processing Systems (NIPS 2008)*, 2009. in print.
- O. Bousquet, S. Boucheron und G. Lugosi. Introduction to statistical learning theory. In: O. Bousquet, U. von Luxburg und G. Rätsch (Hrsg.), *Advanced Lectures in Machine Learning*, Nr. 3176 in LNCS, S. 72–112. Springer, 2004.
- C. Boutilier, R. Dearden und M. Goldszmidt. Exploiting structure in policy construction. In: *Proc. of the 14th Int. Joint Conf. on Artificial Intelligence (IJCAI 1995)*, S. 1104–1111, 1995.
- D. Braziunas und C. Boutilier. Stochastic local search for POMDP controllers. In: *Proc. of the Nineteenth National Conf. on Artificial Intelligence (AAAI 2004)*, S. 690–696, 2004.
- M. Carandini, J. B. Demb, V. Mante, D. J. Tolhurst, Y. Dan, B. A. Olshausen, J. L. Gallant und N. C. Rust. Do we know what the early visual system does? *J. Neurosci.*, 25(46):10577 – 10597, 2005.
- O. Chapelle, V. Vapnik, O. Bousquet und S. Mukherjee. Choosing multiple parameters for support vector machines. *Machine Learning*, 46(1):131–159, 2002.
- O. Chapelle, B. Schölkopf und A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.

- D. A. Cohn, Z. Ghahramani und M. I. Jordan. Active learning with statistical models. *Journal of Artificial Intelligence Research*, 4:129–145, 1996.
- R. Collobert und J. Weston. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: *Proc. of the 25th Int. Conf. on Machine Learning (ICML 2008)*, 2008.
- N. D. Daw, J. P. O’Doherty, P. Dayan, B. Seymour und R. J. Dolan. Cortical substrates for exploratory decisions in humans. *Nature*, 441(15):876–879, 2006.
- T. G. Dietterich, P. Domingos, L. Getoor, S. Muggleton und P. Tadepalli. Structured machine learning: the next ten years. *Machine Learning*, 73:3–23, 2008.
- R. Douglas, T. Sejnowski et al. Future challenges for the science and engineering of learning. Final NSF Workshop Report, <http://www.cnl.salk.edu/Media/NSFWorkshopReport.v4.pdf>, 2008.
- K. Doya. Reinforcement learning: Computational theory and biological mechanisms. *HFSP Journal*, 1(1): 30–40, 2007.
- K. Doya, K. Samejima, K. Katagiri und M. Kawato. Multiple model-based reinforcement learning. *Neural Computation*, 14:1347–1369, 2002.
- K. Doya, S. Ishii, A. Pouget und R. P. N. Rao (Hrsg.). *Bayesian Brain: Probabilistic Approaches to Neural Coding*. MIT Press, 2007.
- B. Draper, K. Baek, M. Bartlett und J. Beveridge. Recognizing faces with pca and ica. *Computer Vision and Image Understanding (Special Issue on Face Recognition)*, 91 (1-2):115–137, 2003.
- S. Droste, T. Jansen und I. Wegener. Upper and lower bounds for randomized search heuristics in black-box optimization. *Theory of Computing Systems*, 39(4):525–544, 2006.
- I. R. Fiete, M. S. Fee und H. S. Seung. Model of birdsong learning based on gradient estimation by dynamic perturbation of neural conductances. *Journal of Neurophysiology*, 98:2038–2057, 2007.
- A. Fischer und C. Igel. Contrastive divergence learning may diverge when training restricted Boltzmann machines. *Frontiers in Computational Neuroscience. Bernstein Conference on Computational Neuroscience (BCCN 2009)*, 2009. doi: 10.3389/conf.neuro.10.2009.14.121.
- D. Floreano, P. Dürr und C. Mattiussi. Neuroevolution: from architectures to learning. *Evolutionary Intelligence*, 1(1):47–62, 2008.
- F. Friedrichs und C. Igel. Evolutionary tuning of multiple SVM parameters. *Neurocomputing*, 64(C):107–117, 2005.
- Z. Ghahramani. Unsupervised learning. In: O. Bousquet, U. von Luxburg und G. Rätsch (Hrsg.), *Advanced Lectures in Machine Learning*, Nr. 3176 in LNCS, S. 72–112. Springer, 2004.
- T. Glasmachers und C. Igel. Second order SMO improves SVM online and active learning. *Neural Computation*, 20(2):374–382, 2008.
- T. Glasmachers und C. Igel. Gradient-based adaptation of general Gaussian kernels. *Neural Computation*, 17(10):2099–2105, 2005.
- F. Gomez, J. Schmidhuber und R. Miikkulainen. Accelerated neural evolution through cooperatively co-evolved synapses. *Journal of Machine Learning Research*, 9:937–965, 2008.
- C. Guestrin und G. Gordon. Distributed planning in hierarchical factored MDPs. In: *Proc. of the 18th Conf. on Uncertainty in Artificial Intelligence (UAI 2002)*, 2002.
- C. Guestrin, D. Koller und R. Parr. Multiagent planning with factored MDPs. In: *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, Bd. 14, S. 1523–1530. MIT Press, 2002.
- C. Guestrin, D. Koller, R. Parr und S. Venkataraman. Efficient solution algorithms for factored MDPs. *Journal of Artificial Intelligence Research (JAIR)*, 19:399–468, 2003.
- C. Guestrin, M. Hauskrecht und B. Kveton. Solving factored MDPs with continuous and discrete variables. In: *Proc. of the 20th Conf. on Uncertainty in Artificial Intelligence (UAI 2004)*, S. 235–242, 2004.
- R. Hadsell, A. Erkan, P. Sermanet, M. Scoffier, U. Muller und Y. LeCun. Deep belief net learning in a long-range vision system for autonomous off-road driving. In: *Proc. Intelligent Robots and Systems (IROS’08)*, 2008.
- B. Hammer. Recurrent neural networks for structured data – a unifying approach and its properties. Co-

- gnitive Systems Research*, 3(2):145–165, 2002.
- B. Hammer. On the approximation capability of recurrent neural networks. *Neurocomputing*, 31(1-4): 107–123, 2000.
- B. Hammer und T. Villmann. Generalized relevance learning vector quantization. *Neural Networks*, 15 (8-9):1059–1068, 2002.
- B. Hammer, A. Micheli und A. Sperduti. Universal approximation capability of cascade correlation for structures. *Neural Computation*, 17:1109–1159, 2005a.
- B. Hammer, M. Strickert und T. Villmann. On the generalization ability of GRLVQ networks. *Neural Processing Letters*, 21(2):109–120, 2005b.
- M. Hauskrecht, N. Meuleau, L. P. Kaelbling, T. Dean und C. Boutilier. Hierarchical solution of Markov decision processes using macro-actions. In: *Proc. of Uncertainty in Artificial Intelligence (UAI 1998)*, S. 220–229, 1998.
- V. Heidrich-Meisner und C. Igel. Neuroevolution strategies for episodic reinforcement learning. *Journal of Algorithms*, 64(4):152–168, 2009a.
- V. Heidrich-Meisner und C. Igel. Hoefding and Bernstein races for selecting policies in evolutionary direct policy search. In: *Proceedings of the 26th International Conference on Machine Learning (ICML 2009)*, S. 401–408, 2009b.
- G. E. Hinton. Learning multiple layers of representation. *Trends in Cognitive Sciences*, 11:428–434, 2007.
- G. E. Hinton, S. Osindero und Y. Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.
- M. Hoffman, A. Doucet, N. de Freitas und A. Jasra. Bayesian policy learning with trans-dimensional MCMC. In: *Advances in Neural Information Processing Systems 20 (NIPS 2007)*. MIT Press, 2008.
- A. Hyvärinen, J. Karhunen und E. Oja. *Independent Component Analysis*. Wiley, 2001.
- C. Igel. Evolutionary kernel learning. In: C. Sammut und G. I. Webb (Hrsg.), *Encyclopedia of Machine Learning*. Springer-Verlag, 2009. In press.
- C. Igel und B. Sendhoff. Genesis of organic computing systems: Coupling evolution and learning. In: R. Würtz (Hrsg.), *Organic Computing*, Kapitel 7, S. 141–166. Springer-Verlag, 2008.
- C. Igel, T. Glasmachers, B. Mersch, N. Pfeifer und P. Meinicke. Gradient-based optimization of kernel-target alignment for sequence kernels applied to bacterial gene start detection. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 4(2):216–226, 2007.
- J. Jägersküpper. How the (1+1) ES using isotropic mutations minimizes positive definite quadratic forms. *Theoretical Computer Science*, 36(1):38–56, 2006.
- T. Jansen und I. Wegener. On the analysis of a dynamic evolutionary algorithm. *Journal of Discrete Algorithms*, 4(1):181–199, 2006.
- J. Jost, N. Bertschinger, E. Olbrich, N. Ay und S. Fraenkel. An information theoretic approach to system differentiation on the basis of statistical dependencies between subsystems. *Physica A*, 378:1–10, 2007.
- M. Kearns und S. Singh. Near-optimal reinforcement learning in polynomial time. *Machine Learning*, 49 (2):209–232, 2002.
- D. Koller und R. Parr. Policy Iteration for factored MDPs. In: *Proc. of the 16th Conf. on Uncertainty in Artificial Intelligence (UAI 2000)*, S. 326–334, 2000.
- D. Koller und R. Parr. Computing factored value functions for policies in structured MDPs. In: *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI 1999)*, S. 1332–1339, 1999.
- B. Kveton und M. Hauskrecht. An MCMC approach to solving hybrid factored MDPs. In: *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI 2005)*, Bd. 19, S. 1346–1351, 2005.
- B. Kveton und M. Hauskrecht. Solving factored MDPs with exponential-family transition models. In: *Proc. of the 16th Int. Conf. on Planning and Scheduling*, S. 114–120, 2006.
- J. Lafferty, A. McCallum und F. Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Int. Conf. on Machine Learning (ICML 2001)*, S. 282–289, 2001.
- T. Lang und M. Toussaint. Approximate inference for planning in stochastic relational worlds. In: *Proc. of*

- the 26rd Int. Conf. on Machine Learning (ICML 2009), 2009.
- T. Latzke, S. Behnke und M. Bennewitz. Imitative reinforcement learning for soccer playing robots. In: *RoboCup 2006: Robot Soccer World Cup X*, Bd. 4434 d. Reihe LNAI, S. 47–58. Springer, 2007.
- Y. LeCun, L. D. Jackel, L. Bottou, A. Brunot, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard und V. Vapnik. Comparison of learning algorithms for handwritten digit recognition. In: *Int. Conf. on Artificial Neural Networks*, S. 53–60, Paris, 1995. EC2 & Cie.
- M. Leo, N. Mosca, P. Spagnolo, P. L. Mazzeo, T. D’Orazio und A. Distanto. Real-time multiview analysis of soccer matches for understanding interactions between ball and players. In: *Proc. of the 2008 international conference on Content-based image and video retrieval (CIVR 2008)*, S. 525–534. ACM, 2008.
- M. L. Littman, R. S. Sutton und S. Singh. Predictive representations of state. In: *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, Bd. 14, S. 1555–1561. MIT Press, 2002.
- S. Lyu und E. P. Simoncelli. Nonlinear image representation using divisive normalization. In: *Proc. Computer Vision and Pattern Recognition*, 2008. To Appear.
- W. Maass, T. Natschlger und H. Markram. Real-time computing without stable states: A new framework for neural computation based on perturbations. *Neural Computation*, 14(11):2531–2560, 2002.
- O. Matan, J. Bromley, C. Burges, J. Denker, L. Jackel, Y. LeCun, E. Pednault, W. Satterfield, C. Stenard und T. Thompson. Reading handwritten digits: A zip code recognition system. *IEEE Computer*, 25(7): 59–63, 1992.
- A. McGovern und A. G. Barto. Automatic discovery of subgoals in reinforcement learning using diverse density. In: *Proc. of the 18th Int. Conf. on Machine Learning (ICML 2001)*, S. 361–368, 2001.
- P. Meinicke, S. Klanke, R. Memisevic und H. Ritter. Principal surfaces from unsupervised kernel regression. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, S. 1379–1391, 2005.
- B. Mersch, T. Glasmachers, P. Meinicke und C. Igel. Evolutionary optimization of sequence kernels for detection of bacterial gene starts. *International Journal of Neural Systems*, 17(5):369–381, 2007. Selected paper from the International Conference on Artificial Neural Networks (ICANN 2006).
- N. Meuleau, K.-E. Kim, L. P. Kaelbling und A. R. Cassandra. Solving POMDPs by searching the space of finite policies. In: *Proc. of Fifteenth Int. Conf. on Uncertainty in Artificial Intelligence (UAI 1999)*, S. 417–426, 1999a.
- N. Meuleau, L. Peshkin, K.-E. Kim und L. P. Kaelbling. Learning finite-state controllers for partially observable environments. In: *Proc. of Fifteenth Conf. on Uncertainty in Artificial Intelligence (UAI 1999)*, S. 427–436, 1999b.
- H. Müller, M. Lauer, R. Hafner, S. Lange, A. Merke und M. Riedmiller. Making a robot learn to play soccer using reward and punishment. In: *Proc. of KI 2007 Advances in Artificial Intelligence*, Bd. 4667, S. 220–234. Springer, 2007.
- T. Ott und R. Stoop. The neurodynamics of belief propagation on binary markov random fields. In: *Advances in Neural Information Processing Systems (NIPS 2006)*, Bd. 19, S. 1057–1064. MIT Press, 2007.
- A. Pellicchia, C. Igel, J. Edelbrunner und G. Schöner. Making driver modeling attractive. *IEEE Intelligent Systems*, 20(2):8–12, 2005.
- M. Pfister, S. Behnke und R. Rojas. Recognition of handwritten ZIP codes in a real-world non-standard-letter sorting system. *Applied Intelligence*, 12(1-2):95–115, 2000.
- P. Poupart und C. Boutilier. Bounded finite state controllers. In: *Advances in Neural Information Processing Systems 16 (NIPS 2003)*, Bd. 16. MIT Press, 2004.
- P. Poupart, N. Vlassis, J. Hoey und K. Regan. An analytic solution to discrete bayesian reinforcement learning. In: *Proc. of the 23rd Int. Conf. on Machine Learning (ICML 2006)*, S. 697–704, 2006.
- L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of IEEE*, 77:257–285, 1989.
- R. P. N. Rao. Bayesian computation in recurrent neural circuits. *Neural Computation*, 16:1–38, 2004.
- C. E. Rasmussen und C. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
- G. Rätsch, S. Sonnenburg, J. Srinivasan, H. Witte, K. Müller, R. Sommer und B. Schölkopf. Improving the

- C. elegans genome annotation using machine learning. *PLoS Computational Biology*, S. e20, 2007.
- M. Riedmiller, T. Gabel, R. Hafner und S. Lange. Reinforcement Learning for Robot Soccer. *Autonomous Robots*, 27(1):55–74, 2009.
- H. Ritter, T. Martinez und K. Schulten. *Neural Networks*. Addison Wesley, 1991.
- H. Ritter, J. Steil, C. Nölker, F. Röthling und P. McGuire. Neural architectures for robotic intelligence. *Reviews in the Neurosciences*, 14:121–143, 2003.
- B. Roark, M. Saraclar und M. Collins. Discriminative n-gram language modeling. *Computer Speech and Language*, 21(2):373–392, 2007.
- R. Rojas. *Theorie der neuronalen Netze*. Springer, 1993.
- S. T. Roweis und L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- J. Schmidhuber. Curious model-building control systems. In: *Proc. of Int. Joint Conf. on Neural Networks*, Bd. 2, S. 1458–1463, 1991a.
- J. Schmidhuber. Adaptive confidence and adaptive curiosity. Technical Report FKI-149-91, Technical University Munich, 1991b.
- G. Schohn und D. Cohn. Less is more: Active learning with support vector machines. In: *Proc. of the 17th Int. Conf. on Machine Learning*, S. 839–846, 2000.
- B. Schölkopf und A. J. Smola. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press, 2002.
- B. Schölkopf, A. Smola und K. R. Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*, 10(5):1299–1319, 1998.
- D. Schuurmans und R. Patrascu. Direct value-approximation for factored MDPs. In: *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, S. 1579–1586, 2002.
- G. Seetharaman, A. Lakhota und E. Blasch. Unmanned vehicles come of age: The DARPA Grand Challenge. *IEEE Computer*, 39(12):26–29, 2006.
- E. P. Simoncelli und B. Olshausen. Natural image statistics and neural representation. *Annual Review of Neuroscience*, 24:1193–1216, 2001.
- F. Sinz und M. Bethge. The conjoint effect of divisive normalization and orientation selectivity on redundancy reduction. In: *Advances in Neural Information Processing Systems (NIPS 2008)*, 2009.
- F. H. Sinz, S. Gerwinn und M. Bethge. Characterization of the p-generalized normal distribution. *Journal of Multivariate Analysis*, 2008. Im Druck.
- I. Steinwart und A. Christmann. *Support Vector Machines*. Information Science and Statistics. Springer-Verlag, 2008.
- S. Still. Information theoretic approach to interactive learning. *Europ.Phys.Letters*, 2009.
- R. Sutton, D. Precup und S. Singh. Between MDPs and semi-MDPs: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112:181–211, 1999.
- B. Taskar, C. Guestrin und D. Koller. Max-margin markov networks. In: *Advances in Neural Information Processing Systems (NIPS 2003)*, Bd. 16. MIT Press, Cambridge, MA, 2004.
- J. B. Tenenbaum, V. Silva und J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- G. Tesauro. Neurogammon wins computer olympiad. *Neural Computation*, 1:321–323, 1989.
- G. Tesauro. Temporal difference learning and TD-Gammon. *Communications of the ACM*, 38(3):58–68, 1995.
- S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann et al. Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, 23(9):661, 2006.
- M. Toussaint. Probabilistic inference as a model of planned behavior. *Künstliche Intelligenz*, 2009a. In press.

- M. Toussaint. Robot trajectory optimization using approximate inference. In: *Proc. of the 26th Int. Conf. on Machine Learning (ICML 2009)*, 2009b.
- M. Toussaint und A. Storkey. Probabilistic inference for solving discrete and continuous state Markov Decision Processes. In: *Proc. of the 23rd Int. Conf. on Machine Learning (ICML 2006)*, S. 945–952, 2006.
- M. Toussaint, S. Harmeling und A. Storkey. Probabilistic inference for solving (PO)MDPs. Forschungsbericht EDI-INF-RR-0934, University of Edinburgh, School of Informatics, 2006.
- M. Toussaint, L. Charlin und P. Poupart. Hierarchical POMDP controller optimization by likelihood maximization. In: *Uncertainty in Artificial Intelligence (UAI 2008)*, 2008.
- I. Tsochantaridis, T. Joachims, T. Hofmann und Y. Altun. Large margin methods for structured and interdependent output variables. *Journal of Machine Learning Research*, 6:1453–1484, 2005.
- V. Vapnik. *Statistical Learning Theory*. Springer-Verlag, New York, USA, 1998.
- M. D. Vose. *The Simple Genetic Algorithm: Foundations and Theory*. MIT Press, Cambridge, 1999.
- C. Watkins. Learning from delayed rewards, 1989. Ph.D. Dissertation, Cambridge University, England.
- J. Weston, F. Ratle und R. Collobert. Deep learning via semi-supervised embedding. In: *Proc. of the 25th Int. Conf. on Machine Learning (ICML 2008)*, 2008.
- S. Whiteson und P. Stone. Evolutionary function approximation for reinforcement learning. *Journal of Machine Learning Research*, 7:877–917, 2006.
- L. Wiskott, J.-M. Fellous, N. Krüger und C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(7):775–779, 1997.
- X. Yao. Evolving artificial neural networks. *Proc. of IEEE*, 87:1423–1447, 1999.
- X. Zhang, D. Aberdeen und S. V. N. Vishwanathan. Conditional random fields for multi-agent reinforcement learning. In: *Proc. of the 24th Int. Conf. on Machine Learning (ICML 2007)*, S. 1143–1150. ACM, 2007.