

# Projektbeschreibung

## Projekt: „Reinforcement Learning for several Environments“

Andreas Matt und Georg Regensburger  
Institut für Mathematik, Universität Innsbruck

Das Projekt „*Reinforcement Learning for several Environments*“ wurde von uns, Mag. Andreas Matt und Mag. Georg Regensburger, Anfang 2001 am Institut für Mathematik initiiert und in enger Kooperation mit dem Departamento de Computacion, Buenos Aires, Argentinien, und Prof. Juan Miguel Santos, entwickelt.

Der wissenschaftliche Schwerpunkt des Projekts ist Reinforcement Learning (Belohnungslernen), ein Teilbereich der Künstlichen Intelligenz. Die Basis für eine erfolgreiche Forschung in diesem Bereich ist die Zusammenarbeit verschiedener Fachgebiete. Mit seinem interdisziplinären Ansatz fügt sich das Projekt somit hervorragend in den zukunftsweisenden Fachbereich Mathematik und Informatik der Universität Innsbruck ein.

Die Publikationen wären ohne den vorherigen Aufbau der Strukturen des Projekts nicht realisierbar gewesen. Daneben tragen internationale Kontakte und die Teilnahme an Kongressen, aber auch Vorträge und Veranstaltungen, wie zum Beispiel im Rahmen der ScienceWeek, zum Projekt bei. Unsere intensive Zusammenarbeit bildet einen weiteren entscheidenden Faktor für den Fortschritt der wissenschaftlichen Arbeit. Zur Zeit arbeiten wir an unseren Dissertationen bei Prof. Ulrich Oberst.

Die Projektbeschreibung enthält nach Informationen zu unseren Personen, eine kurze Beschreibung des Projektinhaltes und der entstandenen Arbeiten. Es folgt eine Zusammenstellung der Publikationen, Kongressteilnahmen, Vorträge und Veranstaltungen. Für weitere Informationen verweisen wir auf die Projekt-Webseite:

<http://mathematik.uibk.ac.at/users/rl/>.

## 1 Personen

### Andreas Matt

Studium der Mathematik, Informatik und Philosophie in Innsbruck, Wien, Sevilla und Buenos Aires. Forschungsstipendium der Universität Innsbruck.

Stipendium für kurzfristige wissenschaftliche Arbeiten im Ausland. Derzeit Lehrbeauftragter am Institut für Informatik und Dissertant am Institut für Mathematik.

### **Georg Regensburger**

Studium der Mathematik in Innsbruck und Madrid. Forschungsstipendium der Universität Innsbruck. Forschungsassistent bei Prof. Otmar Scherzer am Institut für Informatik. Derzeit Lehrbeauftragter am Institut für Informatik und Dissertant am Institut für Mathematik.

## **2 Beschreibung des Projektes**

### **2.1 Einleitung**

Reinforcement Learning ist ein junges Gebiet, das sich in den letzten beiden Jahrzehnten zu einem der aktivsten Forschungsgebiete der Künstlichen Intelligenz entwickelt hat. Markovsche Entscheidungsprozesse und Dynamisches Programmierung bilden die mathematischen Grundlagen der Theorie. Konkrete Anwendungen gibt es vor allem im Bereich mobiler Roboter, industrieller Fertigung, bei kombinatorischen Suchproblemen, Liftsystemen und Spielen, wie zum Beispiel das derzeit weltbeste Backgammonspielprogramm.

Die zugrundeliegende Idee von Theorie und Algorithmen, die unter dem Begriff Reinforcement Learning zusammengefasst sind, ist einfach und überzeugend - Lernen durch Versuch und Irrtum in Interaktion mit der Umgebung. Der Agent (Lernende) befindet sich in einer Situation, hat bestimmte Aktionen zur Auswahl und wählt eine davon aus. Das Ergebnis der Aktion, eine neue Situation, wird von der Umgebung bestimmt. Er erhält nun eine Belohnung oder eine Bestrafung (Reinforcement), einen numerischen Wert, je nachdem ob das Ergebnis der Aktion positiv oder negativ bewertet wird.

Ziel aller Lernalgorithmen in Reinforcement Learning ist es, in jeder Situation Aktionen zu finden, die Belohnungen auf lange Sicht maximieren. Die unter Umständen zufällige Auswahl einer Aktion für jede mögliche Situation einer Umgebung nennt man eine Politik. Bisher wurden in Reinforcement Learning Verfahren betrachtet, die es ermöglichen, Verhalten in einer Umgebung zu lernen. Zahlreiche Verfahren, um optimale bzw. suboptimale Politiken für eine Umgebung zu finden, sind bekannt.

### **2.2 Theorie, Simulation, Experimente**

Ausgehend von einer exakten mathematischen Darstellung der Methoden von Reinforcement Learning war es uns möglich schrittweise die bestehenden Verfahren zu analysieren und in Folge zu erweitern. Das Zusammenspiel von

Theorie, Simulation und realen Experimenten bilden für uns dabei eine zentrale Rolle. Im Zuge des Projektes wurde der Kleinroboter Khepera vom Institut für Mathematik erworben und eine Steuerung programmiert. Darüber hinaus wurde ein umfangreicher Roboter-Simulator in C++/OpenGL erstellt. So können die entwickelten Algorithmen und theoretischen Überlegungen an simulierten und realen Systemen getestet werden.

## 2.3 Kurzbeschreibung der Publikationen

Der Schwerpunkt des Projektes liegt in einer allgemeineren Betrachtungsweise von Verhalten, unabhängig von einer fixen Umgebung. Zum Beispiel soll ein Roboter lernen, Hindernissen auszuweichen. Wir denken dabei an ein allgemeines Verhalten, das nicht, wie bisher an eine bestimmte Umgebung gebunden ist, sondern allgemein in mehreren Umgebungen erfolgreich anwendbar ist.

Die Problemstellung ist also eine Politik zu finden, die in mehreren Umgebungen die erwarteten Belohnungen möglichst maximiert. In der Literatur zu Reinforcement Learning gibt es unseres Wissens nach noch keine allgemeine Theorie zu dieser Fragestellung.

In den vorgelegten Arbeiten beschreiben wir ein mathematisches Modell des Problems mit unterschiedlichen Lösungen und Algorithmen. Mit der neuentwickelten Theorie lassen sich viele aktuelle Fragestellungen in Reinforcement Learning formulieren und lösen. Die Methoden wurden bereits für reale Roboterexperimente adaptiert und erfolgreich getestet.

### 2.3.1 „Policy Improvement for Several Environments“

Naturgemäß kann eine Politik für mehrere Umgebungen im Allgemeinen nicht optimal für jede Einzelne sein. Aus einer Verbesserung der Politik für eine Umgebung resultiert häufig eine Verschlechterung in einer anderen Umgebung. Trotzdem ist es oft möglich eine Politik für alle Umgebungen gleichzeitig zu verbessern.

Ausgehend von einer geometrischen Interpretation der bisherigen Methoden und einer Erweiterung des mathematischen Modells auf mehrere Umgebungen wird ein neuer Algorithmus entwickelt. Das Verfahren berechnet stochastische Politiken, die nicht mehr für alle Umgebungen gleichzeitig verbessert werden können, also Gleichgewichtspunkte.

Das Optimieren einer Politik für mehrere Belohnungsfunktionen, in der Literatur unter multiobjective Markov decision processes bekannt, ist ein Spezialfall des Modells und kann mit derselben Methode umfassend gelöst werden.

Siehe „Policy Improvement for Several Environments“ und „Policy Improvement for Several Environments - extended Version“.

### 2.3.2 „Generalizing over Environments in Reinforcement Learning“

Eine Politik kann in einer Umgebung erlernt und in anderen Umgebungen angewendet werden. Wir diskutieren die Frage, wie Lernumgebungen so gewählt werden können, dass erlernte Politiken auch in anderen Umgebungen die Belohnungen möglichst maximieren.

Um Politiken für verschiedene Umgebungen vergleichen zu können, führen wir den Begriff der „utility“ einer Politik ein. Anschließend beschreiben wir ein theoretisches Modell, bei dem Politiken in verschiedenen Umgebungen angewendet werden können. Die Theorie wird an Hand von Experimenten in verschiedenen „Blockweltumgebungen“ illustriert.

Siehe „Generalizing over Environments in Reinforcement Learning“.

## 3 Publikationen im Rahmen des Projekts

- [1] Andreas Matt and Georg Regensburger. Approximate policy iteration for several environments and reinforcement functions. In Alain Dutech and Olivier Buffet, editors, *Proceedings of the 6th European Workshop on Reinforcement Learning (EWRL-6)*, pages 15–17, Nancy, France, 2003.
- [2] Andreas Matt and Georg Regensburger. Generalization over environments in reinforcement learning. To appear in the *Ibero-American Journal of Artificial Intelligence*, 2003.
- [3] Andreas Matt and Georg Regensburger. Generalization over environments in reinforcement learning. In Juan Miguel Santos and Adriana Zapico, editors, *Proceedings of the 4th Argentine Symposium on Artificial Intelligence (ASAI 2002)*, pages 100–109, Santa Fe, Argentina, 2002.
- [4] Andreas Matt and Georg Regensburger. Policy improvement for several environments. In Marco A. Wiering, editor, *Proceedings of the 5th European Workshop on Reinforcement Learning (EWRL-5)*, pages 30–32, Utrecht, Netherlands, 2001.
- [5] Andreas Matt and Georg Regensburger. Policy improvement for several environments - extended version. URL: <http://mathematik.uibk.ac.at/users/r1/>, 2001.
- [6] Juan Miguel Santos, Andreas Matt, and Claude Touzet. Tuning vector parametrized reinforcement functions. In Marco A. Wiering, editor,

*Proceedings of the 5th European Workshop on Reinforcement Learning (EWRL-5)*, pages 42–43, Utrecht, Netherlands, 2001.

- [7] Andreas Matt. Time variable reinforcement learning and reinforcement function design. Master's thesis, Universität Innsbruck, 2000.

## 4 Kongresse

- 6th European Workshop on Reinforcement Learning (EWRL-6), Nancy, Frankreich, September 2003, <http://www.loria.fr/conferences/EWRL6/>.  
Vortrag „Approximate Policy Iteration for Several Environments“.
- 4th Argentine Symposium on Artificial Intelligence (ASIA 2002), Santa Fe, Argentinien, September 2002, <http://fierro.frsf.utn.edu.ar/investigacion/jaiio2002/>.  
Vortrag „Generalizing over Environments in Reinforcement Learning“. Program committee member.
- 5th European Workshop on Reinforcement Learning (EWRL-5), Utrecht, Holland, Oktober 2001, <http://www.cs.uu.nl/marco/EWRL5>.  
Vortrag „Policy Improvement for Several Environments“. Vortrag „Tuning Vector Parametrized Reinforcement Functions“.

## 5 Vorträge

- 16. November 2002, Vortrag „Introducción al Aprendizaje por Refuerzo para Robots“, Escuela de Roberto Arlt, Tortuguitas, Buenos Aires.
- 6. November 2002, Vortrag „Aprendizaje por Refuerzo, Policy Improvement en conjuntos de entornos y funciones de refuerzo“, Universidad de Buenos Aires.
- 29. Mai 2002, Vortrag „Projekt: Reinforcement Learning“, Universität Innsbruck, Exkursion Vorarlberger MathematiklehrerInnen.
- 17. Jänner 2002, Vortrag „Wie Roboter denken und lernen“, Universität Innsbruck. Veranstaltet vom Tiroler Arbeitskreis für Künstliche Intelligenz.

## 6 Public Understanding of Science

### 6.1 ScienceWeek

- 27. September 2002, Prämierung der Veranstaltung „Mensch vs. Roboter“ im Rahmen des „After Work“ – Treffens der ScienceWeek in Wien.
- 16. Juni 2002, Veranstaltung „Mensch vs. Roboter“, im Rahmen des Festes der Wissenschaft, SOWI, Universität Innsbruck.
- 13. bis 15. Juni 2002, Veranstaltung „Mensch vs. Roboter“, Einkaufspark Sillpark, Innsbruck. Informationen unter <http://fabula.tv/roboter>.

### 6.2 Weitere Veranstaltungen

- 24. Februar 2003, Präsentationsstand „Ein Roboter lernt, sich zu orientieren“, im Rahmen der Veranstaltung „Mathematik – faszinierende Forschung und aktuelle Anwendungen“, Universität Innsbruck.
- 16. November 2002, Ausstellung und Präsentation im Rahmen der Expotortuguitas Robotica 2002, Escuela de Roberto Arlt, Tortuguitas, Buenos Aires, <http://www.expotortuguitas.com.ar/>.
- 16. bis 18. Oktober 2002, Studien- und Berufsmesse Best3, Congress Innsbruck. Präsentation gemeinsam mit dem Verein fabula.
- 21. bis 29. September 2002, Veranstaltung und Vorträge, „Cybervillage“, Herbstmesse, Innsbruck, Präsentationen gemeinsam mit Prof. Juan Miguel Santos, Buenos Aires, und dem Verein fabula.
- 23. Jänner und 13. März 2002, Robotervorführung für Schulklassen, Universität Innsbruck. Berufsschule für EKE (Elektronik, Kommunikation und Energie) und Informatik-Wahlpflichtfachgruppe, BORG Innsbruck.
- 27. September 2001, Vortrag „Intelligentes Verhalten von Robotern“, FIT, Aktionstag Frauen in die Technik, Universität Innsbruck.