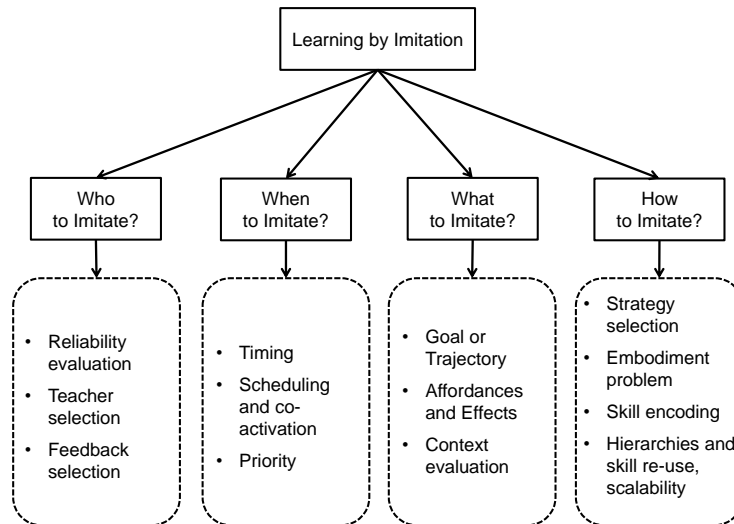# Robust and Incremental Robot Learning by Imitation

Roberto Capobianco

Sapienza University of Rome
capobianco@dis.uniroma1.it

Abstract. In the last years, Learning by Imitation (LbI) has been increasingly explored in order to easily instruct robots to execute complex motion tasks. However, most of the approaches do not consider the case in which multiple and sometimes conflicting demonstrations are given by different teachers. Nevertheless, it seems advisable that the robot does not start as a tabula-rasa, but re-using previous knowledge in imitation learning is still a difficult research problem. In order to be used in real applications, LbI techniques should be robust and incremental. For this reason, the challenge of our research is to find alternative methods for incremental, multi-teacher LbI.

1 Introduction and Relevance of the Problem

Over the last decade, robot *Learning by Imitation* (LbI) has been increasingly explored in order to easily and intuitively instruct robots to execute complex tasks. By providing a human-friendly interface for programming by demonstration, such methods can support the deployment of robotics in domestic and industrial environments. Technical intervention of expert users, in fact, would be not strictly required and, therefore, the costs for (re)programming a robot are drastically reduced. Despite the advantages in terms of flexibility and cost reduction, LbI also brings its own set of problems. For example, understanding the focus of the demonstration ("what to imitate"), adapting the demonstration to the different embodiment of the robots and obtaining good performances in task execution ("how to imitate") are typical challenges of LbI. These problems have been described and addressed in several ways [1][2][3] and a large literature exists on the topic. For example, different representations have been proposed for encoding learned trajectories or goals, and interactive learning techniques have been developed, as in [4], for improving the acquired skills. The common assumption behind a large part of work in literature, however, is that demonstrations are provided by a single teacher, in particular a human. This is not always the case, because not only a robot could learn from other robots [5][6] or animals (e.g., bio-inspired robots), but also multiple teachers could provide to the robot conflicting demonstrations or feedback/advice. Moreover, while only some work focused on the incremental learning problem, it is crucial for achieving robot autonomy. It seems advisable, in fact, that a robot does not start to learn a

```
                         ┌─────────────────────┐
                         │ Learning by Imitation│
                         └─────────────────────┘
          ┌─────────────┬────────┴───────┬─────────────┐
          ▼             ▼                ▼             ▼
   ┌───────────┐  ┌───────────┐   ┌───────────┐  ┌───────────┐
   │   Who     │  │   When    │   │   What    │  │   How     │
   │ to Imitate?│ │ to Imitate?│  │ to Imitate?│ │ to Imitate?│
   └───────────┘  └───────────┘   └───────────┘  └───────────┘
```

- Reliability evaluation
- Teacher selection
- Feedback selection

- Timing
- Scheduling and co-activation
- Priority

- Goal or Trajectory
- Affordances and Effects
- Context evaluation

- Strategy selection
- Embodiment problem
- Skill encoding
- Hierarchies and skill re-use, scalability

Fig. 1. Schema of the main challenges in LbI.

single task every time from scratch, since its knowledge can be augmented for executing more complex tasks or for obtaining increasingly better imitations. The challenge of our research is to propose a set of solutions for improving LbI techniques, by considering both multiple teachers and incremental learning. In contrast to previous work, we will focus our research on learning from multiple categories of teachers (e.g., humans, robots, animals). Moreover, we will consider classical solutions like reliability measurements and teacher selection as well as techniques for strategy co-activation, strategy changing and online refinement via contrasting feedback/advice. Sub-skill co-activation will be also adopted for improving incremental learning, with the underlying idea of extending current non-symbolic approaches to reach higher levels of learning autonomy for hierarchical and complex tasks.

2 Related Work

As already stated in the previous Section, Learning by Imitation provides a high level method for programming a robot which can be easily used by non expert users. However, while the effort for providing prior knowledge to the robot is drastically reduced, new and different issues emerge. A frequently used description of LbI challenges consists of a set of independent problems presented in the form of questions (see Fig. 1): *Who* to imitate? *When* to imitate? *What* to imitate? *How* to imitate?. A huge effort has been done, in previous work, for understanding *what* is relevant for the robot and *how* it should learn a skill, while *who* and *when* to imitate are still open challenges. Indeed, only a small

amount of work has been done in this direction. A detailed overview over the adopted approaches for solving those problems can be found in [7] and [3].

One of the first problems when dealing with imitation is to understand how to encode a learned skill. While spline representations cannot be easily used for encoding a skill, because of their explicit time dependency, many alternatives exist. In detail, Hidden Markov models (HMMs) have often been successfully applied in this context [8]. Billard *et al.* [9], for example, use two HMMs, one to eliminate signals with high variability and the other one, fully connected, to obtain a probabilistic encoding of the task. In the work by Asfour *et al.* [10], a humanoid robot is instructed by using continuous HMMs, trained with a set of key points common to almost all the demonstrations. By detecting also temporal dependencies between the two arms, dual-arm tasks are successfully executed. Calinon *et al.* [11], instead, use HMMs for representing a joint distribution of position and velocity, while generalizing the motion during the reproduction through the use of Gaussian Mixture Regression. The approach is validated on several robotics platforms. Additional improvements in the generalization of movements have been achieved thanks to the use of Gaussian mixture models (GMMs). For example, in [12], the authors propose a LbI framework, based on a mixture of Gaussian/Bernoulli distributions, for extracting relevant features of a task and generalizing the acquired knowledge to multiple contexts. Chernova and Veloso [13], instead, use a representation of the policy based on GMMs in order to address the uncertainty of human demonstration. In particular, they propose an approach which enables the agent to request demonstrations for specific parts of the state space, achieving increasing autonomy in the execution based on the analysis of the learned Gaussian mixture set.

In order to reduce the typically high-dimensional state-action space of those problems, a different category of work focus on the representation of tasks as a composition of motion primitives. Dynamic Movement Primitives (DMPs), in particular, have been proposed by Ijspeert *et al.* [14][15][16] in order to encode the properties of the motion by means of differential equations. These primitives, which can take into account perturbations and feedback terms, have been successfully applied by Schaal *et al.* [17], in the context of learning by demonstration, on several examples. Ude *et al.* [18] present a method for generalizing periodic DMPs and synthesizing new actions in situations that a robot has never encountered before. As an additional example, Stulp and Schaal [19] use DMPs for hierarchical learning via Reinforcement Learning (RL) and apply their approach on a 11-DOF arm plus hand for a pick-and-place task. More recently, an alternative representation such as Probabilistic Movement Primitives has been proposed by Paraschos *et al.* [20], which can be used in several applications and allows for blending between motions, adapting to altered task variables, and co-activating multiple motion primitives in parallel.

In several work, traditional imitation learning techniques have been associated with methods for refining the learned policy, as in the case of Nicolescu and Mataric [21]. More specific approaches based on Reinforcement Learning enable the reduction of the time needed for finding good control policies, while improv-

ing the performance of the robot (when possible) beyond that of the teacher. Guenter and Billard [22], for example, use RL in order to relearn goal-oriented tasks even with unexpected perturbations. More in detail, a GMM is used as a first trial to reproduce the task and, then, RL is used to adapt the encoded speed to perturbations. A limitation of the approach is that the system requires to completely relearn the trajectory every time a new perturbation is added. Kober and Peters [23] use episodic RL in order to improve motor primitives learned by imitation for a Ball-in-a-Cup task. Kormushev *et al.* [24], instead, encode movements with and extension of DMPs initialized from imitation. RL is then used for learning the optimal parameters of the policy, thus improving the learned capability. A different approach in the same direction, instead, has been proposed by Argall *et al.* [4]. Rather than using traditional Reinforcement Learning, in fact, the authors consider the advice of the teacher in order to improve the learned policy, by directly applying a correction on the executed state-action mapping.

Such solutions are well suited whenever a robot needs to learn a task from a single teacher. However, issues emerge if conflicting demonstrations, or rewards in the case of RL, are provided by different teachers ("who to imitate") by means of different sensors and modalities. For non-linear system, in fact, simply averaging the learned trajectories usually results in a new trajectory that is not feasible, since it does not obey the constraints of the dynamic model. Preliminary work in the direction of addressing this problem has been done by Nicolescu and Mataric [21], who propose a topology based method for generalization among multiple demonstrations represented as behavior networks. Argall *et al.* [25] consider the incorporation of demonstrations from multiple teachers by selecting among them on the basis of their observed reliability. More specifically, reliability is measured and represented through a weighted scheme. Babes *et al.* [26] apply Inverse Reinforcement Learning (IRL) [27] to learning from demonstration, by adopting a clustering procedure on the observed trajectory for inferring the expert's intention. This is particularly useful to discriminate among different demonstrations whose underlying goal (and reward function) is not previously or clearly specified. Tawani and Billard [28], instead, propose a method based on IRL for learning to mimic a variety of experts with different strategies. While providing a high adaptability, such an approach enables to bootstrap optimal policy learning by transferring knowledge from the set of learned policies. Most of this approaches, however, neither enable to smoothly switch among different policies, when needed, nor consider the opportunity to prioritize among different strategies which are not incompatible. Moreover, teachers are usually considered to be human beings, while in real applications demonstrations could be provided by arbitrary expert agents, such as other robots [5][6] or even animals. Additional work should be focused on the online version of this problem, in which contrasting feedback is given to the robot by multiple teachers and refinements over different learned policies are required.

Another limitation of the work in the literature is in the assumption that robots need to learn a single task from scratch, without previous knowledge.

Real world applications, instead, are highly demanding for robots which can incrementally acquire new task execution capabilities based on already learned skills. A huge effort for dealing with this problem has been done in the direction of using symbolic representations of tasks, as in [21]. Pardowitz *et al.* [29][30], who follow the general approach described in [31], use a hierarchical representation of complex tasks, generated as a sequence of elementary operators (i.e., basic actions, primitives). The method is applied on a robot servant which has to learn an everyday household task by combining reasoning and learning. A similar approach is used in the work by Ekvall and Kragic [32], who decompose each task in sub-tasks which are then used, together with a set of constraints and the identified goal, for obtaining generalization. Symbolic representations offer, of course, many advantages when dealing with complex tasks, but they require a big effort to provide prior knowledge to the robot, resulting in a loss of flexibility. Conversely, other work is oriented to the achievement of incremental learning from scratch, without the intervention of experts in providing knowledge. Friesen and Rao [33] propose a solution for achieving hierarchical task control, by means of an extended Bellman equation. Starting from the equation used in [34] for "implicit imitation", the authors consider both temporally extended action (called options) and primitives. Such options can execute other options. An interesting evolution towards incremental learning can be noticed in the work by the research group of Jan Peters [35][36][37][38][39][40][41]. In particular, in [39] a general overview of the adopted modular approach is given. The authors describe a method for generalizing and learning several motor primitives (building blocks), as well as learning to select and to sequence the building blocks for executing complex tasks. Even though this technique represents a huge advancement towards incremental learning, the gap between the pure symbolic approach and the "numerical" one is still significant.

3 Methodology and Proposed Solution

The challenge of this research idea consists in addressing both the problems of multi-teaching (robustness) and incremental learning, by starting from the work previously presented. With this purpose, state-of-the-art sensing techniques and off-the-shelf perception modules will be considered to acquire task demonstrations, since they are not directly related to the considered challenges.

The general idea of the proposed approach is based on a mixture of techniques from Artificial Intelligence and Control Theory. In fact, on the one hand Reinforcement Learning has been often explored in combination with traditional LbI for efficient and accurate task reproduction; on the other hand, it has been shown that RL is also effective for obtaining bio-inspired and adaptive controllers able to find optimal policies, in terms of control cost, on-line [42]. Assume that, for each task, n different or contrasting demonstrations are provided to the robot, by k different teachers. Each teacher may have his own strategy or may change his behavior on the basis of the context. Starting from these, a basic step would be the generation of a smaller number $\bar{n}$ of clusters, in order to reduce the dimen-
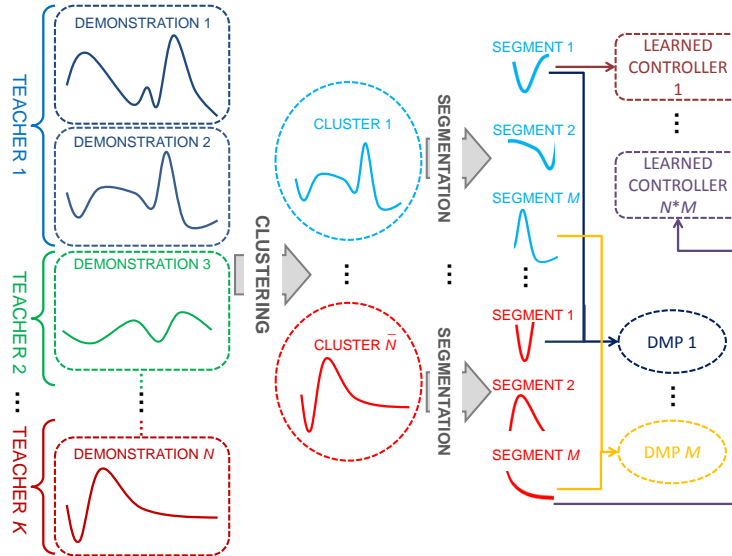
Fig. 2. Graphical presentation of the learning approach.

sionality of the problem. After dividing the obtained clusters into m sub-parts, through a segmentation process, each demonstrated sub-policy $(\bar{n} \cdot m)$ should be learned by applying Inverse Reinforcement Learning techniques. Contextually, in order to produce a more goal-oriented solution, m general DMPs (one for each sub-part) will be continuously refined on the basis of the set of all the n demonstrations. A graphical description of the approach is available in Fig. 2. At task execution time (on-line), for each sub-part, the robot should be able to choose among the different policies and the refined DMPs, on the basis of the context or constraints. The choice will strictly depend on the state of the robot and on the priority (if available) of the tasks to be executed. Interaction with users characterized by different policies will enable a further refinement of the adopted policies, as well as a weighting process among the produced solutions, based on their given reward. Eventually, in case of non contrasting demonstrations, a priority based execution of co-activated policies will be implemented.

Intuitively, such a "motion library" will be useful to address two typical issues of the incremental learning problem: recognizing in the demonstrations the set of already available sub-skills, and reducing the redundancy of task information. Based on this, the approach adopted in [39] for combining the building blocks in the execution of complex tasks, will be extended to consider co-activated non interfering sub-skills, on a priority basis. Moreover, a simple approach, based on the extraction of the most relevant features of each sub-task, will be used in order to try to partially reduce the gap between the numerical and symbolic representations used in LbI. Contextually, higher level planning will be eventually executed by means of *Hierarchical Task Networks*.
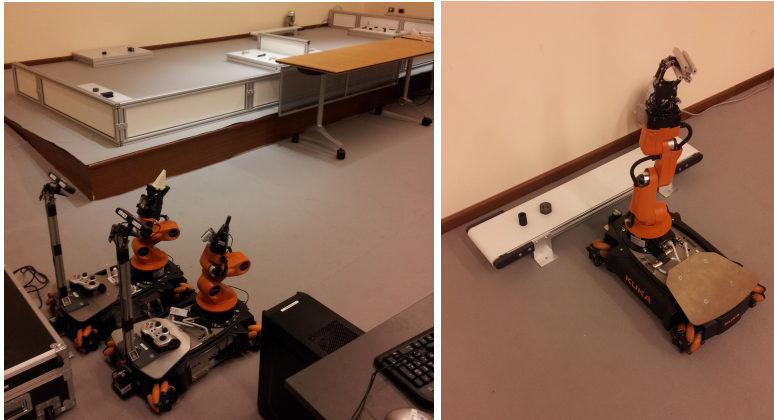
Fig. 3. KUKA Youbot robotic platform at the RoCKIn Camp 2014.

The proposed solutions will be extensively validated on simulated and real robots, as well as both in domestic and industrial domains. In particular, the whole system will be developed on the Robot Operating System (ROS)[1] framework, since it is very popular. This will enable not only an easy integration with realistic simulators like V-REP[2] and Webots[3], but also a simply transferable implementation for a real robot, like the KUKA Youbot (Fig. 3). Such a robotic platform consists of a omnidirectional mobile base and a 5-DOF arm, plus the gripper, and it can be considered a good solution for preliminary experiments in this research. Using the Youbot, in fact, allows to experiment LbI in industrial-like scenarios, as in the case of the RoCKIn[4]@Work competitions. Due to the robot structure, LbI implementations on this platform will have to take into account the correspondence problem [1]. Note, however, that this is a classical issue in the LbI implementation pipeline, since the embodiment of the demonstrator and the one of the robot are usually different, with the exception of humanoid robots. Additional tests will be executed on specific simple tasks (e.g., door opening and ball throwing), as well as in the context of benchmarking activities (e.g., RoCKIn).

4 Conclusions and Potential Impact

Producing a robot which can be easily instructed to perform difficult tasks will open many business opportunities. In the next years, in fact, industrial and general purpose domestic robots will be available to wider communities of non expert users. The use of incremental human inspired learning approaches could

---

[1] http://www.ros.org/

[2] http://www.coppeliarobotics.com/

[3] http://www.cyberbotics.com/

[4] http://rockinrobotchallenge.eu/

enable next generation robots to learn from others as well as from their own experience. For this reason, we strongly believe that an intuitive multi-teaching "interface" for robots could improve not only the overall quality of the user experience and the robot usability, but also the acceptance of robots in our society. We also think that exploring robust and incremental LbI methods could have a long-term positive impact from an economical point of view. Consider, for example, the effort in terms of money spent by big companies for programming robots: industries could save a lot of money, having the possibility to easily reprogram, or improve with the advice of different teachers, a single part of the task that a robot has to execute. For this reason, developed algorithms could be included in ROS Industrial[5], whose goal is to transfer the advances in robotics research to concrete applications, with economical potential. From an academic point of view, the interest towards human movement understanding is increasing[6], and improvements in LbI could have a strong impact in this area, since it is strictly related to natural movement and specific motion dynamics. In conclusion, we believe that research in this area can be further extended towards practical applications and real world scenarios, but we are aware that this document represents only the starting point for a detailed analysis and investigation of the possible techniques for approaching robust and incremental LbI.

References

1. Nehaniv, C., Dautenhahn, K.: Like me? - measures of correspondence and imitation. Cybernetics and Systems **32**(1-2) (2001) 11–51
2. Schaal, S., Ijspeert, A., Billard, A.: Computational approaches to motor learning by imitation. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **358**(1431) (2003) 537–547
3. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robot. Auton. Syst. **57**(5) (May 2009) 469–483
4. Argall, B.D., Browning, B., Veloso, M.: Learning robot motion control with demonstration and advice-operators. In: Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, IEEE (2008) 399–404
5. Hayes, G.M., Demiris, J.: A robot controller using learning by imitation. University of Edinburgh, Department of Artificial Intelligence (1994)
6. Gaussier, P., Moga, S., Quoy, M., Banquet, J.P.: From perception-action loops to imitation processes: A bottom-up approach of learning by imitation. Applied Artificial Intelligence **12**(7-8) (1998) 701–727
7. Billard, A., Calinon, S., Dillman, R., Stefan, S.: Robot programming by demonstration. In: Springer handbook of robotics. Springer (2008) 1371–1394
8. Hovland, G.E., Sikka, P., McCarragher, B.J.: Skill acquisition from human demonstration using a hidden markov model. In: Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference on. Volume 3., Ieee (1996) 2706–2711

---

[5] http://rosindustrial.org/

[6] See, for example, IEEE RAS Technical Committee on Human Movement Understandinghttp://www.ieee-ras.org/human-movement-understanding

9. Billard, A.G., Calinon, S., Guenter, F.: Discriminative and adaptive imitation in uni-manual and bi-manual tasks. Robotics and Autonomous Systems **54**(5) (2006) 370–384

10. Asfour, T., Azad, P., Gyarfas, F., Dillmann, R.: Imitation learning of dual-arm manipulation tasks in humanoid robots. International Journal of Humanoid Robotics **5**(02) (2008) 183–202

11. Calinon, S., D'halluin, F., Sauser, E.L., Caldwell, D.G., Billard, A.G.: Learning and reproduction of gestures by imitation: An approach based on hidden Markov model and Gaussian mixture regression. IEEE Robotics and Automation Magazine **17**(2) (2010) 44–54

12. Calinon, S., Guenter, F., Billard, A.: On learning, representing, and generalizing a task in a humanoid robot. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **37**(2) (2007) 286–298

13. Chernova, S., Veloso, M.: Confidence-based policy learning from demonstration using gaussian mixture models. In: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems, ACM (2007) 233

14. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Trajectory formation for imitation with nonlinear dynamical systems. In: Intelligent Robots and Systems, 2001. Proceedings. 2001 IEEE/RSJ International Conference on. Volume 2., IEEE (2001) 752–757

15. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Learning attractor landscapes for learning motor primitives. Technical report (2002)

16. Ijspeert, A.J., Nakanishi, J., Schaal, S.: Movement imitation with nonlinear dynamical systems in humanoid robots. In: Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on. Volume 2., IEEE (2002) 1398–1403

17. Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A.: Learning movement primitives. In: Robotics Research. Springer (2005) 561–572

18. Ude, A., Gams, A., Asfour, T., Morimoto, J.: Task-specific generalization of discrete and periodic dynamic movement primitives. Robotics, IEEE Transactions on **26**(5) (2010) 800–815

19. Stulp, F., Schaal, S.: Hierarchical reinforcement learning with movement primitives. In: Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on, IEEE (2011) 231–238

20. Paraschos, A., Daniel, C., Peters, J., Neumann, G.: Probabilistic movement primitives. In: Advances in Neural Information Processing Systems. (2013) 2616–2624

21. Nicolescu, M.N., Mataric, M.J.: Natural methods for robot task learning: Instructive demonstrations, generalization and practice. In: In Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems. (2003) 241–248

22. Guenter, F., Billard, A.G.: Using reinforcement learning to adapt an imitation task. In: Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on, IEEE (2007) 1022–1027

23. Kober, J., Peters, J.R.: Policy search for motor primitives in robotics. In: Advances in neural information processing systems. (2009) 849–856

24. Kormushev, P., Calinon, S., Caldwell, D.G.: Robot motor skill coordination with em-based reinforcement learning. In: Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on, IEEE (2010) 3232–3237

25. Argall, B.D., Browning, B., Veloso, M.: Automatic weight learning for multiple data sources when learning from demonstration. In: Robotics and Automation, 2009. ICRA'09. IEEE International Conference on, IEEE (2009) 226–231

26. Babes, M., Marivate, V., Subramanian, K., Littman, M.L.: Apprenticeship learning about multiple intentions. In: Proceedings of the 28th International Conference on Machine Learning (ICML-11). (2011) 897–904

27. Ng, A.Y., Russell, S.J., et al.: Algorithms for inverse reinforcement learning. In: Icml. (2000) 663–670

28. Tanwani, A.K., Billard, A.: Transfer in inverse reinforcement learning for multiple strategies. In: Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on, Ieee (2013) 3244–3250

29. Pardowitz, M., Zöllner, R., Dillmann, R.: Incremental learning of task sequences with information-theoretic metrics. In: European Robotics Symposium 2006, Springer (2006) 51–63

30. Pardowitz, M., Knoop, S., Dillmann, R., Zollner, R.: Incremental learning of tasks from user demonstrations, past experiences, and vocal comments. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on **37**(2) (2007) 322–332

31. Muench, S., Kreuziger, J., Kaiser, M., Dillman, R.: Robot programming by demonstration (rpd)-using machine learning and user interaction methods for the development of easy and comfortable robot programming systems. In: Proceedings of the International Symposium on Industrial Robots. Volume 25., INTERNATIONAL FEDERATION OF ROBOTICS, & ROBOTIC INDUSTRIES (1994) 685–685

32. Ekvall, S., Kragic, D.: Learning task models from multiple human demonstrations. In: Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on, IEEE (2006) 358–363

33. Friesen, A.L., Rao, R.P.: Imitation learning with hierarchical actions. In: Development and Learning (ICDL), 2010 IEEE 9th International Conference on, IEEE (2010) 263–268

34. Price, B., Boutilier, C.: Implicit imitation in multiagent reinforcement learning, Citeseer (1999)

35. Kupcsik, A.G., Deisenroth, M.P., Peters, J., Neumann, G.: Data-efficient generalization of robot skills with contextual policy search. In: AAAI. (2013)

36. Muelling, K., Kober, J., Kroemer, O., Peters, J.: Learning to select and generalize striking movements in robot table tennis. The International Journal of Robotics Research **32**(3) (2013) 263–279

37. Peters, J., Kober, J., Mülling, K., Krämer, O., Neumann, G.: Towards robot skill learning: From simple skills to table tennis. In: Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg (2013) 627–631

38. Kober, J., Peters, J.: Learning prioritized control of motor primitives. In: Learning Motor Skills. Springer International Publishing (2014) 149–160

39. Neumann, G., Daniel, C., Paraschos, A., Kupcsik, A., Peters, J.: Learning modular policies for robotics. Frontiers in Computational Neuroscience **8** (2014) 62

40. Bócsi, B., Csató, L., Peters, J.: Indirect robot model learning for tracking control. Advanced Robotics **28**(9) (2014) 589–599

41. Muelling, K., Boularias, A., Mohler, B., Schölkopf, B., Peters, J.: Learning strategies in table tennis using inverse reinforcement learning. Biological cybernetics (2014)

42. Khan, S.G., Herrmann, G., Lewis, F.L., Pipe, T., Melhuish, C.: Reinforcement learning and optimal adaptive control: An overview and implementation examples. Annual Reviews in Control **36**(1) (2012) 42–59