appropriate mechanisms for trajectory reuse than the current pool-based solution, such as integrating the importance weighting technique with the boosting framework [23].

REFERENCES

[1] P. Abbeel, A. Coates, and A. Y. Ng. Autonomous helicopter aerobatics through apprenticeship learning. *The International Journal of Robotics Research*, 29(13):1608–1639, 2010.

[2] P. Abbeel, A. Coates, M. Quigley, and A. Y. Ng. An application of reinforcement learning to aerobatic helicopter flight. In B. Schölkopf, J. C. Platt, and T. Hoffman, editors, *Advances in Neural Information Processing Systems 19 (NIPS)*, pages 1–8. MIT Press, 2007.

[3] P. Abbeel, V. Ganapathi, and A. Y. Ng. Learning vehicular dynamics, with application to modeling helicopters. In B. S. Yair Weiss and J. Platt, editors, *Advances in Neural Information Processing Systems 18 (NIPS)*, pages 1–8. MIT Press, 2005.

[4] D. A. Aberdeen. *Policy-gradient algorithms for partially observable Markov decision processes*. PhD thesis, Australian National University, 2003.

[5] P. L. Bartlett and J. Baxter. Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15:319–350, 2001.

[6] S. Bhatnagar, R. S. Sutton, M. Ghavamzadeh, and M. Lee. Natural actor-critic algorithms. *Automatica*, 45(11):2471–2482, 2009.

[7] L. Breiman, J. Friedman, and C. J. Stone. *Classification and Regression Trees*. Chapman and Hall/CRC, 1984.

[8] A. Coates, P. Abbeel, and A. Y. Ng. Learning for control from multiple demonstrations. In *Proceedings of the 25th International Conference on Machine learning (ICML)*, pages 144–151, Helsinki, Finland, 2008. ACM.

[9] W. Dabney and A. G. Barto. Adaptive step-size for online temporal difference learning. In *AAAI*, 2012.

[10] Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1):119–139, 1997.

[11] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28(2):337–407, 2000.

[12] J. H. Friedman. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5):1189–1232, 2001.

[13] A. Geramifard, R. H. Klein, C. Dann, W. Dabney, and J. P. How. RLPy: The Reinforcement Learning Library for Education and Research, 2013.

[14] M. Ghavamzadeh and Y. Engel. Bayesian actor-critic algorithms. In *Proceedings of the 24th International Conference on Machine Learning*, pages 297–304, Corvallis, OR, 2007.

[15] E. Greensmith, P. Bartlett, and J. Baxter. Variance reduction techniques for gradient estimates in reinforcement learning. *Journal of Machine Learning Research*, 5:1471–1530, 2004.

[16] M. Kearns and L. G. Valiant. Cryptographic limitations on learning boolean formulae and finite automata. In *Proceedings of the 21st Annual ACM Symposium on Theory of Computing*, pages 433–444, Seattle, WA, 1989.

[17] K. Kersting and K. Driessens. Non-parametric policy gradients: A unified treatment of propositional and relational domains. In *Proceedings of the 25th International Conference on Machine Learning (ICML)*, pages 456–463, Helsinki, Finland, July 2008.

[18] H. Kimura. Reinforcement learning by stochastic hill climbing on discounted reward. In *Proceedings of the 12th International Conference on Machine Learning (ICML)*, pages 295–303, Tahoe City, CA, 1995.

[19] G. Konidaris, S. Osentoski, and P. S. Thomas. Value function approximation in reinforcement learning using the Fourier basis. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2011, San Francisco, California, USA, August 7-11, 2011*, 2011.

[20] R. Koppejan and S. Whiteson. Neuroevolutionary reinforcement learning for generalized control of simulated helicopters. *Evolutionary intelligence*, 4(4):219–241, 2011.

[21] H.-Y. Lo, K.-W. Chang, S.-T. Chen, T.-H. Chiang, and C.-S. Ferng. An ensemble of three classifiers for kdd cup 2009: Expanded linear model, heterogeneous boosting, and selective naive Bayes. *The 2009 Knowledge Discovery in Data Competition (KDD Cup 2009) Challenges in Machine Learning, Volume 3*, page 53, 2009.

[22] L. Mason, J. Baxter, P. L. Bartlett, and M. R. Frean. Boosting algorithms as gradient descent. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12 (NIPS)*, pages 512–518. The MIT Press, 2000.

[23] T. Matsubara, T. Morimura, and J. Morimoto. Adaptive step-size policy gradients with average reward metric. In *Proceedings of the 2nd Asian Conference on Machine Learning*, pages 285–298, Tokyo, Japan, 2010.

[24] T. M. Mitchell. *Machine Learning*. McGraw-Hill, Inc., New York, NY, 1997.

[25] A. Y. Ng and M. Jordan. PEGASUS: A policy search method for large MDPs and POMDPs. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 406–415. Morgan Kaufmann Publishers Inc., 2000.

[26] A. Y. Ng and S. J. Russell. Algorithms for inverse reinforcement learning. In *Proceedings of the 17th International Conference on Machine Learning (ICML)*, pages 663–670, Stanford, CA, 2000.

[27] E. Orate, S. Idelsohn, O. C. Zienkiewicz, and R. L. Taylor. A finite point method in computational mechanics. Application to convectioe transport and fluid flow. *International Journal for Numerical Methods in Engineering*, 39:3839–3866, 1996.

[28] L. Peshkin. *Reinforcement learning by policy search*. PhD thesis, MIT, 2001.

[29] J. Peters. Policy gradient methods. *Scholarpedia*, 5(10):3698, 2010.

[30] J. Peters and S. Schaal. Natural actor-critic. *Neurocomputing*, 71(7):1180–1190, 2008.

[31] N. Roy and J. How. A tutorial on linear function approximators for dynamic programming and reinforcement learning. 2013.

[32] R. E. Schapire. The strength of weak learnability. *Machine Learning*, 5(2):197–227, 1990.

[33] R. E. Schapire and Y. Freund. *Boosting: Foundations and Algorithms*. MIT Press, Cambridge, MA, 2012.

[34] R. S. Sutton and A. G. Barto. *Reinforcement learning: An introduction*. MIT Press, Cambridge, MA, 1998.

[35] R. S. Sutton, D. McAllester, S. Singh, Y. Mansour, et al. Policy gradient methods for reinforcement learning with function approximation. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12 (NIPS)*, pages 1057–1063. 2000.

[36] B. Tanner and A. White. RL-Glue: Language-independent software for reinforcement-learning experiments. *The Journal of Machine Learning Research*, 10:2133–2136, 2009.

[37] P. Viola and M. Jones. Robust real-time object detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[38] R. J. Williams. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, 8(3):229–256, 1992.

[39] I. H. Witten and E. Frank. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, 2005.