

Donut as I do: Learning from Failed Demonstration

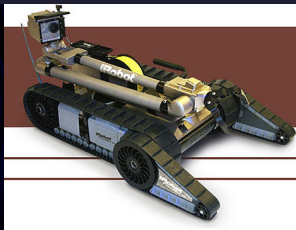
Daniel H Grollman and Aude G Billard

Ecole Polytechnique Fédérale de Lausanne

ICRA 2011 - WeP112.1

Motivation

- ▶ Robots can do great things ...under supervision
 - ▶ Perform surgery
 - ▶ Explore space
 - ▶ Help out in disasters



- ▶ Constant human supervision and/or intense engineering
- ▶ Focus on learning and adaptation to widen applicability

Robot Learning from Demonstration

- ▶ Flexible Skill Acquisition
 - ▶ Acquire skills, flexibly - anywhere / anywhen learning
 - ▶ Acquire flexible skills - adapt to novel situations
- ▶ Intuitive Robot Tasking
 - ▶ Reduce programmatic analysis of task
 - ▶ Untrained users - as interact with human
- ▶ Catchphrase: "If you can do it, you can teach it."
- ▶ But what if you can't?

"He who can, does. He who cannot, teaches."

George Bernard Shaw, Man and Superman (1903) "Maxims for Revolutionists"

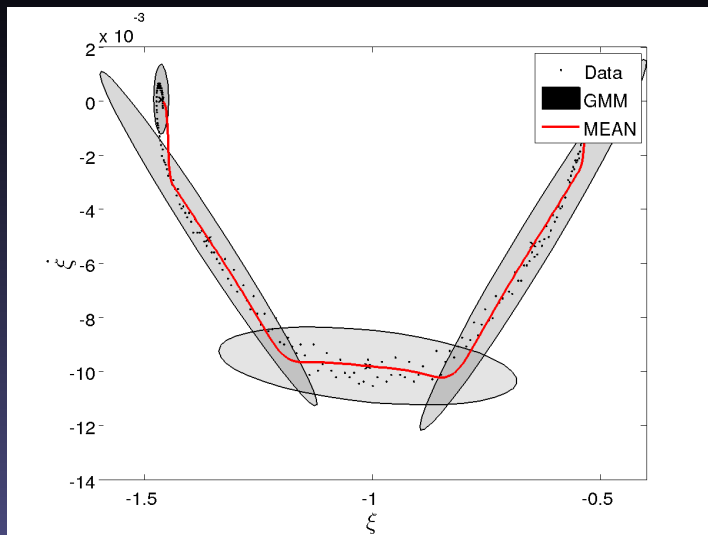
Tasks

- ▶ Discrete, one-stroke movements
- ▶ Velocity important
- ▶ Binary reward - Success/Failure
- ▶ Humans require several attempts

Motion Model

- ▶ Autonomous Dynamical System (ADS): $\dot{\xi} = f(\xi)$
- ▶ Gaussian Mixture Model (GMM)
 - ▶ $P(\dot{\xi}, \xi) = \sum_{k=1}^K \rho^k \mathcal{N}(\dot{\xi}, \xi | \mu^k, \Sigma^k)$
 - ▶ Parameters (θ):
 - ▶ K - Number of components
 - ▶ ρ^k - Prior probability of component k
 - ▶ μ^k - Mean of component k
 - ▶ Σ^k - Covariance of component k
 - ▶ Demonstration data: N state-velocity pairs $\mathbf{D} = \{\xi^n, \dot{\xi}^n\}_{n=1}^N$
 - ▶ Fit $\{\rho^k, \mu^k, \Sigma^k\}_{k=1}^K$ (not K) with Expectation-Maximization (Kmeans initialization)
 - ▶ Find K with Bayesian Information Criterion (Penalized likelihood)
- ▶ Usually, generate $\dot{\xi} = f(\xi) = E\{P(\dot{\xi}|\xi, \theta)\}$

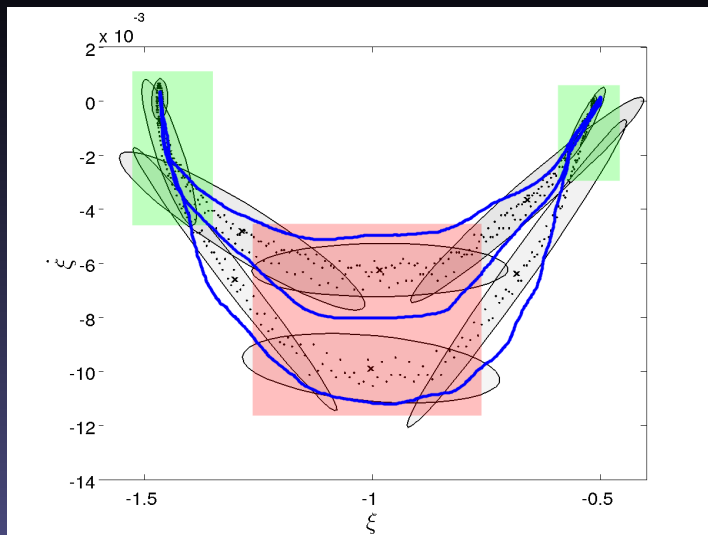
GMM-ADS Illustration



Assumptions

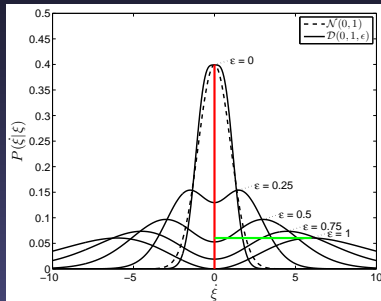
- ▶ (Mostly) Optimal - Humans perform task well, noise is mean-zero Gaussian
- ▶ Sub-Optimal - Humans perform task acceptably, start search here and improve
- ▶ Failed - Humans did not complete the task, do not replicate, explore elsewhere
 - ▶ Failure may be more common than success
 - ▶ Demonstrations = Attempts
 - ▶ Correctness \propto Similarity (Variance⁻¹)

Desired behavior



Donut

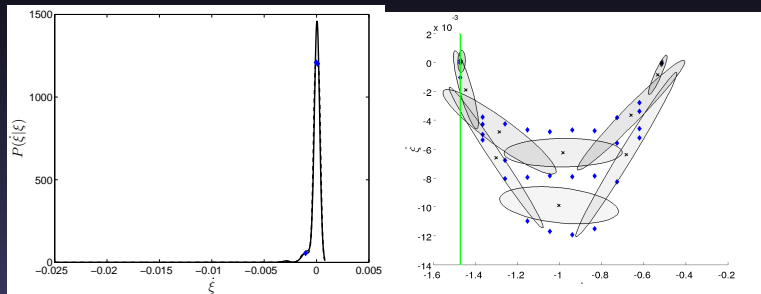
- ▶ Pseudo inverted $\mathcal{N}(\mathbf{x}; \mu, \Sigma)$
- ▶ $\mathcal{D}(\mathbf{x}; \mu, \Sigma, \epsilon) = 2\mathcal{N}(\mathbf{x}; \mu, \frac{1}{r_\alpha^2}\Sigma) - \mathcal{N}(\mathbf{x}; \mu, \frac{1}{r_\beta^2}\Sigma)$
- ▶ r_α, r_β are scalar ratios controlled by exploration (ϵ)



Height: $\eta = 2r_\alpha^D - r_\beta^D$

Width: $\lambda = \sqrt{\frac{2 \log[2(\frac{r_\alpha}{r_\beta})^{D+2}]}{r_\alpha^2 - r_\beta^2}}$

Variance \rightarrow Exploration



Map Similarity (Variance) to exploration: $\epsilon = 1 - \frac{1}{1 + \|V\{P(\dot{\xi}|\xi, \theta)\}\|}$

Model Update

- ▶ Have θ based on $\mathbf{D} = \{\xi^n, \dot{\xi}^n\}_{n=1}^N$
- ▶ New data $\mathbf{D}' = \{\xi^n, \dot{\xi}^n\}_{n=1}^{N'}$
- ▶ Update to θ' based on $\mathbf{D} \cup \mathbf{D}'$
 - ▶ Brute Force
 - ▶ Re-estimate all parameters based on all data
 - ▶ Grows with number of points
 - ▶ Resample
 - ▶ Sample N' points from GMM with θ
 - ▶ Weigh points by $\frac{N}{N'}$
 - ▶ Combine with new data and re-estimate, except K

FlipUp

Basket

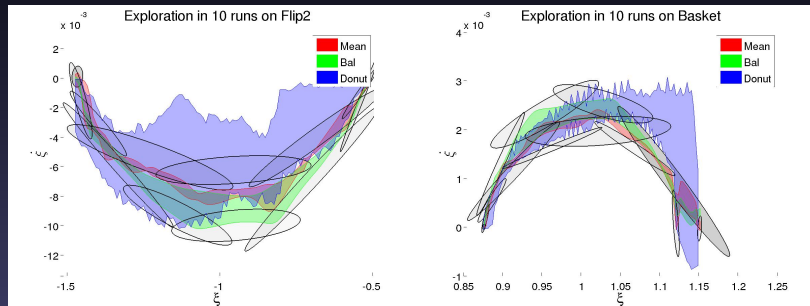
Balanced Mean

- ▶ Partition demonstrations into two groups
 - ▶ Too fast
 - ▶ Too slow
- ▶ Produce mean of each model
- ▶ Average them
- ▶ New trials are assigned to one or the other class (human)

Obvious issues:

- ▶ This approach may not scale well
- ▶ Limited to interior of demonstrations

Exploration



Future Work

- ✓ Learning tasks from failure
- ✓ No explicit reward function¹
 - Only 1 DOF
 - ▶ Motion representation
 - ▶ Donut well defined to arbitrary dimensionality
 - Gradient ascent
 - ▶ Slow
 - ▶ Local optima
 - ▶ Donut on parameters of a parameterized motion
 - ▶ Sampling
 - ▶ Motion is smooth
 - ▶ Arbitrary dimensionality . . . Curse?
 - ▶ Other Limitations (Assumptions)
 - ▶ Extra variance - Irrelevance
 - ▶ Missing variance - Demonstrator bias

¹Recent results with full reward function show smaller variance in # trials to success

Acknowledgments:

Funding provided by FP7-248258 (First-MM) and FP7-ICT-248311 (Amarsi). Florent D'Halluin and Chrisitan Daniel assisted in carrying out this research.