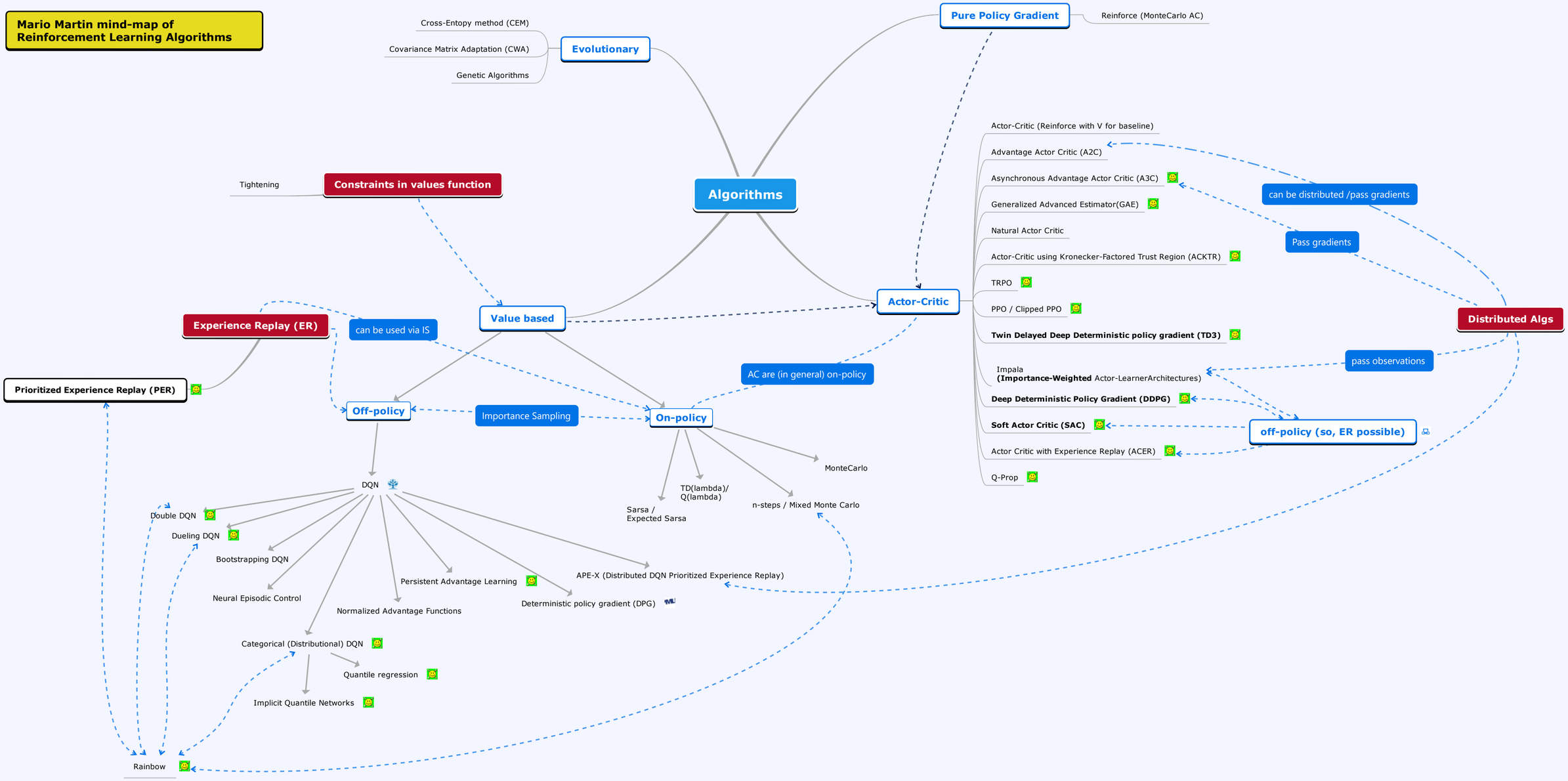
Mario Martin mind-map of Reinforcement Learning Algorithms
Algorithms
Evolutionary
Cross-Entopy method (CEM)
Covariance Matrix Adaptation (CWA)
Genetic Algorithms
Pure Policy Gradient
Reinforce (MonteCarlo AC)
Constraints in values function
Tightening
Value based
Actor-Critic
Actor-Critic (Reinforce with V for baseline)
Advantage Actor Critic (A2C)
Asynchronous Advantage Actor Critic (A3C)
Generalized Advanced Estimator(GAE)
Natural Actor Critic
Actor-Critic using Kronecker-Factored Trust Region (ACKTR)
TRPO
PPO / Clipped PPO
Twin Delayed Deep Deterministic policy gradient (TD3)
Impala
(Importance-Weighted Actor-LearnerArchitectures)
Deep Deterministic Policy Gradient (DDPG)
Soft Actor Critic (SAC)
Actor Critic with Experience Replay (ACER)
Q-Prop
can be distributed /pass gradients
Pass gradients
Distributed Algs
pass observations
off-policy (so, ER possible)
Experience Replay (ER)
can be used via IS
Prioritized Experience Replay (PER)
AC are (in general) on-policy
Off-policy
Importance Sampling
On-policy
MonteCarlo
TD(lambda)/
Q(lambda)
Sarsa /
Expected Sarsa
n-steps / Mixed Monte Carlo
DQN
Double DQN
Dueling DQN
Bootstrapping DQN
Neural Episodic Control
Normalized Advantage Functions
Persistent Advantage Learning
Deterministic policy gradient (DPG)
APE-X (Distributed DQN Prioritized Experience Replay)
Categorical (Distributional) DQN
Quantile regression
Implicit Quantile Networks
Rainbow