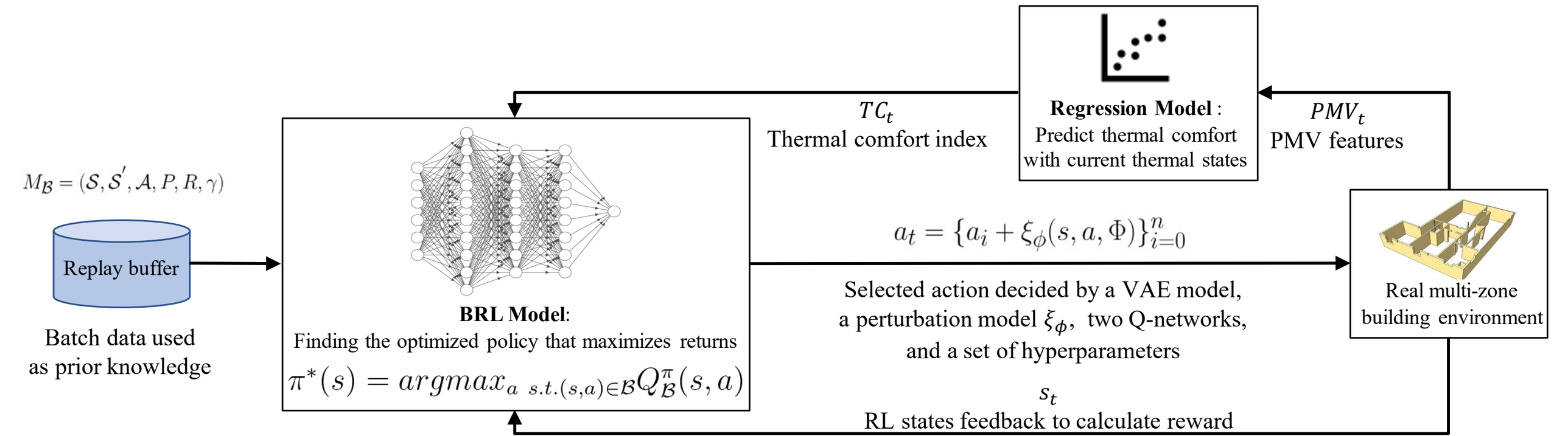
$M_{\mathcal{B}} = (\mathcal{S}, \mathcal{S}', \mathcal{A}, P, R, \gamma)$
Replay buffer
Batch data used as prior knowledge
BRL Model:
Finding the optimized policy that maximizes returns
$\pi^*(s) = argmax_a \ s.t.(s,a) \in \mathcal{B} Q^{\pi}_{\mathcal{B}}(s,a)$
TC_t
Thermal comfort index
Regression Model :
Predict thermal comfort with current thermal states
PMV_t
PMV features
$a_t = \{a_i + \xi_\phi(s, a, \Phi)\}_{i=0}^{n}$
Selected action decided by a VAE model, a perturbation model ξ_ϕ, two Q-networks, and a set of hyperparameters
Real multi-zone building environment
s_t
RL states feedback to calculate reward