**Motivation**
Most robotic systems have a high number of degree-of-freedom, while most tasks in robotics are intrinsically low dimensional, for example, - grasping - walking - human arm movements
Hence, we want to exploit the low dimensional nature of the tasks and, furthermore, use prior structural knowledge to learn movements in an efficient and meaningful way.

**Main Idea**

**Variational Inference**
By following a Variational Bayes approach on a lower bound we can derive the following formula for the estimation of the approximated q-distributions.

\[
\log q_j(\theta_j) = \text{const} + \int \prod_{i\neq j} q_i(\theta_i) \prod_{i=1}^{T} \pi(a_i, \theta_i | s_i) \frac{p(r=1|\tau)}{R} d\tau d\theta_{-j}. \tag{1}
\]

**Group Factor Policy Search**
The model equation for the actions (of a robot) is
\[
a_t^{(m)} = (W^{(m)}z_t + M^{(m)} + E^{(m)}) \Phi(s_t, t). \tag{2}
\]

The probabilistic policy is given by
\[
\pi(a_t \theta, s_t) = \prod_{m=1}^{M} \mathcal{N}(a_t^{(m)} | W^{(m)}z_t + M^{(m)} \Phi, \text{Tr}(\Phi \Phi^T)). \tag{3}
\]

The prior of the transformation matrix incorporating structural information and sparsity is
\[
p(W|\alpha) = \prod_{m=1}^{M} \prod_{k=1}^{K} \prod_{d=1}^{d_m} \mathcal{N}(w_{d,k}^{(m)} | 0, \alpha_{m,k}^{-1}). \tag{4}
\]

The distributions of other parameters are either normal or gamma distributions with
\[
\mathcal{M} \sim \mathcal{N}(M_{old}, \sigma^2 I), \quad z_t \sim \mathcal{N}(0, \text{Tr}(\Phi \Phi^T) I),
\]
\[
\alpha_{m,k} \sim \mathcal{G}(a^2, b^2), \quad \tau_m \sim \mathcal{G}(a^2, b^2).
\]

**Conclusion**
We derived a novel reinforcement learning algorithm that integrates dimensionality reduction and policy search by using sparse structural prior distributions. Resulting factors of the latent space model specific behaviour of (joint) groups in robot arms.

**Algorithm**

**Input**: Reward function \(R(\cdot)\) and choose number of latent dimension \(n\). Set fixed hyper-parameters \(a^2, b^2, a^2, b^2, \sigma^2\) and define groupings.

while reward not converged do
for \(t=1:T\) do
\[
a_t = W, Z\Phi + M \Phi + E \Phi
\]
with \(Z \sim \mathcal{N}(0, I)\) and \(E \sim \mathcal{N}(0, \tau)\), where \(\tau = \tau_m I\)
Observe and store reward \(R(\tau)\)
End while
End for
Initialization of q-distribution
while not converged do
Update \(q(M)\)
Update \(q(W)\)
Update \(q(\alpha)\)
Update \(q(\tau)\)
End while
End while

Result: Linear weights \(M\) for the feature vector \(\Phi\), representing the final policy. The columns of \(W\) represents the factors of the latent space.

**Contact Information**
- Web: www.kevin-luck.net
- Email: ksluck@asu.edu
- ASU Interactive Robotics Lab
- Aalto Intelligent Robotics Group
- Data: lab.engineering.asu.edu/interactive-robotics

**References**