Algorithm : Human-Robot Cross-training

1. Initialize $R(s, a)$ and $T(s, a, s')$ from prior knowledge
2. Calculate initial policy π
3. **while**(number of iterations < *MAX*)
4.   Call Forward-phase(π)
5.   Update $T(s, a, s')$ from observed sequence $s_1, a_1, s_2, \ldots, s_{M-1}, a_{M-1}, s_M$
6.   Call Rotation-phase()
7.   Update $R(s_i, a_i)$ for observed sequence $s_1, a_1, s_2, a_2, \ldots, s_N, a_N$
8.   Calculate new policy π
9. **end while**