

Para entrenar el manager, se emplea la recompensa obtenida, y emplea la similitud coseno entre la dirección en la que se movió el worker y la compara con el objetivo establecido, empleando la similitud coseno como función de

pérdida. Esta pérdida incentiva al Manager a emitir objetivos que maximicen el progreso hacia estados ventajosos.

El vector de objetivos se envía al worker sin propagar gradientes, esto garantiza que los objetivos mantengan un significado semántico independiente, en lugar de ser simples variables latentes optimizadas de manera conjunta.

En el caso del worker, también se emplea una red LSTM para mantener un estado interno y poder capturar información relevante, pero en este caso, el worker recibe tanto la percepción del entorno como el objetivo del manager. El worker emplea esta información para predecir la acción que debe realizar para alcanzar el objetivo. La acción se predice en el espacio de acciones, y se emplea

## **Definición**

## **Estado del arte**

## **Evaluación práctica**

## **Conclusiones**

## **References**

Science, T. D. 2024. Hierarchical Reinforcement Learning: Feudal Networks.