La idea general es:
en vez de darle coordenadas "crudas" a la DQN, le doy información ya interpretada, mucho más alineada con la tarea que tiene que resolver. Eso hace que aprenda con menos datos y con políticas más estables.

Distancia:

Si yo ya le doy la **distancia euclídea al objeto más cercano** y **a la zona de entrega**, le estoy proporcionando un indicador directo de progreso: cada vez que se mueve bien, esa distancia baja; si se equivoca, sube.

En lugar de obligar a la red a inferir eso a partir de coordenadas, le doy directamente cuán lejos está. Así tiene una señal continua de progreso y le resulta mucho más fácil aprender una política buena.

Direccion:

Saber la distancia no es suficiente, también importa **en qué dirección** hay que moverse.

El vector dirección normalizado al objetivo (dx, dy) transforma un problema complejo de navegación en algo muy simple:

- si el objetivo está "hacia arriba y a la derecha", las acciones que mueven arriba/derecha tienden a ser buenas.

De esta forma, la red no tiene que aprender geometría desde cero, solo tiene que asociar direcciones con acciones: si el objetivo está hacia la derecha, moverse a la derecha suele ser bueno.

Ubicación de estanterías:

Si el agente conoce el centro y el tamaño de las estanterías, puede **aprender patrones del tipo "cerca de esta zona suele haber colisiones"** y planificar rutas que rodeen las estanterías.

Esto le ayuda a generalizar: no memoriza trayectorias, sino que entiende la geometría del almacén. En vez de aprender a base de golpes, tiene información explícita de la geometría del almacén y puede planificar rutas más seguras.

➔ Cuando hablamos contigo, Lucía, nos sugeriste añadir explícitamente la información de las estanterías al vector de observación. Antes, el agente no veía las estanterías, solo recibía muchas recompensas negativas cuando se acercaba a esa zona por las colisiones. El resultado es que había 'memorizado' que esa región del mapa era mala y simplemente no se acercaba, aunque el objeto estuviera allí. Esto se veía muy claro en el entorno 3, donde la política se quedaba bloqueada lejos del estante y no generalizaba cuando cambiábamos un poco la configuración.

➔ Al incluir la geometría de las estanterías en la observación, el agente deja de asociar esas coordenadas concretas con "zona prohibida" y pasa a entender que hay un obstáculo que puede rodear. Es decir, ya no evita toda el área, sino que aprende a acercarse al estante sin chocar, y por eso puede recoger objetos en distintas posiciones de la estantería y generalizar mejor a otros entornos."

Segunda:

El enunciado nos daba un estado de **11 variables**: posición del agente, posición de los objetos y tres flags (lleva objeto, colisión y entrega).
Sobre esa base hemos pasado a **31 observaciones**, añadiendo información que es mucho más útil para la DQN.

La 12 es la distancia al objeto más cercano.
Así el agente sabe en todo momento *cuán lejos* está del siguiente objeto que tiene que recoger, sin tener que deducirlo solo a partir de coordenadas. (DISTANCIA EUCLIDEA)

Las 14 y 15 codifican la dirección normalizada hacia el objetivo (X, Y).

Esto responde literalmente a "¿hacia dónde tengo que moverme ahora?". Facilita mucho aprender una política porque la red solo tiene que alinear acciones con esa dirección.

De la 16 a la 31 metemos la información de las estanterías: coordenada X, coordenada Y, ancho y altura de cada una.
Con esto el agente ya no ve solo 'zonas con castigo', sino la geometría real del almacén. Puede aprender a rodear las estanterías, evitar colisiones y seguir acercándose al objeto aunque esté pegado a un obstáculo.

Tercera diapositiva:

En resumen, pasamos de darle al agente solo coordenadas a darle **información útil**.
Por un lado, le damos *información directa del objetivo*: sabe a qué distancia está del objeto, a qué distancia está de la zona de entrega y en qué dirección tiene que moverse ahora mismo.
Por otro lado, le damos *información del entorno*: la geometría de las estanterías, es decir, dónde están los obstáculos.
Con estas dos piezas el estado es mucho más informativo y la DQN puede aprender **políticas más ricas y que generalizan mejor**, en lugar de limitarse a memorizar movimientos concretos