Desarrollo de un Algoritmo de Navegación Autónoma Basado en Técnicas de Aprendizaje por Refuerzo Usando Información Visual

Anteproyecto de Grado

Daniel Felipe Aponte Vargas
Erika Dayanna Martínez Méndez

Director
Juan Manuel Calderon Chavez

Universidad Santo Tomás
Facultad de Ingeniería Electrónica
Bogotá D.C.
2019

Contenido

1 Planteamiento de Problema **2**

2 Justificacion **4**

3 Antecedentes **5**

4 Objetivos **8**
4.1 Objetivo General 8
4.2 Objetivos Específicos 8

5 Presupuesto **9**

6 Cronograma **11**

7 Diseño Metodológico **12**
7.1 Revisión Bibliográfica 12
7.2 Elaboración y Redacción del Anteproyecto 12
7.2.1 Definición y planteamiento del problema 12
7.2.2 Redacción de la justificación 12
7.2.3 Redacción de antecedentes 12
7.2.4 Definición de los objetivos del proyecto 12
7.2.5 Propuesta de la administración del proyecto 13
7.3 Desarrollo Práctico 13
7.3.1 Diseño del algoritmo 13
7.3.2 Simulación del algoritmo diseñado 13
7.3.3 Ejecución y pruebas de simulación 13
7.4 Verificación de Resultados 13

1. Planteamiento de Problema

En el campo de la robótica, existen agentes móviles programados, que suplen las necesidades básicas requeridas como asistentes, sin embargo, estos se ven limitados ante cambios externos. Supóngase un agente para movilizarse en un ambiente cerrado, las decisiones que toma el robot programado están previamente descritas y basadas en un mapeo del escenario, al momento de presentarse una reestructuración o alguna modificación sencilla dentro del escenario, deberá volverse a programar el asistente considerando las nuevas modificaciones, convirtiéndose en una tarea tediosa para el dueño del asistente y para el programador, ya que debe diseñar por sí mismo las rutas alternativas. En el caso de los robots de asistencia personal, se presentan dificultades cuando el dueño quiere movilizarse hacia un espacio que el agente no conoce previamente lo que conduce a hacer uso de él únicamente en su hogar o aquellos lugares reconocidos por el agente. Los robots están provistos de una amplia gama de sensores que le proporcionan la información necesaria acerca de su entorno y su ubicación espacial, esta información puede ser analizada y procesada para definir las acciones que serán ejecutadas por el agente, para ello existen técnicas como el aprendizaje por refuerzo (RL), una técnica que permite enseñarle a un agente a tomar decisiones a partir de la información obtenida, la cual es traducida en estados. La ejecución de estas acciones generan una recompensa que busca ser maximizada a largo plazo, sin embargo, esta se caracteriza por ser escasa, ruidosa y/o demorada, por otro lado, existe el dilema, exploración vs explotación, en el cual se debe evaluar si se quiere que el agente explore nuevas zonas del escenario donde podría encontrar mejores recompensas o por el contrario explotar aquellas opciones que le ofrecen recompensas positivas, esto teniendo en cuenta que el agente se desenvuelve en un ambiente dinámico. Luego de mostrar este conjunto de falencias, se formula la siguiente pregunta de investigación: ¿Cómo enseñarle a un robot móvil a navegar de forma autónoma por medio de aprendizaje

por refuerzo en espacios desconocidos a partir de información visual?

2. Justificacion

Las técnicas de aprendizaje de máquina aplicadas a la robótica son de especial importancia considerando que, el secreto utilizado por varias empresas a nivel mundial para expandirse está resumido en dos áreas de investigación, la inteligencia artificial (IA) y el aprendizaje automatizado. De estas ramas de investigación, buscan pronósticos confiables, sistemas que se adapten a sus intereses y aprendan de las necesidades de sus consumidores. El aprendizaje por refuerzo, al ser un derivado de IA se convierte en un punto destacable e importante para el desarrollo industrial de un país. No es desconocido el hecho de que el emprendimiento en países en desarrollo como Colombia, ha tomado bastante fuerza desde hace algunos años, lo cual se convierte en un desafío para ingenieros e investigadores de estos campos. El uso de las técnicas de aprendizaje es útil cuando, es difícil planificar las acciones correctas, por ejemplo, a la hora de planificar rutas de navegación en entornos desconocidos o cuando las circunstancias del terreno pueden cambiar, como terrenos en áreas de desastres donde es muy importante que el sistema aparte de ser autónomo tenga la capacidad de aprender y logre adaptarse. Así, los robots autónomos y con gran capacidad de adaptabilidad son altamente requeridos para tareas de alto riesgo y/o muy poco deseadas para los seres humanos, tales como búsqueda y rescate de personas en zonas de desastre, cuidado de personas de la tercera edad, inspección y vigilancia en zonas de contaminación. Este proyecto le apunta a las tareas de asistencia doméstica, considerando que, actualmente en Bogotá hay aproximadamente 955 mil personas mayores de 60 años y en 2020 serán más de 1.153.000, de las cuales el 10.9% viven solas, un dato preocupante al observar que un 18% del total de personas mayores tienen al menos una limitación permanente y un 66% padece de una enfermedad crónica, como problemas cardiovasculares, hipertensión, lesiones en los huesos, problemas digestivos y gástricos [1].

3. Antecedentes

Debido al enfoque de este proyecto de grado se revisaron diferentes trabajos con temas relacionados o afines al aprendizaje de máquina usando aprendizaje por refuerzo. En [2] publicado en 1995, proponen el uso de un algoritmo de aprendizaje para el juego de mesa "Backgammon", utilizando una red neuronal como parte de la estructura y haciendo uso del método de TD(Temporal Difference) learning, siendo este una clase de aprendizaje por refuerzo model-free. Aquí mencionan las dificultades que se presentan al implementar la técnica de aprendizaje por refuerzo en casos de la vida real, como lo es la asignación temporal de las recompensas y la limitación del aprendizaje por medio de tablas de búsqueda y funciones de evaluación lineales. Debido a recientes desarrollos se han logrado funciones de aproximación no-lineales como lo son, los árboles de decisión, las funciones base localizadas y los perceptrones multicapa. Acerca del TD-Gammon, tiene una estructura a partir de una arquitectura MLP (Multi layer perceptron) y su algoritmo de aprendizaje consiste en aplicar una fórmula de cambio de pesos para la red a partir de la tasa de aprendizaje definida como $\alpha(alpha)$ y los valores de salida de la red, esto con el fin de optimizar el proceso de selección maximizando la recompensa. El autor concluye exponiendo la calidad de los resultados arrojados por el TD learning y proponiendo el uso de este, en desarrollos de robótica y mercados financieros, resaltando la ventaja que presenta al aplicar una estrategia de control/predicción combinada. A raíz del éxito de este algoritmo, se abrieron campos de investigación en torno al aprendizaje por refuerzo, algunos avances recientes del tema pueden ser evidenciados en [3], proyecto presentado en 2018 como Trabajo de grado de la Facultad de Ingeniería Electrónica de la Universidad Santo Tomás, el autor propone y simula un algoritmo para generar comportamiento de enjambre en robots de esta manera, genera dos posibles soluciones basadas en Q-learning, en una de ellas los estados del agente son generados en función

de distancias entre sus vecinos cercanos, la otra solución que plantea el autor está relacionada con un radio de repulsión y otro de atracción, los cuales determinan la cantidad de individuos o vecinos que pueden estar cerca de los cuadrantes locales del agente. Otra muestra reciente está propuesta en [4], aquí se aborda la navegación autónoma para agentes móviles como un problema a tratar con aprendizaje por refuerzo, con la diferencia de que se asignan unas tareas auxiliares como son la predicción de profundidad y la clasificación de cierre de lazo, las cuales mejoran el tratamiento de datos y basados en la entrada de los datos proporcionados por los diferentes sensores, el agente puede navegar autónomamente. La arquitectura de las redes neuronales está basada en LSTM (Long-Short Term Memory) que poseen una estructura modular y son entrenadas para múltiples tareas, dentro de las tareas, la tarea principal (navegación) es ejecutada, actualizada y supervisada mediante aprendizaje por refuerzo, mientras que las tareas auxiliares son entrenadas con self-supervised learning. Abordando ahora los resultados presentados por Deep Reinforcement Learning (DRL), en [5] los autores buscan analizar el desempeño de una red DQN (Deep Q-Network), que implementa un algoritmo de aprendizaje por refuerzo, en un entorno gráfico virtual, Atari 2600. El objetivo del agente es interactuar con el juego mediante la toma de decisiones que le proporcionarán recompensas, las cuales indican el desempeño actual del agente dentro del juego. Para esto, la entrada de la red son pixeles del ambiente (observada por el agente), pasando por una red profunda, la cual obtiene la información necesaria para que posteriormente la red Q seleccione la acción adecuada. En el proceso de aprendizaje se utiliza la técnica Experience Replay donde se almacenan, en una memoria, N cantidad de experiencias, que serán evaluadas aleatoriamente por el algoritmo para establecer la tasa de aprendizaje y actualizar los pesos de la red. En la evaluación de desempeño del algoritmo evidenciaron que esta sobrepasó con creces la capacidad de un humano experto y algoritmos previos, demostrando así que el uso combinado Deep-learning y Q-learning, proporciona gran utilidad y eficiencia. Hasta este punto se mostró una aplicación específica para el DRL, donde su desempeño es favorable y eficiente, cumpliendo las expectativas. Sin embargo, [6] aborda diferentes técnicas de RL e IA donde se pueden aplicar el DRL, considerando que DRL presenta limitaciones cuando debe desenvolverse en tareas generales, lo que se busca es la solución a problemas complejos, requiriendo un desarrollo general. Para Hierarchical RL, por ejemplo, se implementa DRL para desarrollar multi-timestep "actions" asignándole las acciones primitivas de sub-políticas que permiten conformar

una política, logrando descubrir y alcanzar metas, los cuales son estados específicos del entorno. En Multi-agent RL, el enfoque de DRL ha sido permitir la comunicación entre agentes, con el fin de cooperar entre ellos a través de la identificación espacial. Otro reto propuesto está en Model-base RL, donde se pretende llegar a la navegación de un agente en un entorno desconocido sin la interacción con el entorno real, el agente aprenderá del entorno utilizando uno simulado y para este aprendizaje los algoritmos de DRL aprenderán usando información captada en pixeles. Por último, este trabajo se desarrollará gracias a la colaboración y el apoyo del Grupo de Investigación y Desarrollo en Robótica de la Universidad Santo Tomás (GED). El grupo de investigación ha desarrollado y publicado trabajos relacionados con Navegación Autónoma y Visión Artificial, temas afines al enfoque de este proyecto. En [7] publicado en 2013, se propone un algoritmo de detección de bordes basado en lógica difusa como parte de un sistema de visión local que ayudaría a reconocer marcas visibles en un juego de campo diseñado para la liga humanoide de robots de RoboCup, además en [8] se propone un algoritmo capaz de detectar el camino por donde transita un UGV (Unmanned Ground Vehicle). Por otro lado, en la rama de navegación autónoma, el grupo GED posee publicaciones como [9], [10], donde se enfocan en la navegación de un sistema multiagente a través de la teoría de enjambre donde los robots son los agentes del enjambre, se establecen fuerzas de repulsión y atracción las cuales son usadas para evitar obstáculos, mantener el grupo compacto y navegar objetivamente, el objetivo de lograr esto va orientado hacia operaciones de rescate de víctimas, para lo cual establecieron un algoritmo de consenso en el que cada agente por separado identifica una posible víctima, una vez estén de acuerdo los agentes de que determinada víctima existe, una parte los agentes se separa para formar un sub-enjambre que se encargará de adquirir información de la posible víctima y así facilitar la búsqueda y rescate, como también en [11], se propone la navegación en espacios desconocidos para un agente móvil, en el cual, se usan sensores propios del robot en vez de sensores de localización, esto con el fin de implementar la navegación y la evasión de obstáculos.

4. Objetivos

4.1 Objetivo General

Diseñar un algoritmo de navegación autónoma para un agente móvil, basado en información visual mediante el uso de técnicas de aprendizaje por refuerzo.

4.2 Objetivos Específicos

- Definir las variables que darán origen a los estados del sistema considerando que el robot usa información visual para su navegación.
- Diseñar un sistema de estados basado en redes neuronales que reemplace al clásico sistema basado en la matriz Q.
- Establecer políticas de recompensa para generar el proceso de aprendizaje del agente.
- Implementar el sistema propuesto en un ambiente virtual para la simulación de robots reales (VREP).
- Evaluar el desempeño del algoritmo diseñado, realizando pruebas en diferentes ambientes de simulación, e implementado de acuerdo a la arquitectura establecida para la generación de los estados.

5. Presupuesto

Recursos Técnicos			
Recursos	Cantidad	Fuente	Costo Total(COP)
Equipo de Cómputo	2	Recurso Propio	$ 5.000.000
Material Bibliográfico		Recurso Institucional	$ 1.000.000
Software	Fuente		Costo Total(COP)
Licencia de MATLAB	Recurso Institucional		$ 1.000.000
Licencia de V-REP	Open Source		0
Total			$ 7.000.000

Recursos Bibliográficos			
Fuente	Cantidad	Costo por Artículo(COP)	Costo Total(COP)
IEEE	6	$ 80.000	$480.000
SPRINGER	2	$ 97.000	$194.000
ARXIV	6	Free Access	
SCIELO	1	Free Access	
MDPI Open Access	1	Free Access	
Total			$ 694.000

Recursos Humanos					
Recurso		Tiempo (Horas)	Fuente	Costo por Hora(COP)	Costo Total(COP)
Director	Juan Manuel Calderon Chavez	70	Recurso Institucional	$ 60.000	$ 4.200.000
Estudiante 1	Erika Dayanna Martínez Méndez	240	Recurso Propio	$ 10.000	$ 2.400.000
Estudiante 2	Daniel Felipe Aponte Vargas	240	Recurso Propio	$ 10.000	$ 2.400.000
Total					$ 9.000.000

6. Cronograma

7. Diseño Metodológico

7.1 Revisión Bibliográfica

Se realiza una búsqueda, estudio y clasificación de diferentes documentos (usados como material bibliográfico) relacionados, con temas afines al proyecto contextualizando el problema planteado, verificando y analizando avances previos.

7.2 Elaboración y Redacción del Anteproyecto

7.2.1 Definición y planteamiento del problema

Se expone el asunto o cuestión que se tiene como objeto aclarar, exponiendo el contexto global y local y finalizando con la pregunta de investigación.

7.2.2 Redacción de la justificación

Se plantean las razones de por qué se realiza este trabajo.

7.2.3 Redacción de antecedentes

Se presentan trabajos en temas afines que permitan contextualizar el proyecto.

7.2.4 Definición de los objetivos del proyecto

Se determina un objetivo general, que es la meta y el propósito del proyecto y, y se trazan los objetivos específicos a seguir durante el proyecto, que contribuyen a la realización del objetivo general.

7.2.5 Propuesta de la administración del proyecto

Se propone el cronograma de actividades, en donde se ordena y se estima la duración aproximada en el tiempo para la realización de las actividades propuestas, y se establece el presupuesto, en donde se estima de la cantidad de dinero que se requiere para llevar a cabo el proyecto.

7.3 Desarrollo Práctico

Paralelo a la etapa anterior, en esta se construye el algoritmo y se realizan pruebas y experimentos de funcionamiento en un entorno simulado. Se subdivide en tareas específicas, que son:

7.3.1 Diseño del algoritmo

Esta tarea tiene como objetivo determinar las entradas, las matrices de los estados, acciones y recompensas del algoritmo q-learning para el aprendizaje del agente, y se procede a realizar la primera versión del código.

7.3.2 Simulación del algoritmo diseñado

Con base al algoritmo diseñado se implementa la simulación en el agente utilizando software como Matlab o V–REP.

7.3.3 Ejecución y pruebas de simulación

Se simula el algoritmo para varios agentes, realizando diferentes pruebas con diferentes números de robots y se observa el proceso de aprendizaje con indicadores de escalabilidad.

7.4 Verificación de Resultados

En esta etapa se analizan los resultados obtenidos de las diferentes simulaciones realizadas y según el proceso realizado en la etapa anterior se realiza el respectivo informe.

Bibliografía

[1] Concejodebogota.gov.co, "Concejo de bogotá d.c. - las condiciones de vida del adulto mayor en bogotá siguen siendo preocupantes," 2018. [Online]. Available: `http://concejodebogota.gov.co/las-condiciones-de-vida-del-adulto-mayor-en-bogota-siguen-siendo/cbogota/2018-04-16/154321.php` (visited on 05/04/2019).