Tabla de Contenido

1. Introducción 1
  1.1. Motivación . . . 1
  1.2. Contribuciones de esta tesis . . . 2

2. Preliminares 4
  2.1. Problemas de ruteo de vehículos . . . 4
    2.1.1. Problema clásico . . . 4
    2.1.2. Problema de ruteo de vehículos dinámico . . . 6
    2.1.3. Midiendo el dinamismo . . . 8
    2.1.4. Tipos de soluciones . . . 8
  2.2. Aprendizaje reforzado . . . 9
    2.2.1. Aprendizaje reforzado en problemas de optimización combinatorial . 10
  2.3. Casos de uso en Entel Ocean . . . 11

3. Modelos y formulaciones 13
  3.1. Modelo para el problema de ruteo de vehículos con ventanas de tiempo . . . 13
  3.2. Formulación de programación entera-mixta del VRPTW . . . 15
  3.3. Descripción del problema de ruteo de vehículos dinámico . . . 16

4. El problema desde una perspectiva del aprendizaje reforzado 19
  4.1. Procesos de decisión de Markov: definiciones y propiedades . . . 19
  4.2. Aprendizaje reforzado profundo . . . 22
    4.2.1. Redes neuronales profundas . . . 22
  4.3. Algoritmo Q-learning para el DVRP . . . 23
    4.3.1. Proceso de decisión de Markov . . . 23
    4.3.2. Algoritmo vainilla Deep Q-learning . . . 26
    4.3.3. Una serie de mejoras al algoritmo DQN . . . 27
    4.3.4. Algoritmo Double DQN con experiencias priorizadas . . . 29
  4.4. Algoritmo tipo *Policy Gradients* para el DVRP . . . 30
    4.4.1. Proceso de decisión de Markov . . . 30
    4.4.2. Algoritmos tipo *Policy Gradients* . . . 32
    4.4.3. Algoritmo Actor-Critic . . . 34
    4.4.4. Sub-acciones en cada época de decisión . . . 36
  4.5. Un algoritmo que incorpora información estocástica . . . 37
    4.5.1. Proceso de decisión de Markov para el sub problema estático . . . 38
    4.5.2. Descripción del algoritmo . . . 40

4.5.3. Entrenamiento 41
4.5.4. Ejemplo 42

5. Arquitectura de la red neuronal 44
5.1. Red de codificación 44
5.2. Red de decodificación 45
5.3. Red critic 45
5.4. Diagrama de alto nivel 46

6. Resultados computacionales 47
6.1. Generación de datos y entrenamientos 47
6.2. Entrenamiento Double Q-learning con experiencias priorizadas 49
6.2.1. Configuración 49
6.2.2. Entrenamientos 49
6.3. Entrenamiento algoritmo Actor-Critic 51
6.3.1. Configuración 51
6.3.2. Entrenamientos 51
6.4. Entrenamiento algoritmo estocástico con búsqueda 54
6.4.1. Configuración 54
6.4.2. Entrenamientos 54
6.5. Resultados experimentales 56
6.5.1. Algoritmo DDQN con experiencias priorizadas 57
6.5.2. Algoritmo Actor-Critic 59
6.5.3. Algoritmo estocástico con búsqueda 65
6.5.4. Otros experimentos 66
6.6. Comparación con estrategia reoptimizadora 69
6.6.1. Descripción de la estrategia reoptimizadora 69
6.6.2. Estrategias de espera 70
6.6.3. Comparaciones en diversos escenarios 71
6.6.4. Análisis 76
6.6.5. Visualización de las soluciones en el plano y tiempos de ejecución . . 78
6.7. Comparación con soluciones óptimas 80

Conclusiones 83

Bibliografía 85

A. Descripción detallada de redes neuronales usadas 89
A.1. Funciones de activación usadas 89
A.2. Redes neuronales preliminares 89
A.2.1. Atención y multi atención aplicada a grafos 89
A.2.2. Capa FeedForward 90
A.2.3. Normalización por batch 90
A.2.4. Atención con “glimpses” 91
A.3. Red codificadora 91
A.4. Red decodificadora 92
A.4.1. Red DQN 93
A.4.2. Decodificador con atención 94

A.5. Red Critic 94