UNIVERSITÀ
DEGLI STUDI
FIRENZE

Scuola di Scienze Matematiche, Fisiche e Naturali
Corso di Laurea in Informatica

TITOLO IN ITALIANO

TITLE IN ENGLISH

TERROSI FRANCESCO

BONDAVALLI ANDREA
STRIGINI LORENZO

Anno Accademico 2018-2019

EXECUTIVE SUMMARY

I veicoli autonomi sono sistemi cyber-fisici critici, complessi sotto molteplici aspetti: dalla tecnologia necessaria all'acqusizione di dati esterni come radar, lidar, GPS... all'implementazione del software che si occupera' della guida vera e propria. Con l'enorme progresso avuto nel campo del machine learning nell'ultimo ventennio, la prospettiva di macchine capaci di guidare senza alcuna interazione con l'uomo e' sempre piu' vicina.
Il sistema di controllo del veicolo puo' essere visto in maniera semplificata come formato da una rete neurale, che determina l'azione da eseguire (quanto accelerare/decelerare o l'angolo di sterzata) sulla base dei dati ricevuti dai sensori. Dal momento che nelle predizioni effettuate da una rete vi e' insito un errore (i.e. risulta impossibile avere un'accuratezza del 100% sui risultati prodotti) e' di fondamentale importanza avere un *safety monitor*, il cui compito e' quello di controllare e sanificare gli output dell'intelligenza artificiale.
In questo lavoro abbiamo studiato come varia il rapporto fra un safety-monitor relativamente semplice e una rete neurale addestrata per la guida autonoma, andando a definire dei semplici requisiti di safety e osservando come un continuo training della rete neurale vada a impattare (o meno) sull'utilita' del monitor in questione.
Per poter svolgere questo lavoro sono stati utilizzati molti software open-source: grazie al simulatore CARLA e' stato possibile avere una rappresentazione realistica delle leggi della strada e della fisica dei veicoli. Le reti neurali che sono poi state prese in considerazione sono state addestrate con algoritmi di *reinforcement learning* e *imitation learning*, considerati fra i piu' promettenti in questo campo. CARLA inoltre permette di avere una simulazione realistica dei sensori utilizzati sui veicoli autonomi: questo ha permesso di costruire un semplice (ma efficace) safety monitor che effettua dei controlli di sicurezza sulla base dei dati ricevuti dal lidar e sulla velocita' e direzione del veicolo autonomo; sono stati infine condotti gli esperimenti per studiare l'interazione fra questi due componenti.

Autonomous vehicles are one of the trending topic of the decade. With the new hardware and software technologies such as sophisticated sensors and AIs trained to drive with machine learning techniques, self-driving cars could be seen on the road alonside human drivers very soon. On one hand, neural networks specifically trained to drive seems the only way to accomplish this task, thanks to the networks' ability to perform better in unknown situation, **[i.e. scenarios that were not considere]**, on the other hand, their use in safety-critical systems represent a huge risk for safety.

INDICE

ELENCO DELLE TABELLE

ELENCO DELLE FIGURE

1

INTRODUZIONE

Sistemi informatici ormai ovunque (Cosa sono, esempi)

1.1 CYBER-PHYSICAL SYSTEMS OF SYSTEMS

- Cosa sono i sistemi cyber-fisici
- safety e dependability
- safety-assessment classicamente?

2

STATO DELL'ARTE - O INTRODUZIONE?

2.1 INTRODUZIONE ALLE SELF-DRIVING CARS QUI?

Automotive technology has been one of the hottest topic of the decade. With the continuously growing hardware and software technologies, completely autonomous vehicles don't seem to be unfeasible anymore: multiple sensors can retrieve high quality data from the surrounding environment and new Artificial Intelligence techniques (i.e. neural networks) are capable of working with this data in a manner that outclasses classical statistical models. However, the use of AI to drive cars, requires more focus on safety and the way to assess it.
Autonomous vehicles can be classified in five levels of autonomy:

Level 0 - No Automation

→ The human driver performs all the tasks, such as steering, accelerating, braking... Cars with *forward collision warning systems* and *lane keep assist* fall in this category

Level 1 - Driver Assistance

→ The vehicle assists the human driver in relatively simple tasks (e.g. adaptive cruise control)

Level 2 - Partial Automation

→ At this level the vehicle is capable of performing more complex tasks (e.g. *Parking assistance, Tesla's Autopilot*) but the driver still must be able to correct unexpected behaviours of the car.

Level 3 - Conditional Automation

→ Level three automation means that the vehicle is now in full control under specific conditions (e.g. riding on a highway). However the human driver still must be able to intervene when requested by the system to do so

Level 4 - High Automation

→ At this level the vehicle can drive completely autonomously, *without* any kind of human interaction. However, they are subjected to specific conditions and assumptions that, when not fulfilled, may result in unexpected behaviours (or catastrophic failures!)

Level 5 - Complete Automation

→ True driverless cars. Human intervention is not needed at all and the car can operate in every condition and environment.

********** Fare esempi positivi sui risultati raggiunti oggi?

With today technologies, Level 4 vehicles are not too far from being seen on public roads ([4], [5]) and the scientific community is working hard to make Level 5 cars real.

****

FARE QUA DISCORSO SU SCOPO DELLO STUDIO? RETE-NEURALE -> MONITOR?

****

In this study we are interested at assessing Level 5 cars' safety

****

DA TRADURRE **********

2.2 SELF-DRIVING CARS ARCHITECTURE

Le macchine a guida autonoma rientrano nella categoria dei sistemi informatici cyber-fisici critici. Anche se gia' di per se' assimilabili nella categoria dei sistemi di sistemi (dal momento che vi sono piu' constituent systems che interagiscono fra loro), l'obiettivo della comunita' scientifica e' quello di riuscire ad ottenere un sistema di sistemi risultante dall'interconnessione di piu' veicoli autonomi: ognuno di questi sistemi avra' un obiettivo differente (i.e. diverse destinazioni da raggiungere) ma per ottenerlo e' assolutamente necessaria una cooperazione fra essi. Come

diretta conseguenza, e' indispensabile che il sistema di controllo del veicolo sappia non solo obbedire al codice della strada ma anche essere in grado di riconoscere le situazioni di potenziale pericolo.
Un veicolo autonomo dev'essere in gradio di "osservare" l'ambiente circostante, questo e' reso possibile grazie ai sensori installati su di essa.

Tipicamente vengono utilizzati:

- Telecamere
  → Necessarie per catturare immagini dell'ambiente
- Radar, Lidar, Sonar
  → Utilizzati per creare una mappa dell'ambiente in cui naviga il veicolo e per percepire gli ostacoli
- GPS, sensori inerziali, odometria
  → Indispensabili per pianificare il percorso da seguire e conoscere la posizione del veicolo nell'ambiente operativo

L'architettura software di un veicolo e' composta da piu' moduli interagenti, dove l'errore di uno di questi potrebbe risultare in una minaccia per la safety del sistema.

Possiamo semplificare il modello architetturale come composto da tre moduli separati:

- Environment Mapping
  - La componente che riceve i dati direttamente dai sensori e si occupa di filtrarli (e.g. per ridurne il rumore e scartare valori poco significativi) ed aggregarli per riconoscere il perimetro dell'area circostante ed eventuali ostacoli
- Motion control
  - Questo modulo riceve in input i dati dopo che questi sono stati elaborati dal sistema di *data processing*. Dopo averli osservati, il *motion planner* decide in che direzione debba proseguire il veicolo; questo comando viene quindi inviato al *controller*, composto da un modulo di controllo longitudinale (accelerazione) e uno di controllo laterale (sterzo del veicolo), il quale va effettivamente a interagire con gli attuatori del sistema
- System Supervisor
  - Un supervisore, o monitor, e' il sistema che si occupa di rilevare guasti o fallimenti sia hardware che software. Dal punto di vista hardware, i controlli effettuati sono principalmente sui guasti a componenti hardware e sugli output del *controller* (ad esempio che appartengano al dominio del sistema). Il monitor software invece si occupa di rilevare inconsistenze fra gli output dei due moduli precedenti. Questo e' di fondamentale importanza in quanto permette di controllare che l'output del *controller* non porti il sistema in una situazione di pericolo, o peggio: ad un fallimento catastrofico che coinvolgerebbe anche vite umane

==========

IMMAGINE ARCHITETTURA SISTEMA DI CONTROLLO [SETTIMANA 1 - LEZIONE 3]

==========

Se classicamente venivano implementati in software modelli statistici noti per mappare l'ambiente (e.g. Kalman filter) e modelli fisici o di teoria del controllo per manovrare il veicolo (e.g. PID controller), grazie al progresso nel campo del *machine learning* avuto negli ultimi anni si e' iniziato a utilizzare reti neurali addestrate alla guida [1] [2].
Le reti neurali hanno dimostrato di sapere reagire meglio a situazioni sconosciute rispetto ai meccanismi classici, tuttavia richiedono tanti piu' dati quanto piu' e' complesso il compito da eseguire.
Questi modelli computazionali inoltre soffrono del cosiddetto *black box problem*: e' molto difficile riuscire perche' la rete abbia associato l'output y all'input x. Nonostante siano stati proposti alcuni framework [3] per aiutare a comprendere i meccanismi che regolano le decisioni di un'intelligenza artificiale sotto specifiche assunzioni, non si e' ancora trovata una soluzione universale al problema.

2.3 SAFETY NELL'AUTOMOTIVE

- Intro e standard

For automotive systems, safety becomes a fundamental requirement to guarantee a minimum level of risk.

2.3.1 *Driving Neural Networks*

- Perche' le neural network sono un problema per la safety | citazioni paperz

2.3.2 *Neural network - monitor problem*

Spiegare qui qual e' il problema che vogliamo risolvere

–> dalla letteratura sappiamo che nell'interazione fra 2 software la reliability growth di uno non implica una reliability growth di tutto il sistema. (Paper: Assessing Asymmetric Fault-Tolerant Software)

2.3.3 *Scopo dello studio?*

3

SYSTEM ANALYSIS METHOD

The goal of this work is to develop and to assess the feasability of an experimental method that allows to study the interaction between the AI controller and the Safety Monitor, with particulare attention to these aspectes:

- How much and in what way the benefits given from the use of a safety-monitor can vary the more the neural network learns
- How much vary the effectiveness of the same monitor when applied to two different networks
- What features of the monitor determines an improvement (or worsening) to the safety of the system
- What aspects of the neural network training have an impact on the monitor usefulness

Safety-Monitors are developed in order to detect failures undetected by the network, therefore it is desiderable that the failures covered by the monitor don't overlap with the failures detected by the network. If this will be almost certainly true when the network is in the early stage of training, the lack of scientific results in this topic raises some concerns, while in the industry the long-time usefulness of the safety monitor is not even questioned.
The problem when it comes to neural network is that, since they learn in a way that humans can loosely control, we can't predict what will be the most likely safety-hazard scenarios, nor we can efficiently test them. [assessing ultra-high dependability]

The reasoning behind this study is that it can not be guaranteed the long-time usefulness of the safety-monitor [assessing asymmetric systems EQ. 11].

coverage dei casi coperti dal monitor come cambia

3.1 TOOLS AND SOFTWARES

3.1.1 *Carla Simulator*

In order to have a realistic environment, with accurate physics simulation and data sensors, the open-source simulator Carla was used. This simulator was developed with the purpose of offering an environment where AI agents can be trained to drive.

- CARLA
- Nervana Systems - coach (Intel)
- Reti neurali su git
- Point Cloud Library per filtrare i dati

3.2 ARCHITETTURA DEL SOFTWARE (ESTRAPOLAZIONE DATI, INTERAZIONE RETE-MONITOR)

- Interazione rete-monitor
- Safety Monitor Implementation - obstacle detection
- Come vengono raccolti i dati
- Come vengono preprocessati

3.3 EXPERIMENTS METHODOLOGY

The study consists of several experiments in which we observe how the coverage of the safety-monitor (i.e. the probability of raising an alert if there really is a safety-hazard) vary with respect to a neural network in different stages of training.
The first step to perform the analysis is to define what are the metrics of interest and how these can be measured. This task is harder than it seems because it's unknown *a priori* what the probability distribution function of the hazardous scenarios will be. This means that we don't know whether the probability of observing a failure depends on the *running time* of the experiment (e.g. the more the agent drives, the more likely a failure will happen) or if it depends on other factors.

For this reason we decided to measure the length of an experiment in terms of number of failures: given the same initial scenario, two agents (one communicating with the monitor, the other relying solely on the AI) are let driving until n failures happen. We then observe the elapsed time between the start of the experiment and the moment the n_{th} failure happened.
The time to (the n_{th}) failure provides useful informations on the *efficacy* of the monitor. However, this metric itself can't be used alone to assess the potential safety gain given by safety checking the actions of the neural networks.
****
TO REVIEW:
****
This is why, for each failure, different data were recorded such as:

- Whether or not the monitor raised an alarm
- The change in speed of the car after the alarm was raised
- If a collision with a vehicle V occured, the speed and the direction of V
- Falsi positivi falsi negativi?

In this way it's possible to measure more efficiently the overlapping between the set of safety hazards covered by the AI and the one covered by the Monitor, (Other metrics: velocita' a cui andava la macchina), (direzione da cui veniva l'altro veicolo se incidente) —> per capire le situazioni in cui sbaglia di piu'

- Come vengono effettuati gli esperimenti (scenari? durata fissa? ad oltranza? fino ad un fallimento? . . .)
- Misure scelte - estrapolazione misure

4

RISULTATI DELL'ANALISI

In questa sezione elenchiamo i dati che sono stati raccolti e quali sono i risultati che abbiamo ottenuto (errori ricorrenti, grafici, rapporto monitor-rete neurale)

BIBLIOGRAFIA

[1] Jelena Kocic, Nenad Jovicic, Vujo Drndarevi, *An End-To-End Deep Neural Network for Autonomous Driving Designed for Embedded Automotive Platforms (2019)* (Cited on page 15.)

[2] Qing Rao, Jelena Frtunikj, *Deep learning for self-driving cars: chances and challenges (2018)* (Cited on page 15.)

[3] Carlos Zednik, Otto-von-Guericke-Universitat Magdeburg *Solving the Black Box Problem: A Normative Framework for Explainable Artificial Intelligence* (Cited on page 15.)

[4] *https:waymo.com* (Cited on page 12.)

[5] *https:uber.com* (Cited on page 12.)