Megerősítéses tanulás alapú mozgástervezési módszerek

Tézisfüzet

Fehér Árpád

Budapesti Műszaki és Gazdaságtudományi Egyetem
Közlekedésmérnöki és Járműmérnöki Kar
Közlekedés- és Járműirányítási Tanszék

Témavezető:
Dr. Aradi Szilárd
Közlekedés- és Járműirányítási Tanszék
Budapesti Műszaki és Gazdaságtudományi
Egyetem

2023. március 31.

1. Bevezetés

Az önvezető járművek fejlesztése jelentős előnyöket hozhat a közlekedés terén. Ezek a járművek képesek növelni a biztonságot az emberi hibák okozta balesetek csökkentésével, javítani a hatékonyságot azáltal, hogy kommunikálnak egymással és alkalmazkodnak a változó útviszonyokhoz, valamint javítani a hozzáférhetőséget az idősek és fogyatékkal élők számára [1]. Továbbá az önvezető autók képesek csökkenteni a környezetre gyakorolt káros hatásokat az útvonalak és vezetési szokások optimalizálásával. Összességében az önvezető járművek fejlesztése pozitív hatást gyakorolhat a társadalomra [2].

A magasan automatizált járműfunkciók fejlesztése nagymértékben támaszkodik a mesterséges intelligenciára és a gépi tanulásra [3]. Ezek az eszközök lehetővé teszik a jármű számára, hogy javítsa teljesítményét az adatokból való tanulás révén. A mesterséges intelligencia segítségével a járművek képesek alkalmazkodni a különböző közlekedési szituációkhoz és megbirkózni a környezet bizonytalanságával, emellett szükséges, hogy a rendszerek biztonságos és komfortos utazást biztosítsanak [4]. Fontos megjegyezni azonban, hogy a teljes önvezetés eléréséig még számos kihívást kell megoldani, mielőtt az széles körben elterjedhet.

A járművek mozgástervezése a követendő biztonságos és hatékony útvonal, valamint sebességprofil meghatározásának folyamata. A mozgástervezési folyamat jellemzően több lépésből áll [5]. Elsőként leírjuk a jármű környezetét, beleértve az akadályok és más járművek helyzetét, és valamilyen statikus, vagy dinamikus térképként ábrázoljuk. Ez generálható szenzoradatok, például lidar vagy radar segítségével, klasszikus térképek segítségével, illetve különböző V2X kommunikáció felhasználásával. Majd a járműnek meg kell határoznia egy olyan globális útvonalat a környezetében, amely elkerüli a statikus akadályokat, miközben figyelembe veszi az olyan korlátokat is, mint a jármű kinematikai korlátai és az út geometriája. Erre a feladatra különböző útvonaltervezési algoritmusok léteznek, mint például a Rapidly Exploring Random Trees (RRT), a Probabilistic Roadmap (PRM) vagy az A* [6] algoritmus különböző változatai.

Emellett a járműnek olyan lokális útvonalat kell generálnia és periodikusan frissíteni, amely követi a referencia globális útvonalat, figyelembe véve a pillanatnyi korlátozásokat, a jármű dinamikáját és a dinamikusan mozgó objektumokat. A járműnek követnie kell a lokális útvonalat szabályozó algoritmusok segítségével, ezzel meghatározva a kormányzás, a gyorsítás és a fékezés parancsokat.

Ezeket a lépéseket valós időben kell elvégezni, figyelembe véve a bizonytalanságokat, illetve a jármű folyamatosan változó környezetét. A járművekhez számos mozgástervezési algoritmust fejlesztettek ki. Egy megközelítés szerint vannak geometriai algoritmusok, mintavételezésen alapuló algoritmusok és optimalizáláson alapuló algoritmusok. Mindegyik algoritmusnak megvannak a maga erősségei és gyengeségei, és a módszer kiválasztása az alkalmazás egyedi követelményeitől függ [5, 7].

Egy másik megközelítés szerint a mozgástervező algoritmusok nagyjából két kategóriába sorolhatók: modellalapú és modellmentes. A modellalapú algoritmusok a jármű dinamikájának és a környezetnek a matematikai modelljét használják a pálya megtervezéséhez. Léteznek modellalapú kinodinamikai tervezők, amik figyelembe veszik a jármű fizikai korlátait, például a maximális gyorsulást és a fordulási sugarat, hogy fizikailag megvalósítható pályát tervezzenek. Léteznek modellalapú optimalizálási technikákat használó algoritmusok [8], melyek a legjobb pálya megtalálására költségfüggvényt minimalizálnak, ami olyan tényezőket foglalhat magában, mint az idő, az energiafogyasztás vagy a kockázat.

Ezzel szemben a modellmentes algoritmusok nem használják a jármű vagy a környezet matematikai modelljét. Ehelyett ezek az algoritmusok tapasztalatból tanulnak, és korábbi példák alapján generálnak pályákat. Egy modellmentes megközelítés az utánzásos tanulás (Imitation Learning - IL), ami képes tanulni az emberi járművezetők viselkedéséből és alkalmazkodni a közlekedési szituációkhoz [9]. Továbbá népszerű modellmentes módszer a megerősítéses tanulás (Reinforcement Learning - RL), ami a jutalomérték maximalizálásával tervez pályát. Az ágens tapasztalatokból tanul és a környezethez alkalmazkodva idővel javítja teljesítményét.

A megerősítéses tanulás a gépi tanulás eszköztár része, amely különösen alkalmas a döntéshozatallal és irányítással kapcsolatos problémák megoldására. Egy ágens megtanulja, hogy egy jutalomérték maximalizálása érdekében cselekedjen a környezetben. Ez teszi alkalmassá például az önvezető autókkal kapcsolatos feladatok megoldására, ahol a cél az autó mozgásának irányítása a biztonságos és hatékony navigálás érdekében. A megerősítéses tanulást többféleképpen is alkalmazták már az önvezető autókban a mozgástervezés, döntéshozatal vagy akár alacsony szintű szabályozások területén [10].

A megerősítéses tanulás egyik legnagyobb kihívása az úgynevezett "sim-to-real" probléma [11], vagyis a szimulációban betanított ágens átültése a valóságba, amely a modellezett környezet és a valóság közötti különbségek miatt nem megfelelően viselkedik. Az ágenseket a legtöbbször szimulációs környezetben tanítják, mivel ez biztonságos, gyors és olcsóbb, mint a valóságban történő tanítás. Összességében a "sim-to-real" probléma egy nehéz és aktív kutatási terület, amelynek megoldásához a gépi tanulás, a robotika és az irányításelmélet kombinációjára van szükség. A dolgozatban bemutatott eredmények erre a problémára is keresik a választ, ezért fontos volt, hogy a kidolgozott algoritmusok hatékonyságának bizonyítása érdekében azokat valós járművön teszteljem.

Jelen disszertáció a nagymértékben automatizált járművek mozgástervezési technikáival foglalkozik, a megerősítéses tanulás, mint gépi tanulási megközelítés alkalmazásával.

A megerősítéses tanulás a gépi tanulás izgalmas és dinamikusan fejlődő területe, számos tudományos és alkalmazási területen bizonyult hatékonynak az eddigieknél jobb, sőt az emberi teljesítménynél is jobb eredmények elérésében[12, 13]. Markov-döntési folyamaton alapul, amely eredeti formájában teljesen diszkrét állapot-, esemény- és akcióteret feltételez. Ennek megfelelően az RL kezdeti algoritmusait is ebben a formában definiálták, ami számos járműirányítási probléma esetében megnehezíti az alkalmazásukat. A megközelítés reneszánsza - sok más gépi tanulási algoritmushoz hasonlóan - a 2010-es évek elején történt, a mesterséges neurális hálózatok re-

neszánszával együtt. Ez a változás teret nyitott olyan algoritmusok létrehozására is, amelyek már folytonos állapottérrel és döntési lehetőségekkel látták el a felhasználót.

A disszertációban bemutatott négy tézis egy alapvető megvalósíthatósági tanulmányból kiindulva mutatja be azt az utat, amely egy komplex dinamikus akadályelkerülésre alkalmas hierarchikus ágens létrehozásához vezetett.

Az első tézis egy "end-to-end" pályakövetési feladat megvalósíthatósági elemzését végzi el. A hatékony és gyors tanítás érdekében egy szimulációs környezet került implementálásra, amely tartalmaz egy kinematikai járműmodellt, szög- és távolsághibákat érzékelő ideális szenzormodelleket és egy különböző nehézségű véletlenszerű pályagenerátort. Egy jutalomfüggvényt terveztem, amely kiszámítja az egyenlő távolságra lévő kapukon való áthaladással kapott jutalmat. Emellett kidolgoztam egy módszert a "sim-to-real" problémára, amely lehetővé teszi a Python nyelven programozott ágens tesztelését ipari szimulációs szoftver és egy valós jármű segítségével. Egy virtuális CAN-hálózat valósít meg egy interfészt a szimulációs szoftver és a fizikai jármű, valamint a Python-környezet között.

Bár az "end-to-end" megközelítés helyesnek tűnik, nem kapunk információt egy manőver hosszú távú teljesítményéről, ezért az alkalmazása nem túl célravezető a járműirányítási feladatokban. Ennek megfelelően a második tézisben azt a célt fogalmaztam meg, hogy hogyan lehet az egylépéses irányításnál hosszabb mozgástervezési módszereket megerősítéses tanulással úgy megvalósítani, hogy erre valamilyen megvalósíthatósági mutatót is kapjunk. Ehhez a folytonos megerősítéses tanulási feladatot úgy definiáltam, hogy adott egy jármű pályájának kezdeti és végállapota, amelyek között az ágensnek meg kell határoznia egy folytonos görbe paramétereit, amely a követendő útvonal lesz. Ahhoz, hogy ezt az ágens az öntanulás módszerével megtanulja, egy szabályozónak szimulációban ezen a tervezett útvonalon végig kell vezetnie a járművet, amely során kiértékelhető a jutalomfüggvényt, és megtanítja az ágenst a megvalósíthatóság becslésére is.

Az előző eredményekre alapozva a harmadik tézis egy összetet-

tebb mozgástervezési feladat megvalósítására vállalkozik. Ebben az esetben egy prototípus manőver paramétereinek optimalizálása a cél, melynek demonstrálására a kettős sávváltási problémát használtam fel. Ennek megvalósításához nem egy, hanem több, egymással folyamatosan összefüggő görbe paramétereit kell meghatározni. Ahogy az előző feladat megvalósítása során is felmerült a "sim-to-real" probléma - azaz a szimuláción betanított ágens viselkedése a valóságban - ebben az esetben még fontosabb. A probléma megoldására a domain-randomizációs technikát alkalmaztam, azaz a szimulációs körülményeket véletlenszerűen befolyásoltam úgy, hogy az ágens végül képes legyen alkalmazkodni a valós környezeti feltételekhez. Az előző feladathoz hasonlóan az eredményeket egy tesztpályás teszttel validáltam, amelynek során kiderült, hogy a betanított ágens felülmúlja egy átlagos járművezető képességeit.

A negyedik tézis a statikus objektum-kerülési problémát dinamikus objektumra terjeszti ki. A jármű útvonalát véletlenszerűen keresztező gyalogos modell került implementálásra az algoritmus fejlesztésének illusztrálására. A dinamikus akadály kikerülése már nem egylépéses pályatervezési feladat. Ebben az esetben a feladat megfogalmazása sokkal inkább hasonlít egy modell prediktív szabályozási feladathoz, ahol a jármű állapota és a gyalogos mozgása alapján egy teljes útvonalat tervezünk meg, de csak egy kis részét hajtjuk végre. Ezt követően az új helyzet alapján a teljes tervezési feladatot újra kell kezdeni. A betanított ágenst szimulációs környezetben teszteltem, és összehasonlítottam emberi sofőrök teljesítményével.

2. Új eredmények

1. tézis

Kifejlesztettem egy "end-to-end" módszert, amely alkalmas automatizált közúti járművek útvonalának követésére megerősítéses tanulás segítségével. A módszer a követési és a szöghibát használja, az akciótér pedig a kormányszögértékek diszkrét halmaza. Egy Deep Q-Network ágens határozza meg a kormányszög beavatkozását ezen magas szintű szenzorjelek alapján.

Kapcsolódó publikációk: [Feher1], [Feher2]

Megjegyzések

- A tanításhoz kifejlesztettem egy szimulációs környezetet, amely tartalmazza a jármű modelljét, figyelembe véve annak kinematikai korlátait, az irányszöghiba és a követési hiba ideális szenzormodelljeit, valamint egy paraméterezhető nehézségű véletlenszerű pályagenerátort.

- Kidolgoztam a jutalomfüggvényt, amely az egyenlő távolságra elhelyezett kapukon való áthaladás eredménye alapján határozza meg a jutalmat.

- Kidolgoztam egy módszert a "sim-to-real" problémára, amellyel a python környezetben implementált ágens hatékonyan tesztelhető ipari szimulációs szoftverekben és valós járművekben.

- A kiértékeléshez virtuális CAN-hálózat alapú interfészt fejlesztettem ki a python és a szimulációs szoftver között, amelyet a tesztpályás tesztek során valódi CAN-hálózattal lehet helyettesíteni.

2. tézis

Kidolgoztam egy módszert egy kezdeti feltételekkel és egy előre meghatározott célállapottal rendelkező pályatervezési feladatra, valamint a pálya megvalósíthatóságának becslésére. A tervező egy Deep Deterministic Policy Gradient ágenst használ, amelynek tanítása egy újszerű, egylépéses módszeren alapul. Az ágens tanítása során a tervezett pályát egy belső zárt hurkú szabályozóval értékelem. A módszer egy új jutalomfüggvényt tartalmaz, amely az oldalirányú hiba és a szöghiba mellett a jármű oldalirányú csúszását is figyelembe veszi.

Kapcsolódó publikációk: [Feher3] [Feher4]

Megjegyzések

- A polinomiális trajektória tartópontjainak meghatározása mellett a betanított ágens a dinamikai megvalósíthatóságra is becslését ad.
- A korábban kifejlesztett "sim-to-real" módszerrel az algoritmust szimulációs környezetben és valós járművön teszteltük.
- Bemutattam, hogy a gépi tanulás és a klasszikus módszerek kombinációja alkalmas lehet autonóm járművek mozgásszabályozásának tervezésére.

3. tézis

Kifejlesztettem egy módszert kikerülési manőver tervezésére automatizált közúti járművek számára. A megoldást egy kettős sávváltási manőveren, mint kritikus forgalmi szituáción keresztül mutatom be. A kifejlesztett útvonaltervező olyan dinamikailag megvalósítható útvonal megtervezésére törekszik, amely képes a jármű előtt megjelenő akadályok elkerülésére. A megerősítéses tanulás alapú megoldás bemenetként a jármű dinamikai és a feladat geometriai paramétereit használja, kimenetként pedig egy görbecsoportot határoz meg követendő útvonalként.

Kapcsolódó publikáció: [Feher5]

Megjegyzések

- A tervezést egy Deep Deterministic Policy Gradient ágens végezte, amelynek bemenetei a jármű sebessége és a foglalt területek geometriai paraméterei, amelyeket a környezetérzékelők szolgáltatnak.
- Az ágens kimenetei a pályát leíró görbe szakaszok paraméterei.
- A jutalomértékek a nemlineáris dinamikai járműmodell pályán történő vezetése során kerülnek kiszámításra.
- A tervezést az ISO 3888-2 szabványban meghatározott jávorszarvasteszt végrehajtásával értékeltem.
- Összehasonlítottam a különböző megközelítések eredményeit és az emberi járművezetők teljesítményét.

4. tézis

Kidolgoztam egy lokális pályatervezési és szabályozási módszert közúti járművek számára, amely a jármű dinamikáját és az út geometriai korlátait figyelembe véve megtervezi az útvonalat és a sebességprofilt az úttesten áthaladó mozgó objektumok elkerülésére. A betanított Twin Delayed Deep Deterministic Policy Gradient ágens folyamatosan újratervezi a trajektóriát egy adott horizonton, amit egy modell predictive szabályozó lekövet.

Kapcsolódó publikáció: [Feher6]

Megjegyzések

- A tanítás során a hosszirányú Stanley és az oldalirányú MPC-szabályozók kerültek alkalmazásra a pályakövetésre.
- Az ágens bemenetei a mozgó objektum relatív távolsága és sebessége, amelyet a környezeti érzékelők szolgáltatnak, valamint az ego jármű mozgási állapota.
- Az ágens kimenetei az útvonalat leíró polinomfüggvény és a sebességprofil paraméterei.
- A módszer emberi járművezetők teljesítményével került összehasonlításra szimulációs környezetben.
- Az algoritmus gyalogosokat kikerülő közúti járművekre került kifejlesztésre, de elvei számos más dinamikus objektum-kerülési problémára is alkalmazhatók.

3. További kutatási irányok

A megerősítéses tanulás kutatása, beleértve annak alkalmazását a közlekedési és járműirányítási feladatokban nagyon izgalmas terület, és ennek megfelelően nemzetközi eredményei rendkívüli gyorsasággal fejlődnek. Mind az alapvető tanulási algoritmusok, mind az azokat kiegészítő metaheurisztikák, illetve a mesterséges neurális hálózati struktúrák évről évre megújulnak. A disszertációban bemutatott eredmények felhasználhatók valós idejű járműirányítási feladatok megvalósítására, azonban robusztusságuk még nem garantált, de ez nem meglepő, hiszen ez az egész tudományterület egyik legégetőbb kihívása jelenleg. Ennek megfelelően különösen vonzó kutatási terület az algoritmusok tesztelése és validálása. Ezek mellett fontos fejlesztési szempont lehet még a különböző érzékelési modellek pontosabb implementálása a rendszerbe, a több ágens együttműködésével való tanítás és számos más járműirányítási feladat. További kutatásaim során az integrációs feladatokra, a különböző megvalósítható szabályozások tervezésére, valamint a valóságban tesztelhető rendszerek fejlesztésére tervezek koncentrálni.

A tézisekhez kapcsolódó publikációk

[Feher1] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Q-learning based reinforcement learning approach for lane keeping. In *2018 IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI)*. IEEE, nov 2018.

[Feher2] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Fast Prototype Framework for Deep Reinforcement Learning-based Trajectory Planner. *PERIODICA POLYTECHNICA-TRANSPORTATION ENGINEERING*, ONLINEFIRS:1–6, 2020.

[Feher3] Árpád Fehér, Szilárd Aradi, Tamás Bécsi, Péter Gáspár, and Zsolt Szalay. Proving Ground Test of a DDPG-based Vehicle Trajectory Planner. In *2020 European Control Conference (ECC)*, pages 332–337, 2020.

[Feher4] Arpad Feher, Szilard Aradi, Ferenc Hegedus, Tamas Becsi, and Peter Gaspar. Hybrid DDPG Approach for Vehicle Motion Planning. In *Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics, ICINCO 2019, Vol. 1 Setúbal, Portugália : SciTePress, (2019)*, pages 422–429. Scitepress, 8 2019.

[Feher5] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Hierarchical evasive path planning using reinforcement learning and model predictive control. *IEEE Access*, 8:187470–187482, 2020.

[Feher6] Árpád Fehér, Szilárd Aradi, and Tamás Bécsi. Online trajectory planning with reinforcement learning for pedestrian avoidance. *Electronics*, 11(15):2346, 2022.

Hivatkozások

[1] Daniel Watzenig and Martin Horn. *Automated driving: safer and more efficient future driving.* Springer, 2016.

[2] Tamas Tettamanti, István Varga, and Zsolt Szalay. Impacts of autonomous cars from a traffic engineering perspective. *Periodica Polytechnica Transportation Engineering*, 44:244–250, 10 2016.

[3] Yifang Ma, Zhenyu Wang, Hong Yang, and Lin Yang. Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA Journal of Automatica Sinica*, 7(2):315–329, 2020.

[4] Biagio Ciuffo, Konstantinos Mattas, Michail Makridis, Giovanni Albano, Aikaterini Anesiadou, Yinglong He, Szilárd Josvai, Dimitris Komnos, Marton Pataki, Sandor Vass, and Zsolt Szalay. Requiem on the positive effects of commercial adaptive cruise control on motorway traffic and recommendations for future automated driving systems. *Transportation Research Part C: Emerging Technologies*, 130:103305, 2021.

[5] Brian Paden, Michal Cap, Sze Zheng Yong, Dmitry Yershov, and Emilio Frazzoli. A Survey of Motion Planning and Control Techniques for Self-driving Urban Vehicles. *CoRR*, pages 1–27, 2016.

[6] Michael Montemerlo, Jan Becker, Suhrid Bhat, et al. Junior: The stanford entry in the urban challenge. *Journal of field Robotics*, 25(9):569–597, 2008.

[7] Laurène Claussmann, Marc Revilloud, Dominique Gruyer, and Sébastien Glaser. A review of motion planning for highway autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):1826–1848, 2019.

[8] Ferenc Hegedüs, Tamás Bécsi, and Szilárd Aradi. Dynamically feasible trajectory planning for road vehicles in terms of sensitivity and robustness. *Transportation Research Procedia*, 27:799 – 807, 2017. 20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-6 September 2017, Budapest, Hungary.

[9] Abdoulaye Oumar Ly and Moulay A Akhloufi. Learning to drive by imitation: an overview of deep behavior cloning methods. *IEEE Transactions on Intelligent Vehicles*, 2020.

[10] Szilárd Aradi. Survey of deep reinforcement learning for motion planning of autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):740–759, 2022.

[11] Wenshuai Zhao, Jorge Peña Queralta, and Tomi Westerlund. Sim-to-real transfer in deep reinforcement learning for robotics: a survey. In *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 737–744, 2020.

[12] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[13] David Silver, Thomas Hubert, Julian Schrittwieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144, 2018.