

Veículos e Condução autónoma: Uma revisão

Universidade do Minho, Braga, Portugal

Abstract. Os sistemas multiagente vêm trazer possíveis soluções inovadoras para inúmeros problemas da área dos veículos e condução autónoma, sendo esta uma técnica emergente e promissora, que poderá trazer várias vantagens. No entanto, a comunicação inerente a estes sistemas poderá estar sujeita a problemas de segurança, que podem tornar estas soluções perigosas e incontroláveis. Neste artigo, apresentam-se vários trabalhos realizados nesta área, mostrando-se 4 secções principais, divididas pelos problemas que estes abordam, nomeadamente o combate à congestão de tráfego urbano, combate ao perigo da condução em condições atmosféricas adversas, aprendizagem de manobras de condução e, finalmente, os desafios da comunicação destes sistemas, nomeadamente da segurança. Finalmente, comparam-se os trabalhos em questão, revelando-se os benefícios e lacunas dos mesmos.

Keywords: Sistemas Multi-Agente · Condução Autónoma · Tráfego Urbano · Deep Q-Network · Segurança Vehicle to Everything

1 Introdução

O tema dos veículos e condução autónoma tem sido um dos campos com maiores avanços recentemente e espera-se que a sua evolução revolucione o tráfego rodoviário, atenuando problemas especialmente de acidentes e congestão. No entanto, é uma área que ainda traz muitas incertezas e desafios, não estando inerentes apenas problemas tecnológicos, mas ainda problemas ligados ao comportamento humano, questões éticas, políticas, responsabilidade, segurança, entre outros. Os problemas e desafios tecnológicos no desenvolvimento de sistemas autónomos que possam ser comercializados em ambientes dinâmicos e livres levaram a campos como os Sistemas Multiagente. Este é um campo que interliga de um modo profundo temas como a autonomia, negociação, comunicação, interação e ainda cooperação. O maior objetivo deste campo é o desenvolvimento de sistemas autónomos capazes de coexistir e cooperar com pessoas e outros sistemas do mundo real, sendo este motivado pelo facto de hoje vivermos num mundo digital e interconectado, estando a tecnologia altamente vinculada nas nossas vidas diárias.[1]

2 Estado da arte

Nesta secção serão apresentados algumas soluções do estado da arte, considerando alguns trabalhos desenvolvidos nos últimos 4 anos (2019-2023). O objetivo da estruturação desta secção é mostrar a variedade existente nas possíveis aplicações

de sistemas multi-agente na condução autónoma, resolvendo diversos problemas reais inerentes ao desenvolvimento e implementação de veículos autónomos. Em cada um dos tópicos será abordado pelo menos um artigo com a profundidade necessária para que seja possível explicar o problema que é abordado e a forma como o trabalho pretendeu resolvê-lo.

2.1 Congestão de Tráfego Urbano

O trabalho de Alekszejenkó L. e Dobrowiecki T. [2] foca-se essencialmente no uso de sistemas multiagente em veículos autónomos para combater o problema da congestão no trânsito urbano.

O modelo considerado pelos autores pressupõe a existência de dois tipos de agentes:

- **Um agente *Smart Car***: Estes são veículos totalmente autónomos capazes de efetuar ações como mudar de faixa, obedecer às regras de trânsito, conduzir de um ponto até outro seguindo uma rota. Adicionalmente, possuem dispositivos de comunicação capazes de transmitir toda a informação que for necessária, sem qualquer delay, e conseguem interpretá-la e agir de acordo com a mesma.
- **Um agente "Intersection Manager"**: Estes são dispositivos localizados em zonas de cruzamentos, por exemplo semáforos, que possuirão a capacidade de perceber o ambiente à sua volta, bem como comunicar com os *Smart Cars* que se encontrem na sua proximidade e tomar decisões com base na informação disponível.

Para entender o processo de comunicação entre os vários agentes do sistema é essencial entender dois conceitos mencionados no artigo:

- ***Platoon***: É um conjunto de carros que possui um *platoon leader*, que toma as decisões que serão executadas por esse conjunto.
- ***Judge***: São os "Intersection Manager" que possuem a capacidade de decidir qual dos *platoons* presentes num cruzamento terá prioridade em relação aos restantes.

Todos estes membros possuem protocolos de comunicação uns com os outros e são capazes de agir de acordo com a informação trocada entre eles.

Os Agentes *Smart Cars* são competitivos (competem pelo sinal verde de um semáforo), mas estão dispostos a criarem uma aliança - um *platoon* - de modo a poderem passar de uma forma mais eficaz e segura um cruzamento, se os seus destinos coincidirem. Assim, quando um carro se aproxima de um cruzamento, este precisa de se juntar a um *platoon*. Os carros que constituem um *platoon* possuem exatamente a mesma trajetória: chegam da mesma direção, estão na mesma faixa e saem na mesma saída do cruzamento. No tempo de vida de um *platoon*, os carros são virtualmente introduzidos numa cadeia, mantendo 10 metros de distância uns dos outros. O carro que se encontra à frente deste, é o

platoon leader, e todos os outros são os *platoon members*, que têm de seguir o primeiro. Depois de passar o cruzamento, o líder retira-se do *platoon* e passa as responsabilidades para o carro que se encontra imediatamente atrás, sendo este o novo líder.

Já os *judges* têm como objetivo principal resolver conflitos entre *platoons*, como já foi referido anteriormente. Para que isto seja possível, primeiro os líderes necessitam de anunciar a sua chegada ao "Judge". Depois, este decide que *platoon* estará mais qualificado para passar o cruzamento e comunica a ordem ao líder, que irá depois comunicar a ação aos membros do seu *platoon*, fazendo, em conjunto, a passagem do cruzamento. Depois de um *platoon* anunciar a sua chegada a um cruzamento, caso estes não possam passar de imediato, estarão continuamente a questionar o "Judge" de modo a saber se lhes é dada permissão para atravessar o cruzamento.

Para que sejam tomadas as decisões necessárias mencionadas acima, foram feitos alguns ajustes a algoritmos já existentes, de modo a criar novos algoritmos de decisão adequados ao caso de estudo. Estes basearam-se em dois algoritmos famosos, o Round Robin (RR), um algoritmo justo sem *starvation*, e o Shortest Job First (SJF), que não é tão justo quanto o primeiro. Deste último resultou um algoritmo denominado Minimal Destination Distance First (MDDF).

Depois de modulado o sistema, e codificados os algoritmos de decisão, foi utilizado um ambiente de simulação, na plataforma *Simulation of Urban MObility (SUMO)* [3], de modo a fazer uma recolha de métricas para ajudar na obtenção de resultados do modelo desenvolvido. Para a medição dos valores necessários, a simulação foi feita em dois tipos de infraestruturas diferentes, Americanas e Europeias, Test Track e *BAH intersection* (uma interseção conhecida em Budapeste), respetivamente, uma vez que estas possuem uma organização que as diferencia de forma significativa, no que toca ao número de caminhos possíveis de um ponto até outro. Quanto aos testes, foram feitas quatro simulações. Duas com fluxo de trânsito normal (com trânsito leve e com trânsito mais congestionado) e duas com fluxo irregular (situações de acidente, obras na via, etc...). Quanto às métricas recolhidas, foram consideradas métricas de importância as seguintes:

- **Started vehicles:** número de carros introduzidos na simulação
- **Running vehicles:** número de carros presentes na simulação num dado momento
- **Arrived vehicles:** quantos veículos já chegaram ao destino num dado momento
- **Average time:** tempo médio decorrido desde que um carro foi introduzido na simulação até que chegou ao seu destino
- **Average waiting time:** tempo médio de espera de um carro num semáforo
- **Lost time:** tempo medido enquanto um carro não pôde viajar à velocidade máxima que viajaria se viajasse sozinho na estrada

Das simulações, foram então obtidos os valores para estas métricas que permitiram concluir alguns aspetos, que foram separados, naturalmente, não só pelo tipo de infraestrutura, mas também pelo fluxo de trânsito.

Na **Test Track**, os valores medidos mostraram no geral uma melhoria significativa. O número "Arrived vehicles" aumentou quase 150% (de 200 para 300 carros), o "Average time" diminuiu cerca de 10 a 15%, dependendo do algoritmo de decisão usado pelos "Judges". Finalmente, a melhoria mais significativa foi a redução do "Average waiting time", diminuindo este de 55 segundos para 14 ou mesmo abaixo de 5 segundos, consoante o algoritmo utilizado.

Já no circuito **BAH** as melhorias não foram tão significativas, podendo isto dever-se a uma implementação imperfeita ou até mesmo a uma configuração não otimizada. O número "Arrived vehicles" diminuiu cerca de 10% em relação ao sistema tradicional. Os valores "Average time" e "Lost time" não mostram diferenças significativas, mas o "Waiting time" sofreu uma diminuição de 22 segundos para cerca de 13 ou 6 segundos, dependendo do algoritmo de decisão utilizado.

Os resultados mais interessantes podem ser encontrados em situações de trânsito irregular, não reagindo o sistema tradicional tão bem a estas situações. A simulação mostra então que quanto mais anormal for o fluxo de trânsito, mais congestionada se torna a rede. Destes resultados pode-se concluir que apesar de não terem sido obtidos resultados tão claros nas estradas Europeias, o valor desta solução é inegável, uma vez que no plano geral, em ambos os casos esta organização inteligente é claramente benéfica em situações de trânsito irregular, sendo que nestes cenários não só o tempo de passagem nos semáforos sofre uma diminuição de aproximadamente 15%, em relação aos sistemas tradicionais, como também o tempo médio de espera nos mesmos se mantém aproximadamente constante, entre os 10 e os 15 segundos, independentemente da congestão ou do destino de um carro.

Test Case	Arrived (%)	Waiting Time (s)	Lost Time (s)	Average Time (s)
Traditional	33.81	29.68	51.98	170.55
RR	29.19	12.12	61.97	174.87
MDDF	22.77	12.41	51.45	154.02

Table 1. Resultados da simulação do caso "Irregular 1"

Test Case	Arrived (%)	Waiting Time (s)	Lost Time (s)	Average Time (s)
Traditional	38.48	36.44	67.98	199.38
RR	32.71	11.43	48.59	170.07
MDDF	34.39	10.74	51.23	176.72

Table 2. Resultados da simulação do caso "Irregular 2"

Concluindo, este artigo mostra uma solução eficaz, facilmente extensível para situações de tráfego maiores, sendo a sua implementação simples e pouco custosa.

Assim, mostra uma boa solução para combater a congestão do tráfego urbano no futuro.

Outra abordagem possível, introduzida por Cui J. et al. [4], para o combate à congestão do tráfego urbano sugere aplicar técnicas de Reinforcement Learning para treinar um modelo numa rede fechada e depois fazer transfer learning para uma rede aberta, no mundo real. Adicionalmente, este trabalho mostra ainda uma abordagem distribuída que se revela ser mais realista e prática, já que se apoia apenas na sensorização existente e nas capacidades de atuação, não necessitando da adição de uma infraestrutura de comunicação. Assim, cada agente atua de modo individual com base num modelo treinado em conjunto com outros agentes.

A motivação do uso do paradigma transfer learning, com uma abordagem descentralizada, deve-se ao elevado esforço que é necessário para treinar um modelo de Reinforcement Learning do zero, devido a 3 principais razões:

- À medida que o número de veículos aumenta, o espaço (estado, ação) aumenta exponencialmente, quando é usada uma abordagem centralizada
- A reward de um agente pode-se tornar muito ruidosa, uma vez que esta é calculada com base nas ações de todos os veículos na rede
- Há um grande delay nas rewards devido ao delay do efeito de uma ação feita por um agente na política centralizada

Assim, este paradigma de aprendizagem vem tornar o treino destes modelos possível num tempo razoável, que não seria possível caso os modelos fossem treinados de raiz.

O modelo é então treinado num ambiente simulado, com a framework de treino Flow [5] e no sistema de simulação SUMO [3]. Posteriormente, é lançado numa rede no mundo real, pronto a atuar de acordo com a política treinada. Assim, cada veículo utiliza o modelo para a tomada de decisão na estrada, sem estabelecer qualquer comunicação com os restantes veículos. Estes possuem ainda mecanismos de sensorização do ambiente, podendo avaliar a velocidade e distância dos carros vizinhos, por exemplo, e agir com base na informação percecionada.

O objetivo a atingir é então aumentar a velocidade média de um carro no trânsito, de modo a que este possa fluir melhor, evitando sempre situações de acidentes.

Como reward functions a otimizar pelos agentes de Reinforcement Learning, são usadas 3 funções diferentes:

- **Original Flow Reward:** Encoraja todos os veículos a viajarem o mais perto possível da velocidade pretendida a todos os momentos, mantendo uma distância de segurança considerável
- **Average Speed Reward:** Soma a velocidade média de todos os carros durante um episódio inteiro de treino e relaciona-a com a velocidade máxima permitida numa dada faixa, incentivando os veículos a viajarem mais rápido mas nunca ultrapassando o limite de velocidade

- **Outflow Reward:** Representa o número de veículos que conseguem sair da área de observada em cada instante. Ao ser maximizada, irá resultar numa progressão da aprendizagem do modelo no sentido de maximizar o número de veículos que saem da rede, por instante de tempo.

A política de aprendizagem do modelo é partilhada por todos os agentes, sendo as experiências de todos usadas para treinar uma política centralizada, que será posteriormente a política de atuação final. Assim, cada agente tem apenas acesso à informação local e age independentemente dos restantes agentes, recebendo o seu próprio *reward*.

Cenários de simulação e Resultados:

- **Cenário 1:** Descreve a operação da passagem de um carro numa união de duas vias. A estrada é composta por um troço de autoestrada de 600m antes de uma união e uma via de aceleração de 200 metros. Depois da passagem do carro da via de aceleração para a via principal, os carros ainda viajam mais 100 metros
- **Cenário 2:** Consiste num troço de estrada real, bastante maior do que o anterior. Uma via principal de 3131 metros e uma via de aceleração de 1878.56 metros. Depois de unidas as vias, os veículos ainda viajam mais 5077.7 metros.

Em ambos os cenários são introduzidos tanto veículos autónomos, que agem segundo a política de treino, como veículos conduzidos por humanos, simulados pelo Intelligent Driver Model (IDM) [6].

Para a obtenção de resultados, para cada configuração são executadas 3 políticas e é escolhida a política com maior valor de *reward* e é avaliada a sua performance em 100 simulações. As médias dos valores obtidos na simulação são depois recolhidas para serem analisadas.

Feita a análise dos resultados dos modelos treinados, é possível concluir que todos eles mostraram melhores resultados do que a condução humana simulada. Adicionalmente, é de notar ainda que a *reward function* que produziu melhores resultados foi a *Outflow Reward*.

<i>Reward</i>	Average Outflow (vehs/hr)	Average Inflow (vehs/hr)	Average Speed (m/s)
Human	1558.12±2.99	1725.48±2.89	7.27± 0.15
Original Flow Reward	1724.55±6.98	1769.36±6.60	18.95±0.19
Average Speed Reward	1379.45±2.99	1408.46±3.28	19.34±0.02
Outflow Reward	1804.21±7.17	1864.55±7.24	16.21±0.08

Fig. 1. Resultados das políticas de treino

Adicionalmente, foi avaliada a performance de dois modelos treinados do zero, um na totalidade da área do Cenário 2 (Train From Scratch) e o outro apenas numa zona desse cenário (Train From Scratch Window), e de um modelo pré-treinado também numa zona mais pequena e transferido para aprender na situação do Cenário 2 (Zero-Shot Transfer Window). Todas estas foram treinadas com duas *reward functions*: *Flow Reward* e *Outflow Reward*.

O melhor resultado é evidenciado pela abordagem *Zero-Shot Transfer* integrada com *Outflow Reward*, sendo esta significativamente melhor do que a condução humana simulada. A segunda melhor abordagem foi a *Train From Scratch Window*, não estando a sua performance muito longe da primeira. Por fim, a abordagem *Train From Scratch* foi a pior, tendo um desempenho inferior à condução humana. Ambas as melhores abordagens revelam melhores valores de velocidade média do que a condução humana, tendo ainda valores de *Average Inflow* comparáveis. No entanto, nenhuma das abordagens apresenta valores melhores de *Original Flow Reward*, em comparação à condução humana.

Experiment	Reward	Average Outflow(vehs/hr)	Average Inflow(vehs/hr)	Average Speed (m/s)
Human	None	936.90±5.96	2184.91±0.29	16.27±0.12
Train From Scratch ($N_{AV}=30$)	Outflow	366.98±1.91	561.60±3.12	19.57±0.27
	Flow reward	638.06±10.99	1165.06±10.22	14.95±0.12
Train From Scratch (Window)	Outflow	1012.64±9.23	2178.76±2.81	16.99±0.13
	Flow reward	923.29±5.79	2181.17±1.92	15.98±0.10
Zero-Shot Transfer (Window)	Outflow	1017.32±10.49	2170.55±4.61	17.05±0.16
	Flow reward	928.00±6.06	2181.53±1.67	16.09± 0.11

Fig. 2. Comparação do treino do zero com Transfer Learning

O artigo permite concluir que:

- O modelo treinado numa rede mais pequena e depois transferido para uma pequena janela à volta de uma rede maior apresenta melhores resultados do que uma condução feita por humanos e até mesmo do que uma política treinada do zero.
- A técnica de *Transfer Learning* relevou precisar de muito menos tempo de treino, atingindo reduções até 80%
- Finalmente, mostra que pela primeira vez, até à data de publicação do mesmo, uma política multiagente distribuída pode melhorar a eficiência do trânsito numa pequena rede aberta, confiando apenas em conhecimento local percebido pelos sensores do veículo. Esta abordagem é mais realista do que as abordagens centralizadas, uma vez que a última apoia-se em informação agregada por todos os veículos, o que neste momento ainda não é possível implementar, uma vez que não estamos perto de ter todos os carros conectados uns aos outros.

2.2 Combate aos perigos da condução em condições climáticas adversas

Trazendo uma solução a um outro problema bastante importante para garantir maior segurança na condução, Outay, F. et al. [7] sugerem uma abordagem com recurso ao paradigma de sistemas Multi-Agente em condições atmosféricas adversas.

Segundo os autores, uma das soluções para promover uma condução mais segura em situações de visibilidade reduzida estão relacionadas com a aplicação de:

- Comunicação Vehicle to Vehicle (V2V) e Vehicle to Infrastructure (V2I)
- "Crowd Sensing" para estimar a visibilidade

O domínio dos agentes é composto por agentes que representam as entidades simuladas, ou seja, veículos e objetos conectados. Neste caso de estudo, os objetos conectados são sinais de trânsito que exprimem o limite de velocidade num dado troço de estrada, adaptando a sua informação às condições climáticas do momento.

Assim, para as duas categorias de agentes são caracterizados a sua constituição e comportamento.

Connected Driver: Um veículo conectado é um veículo equipado com dispositivos de comunicação e, normalmente, possui ligação à rede local. Isto permite que este possa comunicar com outros dispositivos com capacidades de comunicação semelhantes. Esta comunicação poderá ser separada em duas categorias:

- **V2V:** Devido à capacidade limitada de sensorização dos sensores que se encontram nos veículos hoje em dia, é necessário compensar a falta de informação, ou mesmo erros, através da comunicação entre veículos, de modo a que a informação ausente num veículo possa ser complementada pela informação percecionada por outros veículos.
- **V2I:** Comunicação de informação wireless entre veículos e as infraestruturas, por exemplo, sinais de trânsito, marcações das linhas da estrada e quaisquer outros dispositivos de infraestrutura inteligentes.

Já a comunicação Vehicle to Everything (V2X) junta estes conceitos e estende a sua aplicação a outros intervenientes. A ideia desta tecnologia é que um veículo com dispositivos de comunicação poderá comunicar com o que o rodeia, incluindo V2I, V2V ou até Vehicle to Pedestrian (V2P) e Vehicle to Network (V2N). As suas funções incluem:

- Informação de veículos fora do campo de sensorização dos sensores ou do campo de visão de um condutor
- Aviso da aproximação de um peão distraído
- Entrega de alertas das condições da estrada ou atmosféricas ao condutor

Sinais de trânsito: V2I captura dados de trânsito gerados pelos veículos, que podem depois ser disponibilizados para informar os condutores acerca de questões de segurança, mobilidade ou até condições do ambiente. Os componentes da infraestrutura possuem 3 capacidades principais:

- **Percecionar o Ambiente:** A informação capturada pode ser usada para alterar o seu estado ou até para ser enviada para outros componentes
- **Comunicar com outros componentes:** Esta comunicação permite a transferência dos dados de um dispositivo da infraestrutura para outro dispositivo qualquer
- **Comunicar com veículos:** É um tipo específico da capacidade anterior, em que o dispositivo recetor é um veículo

Estes dispositivos têm um comportamento bem definido, composto por 5 fases:

- **Extração de dados:** Permite armazenar os dados provenientes da sensorização
- **Filtragem dos dados:** Descarta os dados irrelevantes
- **Mudança do estado computacional:** De acordo com os dados recolhidos, permite tomar uma decisão quanto à mudança de estado a sofrer
- **Estado físico:** Efetua a mudança do estado decidido na fase anterior, por exemplo, alterar a luz de verde para vermelho
- **Emissor de comunicação:** Prepara a comunicação com outros componentes da infraestrutura, rede ou veículos

A aplicação deste último tipo de agentes podia ser feito, como já referido, com inúmeros elementos das infraestruturas das estradas, no entanto, neste artigo foram apenas aplicados a sinais de trânsito com informação sobre o limite de velocidade numa dada zona. O comportamento deste será adaptar o valor contido no sinal de acordo com as condições climáticas percecionadas num dado momento, e comunicar este valor a outros dispositivos do mesmo tipo, para que esta informação seja perceptível não apenas na zona onde se encontra este sinal mas em zonas onde se localizam os sinais adjacentes a este.

A simulação foi feita num ambiente bem definido, um troço de autoestrada, que contém um sub-troço com nevoeiro. Vários cenários foram conduzidos para testar o comportamento dos modelos propostos, sendo dois deles detalhados neste artigo:

- **Situação 1:** Sem sistema inteligente. Os condutores não possuem informação dinâmica sobre a situação do nevoeiro, apenas conhecem o que é possível percecionar no seu campo de visão. É usado como uma representação de uma situação como a atual nas estradas.
- **Situação 2:** Com sistema inteligente. Os condutores percecionam os sinais de trânsito que se adaptam, refletindo a situação atual do nevoeiro, com base em informação que vai sendo passada por uma cadeia de comunicação desde os sinais de limite de velocidade que se encontram no interior da zona com

nevoeiro até aos sinais que se encontram ao longo do troço da autoestrada antes desta zona. Deste modo, a velocidade poderá ser moderada antes e não apenas na entrada da zona com nevoeiro.

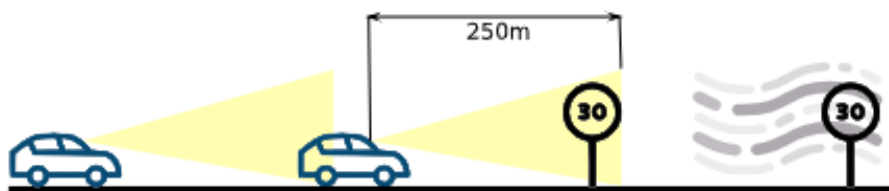


Fig. 3. Cenário 1: Não existe comunicação que envolva infraestrutura ou veículos



Fig. 4. Cenário 2: Há comunicação entre componentes da infraestrutura

Resultados: Para ambos os cenários foram feitas cerca de 100 experiências diferentes. Foram colhidas algumas métricas, entres as quais: Velocidade dos veículos, duração da viagem e acidentes medidos.

No que toca à velocidade média, as melhorias foram encontradas principalmente nas situações de entrada e saída da zona com nevoeiro. À entrada do nevoeiro, no cenário 2 é possível observar que a velocidade média dos veículos diminui uns instantes antes da dos veículos no cenário 1, uma vez que os sinais inteligentes providenciam informação que na situação só é percecionada uns metros depois. Assim, estes podem viajar a uma velocidade ligeiramente mais alta, uma vez que possuem mais informação sobre a situação. Já à saída da zona de nevoeiro, é possível observar que a velocidade média dos veículos do cenário 2 é ligeiramente menor do que a dos veículos do cenário 1, mais uma vez porque esta informação pode ser percecionada mais precocemente através dos sinais inteligentes que comunicam o final da zona de nevoeiro, enquanto que no cenário 1 o final do nevoeiro só é perceptível fisicamente à saída desta zona.

Já os acidentes medidos, a probabilidade de acidentes estimada através dos dados obtidos na simulação desceu cerca de 4% (de 27% para 23%), da situação 1 para a situação 2, podendo esta ser explicada pelo facto de os veículos estarem a abrandar mais cedo no cenário 2, devido à presença dos sinais inteligentes.

Concluindo, apesar de apresentar uma solução interessante no campo do uso de sistemas multiagente para o combate à falta de visibilidade na condução, o artigo em questão não apresenta resultados claros, não sendo a diferença entre as duas situações analisadas muito evidente, pelo que retirar conclusões de alguns dos gráficos se torna uma tarefa complicada. Adicionalmente, os resultados apresentados para a métrica tempo de viagem médio não são conclusivos, sendo o respetivo gráfico difícil de interpretar. Apesar desse facto, a solução apresentada não deixa de ser interessante e extremamente viável, devido à sua simples e barata implementação.

2.3 Aprendizagem de manobras de condução

Nagarajan K. e Yi Z. [8] exploraram a viabilidade da manobra de mudança de faixa, usando uma abordagem inovadora, multi-agent Deep Q-Network. Os autores argumentam que a maioria dos algoritmos existentes até à data da escrita deste artigo usam técnicas de single-agent Deep Reinforcement Learning (DRL) que é apenas uma simplificação do cenário do mundo real, que não considera a negociação multi-agente.

Os trabalhos desenvolvidos mencionados pelos autores, para o problema da manobra de mudança de faixa, com recurso a DRL, apesar de mostrarem resultados positivos quanto ao objetivo de um agente em específico, o ego-agent, fazer esta manobra, estes modelam o problema de tal forma que nenhum agente exceto este possa realizar esta manobra. Isto claramente é algo que não é verdade no mundo real. Como resultado destas suposições menos realistas, resulta uma política que toma decisões pouco seguras, assumindo que os agentes que os rodeiam apenas conduzirão em linha reta. Para além disso, a abordagem single-agent também limita a eficiência dos recursos, uma vez que apenas o ego-agent é usado para explorar, colecionar experiências e aprender a política.

Assim, neste trabalho os autores decidem modelar o problema de mudança de faixa como um problema multi-agente, que irá permitir que os agentes aprendam a negociar e cooperar simultaneamente e a considerar as ações uns dos outros quando estiverem a planear a sua estratégia para atingir os seus objetivos individuais. Para a tomada de decisão é desenvolvida uma política usando uma Deep Q-Network [9].

O objetivo de cada agente é delimitado para atingir a sua faixa destino dentro de uma distância fixa o mais rápido possível, de forma segura e eficiente. As experiências de todos os agentes são colecionadas em simultâneo, de modo que estes aprendam uma política comum, uma abordagem descentralizada, com partilha de parâmetros.

Como plataforma de simulação, é usado o “Carla Simulator” [10]. Para ambiente de simulação, é considerada uma rua em linha reta com 3 faixas. Quanto ao

número de agentes, fizeram-se variar ao longo dos episódios, de modo a garantir a escalabilidade do algoritmo. Para modelar o comportamento dos veículos, foi usado o modelo Intelligent Driver Model (IDM) como módulo Adaptive Cruise Control, que define a velocidade do alvo e evita a colisão frontal [6]. Um episódio termina quando dois veículos colidirem ou quando o tempo da simulação exceder um certo valor temporal.

A forma das *rewards* é um aspeto extremamente importante, já que permite que os agentes aprendam um comportamento cooperativo. O sistema de *rewards* desenvolvido neste trabalho é o seguinte:

- Um agente que é responsável por uma colisão recebe uma *reward* negativa
- Para evitar mudanças de faixa ou abrandamentos desnecessários, qualquer uma destas ações é penalizada
- Se um agente chega à faixa destino, é recompensado com uma *reward* inversamente proporcional ao tempo que demorou a lá chegar, o que o encoraja a fazê-lo no menor espaço de tempo possível
- Se um agente chegar a uma faixa que não era o seu destino, será dada uma *reward* negativa, proporcional à distância desta à sua faixa destino

Devido à ausência de uma forma standard de avaliar o desempenho do comportamento de mudança de faixa, foram usadas 4 métricas informativas:

- Success %
- Failure %
- Timeout %
- Accident %

O score é definido como a média dos scores de todos os agentes num episódio. Um agente que não tenha chegado ao seu destino num episódio é dado como ativo. Um agente que esteja na sua faixa destino na linha final do troço da simulação, a sua missão é considerada como sucesso, caso contrário como falhada. Se um agente ativo colide com outro, o episódio acaba e os resultados dos restantes agentes ativos são ignorados.

Foram feitos 5 testes, cada um com 2500 cenários gerados aleatoriamente. O melhor apresentou os seguintes resultados, em comparação com a abordagem Single-Agent:

Name	Success %	Failure %	Accident %	Timeout %
S-ER106	94.76	0.43	2.83	1.99
ER106	97.32	1.20	0.5	0.98

Table 3. Valores apresentados para Single-Agent (S-ER106) e Multi-Agent (ER106)

Quanto à escalabilidade deste modelo, é mostrado que existe uma degradação da performance com o aumento do número de agentes introduzido na simulação.

No entanto, o modelo porta-se razoavelmente bem sem sequer ser treinado para um ambiente com mais agentes. Assim, é mostrado que o modelo poderá ser facilmente escalado com mais tempo de treino.

n	Success %	Failure %	Accident %	Timeout %
4	97.32	1.20	0.50	0.98
6	93.13	3.24	1.99	1.64
8	87.97	6.19	2.99	2.85
10	82.58	8.06	4.15	5.21

Table 4. Análise da escalabilidade do modelo

Segundo a Tabela 1, o alto valor da % de acidentes, mostra que o comportamento da abordagem Single-Agent é de facto menos seguro. Para provar esta suposição, os autores decidiram introduzir o modelo S-ER106 num ambiente multi-agente. Na simulação, é observado que um agente não está à espera que os restantes abrandem em situações de união de faixa ou cedência de passagem. Isto deve-se ao facto de os agentes tratarem as suas ações individualmente, sem ter em conta as ações dos restantes agentes. Assim, durante os testes era evidente que os agentes mudavam de faixa de forma agressiva, o que levava a mais colisões, refletindo um maior valor de % acidentes.

Assim, este artigo propõe uma abordagem multi-agent DQN com resultados positivos com taxas de sucesso acima dos 97% e mostra alguns comportamentos de negociação sem comunicação explícita, abordando a problemática de mudança de faixa de uma forma mais realista do que os trabalhos com que se compara, que utilizam abordagens single-agent.

2.4 Desafios da comunicação Vehicle to Everything (V2X)

Como é evidente, ao longo de todos os artigos analisados neste artigo de revisão, a comunicação é um ponto essencial na construção de sistemas multiagente, uma vez que a informação tem de ser passada entre os agentes. Com o crescimento desta área, a comunicação V2X é uma tecnologia emergente que irá desempenhar um papel extremamente importante. Os investigadores acreditam que o desenvolvimento de um meio de comunicação wireless através de tecnologias V2X eficientes poderão melhorar de forma significativa o ambiente veicular. Uma das grandes preocupações desta área é, como em qualquer outra tecnologia wireless, os problemas de segurança.

Ghosal, A. e Cont, M. [11] procuram discutir os vários problemas desta área e apresentar possíveis soluções para os combater.

Alguns desafios de segurança e requisitos da V2X:

- **Topologia de Rede Dinâmica:** Os carros movem-se a uma velocidade alta, portanto as comunicações estabelecem-se e cessam num curto período de tempo
- **Escalabilidade da Rede:** O controlo de uma rede tão grande e os problemas de segurança como troca de certificados podem ser uma tarefa muito complicada de atingir
- **Heterogeneidade:** Vários sistemas serão implementados para redes veiculares por todo o mundo. Estes sistemas, desenvolvidos por países diferentes, com políticas diferentes, serão também naturalmente diferentes, podendo ter de ser feita uma sincronização entre todos eles, para que estes possam ainda assim comunicar
- **Latência da comunicação:** É um aspeto crucial da comunicação para que situações que ponham em risco a segurança de um veículo possam ser resolvidas em tempo real
- **Prevenção de ataques:** Para evitar situações que ponham em risco o bom funcionamento das estradas, o correto fluxo do trânsito ou até a segurança dos condutores, é necessário manter a comunicação entre os veículos segura e imune à injeção de informação por ataques.

Requisitos de segurança: Este tipo de comunicação lida com comunicações do tipo V2V e V2I, já mencionadas neste artigo de revisão. Isto implica a intervenção de entidades que são vulneráveis a ataques de redes wireless. Um possível exemplo é um cenário em que um atacante injeta informação falsa num veículo e este age em conformidade com esta informação, podendo colocar em risco a segurança dos viajantes. A injeção de informação falsa poderia também resultar numa grande situação de congestão do tráfego urbano, podendo ter consequências colaterais enormes para o funcionamento de uma metrópole.

Para garantir a segurança destes sistemas, o maior desafio está no desenvolvimento de protocolos seguros para a correta implementação de mecanismos de deteção e defesa contra atacantes. Neste campo, existem vários requisitos de segurança que devem ser cumpridos, tais como:

- **Autenticação:** Implica que o recetor tenha a certeza que a mensagem provém de um remetente genuíno
- **Integridade das mensagens:** Deve ser mantida de tal forma que o conteúdo não seja alterado durante o seu envio
- **Confidencialidade das mensagens:** Não deve ser possível divulgar os conteúdos das mensagens a utilizadores não autorizados
- **Disponibilidade:** Um serviço deve-se manter ativo mesmo que um erro ocorra. Devem ser desenvolvidos protocolos que se consigam reestabelecer após uma falha.
- **Privacidade e anonimidade:** A informação pessoal dos utilizadores da rede deve ser protegida.

Todos estes requisitos, quando cumpridos, irão aumentar a segurança destes sistemas, tornando-os robustos à grande maioria dos ataques que possam ser lançados aos sistemas multiagente, prevenindo situações indesejadas que possam surtir consequências graves para os utilizadores dos mesmos.

3 Discussão e Conclusão

Neste artigo foi evidenciada a larga abrangência da utilização de sistemas multi-agente numa área tão vasta dos veículos e condução autónoma. Foi possível encontrar inúmeros trabalhos, dos quais só alguns puderam ser incluídos e detalhados. Na resolução de problemas de congestão de tráfego urbano foi possível ver duas abordagens contrastantes.

A primeira recorreu apenas à comunicação entre agentes, de um modo simples e intuitivo, utilizando a noção de *platooning* [2]. Já a segunda, explorou o uso de modelos de Reinforcement Learning com uma arquitetura de atuação distribuída, uma abordagem menos intuitiva, no entanto, mais realista nos dias de hoje.

No que toca ao combate dos perigos da condução em condições climáticas adversas foi apresentada uma solução baseada na comunicação Infrastructure to Infrastructure (I2I), focando-se maioritariamente nos componentes da infraestrutura do que nos veículos em si [7]. No entanto, esta solução poderia trazer vantagens adicionais, caso fossem incluídos os veículos como intervenientes na comunicação com a infraestrutura.

Quanto à aprendizagem de manobras de condução, foi sugerida uma abordagem inovadora recorrendo a multi-agent Deep Q-Network, considerando a negociação multi-agente sem negociação explícita [8], em contraste com os restantes trabalhos que utilizam técnicas de single-agent DRL, que vai mais ao encontro dos cenários do mundo real. Este mostrou uma clara evidência de ter desenvolvido um modelo capaz de efetuar a manobra de mudança de faixa de um modo seguro e eficaz. Adicionalmente, este trabalho utilizou a técnica de Transfer Learning que revelou ser altamente eficaz e útil para combater o elevado esforço do treino deste tipo de modelos.

Finalmente, foi apresentado um trabalho cujo objetivo era expor os desafios da comunicação V2X [11], apresentando ainda alguns dos requisitos estritamente necessários ao combate dos mesmos.

Assim, destes trabalhos é possível retirar que, dadas as atuais limitações tecnológicas, uma solução totalmente distribuída será ideal, devido à impossibilidade da comunicação entre 100% dos veículos na estrada, pelo que cada um deverá confiar maioritariamente na sensorização do ambiente. No entanto, considerando que no futuro todos os veículos serão autónomos e farão todos parte de uma rede totalmente conectada, a solução que fará mais sentido nesse momento será uma solução centralizada, uma vez que será possível todos os veículos comunicarem entre si de forma a mitigar a falta de informação dos sensores neles integrados.

References

1. Raul Campos-Rodriguez, Luis González-Jiménez, Francisco Cervantes, Francisco Amezcua-Garcia, and Miguel Fernandez-Garcia. *Multiagent Systems in Automotive Applications*. 09 2017.
2. Levente Alekszejenkó and Tadeusz Dobrowiecki. Intelligent vehicles in urban traffic - communication based cooperation. page 299 – 304, 2019. Cited by: 3.
3. Daniel Krajzewicz, Jakob Erdmann, Michael Behrisch, and Laura Bieker. Recent development and applications of sumo - simulation of urban mobility. 2012.
4. Jiaxun Cui, William Macke, Harel Yedidsion, Aastha Goyal, Daniel Urieli, and Peter Stone. Scalable multiagent driving policies for reducing traffic congestion. volume 1, page 386 – 394, 2021. Cited by: 3.
5. Cathy Wu, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M. Bayen. Flow: Architecture and benchmarking for reinforcement learning in traffic control. *ArXiv*, abs/1710.05465, 2017.
6. Martin Treiber, Ansgar Hennecke, and Dirk Helbing. Congested traffic states in empirical observations and microscopic simulations. *Physical Review E*, 62:1805–1824, 02 2000.
7. Fatma Outay, Stéphane Galland, Nicolas Gaud, and Abdeljalil Abbas-Turki. Simulation of connected driving in hazardous weather conditions: General and extensible multiagent architecture and models. *Engineering Applications of Artificial Intelligence*, 104, 2021. Cited by: 3.
8. Karthikeyan Nagaraajan and Zhong Yi. Lane changing using multi-agent dqn. 2021. Cited by: 2.
9. 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.
10. Alexey Dosovitskiy, German Ros, Felipe Codevilla, Antonio Lopez, and Vladlen Koltun. CARLA: An open urban driving simulator. In Sergey Levine, Vincent Vanhoucke, and Ken Goldberg, editors, *Proceedings of the 1st Annual Conference on Robot Learning*, volume 78 of *Proceedings of Machine Learning Research*, pages 1–16. PMLR, 13–15 Nov 2017.
11. Amrita Ghosal and Mauro Conti. Security issues and challenges in v2x: A survey. *Computer Networks*, 169, 2020. Cited by: 73.