Apprentissage constructiviste à base de systèmes multiagents

Une application au problème complexe de la régulation coopérative du trafic

Maxime Guériau [1], **Frédéric Armetta** [2], **Salima Hassas** [2], **Romain Billot** [4], **Nour-Eddin El Faouzi** [3]

1. *Enable – CONNECT Research Centre, School of Computer Science and Statistics Trinity College, Dublin.*
   maxime.gueriau@scss.tcd.ie
2. *Univ. Lyon, UMR CNRS 5205 LIRIS, F-69622 Villeurbanne, France*
   {frederic.armetta,salima.hassas}@univ-lyon1.fr
3. *LICIT, Univ. Lyon – IFSTTAR, LICIT, F-69675 Bron, France, ENTPE, LICIT, F-69518 Vaulx En Velin, France*
   nour-eddin.elfaouzi@ifsttar.fr
4. *IMT Atlantique, Lab-STICC, Univ. Bretagne Loire, F-29238 Brest, France*
   romain.billot@imt-atlantique.fr

RÉSUMÉ. *Lorsqu'un système autonome évolue dans un environnement complexe, en partie inconnu ou dynamique, il n'est pas possible de fournir une représentation exhaustive a priori facilitant son processus de prise de décision ; cette représentation étant le résultat de l'interaction du système avec son environnement. Pour illustrer ce problème, nous considérons le cas du contrôle décentralisé du trafic coopératif, où une unité d'infrastructure est en charge de réguler localement le flux, en envoyant des consignes aux véhicules connectés. Ce contrôle est le fruit d'une stratégie construite par l'apprentissage d'une représentation précise (états perception-action) des différents états de trafic. Nous proposons un modèle capable, sans connaissance experte, d'utiliser un ensemble de méthodes de classification représentées sous la forme d'une population d'agents et de les combiner dynamiquement pour construire une représentation précise de l'environnement. Cette étude parcourt différents verrous scientifiques à considérer pour qu'un tel système puisse apprendre efficacement. Notre approche s'inscrit dans une démarche*

Revue d'intelligence artificielle – n° 2/2018, 249-277

d'apprentissage constructiviste où la population d'agents construit collectivement une représentation qui exploite, suivant l'usage, les discrétisations possibles de l'espace de perception.

ABSTRACT. *Decision making in autonomous systems is particularly challenging in unknown and changing complex environments, where providing a complete a priori representation is not possible. The so built representation should be the result of the system interactions with the environment. To illustrate the problem, we consider a decentralized control of road traffic, where a control device of the distributed infrastructure locally controls traffic by sending recommendation messages to connected vehicles. We propose an approach able to combine, without prior domain-knowledge, a set of existing traditional unsupervised learning methods that collaborate as a population of agents in order to build an efficient representation. This study addresses the main scientific issues to consider for such a system to efficiently learn. Our approach follows a constructivist learning perspective, where a population of agents is able to collectively build a representation that dynamically combines discretization processes.*

MOTS-CLÉS : *apprentissage constructiviste, intelligence artificielle, système autonome, prise de décision, contrôle.*

KEYWORDS: *constructivist learning, decision-making, control.*

DOI:10.3166/RIA.32.249-277

1. Contexte

La pertinence de la décision prise par un système autonome repose directement sur sa capacité à discriminer ses différentes interactions sensorimotrices, contribuant ainsi à se représenter les informations significatives à considérer. Notons que l'approche présentée dans cet article ne fournit volontairement que peu de connaissances *a priori*, ce qui permet à la représentation de se construire et d'évoluer dans un environnement inconnu, contrairement à ce qui peut être proposé selon une approche plus cognitiviste pour laquelle la représentation est fournie *a priori*. Le problème traité s'attache à la façon de construire une représentation et de la faire évoluer afin de contrôler un système efficacement.

L'aptitude du système à développer une telle autonomie sera corrélée à la qualité des interactions sensorimotrices disponibles ainsi qu'aux spécificités du problème applicatif. Nous verrons dans cet article que les transports intelligents coopératifs (C-ITS) correspondent à des problèmes aux dynamiques complexes pour lesquels la notion d'autonomie et d'adaptation dynamique du contrôle sont essentielles, étant donné la grande diversité des configurations à envisager mais également la diversité des capteurs et effecteurs à considérer. De telles stratégies de contrôle visent des objectifs similaires aux approches classiques de régulation du trafic (voir section 3.1) : fluidifier le flux de véhicules ; mais exploitent le potentiel des technologies impliques dans les C-ITS.

Tel qu'illustré par la figure 1, ces systèmes complexes s'appuient sur les avancées récentes en matière de communication et de technologies de l'information afin d'améliorer l'écoulement du flux de trafic. Dans un futur proche, l'infrastructure routière

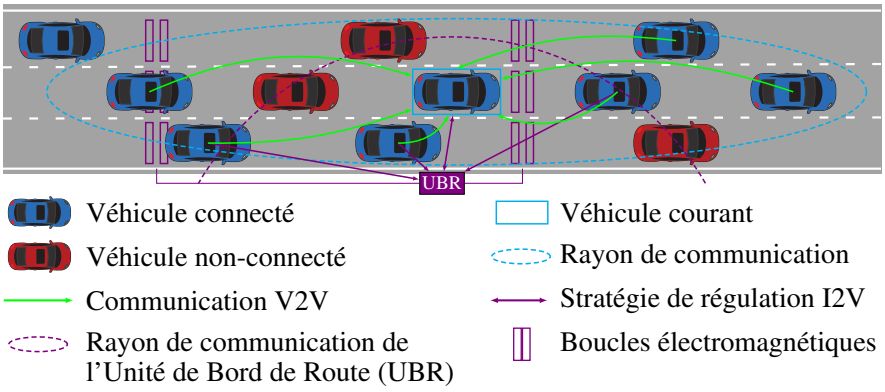


Figure 1. Architecture générale des systèmes de transport intelligents coopératifs

sera partagée par des véhicules connectés et non connectés. Les véhicules connectés peuvent, grâce à un protocole de communication sans fil, échanger des informations entre eux (véhicule à véhicule – V2V) et avec l'infrastructure (infrastructure à véhicule – I2V et véhicule à infrastructure V2I). Dans les C-ITS, les unités de contrôle côté infrastructure, nommées Unités de Bord de Route (UBR), ont pour rôle de collecter des données sur leur section dédiée. Plus que de simples relais d'information, une UBR peut aussi jouer le rôle d'unité de régulation décentralisée du trafic en propageant des consignes aux véhicules connectés par communication I2V.

2. Contributions

Après avoir décrit l'évolution des approches pour la régulation du trafic pour les années à venir dans la section 3, nous nous intéressons en détails au développement de l'autonomie pour ce type d'applications en section 4. Cette analyse synthétise les principales caractéristiques identifiées au cours du travail de thèse de M. Guériau, concernant cet aspect du problème (voir (Guériau, 2016) pour une vue complémentaire concernant l'intégration d'une loi de poursuite multi-anticipative pour la régulation locale du trafic). Après avoir introduit les principes fondateurs du paradigme constructiviste, nous identifions différents verrous scientifiques qui témoignent de la difficulté à développer une telle autonomie d'apprentissage de représentations pour le système, dans un environnement complexe pour lequel l'aspect collectif est prépondérant, et pour lequel au cours de son apprentissage, l'évaluation d'une stratégie de régulation sur la base d'une perception en construction s'avère être particulièrement ambitieuse et doit être réalisée avec beaucoup de précisions. Le modèle proposé est présenté de façon générique dans la section 5. Celui-ci met en œuvre deux processus inter-dépendants : la construction d'hypothèses "d'états perception-action" et l'élaboration d'une "stratégie de contrôle". Ces processus sont réifiés par une population d'agents en compétition afin de construire une représentation précise de l'environnement (ensemble d'états perception-action) sur lesquels se base un processus

de prise de décision (corrélations perception-action). Ce dernier mécanisme est modélisé comme une forme d'apprentissage par renforcement pour lequel le retour sur expérience est fourni par l'environnement et est utilisé pour renforcer les hypothèses d'état perception-action proposées par les agents. L'originalité de notre contribution est d'utiliser un ensemble de représentations concurrentes modélisées sous la forme d'agents autonomes qui s'adaptent dynamiquement grâce à des mécanismes d'intelligence collective, dans le but de produire une représentation émergente d'un contexte dynamique. D'une part, nous sommes attentifs à la capacité à discrétiser (caractériser le périmètre des états à partir des valeurs continues perçues) avec suffisamment de précision les configurations rencontrées, tout en garantissant la convergence de l'apprentissage réalisé. Une instanciation de ce modèle pour le contrôle dynamique du trafic est présenté en section 6. À partir des propriétés observées de ce domaine applicatif, les choix concernant l'instanciation sont présentés et justifiés. On s'intéressera dans cette partie également aux résultats expérimentaux qui permettent d'apprécier la construction de la représentation, l'apprentissage des stratégies de contrôle et leur efficacité. La conclusion et les perspectives sont présentées en section 7.

3. Contrôle et régulation du trafic

L'essor des technologies de l'information et le développement des capteurs côté infrastructure permet aux gestionnaires de surveiller l'état du réseau en temps réel. Le nombre croissant de véhicules évoluant quotidiennement au sein d'un réseau contraint pose la question de la régulation et du contrôle. Ce besoin est motivé par le caractère variable de la demande au sein du réseau (spatialement) et suivant l'heure de la journée (temporellement : les bien connus pics de trafic du matin et du soir).

3.1. Approches classiques de contrôle du trafic

Il existe beaucoup de stratégies de contrôle du trafic routier, qui sont étudiées du côté des gestionnaires d'infrastructure (Papageorgiou *et al.*, 2003). La grande majorité des travaux s'intéresse à un réseau maillé où la cause de la congestion est bien souvent le réseau lui-même. Les informations proviennent de capteurs (boucles électromagnétiques, caméras) permettant d'assurer une surveillance (*monitoring*) du réseau par un opérateur, ou un système intelligent. Le moyen d'application de la stratégie de contrôle choisie passe par la transmission d'informations à l'ensemble du flux (Panneau à Message Variable) ou en modifiant directement les caractéristiques du réseau (fermeture d'accès, régulation du cycle de feux, etc.). La tendance récente oriente les approches vers une gestion dynamique, où la stratégie repose sur une représentation proposée par un expert. C'est par exemple le cas pour la régulation dynamique des vitesses (Khondaker, Kattan, 2015).

La performance de ces approches repose essentiellement sur les connaissances expertes intégrées dans les représentations manipulées par les systèmes de contrôle. Cela permet de proposer des systèmes rapidement opérationnels, mais l'écart observé entre les modèles issus de la théorie du trafic et les comportements hétérogènes des

conducteurs ne garantit pas forcément une stratégie de contrôle optimale. De plus, les moyens d'application des stratégies classiques se limitent principalement à des actions ciblées sur l'ensemble d'un flux d'une section contrôlée. Dans certains cas comme pour le contrôle d'accès (où il s'agit de gérer les flux entrant sur une section), on favorise une partie du flux au détriment des usagers circulant sur d'autres parties du réseau. La difficulté réside dans la propagation des perturbations. L'origine d'une perturbation ne peut pas toujours être maîtrisée et le contrôle d'une section peut avoir des incidences néfastes sur une autre portion du réseau.

3.2. Contrôle et régulation des C-ITS

Face à ces limitations, les C-ITS permettent d'agir sur les deux niveaux : en apportant plus d'informations permettant une description plus fine de l'état de trafic, mais aussi en donnant des moyens d'action plus précis permettant de transmettre des consignes plus ciblées. La complexité de ces systèmes génère des problématiques proches de l'intelligence artificielle distribuée (IAD). C'est pourquoi on retrouve beaucoup d'approches qui utilisent des techniques issues du domaine des systèmes multi-agents (Bazzan, 2009). En parfaite analogie avec les problématiques issues des Systèmes Multi-agents (SMA), la question qui se pose relève du type de contrôle à appliquer. Les travaux récents montrent une tendance allant vers la décentralisation des approches classiques (Farhi *et al.*, 2015). D'autres approches proposent des stratégies de contrôle locales, intégrées à la prise de décision du véhicule (Kesting *et al.*, 2008; Guériau *et al.*, 2015). Ces approches sont cependant plus difficiles à maîtriser pour répondre à des objectifs de haut niveau (*i.e.* à l'échelle du système).

La plupart des travaux existants (Dresner, Stone, 2004; Vasirani, Ossowski, 2009) se focalisent une nouvelle fois sur la gestion des intersections signalisées (*i.e.* carrefours équipés de feux tricolores). Les concepts multi-agents permettent de reproduire des comportements de coalition permettant de favoriser le passage de pelotons homogènes de véhicules. Le moyen d'action (*i.e.* l'application du contrôle) se limite principalement aux limitations de vitesse dynamiques (Zhu, Ukkusuri, 2014). Il est aussi possible de directement modifier les comportements individuels en fonction de l'état de trafic (Schönhof *et al.*, 2007). Conformément à notre hypothèse, le cas d'étude d'un scénario autoroutier est aussi représenté dans les travaux portant sur le contrôle des véhicules connectés (Talebpour *et al.*, 2013) et les approches basées sur une forme d'apprentissage se développent : par exemple Schmidt-Dumont, Vuuren (2015) appliquent un algorithme de type *Q-learning* à un système de contrôle d'accès mais ne tirent pas profit de la communication entre véhicules. L'intérêt de la communauté pour ces approches reste récent mais est en forte croissance depuis quelques années, suivant le rythme de déploiement des véhicules intelligents.

4. Vers une approche constructiviste pour le contrôle des C-ITS

Les limitations des approches de régulation classiques reposent essentiellement sur la dépendance du système à une représentation experte (souvent couplée avec une base

de règles elle-même experte). L'utilisation de formes diverses d'apprentissage semble être une alternative pour donner au système la capacité de construire mais également de faire évoluer en ligne sa représentation. Cette possibilité est en adéquation avec la dynamique de déploiement des systèmes coopératifs, qui nécessite à la fois de nouvelles stratégies de contrôle applicables aujourd'hui, mais qui demanderont aussi une adaptation future. Le mode d'application des stratégies est aussi un axe d'amélioration des approches existantes, qui se limitent actuellement souvent à une seule forme de contrôle même si les consignes propagées sont de plus en plus personnalisées. Enfin, les stratégies décentralisées, qui permettent d'intégrer une dimension d'autorité (l'unité de contrôle) tout en garantissant de mener à bien des objectifs à grande échelle, semblent être la forme de contrôle la plus adaptée aux systèmes coopératifs.

L'approche présentée dans cet article est vouée à être intégrée dans une Unité de Bord de Route, modélisée comme un agent autonome, et dont le processus de décision sera un système décentralisé de contrôle des C-ITS capable, par apprentissage, de construire sa représentation de l'état de trafic courant et d'optimiser les consignes à envoyer aux véhicules (*i.e.* choisir les consignes permettant de fluidifier le trafic, ou simplement maintenir la stabilité du flux). Ces consignes ne se limiteront pas à un seul paramètre (la vitesse) mais permettront aussi aux véhicules d'adapter d'autres variables de comportement (inter-distance, et changement de voie).

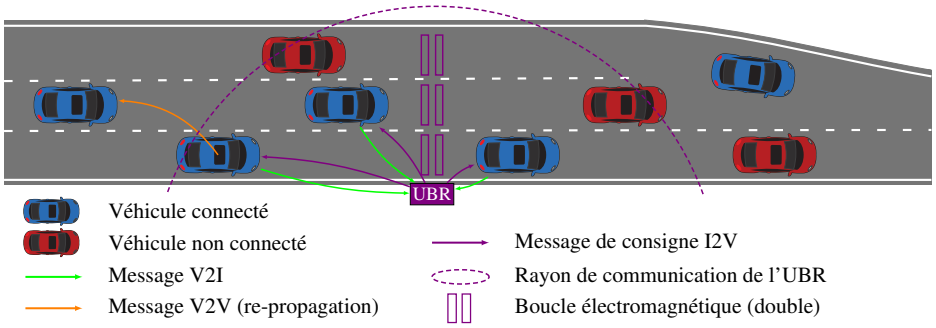


Figure 2. Illustration d'un scénario de contrôle décentralisé des systèmes coopératifs

La figure 2 représente un scénario de contrôle des véhicules connectés par une UBR autonome. L'objectif de ce système de contrôle est d'améliorer l'écoulement du flux. Pour cela, il doit construire une représentation de l'état de trafic courant et apprendre quelles consignes sont les plus pertinentes dans ces différents contextes. Les interactions dans l'environnement sont rendues complexes par l'ajout de la communication entre les véhicules connectés. La proportion de véhicules équipés est une nouvelle variable à prendre en compte dans la décision de l'UBR pour obtenir une stratégie de contrôle robuste. Pour cela, l'UBR dispose de plusieurs capteurs lui permettant de percevoir l'environnement. Les approches classiques ne suffisent plus à décrire précisément l'état de trafic courant : les nouveaux comportements des véhicules

connectés et autonomes doivent être intégrés dynamiquement. L'impact de nouveaux systèmes (comme un nouveau véhicule autonome) à l'échelle du système ne sera pas évalué *a priori* mais observé lors de son déploiement. L'objectif de notre approche est ainsi de permettre au système de construire ces états et les actions associées, qui feront partie de sa représentation de l'environnement. La difficulté de cette tâche résulte fortement des spécificités suivantes de l'environnement des véhicules connectés :

– **continuité spatio-temporelle** : l'environnement physique est continu et le réseau routier complexe, tout en étant le support de dynamiques à la fois court-terme (communication) et plus long-terme (congestion). Le système de contrôle doit être capable de gérer un flot continu de données tout en détectant des événements récurrents ou non, quelle que soit leur dimension temporelle.

– **indéterminisme et évolutivité** : l'environnement physique, notamment l'ensemble des états accessibles, ne sont pas connus *a priori*. Le système de contrôle doit être capable de faire évoluer sa représentation alors que ces objectifs peuvent être mis à jour (de nouveaux usages et besoins vont apparaître).

Ces caractéristiques définissent un type de problème particulier, qui est visé par les approches constructivistes.

4.1. Le constructivisme

La construction de la représentation est un problème fondamental dans l'élaboration de toute stratégie de contrôle. Il s'agit d'un problème cognitif qui fait intervenir toute la chaîne d'interaction de l'agent. Les approches constructivistes visent à reproduire ce comportement itératif d'élaboration de la représentation. Elles s'inspirent des travaux précurseurs de Piaget (1955), en sciences cognitives, décrivant le processus d'acquisition de connaissance lors de la phase de développement chez l'enfant. Ces travaux ont orienté la réflexion portant sur les systèmes autonomes et plus particulièrement les conditions menant à l'autonomie (Zlatev, Balkenius, 2001) :

– le système doit être incarné (*embodiement*) ;

– il doit être situé dans un environnement physique et social ;

– il doit exhiber des mécanismes de développement épigénétiques à travers lesquels des structures plus complexes émergent du résultat des interactions avec l'environnement social et physique.

La notion d'épigénétique a été proposée par Piaget. Elle désigne le développement d'un individu grâce à ses interactions successives avec l'environnement. Cette théorie a donné lieu à la proposition de méthodes permettant de construire une représentation sous la forme de schémas sensorimoteurs (Drescher, 1991), qui représentent les interactions par des triplets contexte-action-résultat. Ce concept a été largement repris par la communauté d' Intelligence Artificielle (IA) (Guerin, 2011). Les applications de ces concepts se retrouvent dans le domaine de l'intelligence ambiante (Najjar, Reignier, 2013; Mazac *et al.*, 2014) ou plus majoritairement sur une variante de la robotique : la robotique développementale (Lungarella *et al.*, 2003; Meeden, Blank, 2006; Asada

et al., 2009; Oudeyer, Smith, 2016). L'approche la plus prometteuse dans le domaine de l'IA développementale semble consister à reproduire le processus d'acquisition de la connaissance dans les différentes phases du développement de l'humain (Mugan, Kuipers, 2007). Ce type d'approche se heurte à une difficulté majeure qui est la complexité de l'environnement, car le système d'apprentissage doit être confronté à un monde physique raisonnablement réaliste (Guerin, 2011).

En repartant du concept de base des schémas sensorimoteurs proposés par Piaget (1955), le processus d'apprentissage est un processus complexe. Le fonctionnement interne d'un système constructiviste (son processus d'apprentissage) peut ainsi être décomposé en deux processus couplés : un processus d'identification du contexte, qui permet au système de discriminer les situations (ou l'état de l'environnement) et un processus de prise de décision, qui permet au système de choisir une action à entreprendre dans le contexte identifié. Ces deux processus sont liés (puisqu'ils reflètent ensemble l'expérience des interactions sensorimotrices du système) et font ainsi partie du processus plus global qui mène à l'élaboration d'une représentation de l'environnement du système. L'ensemble de ce processus global décrit le cadre d'un apprentissage constructiviste développé en détail dans la suite de cet article.

4.2. Problématique et verrous scientifiques

Notre travail est avant tout motivé par des problématiques émanant du contrôle de systèmes complexes, tels que les C-ITS, où le caractère évolutif de l'environnement nécessite de faire appel a un contrôle adaptatif et dynamique, qui pourrait émaner d'un apprentissage en ligne. Nous souhaitons proposer un modèle capable de se substituer à la prise de décision d'un agent autonome (UBR) afin de générer dynamiquement des stratégies de contrôle (une représentation et sa politique d'action) permettant de contrôler un flux perçu par l'intermédiaire de capteurs. En faisant preuve d'un peu d'abstraction, notre problème peut être défini de manière plus générique. Nous souhaitons proposer un processus de prise de décision d'un agent autonome, situé dans un environnement physique, capable de percevoir partiellement cet environnement par ses capteurs, et cherchant une stratégie d'action visant à maximiser des objectifs globaux. L'étude des travaux existants nous a mené à nous intéresser au paradigme constructiviste et à ses propres problématiques.

4.2.1. L'apprentissage dans les systèmes multi-agents (SMA)

L'introduction de mécanismes d'apprentissage dans les Systèmes Multi-Agents peut être réalisée à l'échelle individuelle ou collective. Il est possible de considérer le SMA comme un système d'apprentissage, où chacun des agents est vu comme une entité dotée de capacités d'apprentissage individuelles d'objectifs également individuels. Dans ce cas, les méthodes classiques d'apprentissage peuvent être directement appliquées et permettent d'améliorer le comportement individuel des agents, de catalyser leurs interactions, ou encore de faire émerger des organisations. Les agents peuvent par exemple apprendre à coordonner leurs actions (Wei, 1993). Lorsque c'est le système global qui réalise l'apprentissage, les interactions des agents sont le support de

la construction de la représentation du système. L'objectif est défini pour le système, l'interprétation de l'efficacité des apprentissages est difficile sachant la complexité du système ainsi formé. Il est ainsi possible d'imaginer que les agents eux-mêmes font partie de la représentation du système (Mazac *et al*., 2014), et que leur organisation a une incidence sur le comportement global du système (et donc sa politique). Tous les agents du système ne sont pas forcément dotés de capacité d'apprentissage, le comportement d'un groupe d'agents peut être appris par un seul agent spécialisé. À l'inverse, chaque agent peut être spécialisé dans une sous-tâche de l'apprentissage du système. Panait, Luke (2005) nomment ces types de topologie apprentissage pour l'équipe (*team learning*) et apprentissage concurrent (*concurrent learning*).

Au-delà de l'aspect organisationnel décrit jusqu'à présent, nous attirons l'attention sur la façon de faire évoluer l'apprentissage de façon concurrente. Dans le cadre de l'apprentissage concurrent, le problème peut être projeté dans autant de sous-espaces que le système comporte d'agents apprenants ; ce qui est un avantage manifeste pour les problèmes très complexes ou décrits dans des espaces à grand nombre de dimensions. Cette forme d'apprentissage offre des possibilités intéressantes, telles que l'instanciation naturelle d'une mécanique d'exploration et d'exploitation au sein de la population d'agents, ou encore la possibilité de maintenir des politiques contradictoires, et de les activer suivant les contextes rencontrés. Le problème principal que rencontre ce type d'approches dans le domaine de l'apprentissage multi-agent est qu'il est difficile de s'assurer de la performance du comportement des dynamiques système (Bloembergen *et al*., 2015). Ce verrou reste difficile à lever, malgré les avantages possibles à utiliser les méthodes récentes du domaine des SMA. Par exemple, si l'on revient à la notion de représentation, sa construction et son évolution pourraient être déléguées à une population d'agents dont la prise de décision intègre des mécanismes classiques d'apprentissage (comme la classification non supervisée – *clustering*). L'apport de mécanismes concurrents tels que décrits dans (Graczyk *et al*., 2010) (*stacking, boosting, bagging*) est d'ailleurs un avantage indéniable, qui pourrait être intégré par des solutions multi-agents.

Les méthodes d'apprentissage multi-agent sont séduisantes pour la résolution de problèmes complexes et/ou prenant place dans des environnements avec beaucoup de dimensions. L'apprentissage par renforcement permet de donner un objectif global au système et de guider son processus. Dans le modèle présenté dans cet article, nous proposons de tirer profit d'une forme d'apprentissage concurrent, afin d'améliorer la performance du système. Cela nous confronte donc aux problèmes évoqués précédemment, qui se rapportent à des verrous scientifiques en apprentissage multi-agents, et de façon plus générale relèvent du problème de l'élaboration de toute représentation cognitive.

4.2.2. Le problème du feedback : individuel ou collectif ?

Le problème du *feedback* (ou encore d'attribution de la récompense) est un problème classique en apprentissage. En effet, l'objectif d'un système reposant sur une population d'agents est souvent éloigné des objectifs individuels. Pourtant, l'appari-

tion d'une synergie au sein du système repose essentiellement sur des comportements individuels précis, et exécutés au même moment. En d'autres termes, le problème est de s'assurer que les objectifs individuels des agents seront corrélés avec l'objectif global du système. L'apprentissage par renforcement offre la possibilité au système d'évaluer sa politique en temps réel grâce à une récompense instantanée. Le concepteur peut alors proposer une fonction de récompense qui dépend des objectifs de haut niveau définis pour le problème, ce qui est un avantage pour orienter le comportement du système. Si les mécanismes reposent sur des comportements individuels, cette récompense risque de ne pas guider suffisamment rapidement les agents dans leur propre apprentissage. Dans ce cas, il convient de décomposer le problème de l'attribution de récompense en deux sous-problèmes (Alonso *et al*., 2001) :

– **attribution** de récompenses **inter-agent** : récompenser les agents sur la base de la performance globale du système. Il s'agit d'une tâche difficile car il est nécessaire d'estimer l'apport de chacun des agents dans le fonctionnement global du système.

– **attribution** de récompenses **intra-agent** : récompenser les agents sur la base de leur comportement individuel. Plus simple au premier regard, cela consiste à récompenser les actions individuelles qui peuvent mener à des bonnes performances du système, mais on retrouve le lien complexe interactions-émergence qui n'est pas toujours explicite (Bloembergen *et al*., 2015).

Une approche naïve consisterait à décomposer équitablement les récompenses pour tous les agents. Cela n'est possible que pour des systèmes simples ou les entités sont toutes identiques. Lorsque les agents sont spécialisés, le problème de l'attribution de récompenses est crucial. Il faudra donc proposer des formes de *feedback* à la fois internes aux agents, et aussi entre agents afin d'orienter le comportement du système vers l'objectif global. Pour cela, le système devra être capable de décrire précisément le contexte auquel il est confronté.

4.2.3. Le problème de l'élaboration de représentations concurrentes

L'adaptation est une des composantes les plus importantes de l'autonomie d'un agent. Cette capacité permet à un système autonome de modifier dynamiquement sa représentation en fonction des changements de son environnement. Ce processus peut être réalisé en exploitant des représentations concurrentes. Face aux bénéfices que l'on peut attendre de l'apprentissage multi-agent concurrent, plusieurs verrous restent encore non résolus. Fulda, Ventura (2007) identifient trois facteurs qui peuvent mener à un comportement peu performant du système (équilibre de la sélection, comportements individuels, masquage des actions et activation).

4.2.3.1. Le problème d'équilibre de la sélection

Le problème d'équilibre de la sélection (*equilibrium selection problem*) survient lorsque le système fait face à plusieurs solutions optimales. Dans un cadre multi-agent, cela signifie que plusieurs agents proposent des solutions optimales (ou assimilées) pour le système. Ces solutions peuvent ne pas couvrir l'intégralité de l'espace de perception du système, le système peut se stabiliser sur une solution partielle en préfé-

rant un seul agent au détriment des autres. Dans ce cas, le fonctionnement général du système n'est pas forcément sous-optimal, mais les agents peuvent avoir des problèmes dans leur coordination (s'ils convergent vers des solutions différentes). Ce problème émane directement de la définition des récompenses (et leur attribution) dans les problèmes d'apprentissage par renforcement. Ils peuvent donc être résolus en agissant sur le *feedback* du système (section 4.2.2), à condition de disposer d'un nombre d'interactions suffisant. Le problème reste ouvert si le système souhaite assurer une performance donnée dans un délai qui lui est accessible. Dans le cadre du contrôle des systèmes coopératifs complexes, un grand nombre de cycles d'apprentissage est nécessaire. Il est possible d'augmenter le nombre d'expériences sensorimotrices du système de contrôle en émulant une partie des interactions avec un simulateur (sous réserve que l'environnement réel puisse être simulé fidèlement). La complexité de l'environnement implique néanmoins de s'assurer que les mécanismes concurrents soient suffisamment compatibles pour ne pas altérer la convergence de la représentation du système.

4.2.3.2. Comportements individuels peu performants

Avoir des comportements individuels peu performants (*poor individual behavior*) est possible au sein d'une population d'agents. Il est alors nécessaire de mettre en place des mécanismes permettant de limiter l'apport des agents incriminés ou au contraire d'améliorer leur performance. La tâche la plus critique étant d'identifier les agents nuisant au bénéfice du système (ou au comportement des autres agents). La définition stricte de l'apprentissage par renforcement se base sur un environnement Markovien où l'état futur du système ne doit dépendre que de son état courant et non pas de la séquence d'événements qui l'ont précédé. Cette propriété n'est pas vérifiée puisqu'il est très difficile de proposer un comportement individuel optimal pour chacun des agents (dont le comportement est basé sur celui des autres) et qui couvre tout l'espace des états. Néanmoins, les propriétés des SMA (robustesse, auto-organisation) peuvent être mises à profit pour combler les faiblesses individuelles. Cela ne garantit pas que l'intégralité du problème sera traitée, car le comportement global peut être performant sur une partie du problème seulement. Le problème restant se situe alors au niveau de l'expressivité de la représentation manipulée par le système.

4.2.3.3. Masquage des actions

Le problème du masquage des actions (*action shadowing*) n'apparaît qu'avec les besoins de coordonner un groupe d'agents (avec un seul agent, il n'y a théoriquement pas de conflit). On parle de masquage des actions lorsque le comportement d'un agent (qui propose une action pas forcément optimale) est préféré par le système alors que d'autres actions (proposées par d'autres agents) potentiellement plus performantes sont possibles. Le problème est assez critique puisque dans ce cas, la performance du système est directement impactée sans que le système ne puisse l'améliorer. Une autre façon de voir les choses est d'imaginer que le système fait face à deux problèmes similaires mais dont les politiques optimales sont nécessaires. Garder la même politique entraîne une performance moyenne plus faible. Le système doit donc adapter sa repré-

sentation aux deux sous-problèmes. Ce problème peut être résolu en dotant le système de capacités d'adaptation, lui permettant de réutiliser des connaissances pour d'autres problèmes. Cependant, le nombre de sous-problèmes peut être conséquent, et donc rendre malgré tout difficile l'évaluation des politiques.

4.2.3.4. Problème de l'activation

Grâce à l'utilisation de représentations concurrentes, on peut distinguer deux niveaux de représentation : la représentation individuelle de chacun des agents et la représentation globale, à l'échelle du système, qui est une forme émergente des représentations individuelles. Cela permet plus de robustesse et offre plus de possibilités dans la sélection et l'application de la politique. L'apprentissage peut aussi être introduit aux deux niveaux. L'apprentissage des agents leur permet d'assimiler des connaissances adaptées à de nouveaux contextes ; et l'apprentissage au niveau du système permet de sélectionner la bonne représentation (ou la combinaison des représentations individuelles) qui mènera aux récompenses optimales. Conceptuellement, un tel système pourrait évoluer face à un environnement dynamique, et faire évoluer sa représentation de l'environnement tout au long de son cycle de vie, en cherchant tout le temps à s'améliorer en s'adaptant. Malgré ces avantages attendus pour modéliser les mécanismes d'accommodation (adaptation des représentations aux novelles données) et d'assimilation (intégration de nouvelles données dans les représentations exisantes), il reste un problème à considérer : quels sont les mécanismes minimaux et données minimales nécessaires au système pour réaliser cette forme d'apprentissage ? Comment gérer dynamiquement l'utilisation des représentations concurrentes pour s'adapter au contexte ?

Dans le modèle conceptuel présenté dans la section 5, nous proposons de répondre à ces verrous à travers un mécanisme d'activation des représentations concurrentes. Le challenge est alors de proposer un processus capable de maintenir la performance de la représentation du système, sans retomber dans les problèmes précédents : masquage des actions qui peut survenir dans le cas de comportements individuels peu performants ou en cas de déséquilibre.

5. Construction d'une représentation par apprentissage concurrent

L'objectif à long terme de notre modèle est de concevoir un système de prise de décision dynamique capable de définir les frontières de chaque état perception-action, afin d'identifier précisément le périmètre de chaque action étant donné un ensemble de perceptions. En effet, un tel système serait capable de mieux comprendre les dynamiques de l'environnement et aurait une probabilité plus élevée de sélectionner l'action la plus pertinente dans le contexte actuel estimé. Plusieurs problèmes sont à traiter afin d'arriver à un tel résultat. D'abord, il est nécessaire de choisir quelles sont les connaissances de base à fournir au système sans pour autant incorporer trop de connaissances expertes dans sa représentation. Ensuite, le périmètre des actions peut ne pas être précisément défini et figé selon le problème. La difficulté est de trouver la meilleure façon de modéliser cette notion de périmètre sans pour autant limiter

l'ensemble des états possibles. Enfin, le défi le plus intéressant se situe au niveau du renforcement de la représentation, qui doit mener à un contrôle efficace de l'environnement par le système. Nous proposons un modèle conceptuel générique dont les mécanismes généraux traitent les problématiques précédentes, et qui peut être facilement adapté à d'autres applications.

5.1. Aperçu du modèle

Le modèle proposé s'appuie sur des mécanismes d'intelligence collective afin d'assister le système dans sa tâche itérative de construction de sa représentation de l'environnement. Le système est un agent autonome qui exploite un SMA interne régissant son comportement (résultant de son interaction avec l'environnement). Cette tâche d'apprentissage prend la forme de deux processus couplés, perception et décision, comme illustré dans la figure 3. Les données d'entrée, collectées par les cap-

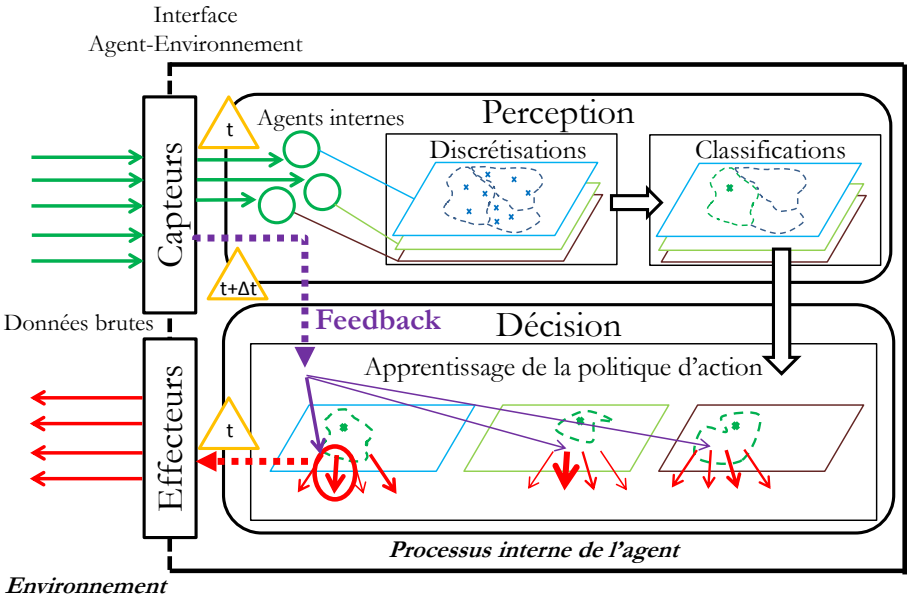


Figure 3. Aperçu du modèle, décrivant le processus global d'apprentissage du système (l'agent) en interaction avec son environnement

teurs, représentent des observations de l'environnement et constituent la perception de bas-niveau du système. Ces données sont continues et non-identifiées dans la mesure où aucune information experte n'est communiquée au système au sujet de la pertinence ou de l'importance de chaque variable perçue. Cette perception est transmise aux agents "discrétiseurs", dont le rôle est de fournir des abstractions à partir de ces informations. Ces abstractions correspondent à des façons alternatives d'interpréter le signal d'entrée. Chaque agent applique une stratégie de discrétisation interne. Les

discrétisations générées peuvent évoluer en temps réel pour adapter le périmètre des (états) classes que l'agent propose. Plusieurs mécanismes peuvent améliorer la précision des discrétisations proposées par les agents. Le processus multi-agent produit un ensemble de partitions concurrentes pour le système correspondant à différentes représentations possibles. La sélection de l'action repose sur un apprentissage par renforcement qui fait intervenir toutes les représentations concurrentes. Le modèle,

Algorithme 1 Processus général du modèle

Instancier l'ensemble A des actions possibles
Instancier l'ensemble D des agents discrétiseurs
Instancier l'ensemble S des états sélectionnés concurrents (vide)
répéter *# boucle qui représente le processus de décision du système*
  si S n'est pas vide **alors** *# bloc non exécuté à la première itération*
    $F_t = feedback(P_{t+1})$
    pour tout $s_i \in S$ **faire** *# récupérer le lien L entre l'état s_i et l'action a^**
      $L = getPercepActionLink(s_i, a^*)$
      $incrementReward(L, F)$
    fin pour
    $reinforce(D, a^*, F)$
  fin si
  $P_t = getPerception()$ *# début du processus de sélection de l'action*
  $S.removeAll()$
  pour tout $d_i \in D$ **faire**
    $p_i = discretizerPerception(P_t)$ *# récupération de la perception de l'agent*
    $D_i.discretize(p_i)$
    $s_i = D_i.classify(p_i)$
    $S.add(s_i)$ *# ajout de l'hypothèse d'état s_i à l'ensemble S*
  fin pour
  $a^* = actionSelection(S)$
  $execute(a^*)$
fin répéter *# fin du processus de décision avant la prochaine perception (et feedback)*

présenté dans l'algorithme 1, vise à combiner dynamiquement les états perception-action précédemment appris afin de construire une représentation encore plus précise de l'environnement.

5.2. Perception de haut niveau

Le système utilise des représentations individuelles proposées par ses agents discrétiseurs pour construire une représentation de plus haut niveau. Chaque discrétiseur exploite son processus de classification exécuté dans son propre espace de perception, qui peut varier d'un agent à l'autre. Cela offre la possibilité d'évaluer des représentations dédiées à des canaux spécifiques de la perception. Ainsi, les discrétiseurs peuvent

percevoir des données à partir des mêmes capteurs ou utiliser des dimensions différentes à l'échelle de la perception du système. La classification permet d'associer une perception (espace continu et multi-dimensionnel) à un état discret, et donc de réduire la taille de l'espace d'entrée du système. Nous proposons de décrire le modèle de façon générique, ce qui permet de repousser le choix de la méthode de classification à l'étape d'implémentation du modèle. Les algorithmes en-ligne ou hors-ligne peuvent être utilisés à la conditions qu'ils soient compatibles avec le type des données d'entrées et qu'ils permettent de générer une représentation sous la forme d'un ensemble d'états discrets. L'ensemble des variables perçues par chacun des agents peut être défini de manière experte ou initialisé aléatoirement, en fonction du problème visé.

Le processus de perception du système peut se résumer en deux étapes consécutives (étapes 1 et 2), telles que représentées dans la figure 4. Chaque fois que le système

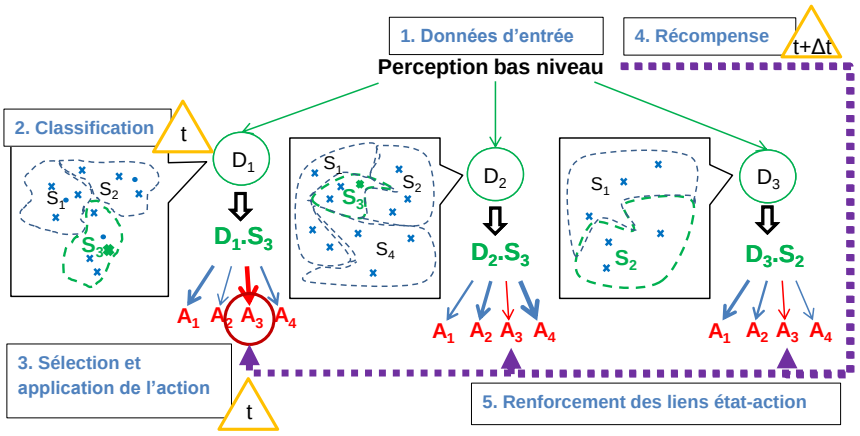


Figure 4. Processus de discrétisation, de classification et de sélection d'action

perçoit son environnement, tous les agents discrétiseurs D mettent à jour en parallèle leur propre espace d'hypothèses d'état, en fonction de leur méthode de classification choisie (K-means, GNG (Fritzke, 1995), etc.). Par exemple, la discrétisation de l'agent D_i est composée de l'ensemble des hypothèses d'état : $\{S_1, S_2, ..., S_j\}$. Pour chaque pas de temps de perception du système, chaque agent D perçoit ses variables d'entrée et classe sa perception comme une hypothèse d'état discret. A partir de l'ensemble des hypothèses d'état (une par agent), le système tente de choisir l'action la plus pertinente dans le contexte défini par les différents agents.

5.3. Contrôle de l'environnement

Les discrétiseurs mémorisent la probabilité de sélection de chaque action lorsqu'ils classent leur perception en tant qu'une hypothèse d'état donnée. Ces probabilités sont modélisées sous la forme de liens état-action qui permettent de garder une trace au niveau du système du résultat des actions consécutives, évalué à travers le retour sur

expérience (*feedback*) fourni par l'environnement. A l'issue de l'étape de perception, tous les discrétiseurs ont sélectionné une de leurs hypothèses d'état en réponse à la perception de bas-niveau propagée. Le système doit désormais choisir une action qui soit la plus pertinente possible en fonction des différentes hypothèses proposées (une par agent). Le système doit alors trouver un compromis entre l'exploitation (appliquer l'action générant le meilleur *feedback*) et l'exploration (tester des actions non utilisées jusque là). Ce choix est illustré par les étapes 3 à 5 dans la figure 4.

Le système fait face à un dilemme classique d'exploration-exploitation (Auer *et al.*, 2002), puisque l'espace des actions est supposé fini et qu'il n'y a pas de corrélation évidente entre des hypothèses d'état distinctes. Cette situation fait écho au problème dit du bandit à n bras, relatif à l'élaboration d'une stratégie optimale de sélection et de test de différentes machines à sous dans un casino. Plusieurs stratégies, disponibles dans la littérature, peuvent ainsi être utilisées dans notre modèle. La plupart utilisent un paramètre objectif à minimiser basé sur la notion de regret, exprimé comme la différence entre la récompense courante et la récompense optimale (théoriquement inconnue). Notre modèle ne nécessite pas que la stratégie soit optimale dans la mesure où la sous-exploration de l'espace des actions risque de nuire aux actions prises par le système. En effet, les agents discrétiseurs travaillent conjointement pour produire une représentation à l'échelle du système mais le système peut agir sur ces représentations individuelles dynamiquement, ce qui limite grandement l'accès à une forme de récompense optimale pour les agents discrétiseurs. Néanmoins, cette hypothèse dépend encore une fois du type de problème visé et pourra être étudiée en comparant différentes implémentations du modèle.

Le retour sur expérience fourni au système fait partie de sa perception de bas-niveau et doit être défini lors de la formalisation du problème (*i.e.* pour l'implémentation). Cette fonction de *feedback* permet une évaluation par le système du résultat de ses actions dans l'environnement. Cette valeur est utilisée comme une récompense pour l'algorithme d'exploration-exploitation et contribue au renforcement de la représentation construite. Par exemple, pour les algorithmes d'exploration-exploitation simples, la somme des récompenses obtenues lors de la sélection de chaque état discret, ainsi que le nombre de sélection respectif de chacune des actions, sont enregistrés par l'agent discrétiseur concerné. Ces informations sont portées par les agents sous la forme de liens perception-action, qui, sous une définition générique, permettent de mémoriser ce genre d'information ou tout autre donnée utile au processus d'apprentissage par renforcement effectivement utilisé lors de l'implémentation du modèle.

Le résultat de la sélection est un lien état-action (étape 3 dans la figure 4) dont chaque état est directement lié à un agent discrétiseur. Ainsi, le processus de renforcement (étapes 4 et 5) contribue à l'apprentissage du système, en l'aidant à identifier les hypothèses d'état les plus précises parmi l'ensemble des discrétisations proposées par la population d'agents discrétiseurs.

6. Application : stratégie I2V de contrôle des C-ITS

Dans cette section, nous proposons un scénario en simulation permettant d'illustrer une utilisation du modèle dans le contexte d'un flux de véhicules connectés sur une section d'autoroute.

6.1. Cadre expérimental

Notre étude se focalise sur un scénario relativement simple d'un point de vue modélisation, mais qui présente un convergent introduisant une grande part d'incertitude dans l'écoulement du flux. L'impact des changements de voies est en effet difficile à mesurer, bien que la communauté s'accorde sur l'effet néfaste sur le flux, observé directement (Geroliminis *et al.*, 2011) ou en simulation (Leclercq *et al.*, 2016). Cette chute de capacité, bien que modélisable, dépend de beaucoup de paramètres individuels auxquels il est difficile voire impossible d'accéder (Zheng *et al.*, 2013) : anticipation, phénomène de relaxation, etc. Les résultats expérimentaux sont menés dans le simulateur MASCAT (Guériau *et al.*, 2015, 2016) : *the Multi-Agent Simulator for Connected and Automated Traffic*. Les expérimentations prennent place dans un scénario représentant une section autoroutière parcourue par un flux mixte composé de véhicules connectés et non connectés. Les véhicules connectés peuvent échanger des informations avec les autres véhicules et l'infrastructure (V2V et V2I), mais notre scénario se concentre surtout sur les interactions de l'infrastructure (messages I2V). L'infrastructure est composée d'unités de bord de route (UBR) en charge du contrôle de leur section dédiée. Une UBR est capable de percevoir des informations afin d'estimer l'état de trafic actuel par l'intermédiaire de ses capteurs. La chaussée est par exemple équipée de boucles électromagnétiques qui agrègent les informations de débits et de vitesse au cours du temps. L'objectif d'une UBR est de propager des messages de recommandations aux véhicules connectés à portée. Cet envoi de message est l'application d'une stratégie de contrôle autonome qui tend à réduire l'ampleur ou retarder l'apparition de congestion.

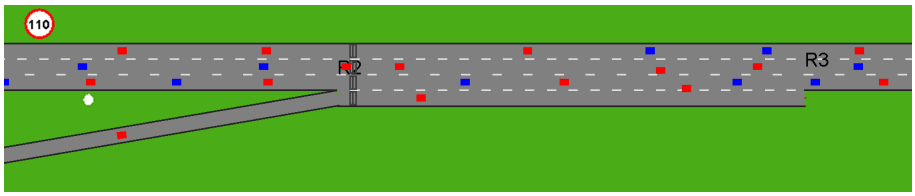


Figure 5. Capture d'écran du scénario de contrôle modélisé dans le simulateur. Le flux est composé de véhicules connectés (bleu) et non connectés (rouge). L'unité de bord de route (disque blanc) récolte les informations des boucles électromagnétiques (rectangles gris) pour construire par apprentissage, sa représentation de cet environnement autoroutier

Comme illustré dans la figure 5, toutes les expérimentations ont lieu sur le même scénario, qui se déroule sur une portion rectiligne de 3 km d'une autoroute à trois voies avec une voie d'insertion à 2,5 km. L'Unité de Bord de Route perçoit des données agrégées de flux (débits) et de vitesses grâce aux boucles électromagnétiques présentes sur les trois voies. Les véhicules connectés partagent à l'UBR leurs vitesses dans son rayon de communication (150 mètres). Le processus de décision de l'UBR se base sur une implémentation de notre modèle de contrôle. Son objectif est, en envoyant des messages de consigne, d'éviter la formation de congestion, ou de réduire son ampleur. Ces messages de consigne sont transmis par communication I2V. La voie d'insertion provoque une chute des vitesses à partir de la voie de droite, qui peut mener à la formation de congestion se propageant à toute la section en condition de trafic dense. Nous proposons d'étudier le comportement du modèle intégré dans une UBR en faisant l'hypothèse d'un taux de pénétration fixe : 30 % de véhicules connectés, mélangés de manière homogène aux autres véhicules.

Afin de générer des contextes de l'environnement différents pour l'UBR, nous avons utilisé des données réelles issues de boucles électromagnétiques. Les débits observés sont reproduits pour générer les flux d'entrée sur la section principale et la voie d'insertion (débits dynamiques). Le jeu de données correspond à plusieurs capteurs situés sur l'autoroute A6, avec un recueil portant sur la période 8h – 9h du mois de juin 2001. Deux capteurs ont été sélectionnés (un sur l'autoroute, l'autre sur une voie d'insertion) et le scénario proposé (figure 5) reproduit la topologie des lieux. Les données utilisées permettent de générer dans le simulateur un flux de trafic (nombre de véhicules par heure) dépendant du temps (date et heure, mis a jour toutes les 6 minutes) sur la section principale et sur la bretelle d'insertion. Les véhicules simulés sont introduits sur la section en fonction d'une loi de probabilité (loi de Poisson) garantissant une distribution correspondant aux flux réellement observés. La proportion (et la distribution au sein du flux) des véhicules connectés et non connectés régit par la même loi. Un ensemble de 30 simulations ont ainsi été exécutées ; chacune se déroulant pendant 60 minutes. La diversité des conditions de trafic (pic du matin, week-ends et jours fériés) permet d'étudier le comportement du modèle pour des contextes très variables (trafic fluide, régime critique ou congestion).

6.2. Implémentation

Afin de décrire précisément les paramètres utilisés dans cette expérimentation, il est nécessaire de décomposer les deux parties des simulations. D'une part, les paramètres du simulateur multi-agent de trafic (Guériau *et al*., 2015) et d'autre part l'implémentation du modèle conceptuel (section 5).

6.2.1. Véhicules et infrastructure

L'architecture multi-agent du simulateur Guériau *et al*. (2016) permet de définir des comportements complexes pour les UBR et les véhicules. Nous proposons d'implémenter notre modèle comme le processus de décision de l'UBR. Afin de reproduire les dynamiques du trafic, le comportement des véhicules est modélisé par une

loi de poursuite. Les consignes envoyées par l'UBR aux véhicules sont intégrées dans leur processus de décision : ils surchargent ce comportement afin de s'adapter aux consignes reçues.

6.2.1.1. L'Unité de Bord de Route

Le résultat de la stratégie de contrôle de l'UBR est la sélection d'une action, à chaque pas de temps de décision (fixé à 120 secondes). Ces actions prennent la forme d'un envoi de message de consigne aux véhicules connectés à portée. Les consignes disponibles sont les suivantes :

- A_1 : pas de consigne envoyée.
- A_2 : consigne de changement de voie (de droite à gauche).
- A_3 : consigne de changement de voie (de gauche à droite).
- A_4 : consigne d'inter-distance (1,8 s).
- A_5 : consigne d'inter-distance (1,2 s).
- A_6 : limite de vitesse (110 km/h).
- A_7 : limite de vitesse (50 km/h).

Les messages de consignes envoyés intègrent les informations relatives aux paramètres des actions de l'UBR. Ils précisent aussi leur zone de pertinence, ici fixée à la section entière, et leur délai d'expiration, ici limité à 120 secondes. La consigne encapsulée dans un message est appliquée par les véhicules qui l'ont reçue tant qu'ils se trouvent dans la zone de pertinence du message et que celui-ci n'a pas expiré.

6.2.1.2. Les véhicules

Le comportement physique de base des véhicules (connectés ou non) est régi par un modèle microscopique (IDM (Treiber *et al.*, 2000), déterministe) couplé à un modèle de changement de voie (stratégie MOBIL (Kesting *et al.*, 2007), déterministe également). Ces modèles, et le calage de leurs paramètres respectifs à partir de donnes réelles (pour obtenir une plus grande variabilité de comportements) sont décrits en détails dans (Guériau *et al.*, 2016). Les véhicules connectés sont équipés d'un dispositif de communication leur permettant de percevoir les messages dans un rayon de 150 mètres. Les messages reçus de l'UBR (I2V) sont re-propagés par communication V2V et couvrent ainsi l'intégralité des 3 km de la section. Les véhicules connectés affichent le contenu des messages de consigne sur une interface embarquée qui permet au conducteur d'adapter son comportement à la consigne courante. Cette modification du comportement est une interprétation du message par les véhicules. A des fins de simplification, tous les conducteurs réagissent de la même manière aux consignes :

– l'action A_1 n'a aucune incidence puisqu'aucun message n'est envoyé par l'UBR.

– les consignes de changement de voie entraînent une modification temporaire de la stratégie de changement de voie ; le modèle MOBIL a été étendu dans Guériau

et al. (2016) pour intégrer un niveau de changement de voie intermédiaire entre les changements obligatoires (imposés par le réseau) et les dépassements.

– pour les consignes d'inter-distance et les limitations de vitesses, il s'agit d'une modification temporaire du paramètre T correspondant dans le modèle longitudinal IDM du véhicule (Treiber *et al*., 2000).

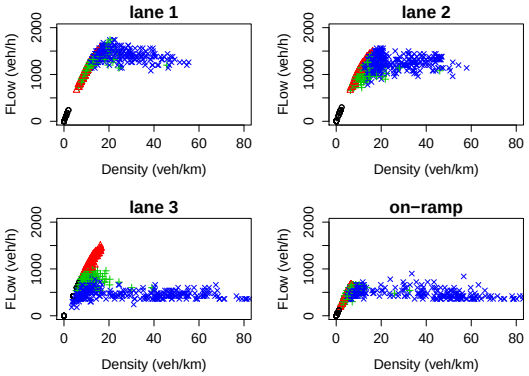
Toutes ces modifications en réaction aux consignes sont maintenues durant la durée de pertinence du message de consigne reçu. Les messages (et leurs effets) sont automatiquement effacés si le message expire ou si le véhicule quitte la zone de pertinence.

6.2.2. Modèle de contrôle

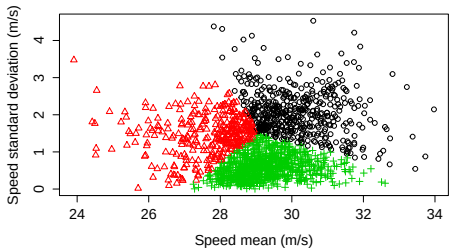
Le modèle est implémenté en lieu et place du processus de décision de l'UBR. Ce système est capable de percevoir les informations des boucles électromagnétiques qui calculent le débit moyen, la concentration et la vitesse de chacune des voies (y compris la voie d'insertion). Il s'agit de données agrégées toutes les 2 minutes (120 secondes). En plus, les véhicules connectés partagent leur vitesse respective lorsqu'ils se trouvent dans le rayon de communication de l'UBR (ce message V2I n'est pas re-propagé). Toutes ces données servent de perception de bas niveau au système. Il dispose donc de deux sources d'informations concurrentes lui permettant de caractériser l'état courant de trafic. La construction de la représentation du système repose sur plusieurs agents discrétiseurs qui manipulent des hypothèses sensorimotrices. Dans cette expérimentation, la perception et la décision du système sont discrétisées temporellement par période de 120 secondes (*i.e.* le cycle d'exécution du système est rythmé par ses capteurs). L'UBR perçoit les données agrégées de ses capteurs durant le dernier pas discret, et choisit immédiatement une action à exécuter (*i.e.* envoie une consigne). Après un nouveau pas de temps de 120 secondes, la consigne expire et le système reçoit un *feedback* pour renforcer sa représentation. Au même moment, l'UBR reçoit sa nouvelle perception par ses capteurs.

6.2.2.1. Les agents discrétiseurs

Afin de montrer les bénéfices apportés par le modèle, nous proposons une implémentation reposant sur l'instanciation de deux agents discrétiseurs. Le but de chaque agent est de générer une représentation individuelle à partir des données qu'il perçoit. Nous avons choisi d'affecter les variables d'un capteur à chacun des agents D_1 et D_2. La différence entre les deux agents se situe donc avant tout dans leurs perceptions. D_1 est lié aux boucles électromagnétiques. Il perçoit donc les données de débit et de concentration pour les 4 voies observées (8 dimensions) et agrégées toutes les 120 secondes. D_2 exploite les données reçues dans les messages V2I et construit un modèle (moyenne et écart-type) de la distribution des vitesses reçues dans les messages provenant des véhicules et qui sont passés devant l'UBR durant le dernier pas de temps de décision. Les discrétisations des agents résultent d'une classification par la méthode des K-moyennes (stabilisé avec 10 000 itérations pour des initialisations aléatoires) utilisant respectivement 4 et 3 classes (pour D_1 et D_2). Le résultat de la classification de chacun des agents sur les 15 premières simulations est représenté dans la figure 6.



(a)



(b)

Figure 6. Discrétisations générées par les agents discrétiseurs (a) D_1 [8 dimensions] et (b) D_2 [2 dimensions] sur le jeu d'apprentissage. Les couleurs (indépendantes entre les deux discrétisations) correspondent à des classes distinctes pour les valeurs observées

L'apprentissage des liens perception-action est réalisé sur les 15 mêmes simulations, grâce à l'exploitation de la récompense fournie au système. Pour cela, chaque agent discrétiseur est intégré seul à un système (donc a une UBR), et réalise l'apprentissage de ces liens perception-action indépendamment. Le résultat de cette phase d'entraînement permet d'obtenir des décisions apprises par les agents discrétiseurs indépendamment afin d'amorcer la construction de leur représentation, utilisée dans la phase d'exploitation.

6.2.3. Renforcement de la représentation

Les récompenses recueillies par les agents sous la forme de liens état-action sont calculées à partir d'une variable de l'environnement choisie. Dans le scénario proposé, nous avons retenu la vitesse moyenne la plus faible (sur les 4 boucles électromagnétiques) en tant que *feedback* du système, puisque maximiser cette valeur correspond à l'objectif global d'amélioration de l'écoulement du flux. Ce retour sur expérience

externe, est utilisé sous la forme de récompenses dans l'algorithme d'exploration-exploitation pour renforcer à la fois la stratégie de contrôle et les représentations proposées. Parmi les algorithmes de bandit à n bras de la littérature, notre choix s'est arrêté sur *Upper Confidence Bound* – UCB – (Auer *et al*., 2002). Son initialisation consiste à essayer chaque machine une seule fois. Ensuite, l'algorithme sélectionne systématiquement la machine j qui maximise $\ddot{x}_j + \sqrt{\frac{2 \ln n}{n_j}}$ où $\ddot{x}_j$ correspond à la récompense moyenne obtenue pour la machine j, n_j est le nombre de fois que la machine j à été jouée jusqu'ici et n est le nombre total d'essais. La mise à jour des récompenses concerne tous les liens perception-action (notés L dans l'algorithme 1), et non pas seulement celui qui a été effectivement sélectionné par l'algorithme. De cette façon, le bénéfice d'une action est également propagé aux liens qui ont contribué à sa sélection. Cette récompense permet d'accélérer la phase d'exploration, en s'appliquant à autant de liens que d'agents discrétiseurs pour chaque itération.

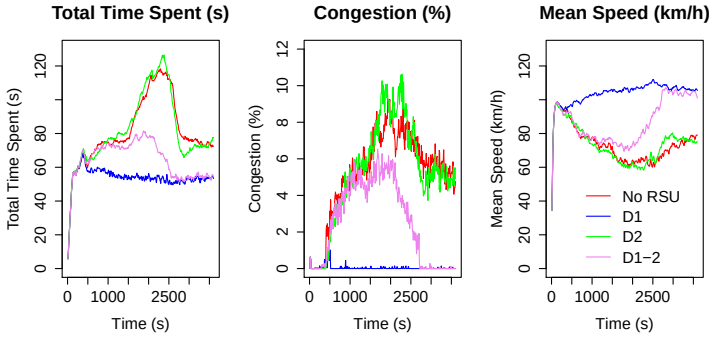
6.3. Résultats

L'expérimentation est menée en trois étapes principales. D'abord, un premier jeu de données d'entrée (perception des agents D) est nécessaire pour générer les classifications à partir de l'algorithme dédié. Durant cette phase, l'apprentissage n'est pas réalisé avec le modèle complet, mais chaque agent discrétiseur apprend la probabilité de sélection des actions de manière individuelle. Ensuite, une partie des simulations est dédiée à l'apprentissage de la représentation par le système, grâce au processus complet du modèle (sur les 15 premières simulations également), pour différentes combinaisons des agents discrétiseurs proposés. Enfin, le système est prêt à être confronté à des scénarios alternatifs, où sa représentation précédemment construite devrait aider le système dans sa stratégie de contrôle. Cette phase a été menée sur la seconde moitié des simulations. Le résultat témoin de chaque simulation est obtenu en observant différents indicateurs dans un scénario dépourvu d'unités d'infrastructure. Cela signifie qu'il n'y a aucune communication entre les véhicules, et qu'aucune consigne n'est propagée. Les autres résultats proviennent de trois implémentations du modèle : le cas avec un seul agent (D_1 et D_2 séparément) et une combinaison dynamique des représentations de D_1 et D_2.

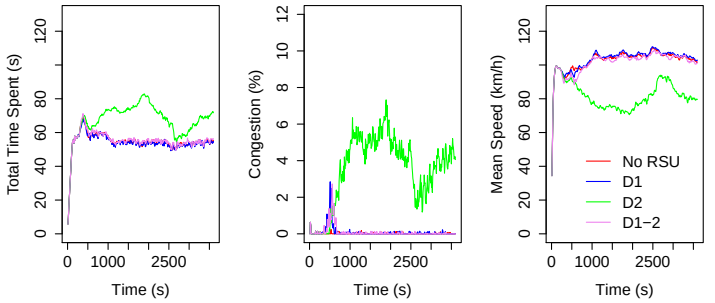
Afin d'évaluer les effets des stratégies de contrôle produites par les différentes implémentations du modèle, nous avons choisi trois indicateurs. Le temps total passé (TTS – somme des temps de trajet de tous les véhicules), la vitesse moyenne sur la section, et le pourcentage de congestion (pourcentage de véhicules évoluant à moins de 30 km/h). Des valeurs de TTS faibles signifient un meilleur écoulement du flux. La vitesse moyenne permet d'observer l'effet sur l'homogénéisation. Et le pourcentage de congestion donne un aperçu de l'état de trafic et de l'ampleur des perturbations. Le comportement du flux contrôlé par le modèle devrait être directement visible grâce à ces indicateurs.

Parmi les 15 simulations, nous avons sélectionné trois résultats qui permettent d'illustrer différentes situations de trafic. Comme les données des débits d'entrée pro-

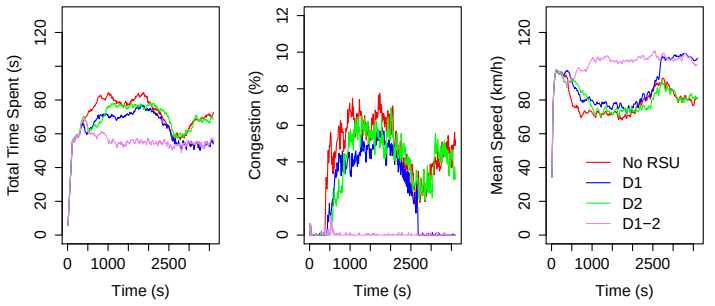
viennent de cas réels, les simulations se répètent pour des jours de même profil. La figure 7 représente un tracé des trois indicateurs pour les 3 implémentations du modèle (et le scénario témoin).



(a) pic de congestion



(b) trafic fluide



(c) trafic dense

Figure 7. Résultat de trois simulations : (a) Lundi 18, (b) Jeudi 21 et (c) Mardi 28 ; pour différentes implémentations du modèle et avec le scénario témoin sans système de contrôle (No RSU – Pas d'Unité de Bord de Route)

Les indicateurs permettent d'étudier les bénéfices de chaque implémentation sur l'amélioration du flux, en les comparant avec les scénarios témoins (courbes rouges). Dans les simulations (a) et (c), sans aucun contrôle, les instabilités produites par la voie d'insertion et les changements de voie associés se propagent à l'ensemble du flux. Les trois simulations montrent des comportements différents du modèle. Dans la simulation (a), la stratégie de contrôle de l'agent D_1 permet d'éviter l'apparition de congestion sur la section, alors que celle de D_2 donne de moins bonnes performances que le témoin. Ces résultats sont clairement visibles en observant l'évolution des valeurs de TTS. Une baisse signifie que le temps de trajet des véhicules est plus faible. La combinaison dynamique des représentations individuelles de D_1 et D_2 produit ici une stratégie équilibrée qui résulte en une amélioration moyenne. La simulation (b) illustre un cas spécifique où l'une des représentations utilisées par le modèle est fausse (celle de D_2). Aucune congestion n'est observée dans la simulation témoin mais la stratégie de D_2 provoque des perturbations. Néanmoins, le modèle est capable de converger vers la meilleure représentation individuelle (D_1) dans ce contexte. Cela ne permet toutefois pas d'améliorer l'état de trafic (résultat équivalent au témoin). Dans la figure 7 (c), l'état de trafic obtenu par les débits du scénario est proche de l'équilibre du flux. Dans le témoin, la congestion est limitée et la vitesse moyenne ne chute que peu (en comparaison à la simulation (a)). Dans ce contexte, les deux stratégies issues des représentations apprises par les discrétiseurs ont un impact positif sur le flux. L'apport du modèle lors de la construction de sa représentation est aussi clairement visible. Le système a su utiliser les récompenses pour combiner les représentations individuelles, ce qui permet d'éviter toute propagation de congestion sur la section.

L'ensemble des résultats (sensiblement similaires) pour les 15 simulations est présenté dans le tableau1, au regard de l'indicateur de temps de trajet.

Tableau 1. Temps de trajet moyen (par véhicule) sur le dernier kilomètre en amont de l'insertion pour l'ensemble des simulations, différentes implémentations du modèle et avec le scénario témoin sans système de contrôle (No RSU)

Simulation	16-Jun	17-Jun	18-Jun	19-Jun	20-Jun	21-Jun	22-Jun	23-Jun
No RSU	74.4	71.26	75.06	74.85	56.57	29.26	76.27	72.92
D1	54.14	54.31	54.03	54.39	54.15	29.3	54.34	54.07
D2	55.53	55.43	81.65	55.57	67.45	55.47	70.8	83.58
D1-2	46.95	49.33	59.16	55.75	55.25	29.88	43.23	42.78

Simulation	24-Jun	25-Jun	26-Jun	27-Jun	28-Jun	29-Jun	30-Jun
No RSU	29.25	72.93	77.43	75.63	75.06	71.15	28.99
D1	54	54.2	63.24	54.05	62.67	54.1	29.87
D2	55.3	65.33	68.27	55.42	73.64	69.2	55.13
D1-2	38.7	36.36	48.26	40.89	52.92	54.48	28.18

La figure 8 est une représentation simplifiée de la contribution de chacun des agents discrétiseurs (en reprenant le même code couleurs que celui de la figure 7).

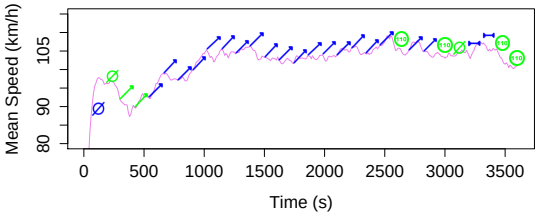


Figure 8. Actions des agents internes au système dans l'implémentation du modèle lors de la simulation (c)

Comme le laissaient présager les résultats obtenus au niveau des indicateurs de trafic, on retrouve l'apport des deux agents dans la politique d'action menée par le système. L'action la plus représentée (A_2) est tout à fait logique puisqu'elle permet de faciliter l'insertion des véhicules arrivant sur le convergent et souhaitant rejoindre la section principale. Cela confirme une nouvelle fois la capacité du modèle à adapter dynamiquement plusieurs représentations concurrentes pour construire une représentation de haut niveau plus précise.

6.4. Discussion

La phase d'implémentation est cruciale lorsque le modèle est exposé à un cadre applicatif concret et complexe. Les deux éléments les plus importants sont le choix de la récompense et la précision des discrétisations utilisées. La difficulté du choix de la perception des agents fait référence à un problème de fusion de capteurs : certains phénomènes ne s'expriment que dans certaines combinaisons de dimensions. Grâce à l'apprentissage mis en place, le système est capable de produire de nouvelles combinaisons de dimensions dynamiquement et d'exploiter celles qui lui permettent d'augmenter l'expressivité et la précision de sa représentation.

Néanmoins, le concepteur doit fournir des représentations individuelles relativement précises afin d'aider le système dans l'amorçage de sa construction. Dans ce travail, les représentations individuelles ont été associées a chaque capteur du système, mais le nombre de classes a été choisi en fonction des résultats empiriques observés. Ce problème pourrait être traité en donnant au modèle la possibilité de générer dynamiquement (ou de faire évoluer) les agents discrétiseurs, en exploitant les expériences d'interactions précédentes. Une perspective intéressante est par exemple d'utiliser une méthode de classification en ligne capable de faire évoluer l'espace d'états d'un agent discrétiseur dynamiquement. Par exemple, l'algorithme TD-GNG proposé par Vieira *et al*. (2013), utilise le *Growing Neural Gas* (Fritzke, 1997) pour faire évoluer les états utilises dans un processus d'apprentissage par renforcement. Une future version du modèle constructiviste proposé pourrait accommoder plusieurs de ces processus conjointement, mais l'on peut s'attendre a ce que la convergence du système devienne difficile à garantir si les états ne cessent jamais d'évoluer.

Dans la version présentée ici, le modèle ne comble qu'une partie de la description d'un système constructiviste (section 4.1). Le problème du *feedback* reste par exemple un verrou important dans l'application proposée. En effet, nous avons proposé une récompense globale: l'objectif du système est de maximiser la vitesse la plus faible sur les 3 voies (et sur la bretelle d'accès). Afin d'accélérer l'apprentissage, un mécanisme de redistribution de la récompense aux agents discrétiseurs proposant les mêmes actions a été introduit (il s'agit donc d'une forme de récompense inter-agents – voir section 4.2.2). Il est très probable que séparer les récompenses du système et des agents discrétiseurs puisse permettre de spécialiser davantage les agents, et donc d'obtenir une représentation plus précise. Il faudra alors proposer un mécanisme de communication (ou de partage de récompense) entre les discrétiseurs.

7. Conclusion et perspectives

Ce travail introduit des processus d'apprentissage dynamiques de façon autonome pour le contrôle de systèmes complexes. Pour cela, nous nous intéressons à la nature distribuée du processus de contrôle ainsi qu'à l'expression de processus complexes tels que la régulation de trafic. Le modèle proposé s'inspire du paradigme constructiviste, et permet à un système de construire de façon générique une représentation de son environnement à partir de ses interactions et de construire ainsi itérativement son processus de prise de décision. Le modèle utilise une population interne d'agents discrétiseurs pour amorcer et faire évoluer la construction de la représentation. La représentation se compose d'états perception-action qui évoluent par renforcement suivant le retour perçu de l'environnement (*feedback*). Après avoir présenté les verrous scientifiques pour le développement d'une autonomie, et introduit le modèle d'un point de vue théorique, nous avons proposé son application à la régulation de trafic coopératif. L'objectif est alors d'utiliser l'infrastructure et les véhicules connectés pour proposer un système d'aide à la décision autonome. Ce cadre applicatif innovant met en valeur les effets de notre approche. Les résultats obtenus en simulation montrent que la combinaison dynamique des discrétisations individuelles permet au système d'adopter une stratégie de régulation efficace.

Les véhicules autonomes remplaceront les véhicules d'aujourd'hui à l'horizon 2030-2050, il sera alors nécessaire d'intégrer progressivement de nouveaux outils de régulation autonome, qui pourront s'adapter aux différents contextes rencontrés. Une extension de ce travail consiste à enrichir l'ensemble des informations mises à disposition pour le système, et donner par exemple la possibilité pour un véhicule de percevoir les véhicules situés tout autour de sa carrosserie. Ceci permettrait de contrôler également le comportement latéral du véhicule d'obtenir de nouveaux gains du point de vue de la sécurité et de la fluidité du trafic. Un autre sujet d'étude pourrait s'intéresser aux systèmes mixtes comprenant des véhicules connectés et non connectés, qui seront nécessaires dans un premier temps. Enrichir ainsi les capacités de perception du système complique encore le problème d'apprentissage, on pourra par exemple s'intéresser aux méthodes de transferts d'apprentissage et identifier les parties de la représentation construite qui peuvent être généralisables et celles qui évolueront sui-

vant les besoins propres à la configuration locale régulée. En complément, il peut être utile de proposer des mécanismes permettant de faire évoluer dynamiquement le périmètre des états associés aux agents afin d'améliorer la finesse et la pertinence des discriminations des configuration rencontrées, selon une perspective constructiviste.

Bibliographie

Alonso E., D'inverno M., Kudenko D., Luck M., Noble J. (2001). Learning in multi-agent systems. *The Knowledge Engineering Review*, vol. 16, n° 03, p. 277–284.

Asada M., Hosoda K., Kuniyoshi Y., Ishiguro H., Inui T., Yoshikawa Y. *et al*. (2009). Cognitive developmental robotics: a survey. *IEEE Transactions on Autonomous Mental Development*, vol. 1, n° 1, p. 12–34.

Auer P., Cesa-Bianchi N., Fischer P. (2002). Finite-time analysis of the multiarmed bandit problem. *Machine learning*, vol. 47, n° 2-3, p. 235–256.

Bazzan A. L. C. (2009). Opportunities for multiagent systems and multiagent reinforcement learning in traffic control. *Autonomous Agents and Multi-Agent Systems*, vol. 18, n° 3, p. 342-375.

Bloembergen D., Tuyls K., Hennes D., Kaisers M. (2015). Evolutionary dynamics of multi-agent learning: a survey. *Journal of Artificial Intelligence Research*, vol. 53, p. 659–697.

Drescher G. L. (1991). *Made-up minds: a constructivist approach to artificial intelligence*. MIT press.

Dresner K., Stone P. (2004). Multiagent traffic management: A reservation-based intersection control mechanism. In *Proceedings of the third international joint conference on autonomous agents and multiagent systems-volume 2*, p. 530–537.

Farhi N., Phu C. N. V., Amir M., Haj-Salem H., Lebacque J.-P. (2015). A semi-decentralized control strategy for urban traffic. *Transportation Research Procedia*, vol. 10, p. 41 - 50. Consulté sur http://www.sciencedirect.com/science/article/pii/S2352146515002410 (18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands)

Fritzke B. (1995). A growing neural gas network learns topologies. In G. Tesauro, D. Touretzky, T. Leen (Eds.), *Advances in neural information processing systems 7*, p. 625–632. MIT Press.

Fritzke B. (1997). A self-organizing network that can follow non-stationary distributions. In W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (Eds.), *Artificial neural networks – icann'97*, vol. 1327, p. 613-618. Springer Berlin Heidelberg.

Fulda N., Ventura D. (2007). Predicting and preventing coordination problems in cooperative q-learning systems. In *Ijcai*, vol. 2007, p. 780–785.

Geroliminis N., Srivastava A., Michalopoulos P. G. (2011). Experimental observations of capacity drop phenomena in freeway merges with ramp metering control and integration in a first-order model. In *Transportation research board 90th annual meeting.*

Graczyk M., Lasota T., Trawiński B., Trawiński K. (2010). Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal. In *Asian conference on intelligent information and database systems*, p. 340–350.

Guériau M. (2016). *Systèmes multi-agents, auto-organisation et contrôle par apprentissage constructiviste pour la modélisation et la régulation dans les systèmes coopératifs de trafic*. Thèse de doctorat non publiée, Université de Lyon I Claude Bernard.

Guériau M., Billot R., Armetta F., Hassas S., El Faouzi N.-E. (2015). Un simulateur multiagent de trafic coopératif. In *23es journées francophones sur les systèmes multi-agents (jfsma'15)*, p. 165–174.

Guériau M., Billot R., El Faouzi N.-E., Monteil J., Armetta F., Hassas S. (2016). How to assess the benefits of connected vehicles? a simulation framework for the design of cooperative traffic management strategies. *Transportation Research Part C: Emerging Technologies*, vol. 67, p. 266 - 279.

Guerin F. (2011). Learning like a baby: a survey of artificial intelligence approaches. *The Knowledge Engineering Review*, vol. 26, n° 02, p. 209–236.

Kesting A., Treiber M., Helbing D. (2007). General lane-changing model mobil for car-following models. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999 / 2007 Traffic Flow Theory 2007, p. 86-94.

Kesting A., Treiber M., Schönhof M., Helbing D. (2008). Adaptive cruise control design for active congestion avoidance. *Transportation Research Part C: Emerging Technologies*, vol. 16, n° 6, p. 668–683.

Khondaker B., Kattan L. (2015). Variable speed limit: an overview. *Transportation Letters*, vol. 7, n° 5, p. 264–278.

Leclercq L., Knoop V. L., Marczak F., Hoogendoorn S. P. (2016). Capacity drops at merges: New analytical investigations. *Transportation Research Part C: Emerging Technologies*, vol. 62, p. 171–181.

Lungarella M., Metta G., Pfeifer R., Sandini G. (2003). Developmental robotics: a survey. *Connection Science*, vol. 15, n° 4, p. 151–190.

Mazac S., Armetta F., Hassas S. (2014). On bootstrapping sensori-motor patterns for a constructivist learning system in continuous environments. In *Alife 14: Fourteenth international conference on the synthesis and simulation of living systems.*

Meeden L. A., Blank D. S. (2006). Introduction to developmental robotics. *Connection Science*, vol. 18, n° 2, p. 93–96.

Mugan J., Kuipers B. (2007). Learning distinctions and rules in a continuous world through active exploration. In *Proceedings of the seventh international conference on epigenetic robotics (epirob-07)*, p. 101–108.

Najjar A., Reignier P. (2013). Constructivist ambient intelligent agent for smart environments. In *Pervasive computing and communications workshops (percom workshops), 2013 ieee international conference on*, p. 356–359.

Oudeyer P.-Y., Smith L. B. (2016). How evolution may work through curiosity-driven developmental process. *Topics in Cognitive Science*, vol. 8, n° 2, p. 492–502.

Panait L., Luke S. (2005). Cooperative multi-agent learning: The state of the art. *Autonomous agents and multi-agent systems*, vol. 11, n° 3, p. 387–434.

Papageorgiou M., Diakaki C., Dinopoulou V., Kotsialos A., Wang Y. (2003). Review of road traffic control strategies. *Proceedings of the IEEE*, vol. 91, n° 12, p. 2043–2067.

Piaget J. (1955). The construction of reality in the child. *Journal of Consulting Psychology*, vol. 19, n° 1, p. 77.

Schmidt-Dumont T., Vuuren J. H. van. (2015). Decentralised reinforcement learning for ramp metering and variable speed limits on highways. *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, n° 8, p. 1.

Schönhof M., Treiber M., Kesting A., Helbing D. (2007). Autonomous detection and anticipation of jam fronts from messages propagated by intervehicle communication. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1999, n° 1, p. 3–12.

Talebpour A., Mahmassani H., Hamdar S. (2013). Speed harmonization: evaluation of effectiveness under congested conditions. *Transportation Research Record: Journal of the Transportation Research Board*, n° 2391, p. 69–79.

Treiber M., Hennecke A., Helbing D. (2000). Congested traffic states in empirical observations and microscopic simulations. *Phys. Rev. E*, vol. 62, p. 1805–1824.

Vasirani M., Ossowski S. (2009). A market-inspired approach to reservation-based urban road traffic management. In *Proceedings of the 8th international conference on autonomous agents and multiagent systems-volume 1*, p. 617–624.

Vieira D. C., Adeodato P. J., Goncalves P. M. (2013). A temporal difference gng-based approach for the state space quantization in reinforcement learning environments. In *Tools with artificial intelligence (ictai), 2013 ieee 25th international conference on*, p. 561–568.

Wei G. (1993). Learning to coordinate actions in multi-agent systems. In *Proceedings of the thirteenth international joint conference on artificial intelligence (ijcai-93)*, p. 311–316.

Zheng Z., Ahn S., Chen D., Laval J. (2013). The effects of lane-changing on the immediate follower: Anticipation, relaxation, and change in driver characteristics. *Transportation research part C: emerging technologies*, vol. 26, p. 367–379.

Zhu F., Ukkusuri S. V. (2014). Accounting for dynamic speed limit control in a stochastic traffic environment: A reinforcement learning approach. *Transportation Research Part C: Emerging Technologies*, vol. 41, p. 30 - 47.

Zlatev J., Balkenius C. (2001). Introduction: Why "epigenetic robotics"? In *Proceedings of the first conference on epigenetic robotics*, vol. 85, p. 1–4.

BON DE COMMANDE D'ABONNEMENT 2018

2018 SUBSCRIPTION FORM

Renvoyer à / *Return to*: Lavoisier SAS, Abonnements Revues
14, rue de Provigny – 94236 Cachan cedex – France
tel : (33) 01-47-40-67-00 – Fax : (33) 01-47-40-67-02 – abonne.ria@lavoisier.fr

REVUE D'INTELLIGENCE ARTIFICIELLE		
RIA – VOLUME 32/2018	**6 N°/AN** *(3 issues/year)*	
Tarif d'abonnement	TTC FRANCE	HT ÉTRANGER (*)
Version imprimée *incluant la version on line*	415 €	478 €
Version on line + archives	378 €	378 €
ABONNEMENT AUX 4 REVUES RSTI	**6 TSI + 6 RIA + 6 ISI + 3DN = 21 N°/AN**	
Tarif d'abonnement	TTC FRANCE	HT ÉTRANGER (*)
Version imprimée *incluant la version on line*	1 325 €	1 494 €
Version on line + archives	1 206 €	1 206 €

CONDITIONS D'ABONNEMENT / *CONDITIONS OF SUBSCRIPTION*

Les abonnements sont enregistrés à réception de leur règlement et sont acceptés pour l'année civile uniquement. / *Subscriptions are entered upon receipt of payment and are accepted for a calendar year only.*

(*) Pour les tarifs TTC étranger, merci de nous contacter / *Other countries rates are available on our web site: http://www.revuesonline.com or on request (revues.abo@lavoisier.fr)*

Nom / *Name* . . . . . .

Organisation / *Organization* . . . . . .

Adresse / *Address* . . . . . .

. . . . . .

Code postal – Ville / *ZIP – City* . . . . . .

Pays / *State* . . . . . .

☐ Règlement par chèque joint à l'ordre de Lavoisier / *Cheque enclosed payable to Lavoisier*

☐ Règlement par carte VISA /*Payment by VISA card*

N°carte / *Card No* ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐

Date d'expiration / *Expiry Date*

3 derniers chiffres du cryptogramme au dos de votre carte ☐☐☐

The last 3 digits of the cryptogram on the reverse of your card

Date et signature / *Date and signature*