Méthodes formelles et développement de systèmes émergents

Porteurs du défi : groupe MFDL

December 2019

Les méthodes formelles ont été élaborées depuis de nombreuses années afin d'assurer un niveau aussi élevé que possible en matière de précision et de fiabilité et ont ainsi montré que l'objectif du zéro-faute est réalisable pour des systèmes dits "fermés" (qui fonctionnent dans un environnement complètement maîtrisé). Cependant, les systèmes informatiques qui émergent aujourd'hui sont de plus en plus ouverts, donnant lieu à de nombreux défis tant pour ces systèmes que pour l'usage et l'application des approches formelles. Dans ce document nous avons choisi d'aborder ces défis au travers de trois types de systèmes émergents : les systèmes cyber-physiques, les systèmes (véhicules) autonomes et les systèmes intelligents. La complexité de ces systèmes est évidente et la maîtrise des risques inhérents à leur utilisation est ainsi de plus en plus pressante. Il est clair que pour garantir leur sûreté et leur sécurité, l'adoption de moyens d'investigation indubitables et des techniques sûres et fiables reposant sur des fondements mathématiques s'impose. Aussi, les méthodes formelles ont-elles vocation à jouer un rôle d'envergure dans ce cadre avec des bénéfices indéniables. Ce faisant, pour être efficaces, les méthodes formelles devront inévitablement s'adapter au caractère "ouvert", "imprécis" et "intelligent" de ces systèmes. Ce défi, porté par le groupe MFDL, vise à apporter une vision d'actualité aux nouveaux usages et applications des méthodes formelles.

1 Méthodes formelles et systèmes cyber-physiques : faire interagir des modèles à temps discret et des modèles à temps continu

1.1 Verrous scientifiques et techniques à lever

Un système cyber-physique est un système dont le comportement ou la sémantique combine un aspect physique décrit typiquement par des équations différentielles (ordinaires ou algébriques) avec un calculateur contrôlé par un programme informatique. La physique Newtonienne décrit l'évolution du système physique par des équations qui dépendent du temps. Certains paramètres du modèle, comme justement le temps, évoluent dans des domaines continus ou denses.

Le composant informatique, bien que réalisé par un mécanisme physique à base de transistors, a un comportement qui peut être décrit dans un monde plus discret. Le temps n'est plus dense mais cadencé par l'horloge de l'ordinateur. Les autres variables ou paramètres manipulés sont codés par des séquences de bits.

Les outils mathématiques pour la modélisation et la vérification des systèmes discrets, ou à temps discret par rapport à ceux continus, ou à temps continu, sont très différents. Les premiers s'appuient plutôt sur l'algèbre générale et la logique, avec des raisonnements basés sur le principe d'induction; les seconds sur la topologie et l'analyse.

Même lorsque les systèmes à temps continu sont discrétisés, les modèles obtenus sont souvent trop complexes et peu intuitifs pour être traités efficacement par les méthodes des systèmes discrets. Par exemple, la représentation des réels par des flottants ou des encodages à base d'entier, rend les analyses difficiles. La complexité entraîne des problèmes de passage à l'échelle des outils de vérification automatique et le côté peu intuitif de certains modèles ne permet pas aux utilisateurs des méthodes automatiques ou semi-automatiques de construire des preuves rapidement.

L'étude des systèmes hybrides désigne parfois l'étude de systèmes physiques dont le comportement est décrit par une combinaison discrète (typiquement finie) de dynamiques à temps continu. Même si plusieurs

travaux ont montré la faisabilité de développement formel de systèmes hybrides, les méthodes et outils actuellement disponibles manquent encore de maturité par rapport aux méthodes et outils dédiés aux systèmes exclusivement discrets ou continus. Ainsi il est difficile de plonger un système cyber-physique même basique, comme par exemple une boucle fermée entre un balancier inversé et un contrôleur linéaire avec saturations, dans ces formalismes.

Le défi proposé consiste donc à proposer un cadre théorique outillé pour raisonner sur ces systèmes cyber-physiques. Il s'agira donc de proposer des méthodes ou outils permettant de mieux appréhender ses systèmes en:

- aidant à leur définition;
- proposant des simulations fiables;
- construisant des analyses sur couplages temps discret/temps continu;
- fournissant un cadre théorique pour formaliser les propriétés d'intérêt comme des éléments des modèles couplés;
- synthétisant du code et des preuves à partir de ces modèles;
- intégrant les approches proposés au cycle de développement de ces systèmes cyber-physiques.

1.2 Usages et impacts sociétaux qu'il aborde

La plupart des systèmes critiques combinent des parties continues en interaction avec l'environnement (capteurs, actuateurs) et des parties discrètes principalement à base de logiciel. Le défi proposé et les méthodes et outils qui pourront être définis auront un impact majeur dans le développement et la certification de systèmes cyber-physiques critiques, comme les drones, l'IoT, les véhicules autonomes, le développement du transport spatial, les dispositifs médicaux (appareils de mesure, de soin, prothèses informatisées, ...)

1.3 Projet nationaux et internationaux sur le thème

- Projet ANR DISCONT - Correct Integration of Discrete and Continuous Models
- Projet ANR EBRP PLus
- Projet ANR JCJC FEANICSES - Formal and Exhaustive Analysis of Numerical Intensive Control Software for Embedded Systems
- Projet IPL INRIA ModeliScale
- Projet ARTEMIS UnCoVerCPS - Unifying Control and Verification of Cyber-Physical Systems

2 Méthodes formelles et systèmes ouverts : les véhicules autonomes

La composante logicielle domine le véhicule de demain, c'est pourquoi l'usage de méthodes formelles est un enjeu majeur pour leur sûreté. Les défis relevés par ces méthodes durant les années à venir porteront sur plusieurs axes de recherche : la conception de systèmes temps réels embarqués, la correction des algorithmes décisionnels, la tolérance aux fautes dans un environnement incertain, la sécurité, etc. Dans la suite nous focalisons notre réflexion principalement sur deux verrous scientifiques majeurs.

2.1 Verrous scientifiques et techniques à lever

1. **Preuve de sûreté des actions engagées vis-à-vis d'informations environnantes peu précises**: Bien que les techniques de preuve (theorem proving et model-checking) et de vérification à l'exécution (runtime verification) - largement adressées par la communauté des méthodes formelles - soient incontournables, elles s'avèrent insuffisantes pour ces systèmes étant donné le côté imprévisible du comportement du véhicule. Ce dernier, ayant la responsabilité de reconnaître son environnement, est amené à engager en conséquence des actions qui se doivent d'être sûres. Ceci n'est pas sans failles car la reconnaissance de l'environnement se base sur des techniques qui manquent parfois de précision tels que l'apprentissage, la prédiction ou encore l'approximation. À titre d'exemple, un rapport public dressé par le département des transports américain [1] désigne des failles de sûreté observées sur des véhicules Tesla et qui indiquent des comportements non conformes à ceux normalement attendus: "*The Automatic Emergency Braking or Autopilot systems may not function as designed, increasing the risk of a crash*". Il est important de distinguer la précision des informations issues du monde réel, et la justesse des actions engagées suite à ces informations. L'application d'une approche formelle donne aujourd'hui des solutions quant à la sûreté des actions entreprises par un automatisme ; en revanche, cela est souvent fondé sur l'hypothèse que les données en entrée sont bien précises. Ceci n'étant pas le cas du véhicule autonome, des techniques de supervision, elles-mêmes prouvées correctes, permettant d'évaluer la précision et la justesse des informations environnantes au système, sont devenues de mise.

2. **Prise en compte des interactions et de leur complexité**: La conduite est un processus qui se veut social car il implique des interactions parfois intenses et complexes avec des humains : autres conducteurs, cyclistes, piétons, etc. De nombreux travaux montrent qu'il est très difficile de remplacer les décisions de nature humaine par des décisions issues d'algorithmes vu que les humains s'appuient sur une intelligence généralisée et sur le bon sens. Dans un monde où le véhicule interagit très peu (voir pas du tout) avec des humains lors de la prise de décision, comme dans le cadre d'un train sans conducteur ou d'un pilote automatique d'avion, les techniques formelles ont montré leur efficacité. Néanmoins, dès lors que la prise de décision est collective et est le fruit d'interactions sociales, il devient nécessaire d'envisager une multitude de scénarios. Google par exemple, a mené des travaux pour que le véhicule puisse reconnaître un cycliste, interpréter ses signaux gestuels et prédire son intention (*e.g.* dans quel sens il envisage de tourner). Pour garantir la sûreté d'un tel système, les approches formelles se doivent de proposer des mécanismes qui couvrent la modélisation du comportement humain en plus de celui du système ainsi que les liens entre eux. Cela aura un impact direct sur les techniques de vérification actuelles (tel que le model-checking borné et/ou symbolique, la simulation, etc) qui devront être étendues non seulement pour exhiber les failles possibles, mais aussi pour élaborer des niveaux de confiance et/ou de tolérance, et identifier les responsabilités qui pourraient découler de ces failles.

2.2 Usages et impacts sociétaux qu'il aborde

L'industrie du véhicule autonome est un cadre d'application où les approches formelles sont incontournables. Cette industrie a pris de l'ampleur durant les dernières années notamment dans le domaine des transports routiers (General Motors, BMW, Tesla, Uber, etc), du ferroviaire (Alstom, SNCF, Bombardier, etc), du nautisme (projet Roboat du MIT), etc. Cependant, les expériences des systèmes désormais opérationnels (voire commercialisés) montrent qu'ils sont sujets à de nombreux challenges. En effet, outre les obstacles techniques et réglementaires qui entravent la création de véhicules totalement autonomes, plusieurs interrogations restent soulevées concernant leur sûreté. L'essor des véhicules autonomes aura un impact sur la façon dont les villes seront structurées ainsi que sur les habitudes des personnes au quotidien. Cela ne peut se faire sans répondre aux diverses exigences de sûreté de manière formelle et approuvée. En outre, l'hostilité de l'environnement dans lequel ces véhicules seront plongés, donnera une nouvelle vision à la cyber-sécurité et à la surveillance des usagers, et impactera de fait le cadre juridique et éthique de la société d'aujourd'hui.

2.3 Projet nationaux et internationaux sur le thème

Il existe une foultitude de projets sur le thème du véhicule autonome avec un intérêt manifeste de la part des secteurs publics et privés. Nous donnons quelques pointeurs vers des projets qui font usage de méthodes formelles dans le développement de ces systèmes:

- TrustMeIA (Toulouse) `https://www.laas.fr/projects/trustmeia/`
- Programmes "Train Autonome" de l'IRT Railenium et Tech4Rail de la SNCF
- SynC Contest: Automatic Driving Challenge using Model-based Design and Synchronous Programming
- TraCE-IT : Train Control Enhancement via Information Technology (`http://traceit.isti.cnr.it`)

3 Méthodes formelles et IA : certification des algorithmes d'IA ou certification des sorties des algorithmes d'IA

Il est difficile aujourd'hui d'ignorer l'engouement pour l'intelligence artificielle et sa diffusion dans tous les domaines applicatifs, y compris les domaines critiques. Il est donc indispensable d'avoir des garanties sur les algorithmes d'IA pour pouvoir bénéficer des avancées de ce domaine en évitant ses dangers. Pour y arriver, plusieurs disciplines doivent cohabiter : économie, éthique, juridique, sécurité, méthodes formelles,... [2].

3.1 Verrous scientifiques et techniques à lever

Il existe de nombreux verrous pour la vérification des algorithmes / systèmes d'IA liés aux caractéristiques de ceux-ci :

1. **Spécification** : les algorithmes / systèmes d'IA sont souvent peu ou mal spécifiés. Ils consistent à reconnaître des patterns dans les données d'entrée. L'algorithme d'IA est performant justement car on ne sait pas caractériser algorithmiquement / formellement les données d'entrée.
2. **Explicabilité** : bien souvent les spécialistes de l'IA eux-mêmes ont du mal à expliquer (en langage humain ou mathématiques) pourquoi les algorithmes donnent une réponse.
3. **Reproductibilité** : un même apprentissage répété plusieurs fois peut construire des IA ayant des comportements différents (part d'aléatoire dans l'apprentissage).
4. **Prédictabilité et robustesse** : face à une même (ou presque) configuration, il n'y a pas (toujours) de garanties d'avoir la même réponse en sortie.

La preuve de correction d'un algorithme est une preuve par rapport à une spécification. Sans spécification, pas de preuve. Il faudra donc trouver une façon / un langage / un formalisme /... pour décrire la spécification d'un algorithme d'IA.

Il est déjà difficile d'assurer la correction de programmes que l'être humain écrit en se basant notamment sur des algorithmes. Comment garantir la correction de programmes qui ne sont plus écrits mais appris ? Pour contourner ce problème d'explicabilité, ainsi que ceux liés à la reproductibilité, la prédictabilité et la robustesse, une certification à postériori peut être plus adaptée : au lieu de certifier l'algorithme lui-même, chaque sortie est accompagnée d'un certificat de validité qui peut être rejoué.

3.2 Usages et impacts sociétaux qu'il aborde

Aujourd'hui, nombreux sont ceux qui envisagent que les algorithmes d'IA remplacent l'intervention humaine dans bon nombre de domaines critiques : transport (véhicule autonome par exemple), médical (aide au diagnostique médical par exemple), aviation (reconnaissance vocale pour le contrôle aérien par exemple), armement, finance, ... Même si ces algorithmes éviteront des erreurs liées à l'humain (fatigue, inattention,...), ils en introduiront de nouvelles liées à une mauvaise conception / maîtrise des algorithmes / systèmes d'IA.

Jusqu'à quel point l'Homme est-il prêt à accepter des pertes humaines liées à une défaillance informatique? Weld et Etzioni [3] ont déclaré que *"society will reject autonomous agents unless we have some credible means of making them safe."*. Pour bénéfier pleinement des progrès apportés par l'IA, une garantie sur le comportement des algorithmes / systèmes est indispensable.

3.3 Projet nationaux et internationaux sur le thème

Au niveau national

- Une recherche sur le site de l'ANR permet de se rendre compte de la difficulté du problème : aucun projet n'est financé sur ce sujet, probablement car le domaine n'est pas assez mûr.
- Dans le cadre du Labex DigiCosme, une école d'été (ForMaL) a été organisée sur le thème des méthodes formelles et du machine learning pour aider à la création de synergies entre ces deux mondes : `https://formal-paris-saclay.fr/`

Au niveau international

- "Summit on Machine Learning Meets Formal Methods", 13 Juillet 2018, `http://www.floc2018.org/summit-on-machine-learning/`, en marge de la conférence FLoC.
- En 2017, l'école d'été de Dagstuhl était sur le thème des méthodes formelles et du machine learning : `https://www.dagstuhl.de/en/program/calendar/semhp/?semnr=17351`

4 Conclusion

Les équipes associées au groupe MFDL ont développé de nombreux travaux autour de l'usage des approches formelles pour le développement de logiciels. Les recherches menées dans ce cadre couvrent un large spectre d'applications. Dans ce document, nous avons identifié trois défis majeurs intéressant la communauté MFDL et qui méritent, à notre sens, une grande attention pour les années à venir. En effet, l'informatique est de plus en plus présente dans les domaines critiques, les systèmes sont de plus en plus complexes et de nouveaux paradigmes apparaissent. Les méthodes formelles vont devoir s'adapter à ces nouveaux systèmes pour pouvoir en garantir le bon fonctionnement.

References

[1] National Highway Traffic Safety Administration. Automatic vehicle control systems - ODI Resume. Technical report, U.S. Department of Transportation, 2017. `https://static.nhtsa.gov/odi/inv/2016/INCLA-PE16007-7876.PDF`.

[2] Stuart J. Russell, Daniel Dewey, and Max Tegmark. Research priorities for robust and beneficial artificial intelligence. *AI Magazine*, 36(4), 2015.

[3] Daniel S. Weld and Oren Etzioni. The first law of robotics (A call to arms). In Barbara Hayes-Roth and Richard E. Korf, editors, *Proceedings of the 12th National Conference on Artificial Intelligence, Seattle, WA, USA, July 31 - August 4, 1994, Volume 2*, pages 1042–1047. AAAI Press / The MIT Press, 1994.