

04 novembre 2019

Alain DUTECH, CR INRIA
Loria,
Campus Scientifique, BP239,
54506 Vandœuvre-lès-Nancy

Rapport sur le mémoire présenté par **Nicolas CARRARA**, intitulé
“*Reinforcement learning for Dialogue System optimization with user adaptation*”
en vue de l’obtention du
Doctorat de l’Université des Sciences et Technologies de Lille.

Le manuscrit de Nicolas CARRARA nous présente ses travaux en intelligence artificielle, et plus précisément en apprentissage par renforcement et apprentissage par transfert, dans le domaine du dialogue homme-machine autonome. L’idée principale de ce travail est d’améliorer la performance d’un système de dialogue quand il est confronté à un nouvel utilisateur, en s’appuyant fortement sur l’apprentissage par transfert : quelles connaissances et savoir faire appris avec d’autres utilisateurs doivent être réutilisés, et comment, pour mieux répondre aux attentes de ce nouvel utilisateur.

Au sein du pipeline complexe qui permet à un système de dialoguer avec un utilisateur, Nicolas se concentre donc sur le *Dialogue Manager* qui est modélisé dans le cadre des Processus Décisionnels de Markov et confronté à un problème d’apprentissage par renforcement où l’utilisateur est considéré comme un élément de l’environnement. Dans ce cadre, Nicolas postule que l’apprentissage par transfert peut résoudre le problème posé, position avec laquelle je suis d’accord même si j’aurais préféré que cette affirmation soit un peu plus étayée et argumentée (j’y reviendrai plus tard). Nicolas propose alors des contributions qui s’articulent sur plusieurs axes :

- comment passer à l’échelle pour choisir des systèmes de dialogues sources qui seront réutilisés, et adaptés, pour un nouvel utilisateur ;
- explorer des pistes originales pour incorporer cette notion d’apprentissage par transfert dans des outils d’apprentissage profond qui sont maintenant largement utilisés ;
- une approche originale d’apprentissage par renforcement budgétisé pour construire des politiques de dialogue “sûres” (*safe*) pour optimiser un critère (performance du dialogue) tout en s’assurant que d’autres critères (minimiser le risque d’arrêt prématuré du dialogue par l’utilisateur) sont respectés ;
- une approche préliminaire pour transférer ces politiques “sûres” dans le cadre de l’apprentissage par renforcement.

Ces contributions sont toutes très motivantes, d’autant que la thématique d’apprentissage par transfert est actuelle et pertinente, mais elles n’ont pas toutes été développées et présentées avec la même rigueur scientifique, ce qui rend certaines d’entre elles moins convaincantes, d’autant que les problématiques auxquelles elles répondent sont exposées à un niveau général mais souvent peu ou pas vraiment formalisées en détail.

Le manuscrit de Nicolas CARRARA est écrit en anglais et comporte 140 pages, dont 12 pages d'annexes et 20 pages de références bibliographiques. Le cœur du mémoire (85 pages) est composé de 9 chapitres (numérotés de 2 à 10), regroupés en 3 grandes parties précédées d'une introduction (chap. 2) et ponctué d'une conclusion (chap. 10). Il est agréable à lire et révèle l'aisance d'écriture de Nicolas CARRARA et sa structuration est bien adaptée à la présentations des différentes contributions de Nicolas, avec un petit bémol cependant : l'utilisation trop fréquente de nombreux acronymes, même s'ils sont listés en début de manuscrit, ne facilite pas la lecture du document.

Synthèse du manuscrit

Après le chapitre 1 qui fait la liste des notations et des acronymes utilisés dans le manuscrit, le chapitre 2 est une **introduction à la problématique**. Avec une écriture fluide et aisée, Nicolas présente la problématique générale des systèmes de dialogue artificiels automatiques ainsi qu'un historique de leurs évolutions. On y apprend que les techniques actuelles sont guidées par les données et que trois grandes sortes de problèmes (conversation sociale, répondre à des questions et compléter une tâche) focalisent les recherches. Quand Nicolas CARRARA présente ensuite ses contributions, sa problématique principale (adaptation à un nouvel utilisateur), et les difficultés théoriques, techniques ou pratiques qu'elle soulève, n'a été que trop rapidement évoquée par quelques lignes au début du chapitre.

Néanmoins, cette introduction joue bien son rôle en donnant un bon aperçu informel et de haut niveau de l'état de l'art et de l'idée des contributions qui seront développées dans le manuscrit.

La **première grande partie** du manuscrit ("*Task-oriented Dialogue Systems*"), composée de trois chapitres, permet d'entrer plus en détail dans la problématique de ce travail de thèse. Le chapitre 3 s'intéresse d'abord à l'architecture globale d'un système de gestion de dialogue et détaille l'enchaînement des modules qui permettent de passer de la parole d'un utilisateur à une réponse adaptée du système, toujours sous forme orale, en passant par ce qui focalisera la suite du manuscrit : le gestionnaire de dialogue. Le chapitre se termine par deux paragraphes qui indiquent que des approches statistiques s'appuyant sur le formalisme de l'apprentissage par renforcement sont maintenant les plus appropriées pour appréhender le développement de tels systèmes. L'utilisateur est donc considéré comme faisant partie de l'environnement et le système traite des séquences d'actes de dialogue. Le but est alors "d'optimiser" ce module en utilisant des corpus de données et des interactions directes avec l'utilisateur. Mais, comme on le verra un peu par la suite, et cela aurait pu être plus détaillé par Nicolas, on ne sait pas encore très bien ce que l'on veut optimiser, quels sont les critères, les enjeux.

C'est donc naturellement que le chapitre suivant (chapitre 4) traite de l'apprentissage du gestionnaire de dialogues en utilisant l'apprentissage par renforcement. Ce chapitre est plus technique et plus formel et Nicolas y présente avec clarté et pédagogie la théorie sous-jacente des Processus Décisionnels de Markov, ce qui lui permet d'exposer les problèmes génériques (dilemme exploration/exploitation, estimation de la fonction de valeur, apprentissage en ligne ou hors-ligne) de cet apprentissage. Si quelques petits détails m'ont interpellés (un processus *Markov Decision Process* en page 34 ou l'équation (4.3) qui est étrange), et si je pense que les exemples proposés auraient été plus clairs si les états avaient été bien définis, le tout est rigoureux et montre une réelle maîtrise de l'outil théorique.

Il n'en reste pas moins que ce chapitre me laisse un peu sur ma faim du point de vue de l'analyse de l'adéquation, ou non, de l'apprentissage par renforcement avec la problématique visée par Nicolas, et de la pertinence de l'apprentissage par transfert. Par exemple, Nicolas présente quelques limites ou problèmes de l'apprentissage par renforcement sans faire de liens explicite avec sa problématique. Prenons le dilemme exploitation/exploitation, Nicolas ne nous explique pas en quoi ce problème se traduit dans le cas d'un nouvel utilisateur, il nous dit juste que ce problème existe. "Et alors ?", serait-on tenté de demander. De plus, quand Nicolas aborde, malheureusement trop rapidement, la pertinence de son approche (section "*Learning from scratch*"), la logique de son argumentation est maladroite. Il commence par nous dire qu'il n'est *pas difficile* d'apprendre une politique de dialogue à partir

1-2

2 notes:

03

Nicolas Carrara

de rien en utilisant l'apprentissage par renforcement en ligne. Où est le problème alors? Mais, juste après, il nous dit que les premières interactions sont cruciales, et que pour gérer *ce problème* (donc c'est plus difficile que prévu), plusieurs options s'offrent à nous, dont le transfert. Soit. Mais l'argumentation est en plus un peu courte. D'autant que le chapitre suivant reste assez descriptif et manque aussi un peu d'analyse.

Le chapitre 5 présente en effet l'apprentissage par transfert et son utilisation pour adapter un système de dialogue à un utilisateur. Comme précédemment, Nicolas y expose avec clarté, précision et aisance la nature de ce type d'apprentissage et les problèmes qu'il pose, notamment le choix de ce qui est transféré, de la manière de le transférer et de quand cela doit être transféré. Le tout reste descriptif. Nicolas décrit ensuite quelques travaux de la littérature sans nous donner une analyse un peu poussée des limites de ces travaux et des liens avec ses objectifs.

La fin de ce chapitre se concentre sur le problème de l'adaptation à l'utilisateur, problème effectif et important comme le montre Nicolas de manière rapide, mais convaincante, en s'appuyant sur plusieurs références de littérature. Nicolas détaille ensuite 5 travaux de l'état de l'art. Je suis un peu surpris d'y trouver 2 références aux propres contributions de Nicolas, ce qui fait que **seulement** trois travaux sont décrits, l'un étant un travail qui a été prolongé par Nicolas au cours de sa thèse. **Or, même pour ce travail qui sera prolongé, Nicolas ne nous dit pas explicitement pourquoi il faut prolonger ce travail.** Quels en sont les limites? Sont-elles vraiment contraignantes? J'appuie sur ce point mais je trouve vraiment que cette première partie, mélange d'état de l'art et d'apports théoriques, reste trop descriptive, Nicolas n'y fait pas ressortir son esprit d'analyse et son sens critique.

Dans la **deuxième grande partie** du manuscrit, Nicolas expose ses contributions pour l'adaptation par transfert en se focalisant sur l'efficacité du dialogue. Le chapitre 6 commence par rappeler les motivations de ce travail (mais sans dire en quoi le travail qu'il prolonge est limité). Le principe de la proposition de Nicolas est de choisir d'utiliser des ressources préexistantes (dialogues avec des utilisateurs précédents) pour augmenter le corpus avec lequel le système va apprendre un système de dialogue adapté. Comme la base de donnée préexistante est conséquente, l'idée principale est de regrouper (faire du "*clustering*") ces données et d'en utiliser des échantillons représentatifs, en utilisant une méthode de *k-means* ou une méthode de *k-monoïde*, ce qui nécessite de définir une distance inter-utilisateur originale. Le choix des bons représentants est ensuite vu comme un problème de bandit multi-bras. L'idée est intéressante et j'aurais juste aimé un peu plus de détails, ou de formalisme, sur ce que sont concrètement ces "*source representatives*" car en l'état actuel du texte, cette information est distribuée et je ne suis pas sûr d'avoir pu en saisir la teneur exacte.

Les approches proposées par Nicolas sont ensuite validées par des expériences intelligemment agencées. Dans un premier temps, il vérifie qu'un système adapté à un utilisateur est plus performant qu'un système adapté sur d'autres usagers. Ensuite, la sélection de représentants pertinents est testée en comparant avec une méthode agglomérante (qui ne fait pas de choix) et une méthode qui fait un choix mais qui ne fait pas d'apprentissage ensuite. Les résultats donnés montrent que la contribution de Nicolas donne de bons résultats sur des utilisateurs (virtuels) fait à la main mais ces résultats sont moins probants sur des utilisateurs conçus comme des modèles d'utilisateurs humains. Ces résultats, peu nombreux, sont à peine commentés par Nicolas. Il y aurait pourtant matière à discuter ces expériences et ces résultats. On pourrait par exemple discuter de la validité des modèles d'utilisateurs utilisés, se demander si la génération de nombreux utilisateurs virtuels en faisant varier les paramètres donne un ensemble homogène, proche des utilisateurs les générant ou permet, au contraire, de générer des utilisateurs totalement différents. De même, si la figure 6.2 est une bonne idée pour comparer des politiques et permet de faire un début d'analyse de certains résultats, ces comparaisons restent qualitatives et subjectives ("*quite similar*", p. 63) et on se demande s'il n'est pas possible de quantifier ces différences (par exemple avec la distance proposée), ce qui serait aussi utile en complément des résultats de la figure 6.3.

Une deuxième contribution de Nicolas CARRARA est exposée dans le chapitre 7 où il explore le concept de transfert autour de l'algorithme d'apprentissage par renforcement profond DQN ("*Deep Q-Learning Network*"). Deux questions sont rapidement abordées, à savoir comment choisir la source à transférer et que transférer.

Honnêtement, ce chapitre n'apporte pas beaucoup à la thèse. D'une part, les expériences réalisées avec des auto-encodeurs ne sont que résumées. Pour leur donner plus de poids, il aurait été préférable d'en donner les détails et de les approfondir. Ce n'est pas parce que les résultats sont négatifs qu'ils ne doivent pas être développés, au contraire, une thèse est un bon support pour ce genre de travail. De plus, la réflexion autour de l'utilisation ou non de la *TD-error* pour guider le choix de la source, bien que rapide, semble indiquer à l'avance que les expériences vaguement évoquées en section 7.2 ne sont pas construites sur des bonnes bases, ce qui questionne leur évocation même. C'est dommage car les questions abordées sont des plus intéressantes, comme le signale avec pertinence Nicolas, mais le traitement qui en est réalisé ici n'apporte pas grand chose.

La **troisième grande partie** du manuscrit de Nicolas CARRARA se focalise sur des politiques de dialogues “safe” ou “sûres”. Plusieurs approches de la littérature sont exposées pour appréhender ce concept et Nicolas se concentre sur les Processus Décisionnels de Markov Contraints qui s'apparentent à une optimisation multiobjectifs (récompense et coût) avec comme idée originale de rechercher des politiques budgétisées dans des environnements continus pour des horizons infinis. L'idée est intéressante et je regrette seulement que la notion de “politique sûre”, et surtout les critères de cette sûreté, n'aient pas été plus formalisés dans le cadre des systèmes de dialogue avant de passer à la description formelle des outils disponibles.

Dans le chapitre 8, les développements formels, théoriques et algorithmiques sur l'utilisation d'un nouvel opérateur de Bellman budgétisé constituent un morceau de choix de cette thèse. La démarche est rigoureuse, formelle et montre une analyse en profondeur de la solution proposée et des propriétés que l'on peut exploiter d'un point de vue algorithmique. L'intuition concernant la non-contractivité de l'opérateur \bar{T}^* est non seulement démontrée par un contre-exemple mais aussi prise en compte en proposant des hypothèses plus contraignantes sur l'évolution de la fonction de coût pour donner un cadre théorique assurant la convergence des algorithmes. C'est un travail sérieux, conséquent, s'appuyant sur des preuves claires et validés par des expériences intelligemment menées. La présentation de cette belle contribution aurait pu être améliorée en expliquant un peu plus certains points et en donnant quelques exemples pédagogiques supplémentaires, pour en faciliter la compréhension et faciliter son appréhension. Ainsi, je ne suis pas sûr d'avoir bien saisi l'importance de pouvoir agir sur le budget à chaque action, ou de bien comprendre les implications de borner l'espérance du budget d'action (β_a) lors de l'exploration. De même si l'idée de s'appuyer sur la “coque” (*hull*) de la politique est appropriée et judicieuse, le concept lui-même n'est pas aisé à bien saisir.

Le chapitre 9, bien trop court, propose une contribution élémentaire à l'utilisation d'une politique “sûre” dans le cadre de l'apprentissage par transfert. En s'inspirant des politique ϵ -gourmandes, Nicolas propose de choisir entre l'action de la politique “sûre” et l'action d'une politique d'exploration avec un facteur de probabilité ϵ . Nous devons nous contenter de deux courbes à peine discutées montrant les résultats sur un problème très simple. Je comprend la volonté de Nicolas CARRARA d'aller jusqu'au bout de sa démarche de transfert, mais ici on reste vraiment sur sa faim.

Le manuscrit de Nicolas CARRARA se termine par une conclusion dans le chapitre 10. Il y fait un bilan de ses contributions en ouvrant sur quelques perspectives de haut niveau assez générales. C'est synthétique et rapide.

L'annexe qui suit est intéressante à plusieurs points de vue. D'une part, on y trouve les preuves détaillées et rigoureuses des différents lemmes et théorèmes du chapitre 8. Avoir relégué ces preuves en annexe est une bonne idée qui allège la lecture du manuscrit sans nuire à sa qualité scientifique. D'autre part, Nicolas s'attache à la reproductibilité de ses expériences, en listant la valeurs des paramètres de ses algorithmes et de ses modèles, en communiquant son code via un dépôt *git* et s'appuyant sur une charte de reproductibilité en apprentissage machine. C'est juste dommage que cela ne soit pas le cas pour les expériences, même préliminaires, du chapitre 7.

Avis sur le manuscrit et le travail de thèse

L'ensemble des contributions présentées par Nicolas CARRARA dans son manuscrit démontre un travail scientifique intéressant, original et pertinent sur une problématique actuelle de l'intelligence artificielle. Toutes ces contributions ne sont pas présentées avec le même souci du détail et la même profondeur d'analyse, ce qui est dommage. Le manuscrit bénéficie de l'aisance d'écriture de Nicolas qui s'exprime clairement et agréablement. Néanmoins, je trouve que ce manuscrit, et le choix qui est fait de s'appuyer trop largement sur la "simple" réutilisation des articles qu'il a publiés, dessert Nicolas car il ne lui permet pas de mettre en avant ses capacités à analyser et prendre du recul, capacités qui sont alors difficiles à évaluer. De plus, la contribution la plus intéressante de ce thèse qui concerne son travail autour de la recherche de politiques contraintes (chapitre 8) est obscurcie par le fait qu'on ne sait pas exactement la part de Nicolas dans cette contribution car Nicolas reprend le travail décrit dans un article avec plusieurs coauteurs.

Je veux aussi souligner l'importance accordée par Nicolas CARRARA à la reproductibilité de ses expériences, ce qui renforce la qualité de sa démarche scientifique.

Au final, j'ai un avis assez positif sur le manuscrit et le travail de Nicolas CARRARA qui m'a convaincu qu'il savait mettre en œuvre une démarche scientifique pour aborder des problèmes complexes et intéressants. Je donne donc un avis favorable à la soutenance de thèse de Nicolas CARRARA en vue de l'obtention du Doctorat de l'Université des Sciences et Technologies de Lille.

Nancy, 04 novembre 2019
Alain DUTECH



Thesis rapport Alain

Scientifique, Campus

- | | | |
|-----------------|-----------------|--------|
| 01 | Nicolas Carrara | Page 2 |
| 18/12/2019 8:51 | | |
| 02 | Nicolas Carrara | Page 2 |
| 18/12/2019 8:51 | | |
| 03 | Nicolas Carrara | Page 2 |
| 18/12/2019 8:52 | | |
| 04 | Nicolas Carrara | Page 3 |
| 18/12/2019 8:54 | | |
| 05 | Nicolas Carrara | Page 4 |
| 22/12/2019 6:38 | | |
| 06 | Nicolas Carrara | Page 5 |
| 22/12/2019 6:38 | | |