

Au directeur de l'ED EDSPI

Avignon, le 27 novembre 2019

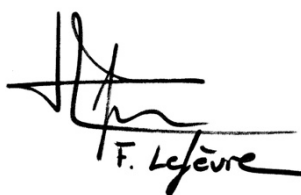
Objet : rapport de thèse Nicolas Carrara
N/ Réf. :

Monsieur le Directeur, Cher collègue,

Je vous joins ci-après mon rapport sur le manuscrit de thèse de Nicolas Carrara dont vous m'aviez confié la lecture et l'appréciation. Le travail présenté offre tous les critères me permettant de donner un accord à la soutenance prévue.

Je reste à votre disposition pour toute information supplémentaire.

Bien cordialement,



Fabrice Lefèvre
Full Professor in Computer Science

Equipe
Vocal Interactions Group

Affaire suivie par
Fabrice LEFEVRE

Téléphone
+33(0)490843563

Courriel
fabrice.lefevre@univ-avignon.fr

AVIGNON UNIVERSITE

Campus J.-H. Fabre
74 rue Louis Pasteur
84 029 Avignon cedex 1

Tél. +33 (0)4 90 16 25 00
Fax. +33 (0)4 90 00 00 00
courriel@univ-avignon.fr
univ-avignon.fr

Rapport sur le manuscrit de thèse de Nicolas Carrara

A l'attention du comité d'évaluation de l'Ecole Doctorale,

Ce document propose un rapport initial sur le manuscrit de thèse soumis par Nicolas Carrara. La thèse a été supervisée par le Prof. O. Pietquin à l'Université de Lille au sein du laboratoire Cristal dans l'équipe SequeL, et réalisée dans le contexte d'une collaboration CIFRE avec Orange Labs à Lannion et l'équipe NADIA, encadrée par Drs R. Laroche et T. Urvoy.

La thèse présentée a pour objectif principal l'optimisation des systèmes de dialogues appris par renforcement à l'aide d'une adaptation au locuteur. Le moyen principal envisagé pour atteindre cet objectif est le transfert. Il s'agit de détecter le locuteur pour lequel nous disposons déjà de données, le plus proche du nouveau locuteur afin de transférer les connaissances disponibles dans son modèle de dialogue, représentées ici par la stratégie de dialogue optimale lui correspondant.

Paradigme général en apprentissage automatique, l'apprentissage par transfert, habituellement envisagé à travers des domaines ou tâches applicatives, est utilisé ici comme moyen de fournir des données initiales à l'apprentissage d'une stratégie de gestion du dialogue pour un nouveau locuteur. Du fait de son importance dans la direction des travaux présentés, je suggère d'ailleurs que la notion de transfert soit intégrée explicitement dans le titre de la thèse.

D'emblée on notera que ce positionnement implique plusieurs objections. D'abord sur l'hypothèse que les stratégies de dialogue sont transférables entre locuteurs, considérer qu'il existerait des paires de locuteurs dont les stratégies optimales seraient plus proches que de n'importe quelle autre stratégie ('générique' par exemple) mériterait d'être étayé, au moins conforté par des observations. Mais aussi qu'il existe une stratégie de dialogue représentative d'un locuteur. Il semble que cette hypothèse n'est pas très réaliste si nous tenons compte de l'effet de co-adaptation entre système et humain, qui va conduire ces

derniers à faire évoluer leur stratégie d'une façon et dans une mesure que nous sommes encore loin d'appréhender correctement (malgré quelques travaux initiaux, entamés notamment par le directeur de thèse du candidat, et dument référencés). Toutefois ceci ne remet pas en cause la démarche de la thèse. Même si l'on peut donc questionner le mode d'application du transfert tel que conçu dans la thèse le principe d'un transfert initial de connaissance pour l'amélioration de l'apprentissage des stratégies de dialogue reste valide (surtout que ces hypothèses initiales sont remises en cause progressivement dans la thèse, par exemple par l'usage de prototypes de regroupement d'utilisateurs).

Le corps de la thèse consiste en 84 pages dans 3 parties principales, subdivisées en 7 chapitres, que viennent compléter : une introduction (10p), une conclusion (2p), des annexes (11p) et enfin une bibliographie (20p). L'ensemble du document, et ses nombreuses tables (index, sommaire, acronymes...) atteint 140p. La bibliographie est bien formatée et, ce qui est remarquable, elle contient peu de références de pré-publications. Toutefois il reste des références sans indication du support de publication (il est fort possible qu'il s'agisse vers d'occurrences en provenance d'arXiv ?, mais il faudrait le préciser).

L'introduction aux systèmes de dialogue est un survol assez large du domaine, qui a le mérite d'une qualité pédagogique évidente pour des lecteurs non-spécialistes. Les quelques imprécisions et lacunes seront dans un certain nombre de cas rattrapées dans les compléments offerts ensuite dans chacune des parties décrivant les contributions.

La première partie de la thèse est consacrée à mettre en place le protocole expérimental global avec l'ensemble des composants que le candidat entend traiter ensuite : un système de dialogue, un gestionnaire de dialogue, l'apprentissage par renforcement et le transfert. L'apprentissage en ligne étant un thème a priori perçu comme fondamental de la thèse, on pourra regretter un état de l'art peu clair sur le sujet et ses enjeux (« In dialogue system applications, it is *not* usually hard to learn a DP from scratch using Online RL algorithms. » ??) et lacunaire dans sa présentation. Ainsi il m'est malheureusement facile de détecter que les travaux menés dans mon propre laboratoire, dès 2012, ne sont pas mentionnés (par exemple : Emmanuel Ferreira and Fabrice Lefèvre. 2013. Social signal and user adaptation in reinforcement learning-based dialogue management. In Proceedings of 2nd Workshop on Machine Learning for Interactive Systems (MLIS '13)), alors que certaines références indiquées ne sont pas en rapport avec la notion d'*apprentissage direct* avec des usagers. Mais plus fondamentalement ce sont les différents impacts sur l'apprentissage des stratégies qui ne sont pas explicités, comme la vitesse d'apprentissage en relation avec la quantité de données nécessaires ou la stabilité des solutions (convergence, tracking...).

La deuxième partie de la thèse propose de s'attaquer aux solutions nécessaires pour un passage à l'échelle des techniques de transfert dans le contexte des systèmes de dialogue. Le chapitre 6 s'intéresse à l'amélioration d'une solution existante pour le regroupement de stratégies de locuteurs. Il est implémenté selon deux critères (kmedoids et kmeans) qui permettent de sélectionner (à l'aide d'un bandit multi-bras) des échantillons d'apprentissage source utilisés lors de l'évaluation. Le protocole est évalué sur le « negotiation dialogue game. »

Ce jeu de dialogue présente l'intérêt d'être suffisamment simple pour permettre une simulation. On notera le recours à une modélisation des erreurs de transcriptions bien conçue. Par ailleurs une partie de l'expérience a pu être réalisée avec des locuteurs humains. Mais dans tous les cas les résultats n'affichent pas de différences réellement notables par comparaison à une simple agrégation (aléatoire) de données d'apprentissage. Du moins c'est l'impression que l'on peut s'en faire en l'absence d'une analyse en signification statistique des résultats.

Puis le chapitre 7 discute des potentialités de l'application du transfert dans le contexte du renforcement profond. Ce chapitre pose un vrai problème il s'agit essentiellement de suggestions sur la possibilité d'appliquer le transfert au DQN. Les conclusions proposées ne sont pas le résultat d'expériences dument décrites et rapportées. Dès le début il apparaît assez évident que l'erreur TD ne permet pas d'établir une bonne métrique pour évaluer la proximité entre les réseaux Q. En l'absence d'évolution de cette constatation au cours du chapitre, il n'est pas nécessaire d'avoir à y revenir aussi souvent. Si les suggestions ont réellement donné lieu à des expérimentation, même avec de mauvais résultats, elles devraient être résumées dans le document pour supporter les conclusions principales du chapitre. Aussi je proposerais que dans la version finale du document un remaniement consistant à le décaler en Annexe en association avec un court paragraphe dans le document (attaché au chapitre précédent par exemple) présentant cette déclaration d'intention et sa conclusion, et renvoyant à l'annexe pour une discussion plus longue sur le sujet.

La troisième et dernière partie de la thèse va décaler légèrement le problème en proposant de le revoir sous l'angle de la sûreté (*safety*). Cette notion même n'est pas très bien définie dans notre communauté, et présente donc un grand facteur de nouveauté. Généralement les effets positifs et négatifs d'une stratégie de dialogue sont vus globalement et matérialisés dans une fonction de récompense qui regroupe arithmétiquement leur influence. Il s'agit ici d'extraire des résultats possibles de l'application d'une stratégie certains effets considérés comme dangereux, dommageables du point de vue applicatif. Appliquée au dialogue humain-machine, le candidat considère principalement la dangerosité du point de vue de

l'acceptation de la machine par l'utilisateur (danger = cela fait raccrocher l'utilisateur), on pourra aisément étendre la notion à la possibilité pour la machine d'induire l'utilisateur en erreur sur des sujets graves.

Le chapitre 8 représente certainement la contribution la plus originale et conséquente de la thèse. Il s'agit de proposer un algorithme permettant d'apprendre un modèle de décision markovien avec budget fonctionnant sur un espace d'états continus. La notion de MDP avec budget, proposée par Boutilier et al, permet de compléter la fonction de récompense avec une notion de risque a priori matérialisée par une fonction de dépense (ou signal de coût) à ne pas dépasser. Evidemment cela complexifie grandement l'apprentissage d'un tel modèle (principalement deux fonctions Q doivent être apprises de manière cohérente, l'une classique associée à la fonction de récompense l'autre à un signal de coût, avec la contrainte de rester en-deçà d'une borne supérieure, aka *le budget*).

Un effort très important est fait ici pour rendre le modèle compatible avec des états continus, et permettre ainsi de prendre en compte des variables continues de l'état du dialogue (comme le score de confiance des modules de transcriptions ou de compréhension). Même si cette direction de recherche paraît pertinente il aurait été appréciable qu'elle soit aussi comparée aux performances obtenues avec l'approche de base appliquée à des états obtenus par quantification discrète.

De façon assez complète la démonstration de la solution proposée est fournie en annexe. N'étant pas expert de l'apprentissage par renforcement je laisserai mon collègue rapporteur confirmer son exactitude (la publication de la méthode à NeurIPS'19 plaidant déjà en sa faveur). Par contre étant informaticien je me suis empressé de chercher à tester moi-même l'algorithme à l'aide des informations fournies dans l'annexe 3 Reproducibility. Malheureusement les adresses sont incorrectes, je n'ai donc pas pu tester l'algorithme au moment d'écrire ces lignes. Comme je ne doute pas que l'outil soit bien disponible en ligne il faudra corriger cela dans la version finale du manuscrit. Par ailleurs les expériences portant sur la conduite autonome et les couloirs semblent être le fruit d'une collaboration avec un autre thésard (liste d'auteurs du papier correspondant). Si c'est le cas il est impératif de le préciser et de mentionner la part de chacun dans le travail. Une thèse de doctorat est un travail fondamentalement individuel, hors la supervision du directeur de thèse et sauf mention contraire explicite.

Un autre aspect qui est peu discuté dans la thèse et qui est pourtant d'une grande importance est la quantité de données nécessaire pour l'apprentissage. En effet l'apprentissage de stratégies en ligne suppose l'utilisation du système par des humains, ce qui ne peut donc impliquer des milliers d'heures de données comme dans le cas des corpus

collectés (WoZ...) ou avec simulateurs. Cette dimension est bien sûr à mettre en relation avec la proposition d'exploration sensible au risque (§8.3.2). Toutefois la création de nouvelles transitions (Algorithme 5 ligne 11) ne me paraît pas complètement claire. Plus d'explications, et un plus large rappel de l'épsilon-greedy du FTQ, seraient surement profitables.

De même la comparaison faite avec la possibilité d'introduire les précautions directement dans la fonction de récompense FTQ(λ) ne semble pas concluante. Si de bonnes raisons existent pour envisager de dépasser le cadre simple des MDP (l'exemple en §8.1.1 relève d'une excellente démarche mais n'est finalement pas facilement interprétable), il reste que la complexification de l'apprentissage de BMDP n'est pas négligeable (notamment en termes de temps de calcul, même si des solutions efficaces pour la parallélisations des opérations sont proposées dans la thèse) et l'absence d'amélioration des performances semble donc défavorable. Du moins dans le cadre des expériences réalisées ici qui, on en discute plus loin, ne sont peut-être pas les plus à même de mettre en avant les propriétés intéressantes des BMDP. Par ailleurs la description des expériences et de leurs résultats pourrait être complétée, la façon dont le signal de coût est implémenté, le nombre de runs effectués (Nseeds, Ntrajs...), la taille des batchs intermédiaires... et les figures 8.4 et 8.4 sont complexes à interpréter.

Finalement, devant la difficulté de mise en œuvre de la solution du chapitre précédent, le dernier chapitre (9) envisage de se restreindre à une solution moins globale au problème mais pratiquement plus exploitable : l'épsilon-safe. Il faudrait à cet égard corriger la dernière phrase de conclusion (p. 91) qui laisse entendre que le cadre des BMDP (chap. 9) est utilisé dans le chapitre suivant.



En effet il s'agit alors d'obtenir une certaine sécurité en contrôlant plus sévèrement l'exploration durant le transfert, notamment en la biaisant en direction d'une solution existante connue pour être sûre (extraite du corpus de stratégies disponibles des locuteurs précédents). Cette solution est à rapprocher d'études précédentes où l'exploration était contrainte par des règles expertes, qui permettait tout à la fois de guider la stratégie vers des couples états-actions permettant de progresser dans la tâche mais aussi d'éviter les non-sens (eg E. Ferreira and F. Lefèvre, "Expert-based reward shaping and exploration scheme for boosting policy learning of dialogue management," IEEE ASRU, Olomouc, 2013). Dans le cadre des expériences slot-filling reportées la méthode ne permet pas d'amélioration notable de la courbe d'apprentissage de la stratégie. Ce qui est finalement assez attendu vu le peu d'écart que représente la notion de sûreté dans le cas présent sur les trajectoires engendrées (où elle se cantonne à : éviter l'action « repeat-numpad »).

Enfin la thèse s'achève par une conclusion assez courte reprenant les points marquants développés dans le document, suivie d'une discussion (très rapide aussi) des perspectives à moyen et long terme du domaine. La conclusion est très lucide sur les résultats des expériences, la discussion aurait comparativement mérité d'être plus précises sur les voies d'amélioration possibles des propositions de la thèse.

De cette conclusion on extraira une remarque qui si elle souligne bien l'esprit critique de l'auteur révèle aussi une légère contradiction du travail : "Despite good results with handcrafted users, our experiments suffer from the fact that the toy game we experiment on and the models used are too simple to extract discriminative behaviours among human-model users. » En effet les « toy games » ont pour but de favoriser la démonstration des bonnes propriétés des approches proposées. Aussi ils doivent être *construits* dans ce sens ! Il est donc regrettable de ne pas avoir su introduire dans ces exemples jouet les caractéristiques nécessaires à la bonne exploitation des propositions. Car il est de même regrettable que dans un contexte impliquant pourtant un partenaire industriel (Orange) aucune donnée réelle ne soit utilisée. En général on peut reprocher à ceux qui évitent la confrontation avec les données réelles de vouloir préserver le doute sur l'efficacité réelle de leur méthode. Ici au contraire, des données et des modèles plus réalistes, moins lisses, auraient pu mieux convaincre de l'intérêt des solutions.

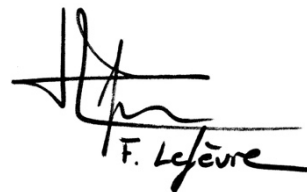
Le texte est globalement clair et bien rédigé. L'iconographie est de bonne qualité, mais pourrait être mieux homogénéisée et un peu complétée (des graphes pour illustrer les algorithmes complexes sont toujours bienvenus pour mettre en place des flux complexes et leur notation, et aider les lecteurs les moins familiers du domaine). Les références sont nombreuses et généralement appropriées. Chacune des sous-parties traitées est relativement autonome, et re-propose ses motivations avec un rappel rapide de l'état de l'art. De plus, plusieurs publications dans des conférences internationales résultent de ce travail de thèse (5, dont une NeurIPS cette année). Elles contribuent à appuyer l'intérêt perçu par la communauté scientifique des travaux présentés dans le document. Enfin il est notable que la thèse présentée a une réelle cohérence de bout-en-bout, contrairement à de nombreuses thèses de type agrégats de papiers ayant seulement le domaine comme lien (tenu) entre les parties (et bien que le découpage selon les publications de l'auteur soit évident).

Ainsi, les quelques remarques et commentaires précédents ne sauraient occulter que le candidat nous présente un ensemble de travaux conséquent qui a toutes les caractéristiques permettant de poursuivre jusqu'à la soutenance. La thèse est notamment riche en contributions originales et expérimentations, dont les analyses proposées seront d'un apport

certain pour les recherches en systèmes d'interactions humain-machine, ainsi que pour leurs applications industrielles.

Je donne un avis favorable à l'autorisation de soutenance pour l'obtention du titre de docteur de l'Université de Lille.

Cordialement,



Pr. Fabrice Lefèvre
Professeur en Informatique

Thesis rapport Fabrice

Universite, Avignon

01	Nicolas Carrara	Page 3
22/12/2019 6:33		
02	Nicolas Carrara	Page 3
22/12/2019 6:40		
03	Nicolas Carrara	Page 4
22/12/2019 6:41		
04	Nicolas Carrara	Page 5
22/12/2019 6:42		
05	Nicolas Carrara	Page 5
22/12/2019 6:43		
06	Nicolas Carrara	Page 6
22/12/2019 6:43		