

Introduction générale

“Nous entrons dans un nouveau monde. Les technologies d’apprentissage automatique, de reconnaissance de la parole et de compréhension du langage naturel atteignent un niveau de capacités. Le résultat final est que nous aurons bientôt des assistants intelligents pour nous aider dans tous les aspects de nos vies.”

– Amy Stapleton, Opus Research

Plus de quarante ans se sont écoulés depuis la présentation du premier assistant virtuel commandé vocalement par la compagnie IBM (VanDijck, 2005). Déjà à cette époque là, ce fut présenté comme une révolution technologique. Un programme qui pouvait reconnaître 16 mots et les chiffres de 0 à 9. Quelques générations plus tard, nous nous retrouvons avec des assistants capables de reconnaître, comprendre et parler plusieurs langues. Ces assistants intelligents sont la nouvelle génération d’intelligences artificielles capables de s’intégrer dans nos vies personnelles et professionnelles.

Un domaine qui a particulièrement émergé est celui de l’aide à la manipulation d’un ordinateur personnel. Selon certains experts (Milhorat et al., 2014 ; Trappl, 2013 ; Knote et al., 2018), l’époque où nous utilisons encore le clavier, la souris et l’écran est une étape de transition. Le futur se trouve dans l’exploitation de la parole comme moyen de communication principal avec les machines. Et ce futur est proche. La course à la réalisation d’assistants qui excellent dans plusieurs domaines a commencé il y a quelques années avec l’entrée de grandes compagnies comme Google et Apple dans le secteur (Tulshan and Namdeorao Dhage, 2019). Par la suite, de très grands efforts ont été fournis dans le but d’améliorer l’expérience d’utilisation de ces assistants. Les chercheurs ainsi que les industriels se sont orientés vers cette solution très rapidement.

Nous sommes sans doute actuellement entrain de vivre une époque importante de l’intelligence artificielle. Le temps où nous rêvions encore de converser avec une machine semble à la fois proche en terme de temps, mais loin en terme de progrès. Les utilisateurs réguliers de produits technologiques sont confrontés à une technologie nouvelle mais prometteuse. Ce secteur d’activités peut donc s’avérer très prolifique si les efforts fournis sont suffisamment conséquents.

En tant que novice dans ce secteur qu'est la personnalisation des services électroniques, l'Algérie devra rapidement se positionner pour s'incorporer dans l'évolution de ces technologies. Nous avons donc été motivés par l'envie de nous initier à ce domaine, ainsi que de contribuer aux travaux d'autres compatriotes scientifiques qui traiteraient de ce sujet. Nous pensons que les plus ambitieux des projets commencent avec des contributions à petites échelles. Notre assistant aura pour but d'améliorer l'expérience d'utilisation d'un ordinateur. Ceci en effectuant des tâches rudimentaires, efficacement et sans réel effort hormis l'énonciation de la requête. En utilisant des techniques d'intelligence artificielle d'actualité comme la reconnaissance automatique de la parole, la compréhension du langage naturel et l'apprentissage par renforcement, ce projet se veut assez ambitieux et vise à faire gagner du temps à tout utilisateur d'un ordinateur de bureau ou portable.

Dans cette optique, nous sommes inspirés de notre étude des travaux de la littérature sur les assistants personnels intelligents. Nous passerons en revue les aspects théoriques qui sont utilisés dans des solutions considérées comme état de l'art du domaine. Nous nous baserons sur ces techniques pour la conception des modules de notre système tout en les adaptant à nos besoins.

Ce mémoire se constitue de quatre chapitres. Le premier chapitre sera consacré à la présentation des assistants virtuels intelligents. Le deuxième chapitre se focalisera sur l'étude des travaux existants liés à la thématique de notre sujet. Le troisième chapitre traitera de l'étude conceptuelle. Le quatrième et dernier chapitre présentera notre système Speech2Act avec son évaluation, une synthèse des résultats obtenus et notre interface pour l'application. Enfin nous présenterons notre interface pour l'application et nous conclurons ce travail avec une conclusion générale et les perspectives envisageables.

Conclusion générale

La création d'un nouveau besoin qu'est la personnalisation des services, a été la principale source de motivation pour ce projet. Cependant, et par faute de temps ainsi que de moyens techniques (surtout en ce qui concerne le côté matériel pour l'apprentissage automatique), certains modules n'ont pas été exploités à leur maximum. Cela ne nous a pas empêché de réaliser un travail dont nous sommes particulièrement fières. Mais, nous gardons toutefois un esprit critique, ainsi qu'une objectivité envers le travail fourni.

Tout au long de la réalisation de ce mémoire, nous avons étudié l'état actuel des assistants personnels intelligents. Nous avons dû passer une majeure partie de cette étape à comprendre les fondements théoriques et conceptuelles de chaque composants de ces systèmes, principalement à cause de la grande densité de techniques, concepts et théories qui sont nouvelles pour nous.

Après avoir assimilés la totalité des concepts, et qui font office d'état de l'art du domaine, nous sommes arrivés à certaines conclusions. Tout d'abord, développer un système en partant de rien était un travail assez massif. Dépassant de loin le cadre d'un projet de fin d'études de master. Nous avons donc mis l'accent sur certains modules, le module de compréhension du langage naturel et le module de gestion du dialogue. Ces derniers dépendaient énormément de notre problématique. Le module de reconnaissance automatique de la parole a été sujet à une amélioration spécifique à nos besoins tout en exploitant un système de base déjà existant (à savoir DeepSpeech).

Avec une idée claire du travail à réaliser, nous avons pu entamer la conception de chaque module en y incorporant nos ajouts et modifications. Beaucoup de ces modifications sont le fruit de longues séances de débat et de discussions.

L'étape de réalisation est celle qui a pris le plus de temps. Entre la découverte de nouvelles technologies à utiliser, la collecte des données et l'apprentissage des différents modèles.

En ce qui concerne le module de reconnaissance automatique de la parole, l'ajout de ce modèle de langue a amélioré les résultats. Cela s'accordait avec nos prédictions théoriques.

La partie de construction du corpus pour ce modèle a été soumise à beaucoup d'optimisations incrémentales, en tombant à chaque fois sur un nouveau problème, ou bien un obstacle matériel (manque de puissance de calcul). En conséquent, les résultats n'étaient pas assez encourageants. Surtout si le but est de détrôner les systèmes propriétaires comme celui de Google qui réalise un score quasi-parfait sans apprentissage supplémentaire.

Pour le module de compréhension automatique du langage naturel, les ajouts faits au modèle d'apprentissage ont été expérimentalement validés dans le chapitre Réalisations et résultats. L'ajout de l'information morphosyntaxique a permis de donner plus de valeur sémantique à chaque mots de la requête. L'introduction d'erreurs aléatoires a quad à elle permit la gestion d'éventuelles erreurs que pourrait engendré le module de reconnaissance automatique de la parole. La construction de l'ensemble d'apprentissage à partir de zéro était l'étape la plus longue de la réalisation de ce module. Plus l'ensemble grandissait, plus il était difficile de maintenir sa validité sans l'intervention d'un soutien externe. De plus, vu que la tâche à accomplir était relativement simple et limitée, le réel impacte de cet ajout ne peut pas être certifié et validé dans un cadre plus général. Une autre problématique est celle du manque de données d'apprentissage consacrées au domaine de la manipulation d'ordinateurs. Ces données sont généralement construites manuellement par les développeurs du système. Un autre point à soulever est celui du manque de diversité dans les tâches réalisable par l'assistant. Cet ensemble de tâche reste facilement extensible. Il suffit d'ajouter des exemples assez exhaustifs à l'ensemble de données. Cela reste néanmoins une tâche lourde et manuelle à plus grande échelle.

En ce qui concerne le gestionnaire de dialogue, nous l'avons conçu pour qu'il soit facilement ajustable et mis à l'échelle. L'utilisation d'une architecture hiérarchisée d'agents de dialogue permet de grandement réduire la complexité de développement et d'ajout d'un nouveau gestionnaire de tâches dans le système. L'utilisation d'un système de gestion d'état du dialogue plus riche que ce qui est proposé en temps normal a permis à l'agent apprenant de mieux comprendre les échanges effectués avec l'utilisateur. L'utilisation d'un graphe de connaissances au lieu de trames sémantiques est justifié par le fait que ces derniers offrent une plus grande flexibilité dans la représentation de l'état du dialogue. Pour ce qui est des agents de dialogue, il sont entraîné en utilisant des techniques d'apprentissage par renforcement. Ils interagissent avec le simulateur d'utilisateur dans le but d'atteindre le but de l'utilisateur à travers l'optimisation d'une politique d'actions basée sur un système de récompenses.

Cependant, l'utilisation du graphe de connaissances et de l'apprentissage profond a un prix. En effet, la codification de la totalité du graphe en un seul vecteur de taille fixe était une tâche assez difficile. La taille de ce dernier augmente exponentiellement avec l'ajout de nouvelles connaissances. Ce qui rend le processus d'encodage très compliqué. De plus, la codification choisies pour les nœuds du graphe a eu pour effet de faire perdre de l'information sémantique. Le décodage du graphe s'en est trouvé grandement affecté.

Pour le dernier module, à savoir le module de génération du langage naturel. Il fût assez simple à réaliser. Son fonctionnement est très rudimentaires et intuitif. Il s'agissait d'uti-

liser un ensemble de phrases modèles et de remplacer les valeurs manquantes avec des valeurs réelles.

Pour ce qui est de l'application, nous avons fait le choix d'utiliser une architecture trois tiers basée web. Ce choix fût motivé par le fait que l'utilisation d'un serveur offre une puissance de calcul considérablement plus élevée que celle d'une machine personnel en local. L'interface reste assez simple et épurée. Le but est de minimiser l'interaction avec l'utilisateur avec tout autre moyen de communication que la parole. Cependant cette interface à aussi pour but de montrer les fonctionnalités du système. Elle est donc plus orientée vers le développement.

En ce qui concerne les perspectives envisageables pour ce travail, nous avons longuement réfléchi à des alternatives possibles pour certains modules.

Premièrement, l'avenir de systèmes Open source pour la reconnaissance automatique de la parole est très prometteur. Ces derniers offrent un moyen libre de mener des études et contribuer au développement à grande échelle de cette discipline. Une perspective future pour ce module serait de lancer notre propre plateforme de collecte de données. Cette idée a déjà été discutée dans la partie de l'étude de l'état de l'art. Malheureusement le temps a cruellement manqué. L'investissement de la communauté dans de telles initiatives n'en reste pas moins indéniable. Comme l'a démontré le projet CommonVoice de Mozilla.

Deuxièmement, pour le module de compréhension automatique du langage naturel, faire appel à une collecte massive de données est une solution explorable. Le développement d'un outil d'aide à l'annotation d'un corpus était elle aussi le sujet d'un long débat. Le manque de temps nous a poussé à retarder son développement. La mise à l'échelle d'une telle plateforme pourrait grandement faire avancer la tâche fastidieuse qu'est la collecte de données. L'automatisation de l'ajout de nouvelles fonctionnalités, comme de nouvelles intentions ou nouvelles étiquettes, n'a pas été discutée pendant la réalisation de ce travail. Mais, une intégration à la plateforme de collecte de données précédemment citée serait une solution de départ.

Ensuite, en ce qui concerne le module de gestion du dialogue, il est envisageable d'utiliser une méthode d'apprentissage semi-supervisée pour l'encodage des nœuds du graphe. Cette ajout pourrait permettre d'enrichir la valeur sémantique de ces nœuds, et facilitera la tâche au réseau de neurones de l'agent apprenant pour le décodage du graphe. Principalement grâce au fait que les nœuds dont les sens sont arbitrairement proches auront des codifications similaires.

En ce qui concerne le module de génération du langage naturel, une méthode plus sophistiquée comme l'utilisation de modèles d'apprentissage automatique basés encodeur-décodeur, ou bien basés convertisseur de graphes de connaissances en texte pourraient être utilisées. Cependant, ces architectures requièrent un très grand volume de données d'apprentissage annotés et spécifique à notre problématique.

Enfin, pour ce qui est de l'application principale, une amélioration possible serait le déploiement du serveur dans un service de Cloud Hosting. Cela permettrait de minimiser les

temps d'inférence et de post-traitement. Le développement d'une interface plus légère et plus orientée vers les cas pratiques est une amélioration possible. Garder les deux cas de figures, c.à.d. utilisation et développement, est aussi possible.

Pour conclure, nous estimons que la totalité du projet était une énorme occasion d'approfondir nos connaissances. Que ce soit celles qui nous ont été enseignées durant notre cursus, comme l'apprentissage automatique, le traitement automatique du langage naturel, le web-sémantique et la représentation de connaissances. Ou bien celles que nous avons apprises au cours de notre étude de la littérature, comme l'apprentissage par renforcement, le traitement automatique de la parole, le nettoyage des données, etc. Ce projet nous a aussi initié au travail en équipe pour la réalisation d'un projet assez conséquent. Tout en étant encadrés par nos supérieurs. Finalement, nous pensons que la plus grande satisfaction vient du fait que nous avons réalisé, dans un certain délai restreint, un travail qui traite d'un sujet récent et ambitieux. Et ainsi, de poser la première pierre à l'édifice pour, idéalement, encourager les chercheurs en Algérie à s'investir dans ce domaine.

Table des figures

Bibliographie

- Knote, R., Janson, A., Eigenbrod, L., and Söllner, M. (2018). The what and how of smart personal assistants : Principles and application domains for is research. In *Multikonferenz Wirtschaftsinformatik (MKWI)*, pages 1083–1094. Lüneburg, Germany.
- Milhorat, P., Schlögl, S., Chollet, G., , B., Esposito, A., and Pelosi, G. (2014). Building the next generation of personal digital assistants.
- Trappl, R. (2013). *Your Virtual Butler*. Springer Berlin Heidelberg.
- Tulshan, A. and Namdeorao Dhage, S. (2019). *Survey on Virtual Assistant : Google Assistant, Siri, Cortana, Alexa : 4th International Symposium SIRS 2018, Bangalore, India, September 1922, 2018, Revised Selected Papers*, pages 190–201.
- VanDijck, J. (2005). From shoebox to performative agent : the computer as personal memory machine. *New Media & Society*, 7:311–332.