



UMR 7503

LABORATOIRE LORRAIN DE RECHERCHE
EN INFORMATIQUE ET SES APPLICATIONS

Nancy le, 30 octobre 2008

Rapport de thèse de Mr CHRISTOPHE SERVAN

Apprentissage Automatique et Dialogue Homme-Machine Téléphonique à initiative mixte

La thèse de M. Christophe Servan concerne le dialogue oral Homme-Machine et s'intéresse plus particulièrement à une des composantes les plus difficiles : La compréhension de la parole. Il y a une littérature abondante sur le sujet mais aucune ne propose une solution satisfaisante.

Il est possible de classer les méthodes utilisées en compréhension en deux catégories. La première consiste à reconnaître la parole et ensuite à tenter de comprendre le contenu du signal de départ à partir des n-best. Cette solution est, en général, facile à mettre en œuvre à cause de la linéarité de fonctionnement de ses composantes. En revanche, elle présente l'inconvénient de ne pas exploiter dans la chaîne de compréhension, l'ensemble des informations contenues dans le signal d'entrée. Quant à la méthode utilisée dans la 2^{ème} catégorie, elle considère qu'il est judicieux de voir la compréhension comme un processus de « matching » entre un signal d'entrée et un signal composé d'unités de compréhension. L'auteur de cette thèse positionne ses travaux dans cette seconde classe.

Les 47 premières pages (chapitre 1 et 2) sont consacrées à l'étude de l'état de l'art allant de l'acoustique à la compréhension en passant par les modèles de langage. On y présente également la campagne MEDIA et les corpus qui ont servi aux différentes expérimentations du doctorant. Ce corpus est composé de 1250 dialogues enregistrés par ELDA selon un protocole du magicien d'Oz. Ce corpus est annoté avec des concepts sémantiques (au nombre de 83) développés dans le cadre du projet MEDIA. Le corpus audio est issu de parole téléphonique (fixe et cellulaire). Il est surprenant de trouver les résultats du LIA à ce niveau du document alors que le doctorant n'a pas encore présenté les principes de sa méthodologie, mais cela ne nuit aucunement à la clarté du document.

Le chapitre 3 est consacré au décodage conceptuel, il s'agit du cœur de la thèse du travail de Christophe. En complément de la campagne MEDIA, le doctorant présente ses travaux sur le corpus audio MEDIA.

Dans la section 3.2.2, le doctorant stipule, je cite «dans l'approche intégrée, le processus de compréhension est intégré dans le processus de décodage en remplaçant le modèle de langage par le modèle conceptuel ». Cette assertion n'est pas avérée, en effet, le décodage de la parole ne pourrait pas se faire que sur la base de concepts seulement, cela donnerait

Campus scientifique - B.P. 239 - 54506 VANDŒUVRE-LÈS-NANCY Cedex - France
Tél. +33 3 83 59 20 00 - fax +33 3 83 27 83 19 - +33 3 83 41 30 79

de très mauvais résultats et par ailleurs la formule 3.2 montre bien que le modèle de langage est pris en compte dans le processus de transcription.

La section 3.4 est consacrée à la formalisation du modèle de langage conceptuel, il est vu comme un HMM où chaque état émet plusieurs mots (un segment). Autrement dit, il est question d'étiqueter le corpus avec des étiquettes conceptuelles. La réécriture des formules aboutit à la formule 3.6 dans laquelle la probabilité de générer des étiquettes pour un ensemble de mots est le produit, sur l'ensemble des mots de la chaîne à étiqueter, de deux probabilités. La première est une sorte de modèle de langage « 3-étiquettes » utilisant un 3-grammes. La deuxième fait dépendre le mot w_i d'un bigramme et d'un « 3-étiquettes ». Je m'interroge sur le degré d'influence de la « 3-étiquettes » sur la génération de w_i . Je pense que son influence est minime voire nulle, le candidat aurait pu faire des expériences affirmant ou infirmant ce que j'avance. Ce qui est surprenant est que l'on attribue à chaque mot une étiquette même si ce mot est un mot non porteur de sens. Je crains que cette façon de faire n'introduise une ambiguïté d'étiquetage. Comment peut-on faire la différence entre : « Oui à midi » et « Oui à Marseille » ? L'étiquetage du mot outil « à » ne serait-il pas influencé par la dominance statistique du concept « LocalisationVille » ou de celle du concept « Temps » ? Une utilisation des séquences de mots récurrents permettrait de résoudre partiellement ce problème. Une autre solution aurait été de considérer le problème d'étiquetage comme un problème d'alignement entre deux chaînes de mots de longueurs différentes. Cela ne remet pas en cause la formulation du modèle conceptuel de Christophe, son approche est estimable et adaptée au problème.

Le reste du chapitre est consacré à la transformation d'un graphe de mots (produit par un système de reconnaissance de la parole) en une liste d'interprétation. Une grammaire de concepts est définie semble-t-il pour réévaluer les scores donnés par l'étiqueteur à base de HMM. Il n'est pas très facile à comprendre le fondement scientifique de cette approche. C'est comme si on accordait peu de crédits à l'étiqueteur stochastique et que l'on avait besoin de contrôler son résultat par un module supplémentaire. J'aurais voulu que le candidat s'attarde sur cette partie et justifie un tel choix. Par ailleurs, je me demande dans ce cas si une grammaire de concepts seule n'aurait pas été suffisante pour l'étiquetage.

L'étude de la variation du WER est faite selon deux protocoles de tests. Je trouve le second plus intéressant et plus rigoureux. La combinaison de plusieurs systèmes a permis de faire baisser le WER de plus de 13%.

Le chapitre 4 est consacré à l'étude et à l'analyse des résultats obtenus par le doctorant. Plusieurs tests sont menés, j'ai apprécié plus particulièrement celui qui permet de mesurer l'influence du WER sur le CER. Le résultat est prévisible mais les expériences l'attestent et montrent une relation linéaire entre le WER et le CER. Les expériences du tableau 4.3 donnent une idée sur la taille minimale du corpus d'apprentissage nécessaire à l'obtention de performances acceptables. Christophe ne dit pas ce qu'on entend par performance acceptable. Il aurait été intéressant d'étudier l'acceptation d'une interprétation par un humain en fonction de la valeur du CER. L'utilisation des courbes ROC comme il est de coutume de faire dans ces cas aurait apporté un peu plus d'éclairage aux résultats de Christophe.



UMR 7503

D'autres tests sont effectués dans l'objectif de comprendre les problèmes liés à la compréhension dans le dialogue et d'y apporter des solutions.

Le dernier chapitre est consacré au routage d'appels qui est en fait une tâche de classification ou d'identification thématique. Si l'on réussit à router vers le bon service, on peut considérer que le contenu du message a été compris. Ce chapitre a donc tout à fait sa place dans cette thèse.

La partie expérimentale utilisée dans cette étude est un système de routage d'appels appliqué aux renseignements téléphoniques du conseil général du Vaucluse.

La mise en œuvre de ce système a nécessité le développement de toutes les ressources nécessaires : système de reconnaissance (modèle de langage, modèle acoustique), un modèle conceptuel et un corpus de référence. Ces ressources ont été calculées ou adaptées de ressources existantes au LIA afin d'obtenir un système opérationnel. Malgré un WER important quasi un mot sur deux erroné, le système de classification obtient des résultats corrects. Cela peut être interprété de deux manières : la méthode mise en place par Christophe est robuste ou bien on n'a pas besoin de reconnaître toute la phrase mais seulement les unités porteuses de sens. Dans l'identification de thèmes, on utilise souvent une liste de mots caractéristiques du thème que l'on cherche à modéliser.

En conclusion, Mr Christophe Servan a présenté une thèse qui se défend sur un sujet difficile : la compréhension et le routage de la communication téléphonique. Christophe a mené plusieurs expériences intéressantes pour répondre aux obligations de la campagne d'évaluation MEDIA. Le travail informatique est sérieux, significatif et conséquent. Le système développé est fondé sur des principes scientifiques corrects. Son système est opérationnel même si au niveau fondamental il aurait pu mener une recherche moins conventionnelle.

Pour finir, Monsieur Christophe Servan a présenté une thèse respectable, c'est pourquoi je donne un avis favorable à la publication de sa thèse et à l'obtention de son diplôme de docteur en informatique.

K. Smaïli
Professeur des universités
Nancy Université

Campus scientifique - B.P. 239 - 54506 VANDŒUVRE-LÈS-NANCY Cedex - France
Tél. +33 3 83 59 20 00 - fax +33 3 83 27 83 19 - +33 3 83 41 30 79



UMR 7503

LABORATOIRE LORRAIN DE RECHERCHE EN INFORMATIQUE ET SES APPLICATIONS

Campus scientifique - B.P. 239 - 54506 VANDŒUVRE-LÈS-NANCY Cedex - France
Tél. +33 3 83 59 20 00 - fax +33 3 83 27 83 19 - +33 3 83 41 30 79

