TALN 2010, Montréal, 19–23 juillet 2010

Processus de décision à base de SVM pour la composition d'arbres de frames sémantiques

Marie-Jean Meurs Fabrice Lefèvre
Université d'Avignon et des Pays de Vaucluse
Laboratoire Informatique d'Avignon (EA 931), F-84911 Avignon, France.
{marie-jean.meurs,fabrice.lefevre}@univ-avignon.fr

Résumé. Cet article présente un processus de décision basé sur des classifieurs à vaste marge (SVMDP) pour extraire l'information sémantique dans un système de dialogue oral. Dans notre composant de compréhension, l'information est représentée par des arbres de frames sémantiques définies selon le paradigme FrameNet. Le processus d'interprétation est réalisé en deux étapes. D'abord, des réseaux bayésiens dynamiques (DBN) sont utilisés comme modèles de génération pour inférer des fragments d'arbres de la requête utilisateur. Ensuite, notre SVMDP dépendant du contexte compose ces fragments afin d'obtenir la représentation sémantique globale du message. Les expériences sont menées sur le corpus de dialogue MEDIA. Une procédure semi-automatique fournit une annotation de référence en frames sur laquelle les paramètres des DBN et SVMDP sont appris. Les résultats montrent que la méthode permet d'améliorer les performances d'identification de frames pour les exemples de test les plus complexes par rapport à un processus de décision déterministe ad hoc.

Abstract. This paper presents a decision process based on Support Vector Machines to extract the semantic information from the user's input in a spoken dialog system. In our interpretation component, the information is represented by means of trees of semantic frames, as defined in the Berkeley FrameNet paradigm, and the understanding process is performed in two steps. First Dynamic Bayesian Networks are used as generative models to sequentially infer tree fragments from the users' inputs. Then the context-sensitive SVMDP introduced in this paper is applied to detect the relations between the frames hypothesized in the fragments and compose them to obtain the overall semantic representation of the user's request. Experiments are reported on the French MEDIA dialogue corpus. A semi-automatic process provides a reference frame annotation of the speech training data. The parameters of DBNs and SVMDP are learned from these data. The method is shown to outperform an ad-hoc deterministic decision process on the most complex test examples for frame identification.

Mots-clés : système de dialogue oral, compréhension de la parole, composition sémantique, frame sémantique, séparateur à vaste marge.

Keywords: spoken dialogue system, spoken language understanding, semantic composition, semantic frame, support vector machines.

1 Introduction

L'obtention de la représentation sémantique des propositions orales a été l'objet de nombreux travaux au cours des vingt dernières années. L'introduction de composants stochastiques dans les systèmes de dialogue oral améliore leurs performances globales en les rendant plus robustes aux variabilités de la parole (Lefèvre, 2007; He & Young, 2006). A partir des transcriptions fournies par le module de reconnaissance automatique de la parole, le module de compréhension construit une représentation sémantique de haut niveau des propos du locuteur qu'il transmet au gestionnaire de dialogue. Dans un travail précedent (Meurs *et al.*, 2009), nous avons proposé un modèle entièrement stochastique basé sur des réseaux bayésiens dynamiques (DBN) extrayant les concepts de base de la requête d'un utilisateur puis générant les sous-arbres de frames sémantiques à partir de tous les niveaux d'annotation disponibles (mots et concepts). La génération des fragments sémantiques par les DBN étant séquentielle, elle ne tient pas compte des dépendances longue-distances. Cet article décrit l'algorithme de recomposition que nous appliquons aux fragments pour obtenir la représentation sémantique globale de la requête de l'utilisateur. Un apprentissage strictement statistique des fragments en contexte (voir (Zettlemoyer & Collins, 2009) pour une approche utilisant des grammaires) permet à l'algorithme proposé de s'appuyer sur un processus de décision par classification binaire. L'article est organisé comme suit. La section 2 présente la génération des frames sémantiques sur le corpus MEDIA. La section 3 introduit ensuite notre algorithme de composition des fragments sémantiques. Enfin, la section 4 détaille expériences et résultats.

2 Génération de frames sémantiques sur le corpus MEDIA

Issu de la simulation d'un serveur téléphonique d'informations touristiques et de réservation d'hôtels, le corpus MEDIA (Bonneau-Maynard *et al.*, 2005) est composé de 1.257 dialogues en français collectés en utilisant le protocole du *Magicien d'Oz* (un opérateur humain simule les réponses du serveur). Transcription et annotation sémantique manuelles ont été réalisées par deux experts. A partir du dictionnaire sémantique MEDIA (83 concepts) des segments de mots sont associés à une paire *concept-valeur*. Le choix dans nos travaux d'annoter le corpus MEDIA en frames sémantiques (Lowe *et al.*, 1997) est motivé par l'aptitude des frames à représenter les dialogues de négociation. Une frame décrit une situation concrète ou abstraite impliquant ses rôles, les frame-éléments (FE). Nous avons défini, selon le paradigme du projet FrameNet (Fillmore *et al.*, 2003), des frames couvrant le domaine du corpus MEDIA et adaptées à la nature particulière du support textuel. L'ontologie MEDIA est composée de 21 frames et de 86 FE définis par des modèles composés d'unités lexicales et de concepts de base. Les données d'entraînement sont automatiquement annotées par un système à base de règles (voir (Meurs *et al.*, 2008)). Une annotation de référence en frames et FE est ainsi obtenue permettant l'apprentissage des paramètres des modèles stochastiques utilisés pour générer les fragments sémantiques de frames-FE associés à la proposition du locuteur.

Ces modèles à base de DBN (Bilmes & Zweig, 2002) sont exposés et évalués dans (Meurs *et al.*, 2009). L'apprentissage de leurs paramètres nécessite la détermination des fragments sémantiques associés aux concepts (et aux mots) de la requête du locuteur. La représentation en frames étant hiérarchique, des situations de recouvrement peuvent se produire lors de la détermination des frames et FE associés à un concept. Pour résoudre ce problème, un algorithme de projection d'arbre est appliqué sur l'annotation en frames et FE de la phrase complète. Il permet de définir les fragments sémantiques (sous-branches) associés à un concept. Partant d'une feuille de l'arbre, un fragment de frames-FE est obtenue en agrégeant les valeurs de nœuds pères aussi longtemps qu'ils sont associés au même concept (ou à aucun). Par exemple, la séquence de mots *réserver un hôtel à Paris*, illustrée figure 1, entraîne la création des branches projetées de frames et FE `HOTEL-lodging_hotel-LODGING-reserve_theme-RESERVE` et

`location_town-LOCATION-lodging_location-LODGING`. La séquentialité du décodage des fragments entraîne la perte d'une partie des liens entre frames et FE.

3 Composition d'arbres

L'algorithme de projection réalise deux types d'opérations. Les *séparations* rompent des liens entre frames et FE selon les concepts qui leurs sont associés. Les *duplications* des objets sémantiques (frames ou FE) sont nécessaires lorsque ces objets sont présents dans plusieurs sous-branches distinctes. L'algorithme de recomposition est développé pour rassembler les fragments produits par les DBN et rétablir l'arbre sémantique associé à la globalité du message. Il décide des opérations réciproques de celles effectuées lors de la projection, soit des opérations de *liaison* entre frames et FE et des opérations d'*identification* entre frames ou FE.

Les liaisons potentielles inter-fragments s'appuient sur l'ontologie développée pour le domaine du corpus MEDIA : deux objets sémantiques ne peuvent être reliés que s'ils le sont dans l'ontologie. Les *liaisons* consistent donc en l'ajout d'arêtes entre des nœuds de fragments sémantiques distincts pour les rassembler sous un arbre sémantique unique. Les *identifications* potentielles concernent les objets sémantiques semblables présents au sein de plusieurs fragments associés à un même message. L'algorithme de recomposition considère ces objets et décide de la pertinence de leurs présences multiples. Les *identifications* suppriment ainsi les objets sémantiques redondants produits par les DBN. Lorsque deux nœuds de sous-branches sont *identifiés*, un seul est conservé dans l'arbre sémantique global et les nœuds fils du nœud supprimé sont reliés au nœud conservé.

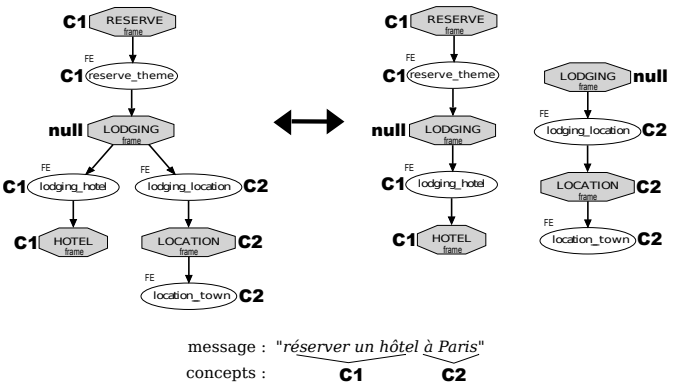

FIG. 1 – Projection et recomposition de l'arbre sémantique associé à la requête *réserver un hôtel à Paris*

L'exemple de la figure 1 illustre ces processus. L'arbre sémantique associé au message *réserver un hôtel à Paris* est reproduit sur la gauche de la figure. Les branches de l'arbre décomposé sont reproduites à droite. Lors de la décomposition pour l'apprentissage des paramètres des modèles DBN, la frame `LODGING` est dupliquée et le FE `reserve_theme` est séparé d'une des deux instances de cette frame. L'algorithme de recomposition doit être à même de recomposer l'arbre complet en disposant du message, des concepts associés et des branches générées par le DBN à partir de ces connaissances.

La pertinence de l'arbre sémantique recomposé est donc directement dépendante de la pertinence des décisions de liaisons et d'identifications. Deux stratégies de décision sont évaluées dans ce travail, présentées ci-après.

3.1 Méthode de connexion forte

La première stratégie évaluée est une heuristique visant à obtenir pour chaque message une représentation sémantique compacte dans le cadre autorisé par l'ontologie. Dans cette méthode de **connexion forte** (CF), toute liaison ou identification, possible selon l'ontologie, est réalisée. Cette approche est a priori efficace pour les messages simples contenant des phrases courtes et peu ambigües. En revanche, elle ne prend pas en compte les mots et les concepts associés au message. Elle n'est donc pas très adaptée aux messages complexes dont la représentation sémantique peut contenir de nombreuses sous-structures non connectées.

3.2 Processus de décision à base de séparateurs à vaste marge (SVMDP)

La seconde méthode de connexion évaluée est basée sur l'apprentissage de classifieurs SVM. Le choix du type de classifieurs linéaires employés est dicté par plusieurs considérations : la quantité de données disponibles, la rapidité de réponse ou encore les performances obtenues sur des données comparables. En raison de leurs propriétés, les classifieurs SVM s'adaptent parfaitement au contexte applicatif de ce travail.

Apprentissage

Les *séparations* et les *duplications* réalisées lors de la projection des arbres sont recensées. A chaque opération est associé l'ensemble des exemples du corpus d'entraînement contenant les objets sémantiques qu'elle fait intervenir. Ces messages sont répartis en deux classes selon qu'ils ont ou non *déclenché* l'opération. On dispose de $\mathcal{T}$, ensemble des exemples d'apprentissage annotés en arbres sémantiques par le système à base de règles évoqué en section 2. Soit $\mathcal{A}$ l'ensemble construit à partir de $\mathcal{T}$ tel que tout élément de $\mathcal{A}$ est composé des mots, des concepts et de l'arbre sémantique associés à un exemple de $\mathcal{T}$. Soit $\mathcal{A}^p$ l'ensemble construit à partir de $\mathcal{A}$ tel que tout élément de $\mathcal{A}^p$ est composé des mots, des concepts, des fragments sémantiques obtenus après projection de l'arbre sémantique et des opérations de projection réalisées lors de la projection de l'arbre sémantique d'un exemple de $\mathcal{A}$. Soient $\mathbb{O}$ l'ensemble des opérations observées dans $\mathcal{A}^p$. Par souci de simplification, une opération de projection et sa réciproque de recomposition (ou *regroupement*) seront également notées $\mathcal{O}_i$, le contexte d'application levant toute ambigüité.

Chaque opération de projection $\mathcal{O}_i \in \mathbb{O}$ met en relation deux objets sémantiques f_{i1} et f_{i2} et on notera $\mathcal{O}_i = f_{i1}\mathcal{R}f_{i2}$. Pour chaque paire $\{f_{i1}, f_{i2}\}$ associée à une opération de $\mathbb{O}$, on construit l'ensemble $\mathcal{A}_i^p$ des exemples de $\mathcal{A}^p$ contenant f_{i1} et f_{i2}. Les exemples de $\mathcal{A}_i^p$ pour lesquels l'opération $\mathcal{O}_i$ s'est appliquée lors de la projection sont dits "positifs" pour $\mathcal{O}_i$. Les exemples $\mathcal{A}_i^p$ contenant f_{i1} et f_{i2} pour lesquels $\mathcal{O}_i$ n'a pas été appliquée sont "négatifs" pour $\mathcal{O}_i$. On dispose pour chaque opération $\mathcal{O}_i$ de la partition $\{\mathcal{A}_i^{p+}, \mathcal{A}_i^{p-}\}$ de $\mathcal{A}_i^p$ où $\mathcal{A}_i^{p+}$ et $\mathcal{A}_i^{p-}$ sont respectivement les sous-ensembles d'exemples positifs et négatifs de $\mathcal{A}_i^p$.

Pour appliquer la méthode de classification SVM, il est nécessaire de plonger les données dans $\mathbb{R}^n$. Un exemple $\mathcal{E}$ est représenté dans $\mathbb{R}^n$ par un point E dont les coordonnées sont les index numériques des mots et trigrammes de mots de l'exemple, de la séquence de concepts associée à l'exemple et des frames et FE présents dans les fragments sémantiques associés à cet exemple. L'introduction des n-grammes de mots dans le point caractérisant un exemple permet de prendre en compte une information séquentielle. Les paramètres d'un classifieur linéaire binaire à base de SVM sont appris sur les points représentant les exemples de chaque ensemble $\mathcal{A}_i^p$. A l'issue de cette procédure, on dispose d'un classifieur $\mathcal{S}_i$ par opération $\mathcal{O}_i$ et on a $|\mathbb{O}| = |\mathbb{S}| = I$, avec $\mathbb{S}$ l'ensemble des classifieurs entrainés.

Application aux exemples de l'ensemble de test

Pour chaque exemple $\mathcal{E}$ de l'ensemble de test, annoté en fragments sémantiques, on construit l'ensemble des opérations pouvant le concerner en fonction des paires d'objets sémantiques contenues dans ses fragments. Soit $\mathbb{O}^{\mathcal{E}}$ cet ensemble, on a :
$\mathbb{O}^{\mathcal{E}} = \{\mathcal{O}_{i\in I} = f_{i1}\mathcal{R}f_{i2}$ tel que f_{i1} et f_{i2} appartiennent aux fragments sémantiques associés à $\mathcal{E}\}$ et $\mathbb{O}^{\mathcal{E}} \subset \mathbb{O}$. Pour toute opération $\mathcal{O}_i \in \mathbb{O}^{\mathcal{E}}$, le point E représentant l'exemple $\mathcal{E}$ est soumis au classifieur S_i. La réponse de S_i quant à la classe de E détermine la pertinence de la réalisation de $\mathcal{O}_i$ sur les objets sémantiques de l'exemple. A l'issue de ce processus, la phase de composition sémantique est achevée par la réalisation sur les objets sémantiques de $\mathcal{E}$ de toutes les opérations jugées pertinentes par les $S_{i\in I}$. Ces deux méthodes sont évaluées sur les fragments sémantiques produits par le DBN.

4 Expériences et résultats

Les expériences sont menées sur l'ensemble de test MEDIA (3005 tours de parole) dans trois conditions différentes, fonctions de la nature des données utilisées. Les données de type MAN rassemblent les tours de parole du locuteur, manuellement transcrits et annotés en concepts. Pour celles de type SLU, les concepts de base sont décodés à partir des transcriptions manuelles des tours de parole locuteur en utilisant le modèle à base de DBN décrit dans (Lefèvre, 2007) (taux d'erreurs concepts : 10,6%). Pour les données de type ASR+SLU, les concepts sont décodés par le modèle de compréhension en utilisant la meilleure hypothèse de séquence de mots générée par un système conforme à (Barrault *et al.*, 2008) (taux d'erreurs : mots 27%, concepts 24,3%).

Quatre niveaux d'évaluation sont considérés. Au niveau **Frames**, les hypothèses de frames sont correctes dès lors que les frames correspondantes sont présentes dans la référence, idem pour le niveau **FE**. Au niveau **FE{Fr}**, seules les hypothèses de FE appartenant à des hypothèses de frames correctes sont examinées et au niveau **Liens{Fr}**, seules les hypothèses de liens reliant des hypothèses de frames et FE correctes sont examinées.

Toutes les expérimentations ont été réalisées en utilisant la librairie `libSVM` (EL-Manzalawy & Honavar, 2005) pour WEKA (Bouckaert *et al.*, 2008). Le tableau 1 regroupe les résultats issus de l'application des méthodes CF et SVMDP sur les fragments sémantiques issus des 3005 tours de parole de test (5,5 frames ou FE par tour en moyenne) ; des 515 tours contenant plus de 6 segments conceptuels (17,5 frames ou FE par tour en moyenne) et des 223 tours contenant plus de 10 segments conceptuels (24,2 frames ou FE par tour en moyenne). Les résultats sont donnés en termes de F-mesure pour les trois types de données (MAN, SLU, ASR+SLU).

Trs de parole	**test complet**		**+ de 6 concepts**		**+ de 10 concepts**	
Données	F-m		F-m		F-m	
MAN	CF	SVMDP	CF	SVMDP	CF	SVMDP
Frames	0,90	0,91	0,85	0,87	0,84	0,87
FEs	0,87	0,88	0,73	0,74	0,72	0,74
FE{F}	0,95	0,94	0,85	0,83	0,84	0,81
Liens{F}	0,88	0,87	0,67	0,64	0,66	0,61
SLU	CF	SVMDP	CF	SVMDP	CF	SVMDP
Frames	0,89	0,90	0,83	0,85	0,82	0,85
FEs	0,82	0,83	0,65	0,67	0,66	0,68
FE{F}	0,91	0,91	0,79	0,78	0,79	0,77
Liens{F}	0,84	0,84	0,61	0,60	0,60	0,58
ASR+SLU	CF	SVMDP	CF	SVMDP	CF	SVMDP
Frames	0,80	0,81	0,77	0,79	0,78	0,80
FEs	0,77	0,77	0,61	0,62	0,62	0,63
FE{F}	0,91	0,90	0,78	0,76	0,77	0,75
Liens{F}	0,84	0,83	0,61	0,59	0,59	0,57

TAB. 1 – F-mesures sur le test MEDIA, les tours à + de 6 concepts et les tours à + de 10 concepts après application des méthodes CF et SVMDP sur les fragments sémantiques générés par le modèle DBN.

La méthode SVMDP utilise 74 classifieurs appris sur le corpus d'entraînement : 18 classifieurs sont dédiés à l'identification de frames (10) ou de FE (8) et 56 classifieurs sont dédiés à la liaison entre frames et FE.

Les résultats confirment l'aptitude de ces algorithmes à composer les fragments d'arbre sémantiques de grande taille pour former une représentation complète consistante du message de l'utilisateur. La structure de la base de connaissance et le contexte relativement fermé des messages de test peuvent expliquer les performances comparables des deux méthodes. En effet, les opérations d'identification et de liaison des objets sémantiques contenus dans les fragments étant presque toujours justifiées, la méthode de connexion forte commet finalement peu d'erreurs mais on notera que l'emploi de la méthode SVMDP permet une sélection plus fine des frames et FE.

5 Conclusion

Cet article présente et évalue un processus de décision à base de SVM pour la composition de fragments sémantiques (sous-arbres de frames). Les fragments sont séquentiellement décodés par un modèle entièrement stochastique à base de réseaux bayésiens dynamiques. La composition de ces fragments est réalisée par un processus de décision SVM dépendant du contexte. Les expériences menées sur le corpus MEDIA attestent la validité de notre approche. Les performances du système proposé confirment son aptitude à produire automatiquement une représentation sémantique consistante de la requête d'un utilisateur.

Références

BARRAULT L., SERVAN C., MATROUF D., LINARÈS G. & DE MORI R. (2008). Frame-based acoustic feature integration for speech understanding. In *IEEE ICASSP*, Las Vegas.

BILMES J. & ZWEIG G. (2002). The graphical models toolkit : An open source software system for speech and time-series processing. In *IEEE ICASSP*, Orlando, Florida.

BONNEAU-MAYNARD H., ROSSET S., AYACHE C., KUHN A. & MOSTEFA D. (2005). Semantic annotation of the french media dialog corpus. In *Eurospeech*, Lisboa, Portugal.

BOUCKAERT R. R., FRANK E., HALL M., KIRKBY R., REUTEMANN P., SEEWALD A. & SCUSE D. (2008). *WEKA Manual for Version 3-6-0*. User manual, The University of Waikato, New Zealand.

EL-MANZALAWY Y. & HONAVAR V. (2005). *WLSVM : Integrating LibSVM into Weka Environment*. Software available at `http://www.cs.iastate.edu/~yasser/wlsvm`.

FILLMORE C. J., JOHNSON C. R. & PETRUCK M. R. (2003). Background to framenet. *International Journal of Lexicography*, **16.3**, 235–250.

HE Y. & YOUNG S. (2006). Spoken language understanding using the hidden vector state model. *Speech Communication*, **48 (3-4)**(3-4), 262–275.

LEFÈVRE F. (2007). Dynamic bayesian networks and discriminative classifiers for multi-stage semantic interpretation. In *IEEE ICASSP*, Hawaii, USA.

LOWE J., BAKER C. & FILLMORE C. (1997). A frame-semantic approach to semantic annotation. In *SIGLEX Workshop on Tagging Text with Lexical Semantics : Why, What, and How ?*, Washington D.C., USA.

MEURS M.-J., DUVERT F., BÉCHET F., LEFÈVRE F. & DE MORI R. (2008). Semantic frame annotation on the french media corpus. In *LREC*, Marrakech, Maroc.

MEURS M.-J., LEFÈVRE F. & DE MORI R. (2009). Learning bayesian networks for semantic frame composition in a spoken dialog system. In *NAACL-HLT*, Boulder, CO, USA.

ZETTLEMOYER L. S. & COLLINS M. (2009). Learning context-dependent mappings from sentences to logical form. In *ACL-IJCNLP*.