



EDF – Direction des Systèmes
d'Information et du Numérique

Segmentation de phases de dialogue dans des retranscriptions de conversations de centres d'appels

Conférence APIA 2023

Auteurs : Guillaume Dubuisson Duplessis, Manon Richard (externe),
Anne-Laure Guénet

DSIN - CSC Datascience & IA
Responsable CSC : Sonia Audheon
Responsable déléguée : Laura Rouhier

Responsable opérationnelle équipe « données non-structurées » :
Anne-Laure Guenet

Accessibilité : limitée aux participants de la conférence APIA 2023

Entité émettrice : CSC datascience & IA

Auteur : Guillaume DUBUISSON DUPLESSIS

Destinataires : participants à la conférence APIA 2023

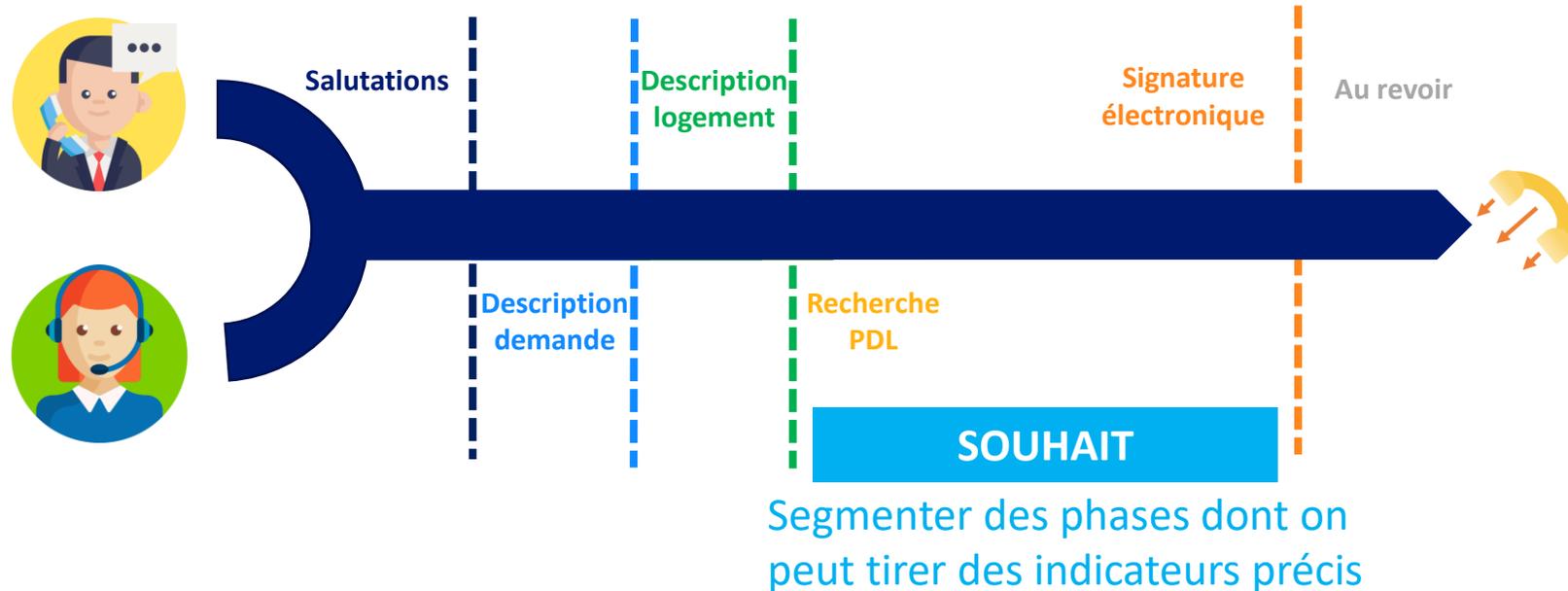
Intérêt métier

OBJECTIF GÉNÉRAL

Segmenter les phases dans les retranscriptions pour effectuer des analyses plus fines

A DISPOSITION

- Un appel retranscrit de bout en bout
 - Des indicateurs temporels
 - Des métadonnées comme la date, des tags, la durée d'appel

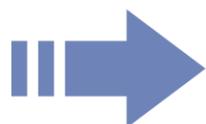


Intérêt du POC technique



Un algorithme par règles a été construit et proposé au métier pour répondre à des questionnements sur la phase de signature

ALGORITHME PAR REGLES CRÉÉ	
Atouts	Limites
<ul style="list-style-type: none">• Possible avec une petite volumétrie• Expertise humaine• Création d'une baseline	<ul style="list-style-type: none">• Subjectivité des choix : hyperparamètres arbitraires• Fragilité : pas de pattern complexe• Difficile à maintenir / faire évoluer l'ensemble des règles



Objectif: passage à une approche « générique » par apprentissage automatique

CHALLENGE TECHNOLOGIQUE

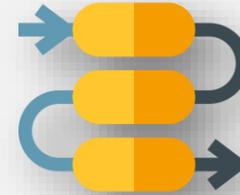
- **Données de natures différentes** : texte, datetime, interlocuteur, durée, etc.
- Prise en compte du **contexte de l'intervention** ou non : blancs autour, texte des interventions autour, etc.
- Longueur des documents

Démarche

1 REFORMULATION DU PROBLEME

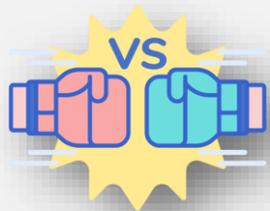


2 MODÉLISATION



Le modèle prédit pour chaque intervention si elle fait partie ou non de la phase de signature

4 ÉVALUATION ET COMPARAISON



Comparaison avec l'algorithme par règle construit avec le métier en 2021

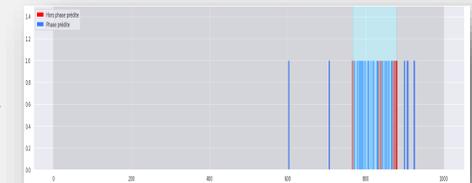
3 RECONSTRUCTION DES PHASES DANS LES APPELS

Le modèle a prédit chaque intervention indépendamment, il faut reconstruire les appels :

INTERVENTIONS



Reconstruction des phases dans les appels



Sommaire

1. Démarche de modélisation

2. Résultats

3. Retour à la segmentation

1

Démarche de modélisation

Formalisation du problème

DONNÉES À DISPOSITION

700 appels anonymisés annotés manuellement par deux personnes



Données désidentifiées

"je suis monsieur [person] je vous appelle pour mon logement situé à [localisation]"

Dubuisson Duplessis, G.; Bartholme, E.; Kerroua, S.; Poulain, M.; Roulier, A.; Guénet, A.-L., **Désidentification de données texte produites dans un cadre de relation client**, 27ème Conférence sur le Traitement Automatique des Langues Naturelles, 2020

ECHANTILLONS TRAIN / DEV / TEST

Découpage à la maille appel et non uip de façon à garder les conversations intactes

TRAIN - 60%

420 appels
257K UIPS

DEV - 20%

140 appels
85K UIPS

TEST - 20%

140 appels
90K UIPS



Echantillon déséquilibré : 15% des UIPS font partie de la phase de signature électronique



REFORMULATION DU PROBLÈME

On ne traite pas une retranscription en entier, mais on découpe en UIP (permet d'avoir + de volumétrie)



Découpage en uips

UIPS



Définition de la tâche

L'UIP appartient-elle à la phase de signature oui ou non?

CHOIX DE MODELISATION

Les conversations sont des documents de taille très grande (~ 960 UIPS par appel – écart type : 410)

→ modélisation de type RNN non testée

Exemple de données

Exemple de retranscription avec les différentes features :



UIP	Caractéristiques de l'UIP			Contexte avant - fenêtre 3	Contexte après -fenêtre 3	Caractéristiques du contexte		Variable à prédire
Retranscription de l'intervention	durée	datetime normalisé (0-1)	interlocuteur	Retranscription des interventions avant l'UIP	Retranscription des interventions après l'UIP	Durée blancs avant	Durée blancs après	L'uiip fait-elle partie de la phase de signature
euh alors du coup euh tout est mis en place hein au niveau des prélèvements vous n'avez rien à nous retourner	4.47	0.6374	conseiller	j_out ai_out besoin_out encore_out de_out madame_out fixe_out ok_in je_in suis_in là_in c_in est_in bon_in oui_out merci_out beaucoup_out	on_out va_out valider_out ok_in ensemble_out le_out contrat_out donc_out pour_out du_out coup_out je_out vais_out vous_out envoyer_out vers_out votre_out boîte_out mail_out est_out ce_out que_out ça_out va_out être_out possible_out pour_out vous_out de_out le_out valider_out en_out ligne_out avec_out moi_out	0.39	1.45	0
on va valider ensemble le contrat donc pour du coup je vais vous envoyer vers votre boîte mail est-ce que ça va être possible pour vous de le valider en ligne avec moi	7.01	0.6375	conseiller	euh_out alors_out du_out coup_out euh_out tout_out est_out mis_out en_out place_out hein_out au_out niveau_out des_out prélèvements_out vous_out n_out avez_out rien_out à_out nous_out retourner_out on_out va_out valider_out ok_in	alors_out oui_in oui_in oui_in je_out vais_out également_out	1.24	2.27	1
ok	0.42	0.6375	client	oui_out merci_out beaucoup_out euh_out alors_out du_out coup_out euh_out tout_out est_out mis_out en_out place_out hein_out au_out niveau_out des_out prélèvements_out vous_out n_out avez_out rien_out à_out nous_out retourner_out on_out va_out valider_out	ensemble_out le_out contrat_out donc_out pour_out du_out coup_out je_out vais_out vous_out envoyer_out vers_out votre_out boîte_out mail_out est_out ce_out que_out ça_out va_out être_out possible_out pour_out vous_out de_out le_out valider_out en_out ligne_out avec_out moi_out alors_out oui_in oui_in oui_in	0.61	2.30	1

Dans les contextes, on a gardé l'information de l'interlocuteur :

in → client
out → conseiller

Modéliser au mieux sans se perdre

Nous avons cherché à améliorer la baseline en testant plusieurs features, vectorisations, modèles en retraçant les itérations :

FEATURES CANDIDATES

MODÈLE UNIQUEMENT UIP

UIP



MODÈLE UIP + INFOS UIP

MODÈLE
UNIQUEMENT
UIP



MODÈLE UIP + INFOS UIP + CONTEXTE

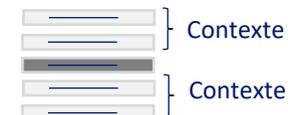
MODÈLE UIP
+ INFOS UIP



DURÉE BLANC Avant / après



CONTEXTE (choix de fenêtre)



VECTORISATIONS CANDIDATES

TFIDF

TFIDF + SVD

50% des uips ont uniquement 1-2 tokens
Le mot le + utilisé apparaît dans 19% des uips : hyperparamètres non standards

GLOVE

Modèle Glove dimension 100
Maille document :
meanpooling / maxpooling

MODÈLES CANDIDATS

Nous avons testé plusieurs types de modèles simples avec scikit-learn :

Régression logistique

SVM linéaire

MLP non linéaire

2

Les enseignements de la phase de modélisation

Tout dépend du contexte

Quel est l'apport du contexte ?

On compare le modèle de la strong baseline sans contexte avec des modèles avec contexte global :

TESTS FEATURES

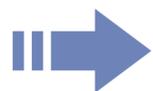
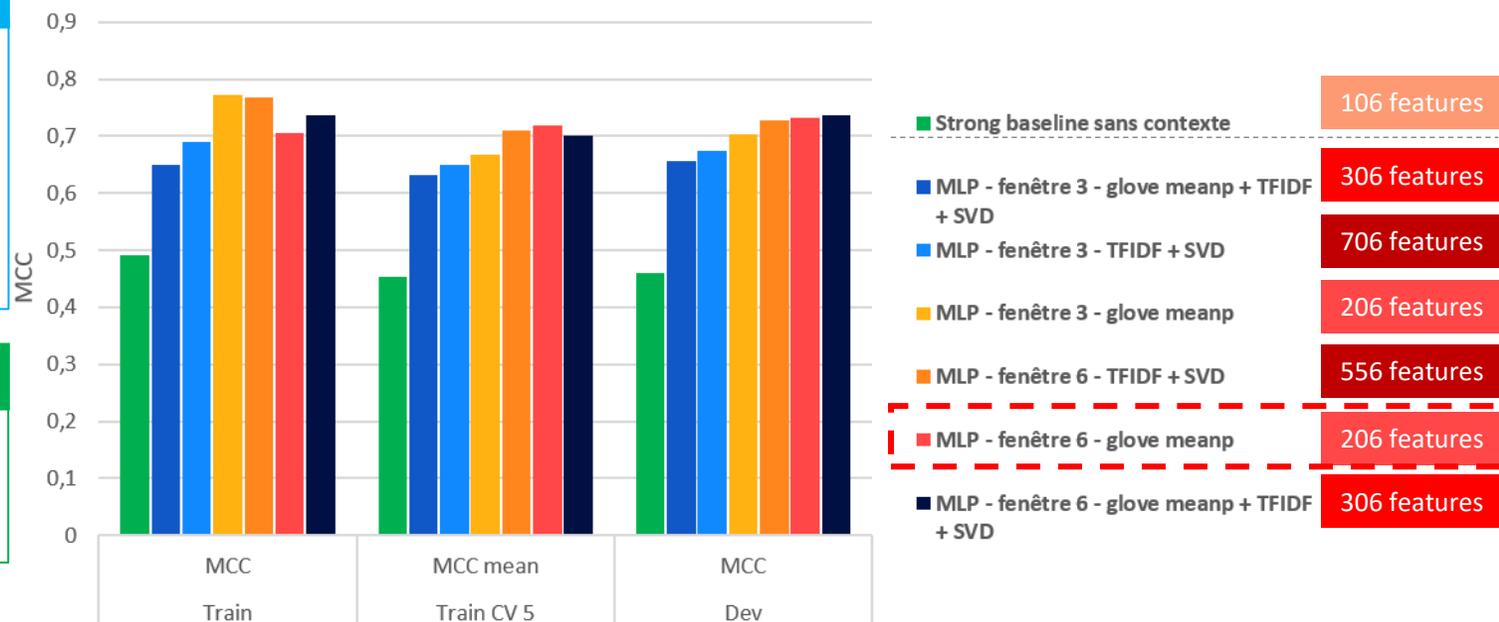
Le rajout du contexte augmente fortement les performances : on passe d'un MCC de 0.45 à une moyenne de 0.65.

La taille de la fenêtre du contexte joue également : on obtient de **meilleures performances avec une fenêtre de 6** qu'une fenêtre de 3.

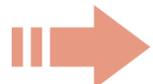
TESTS VECTORISATION

Les performances avec **glove meanpooling** sont équivalents voire **meilleurs qu'avec TFIDF + SVD** pour un nombre de features inférieur.

MCC sur échantillon train, dev, et MCC moyen sur CV



Le contexte apporte beaucoup d'information aux modèles qui performant bien mieux



Sélection du meilleur modèle : fenêtre de 6 – glove – MLP avec un MCC de 0.73

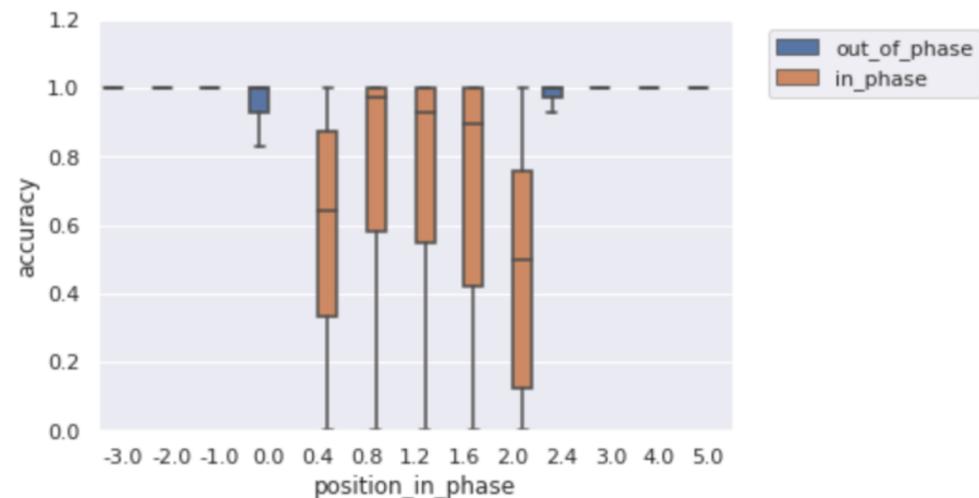
Performances du meilleur modèle



Modèle sélectionné : vectorisation glove avec uip + contexte fenêtre de 6 / modélisation MLP

PERFORMANCE VISUELLE

Proximité de l'uiip avec la phase :



Le modèle performe bien sur :

- les UIPS hors phases
- les UIPS au milieu de la phase

En revanche, il a beaucoup plus de difficultés à capter le début et la fin de la phase de signature.

PERFORMANCES CHIFFRÉES

Accuracy sur l'ensemble des échantillons : ~ 0.94

MCC			
TRAIN	CV	DEV	TEST
0,706	0,719	0,732	0,736

Le modèle performe aussi bien sur l'échantillon test : MCC ~ 0.72 .

PRECISION			RECALL		
TRAIN	DEV	TEST	TRAIN	DEV	TEST
0.94 / 0.90	0.94 / 0.91	0.95 / 0.90	0.99 / 0.61	0.99 / 0.65	0.99 / 0.66

On observe les mêmes performances en termes de précision / recall

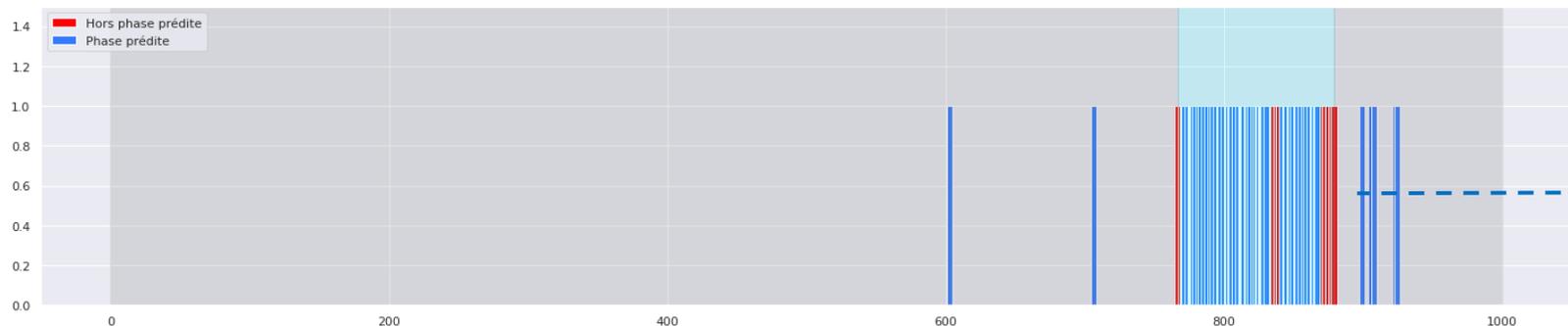
3

Retour à la segmentation

Reconstruction des appels

La prédiction se fait à la maille UIP mais ne prend pas en compte l'ordre des UIPS.

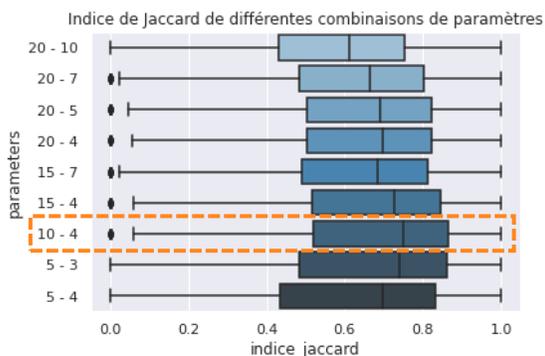
Exemple de retranscription :



Des UIPS isolés sont prédites dans la phase alors que leurs voisins ne le sont pas

→ Nécessité de reconstruire la phase à partir des prédictions des UIPS

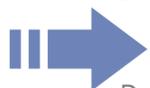
Construction d'un algorithme simple



→ Création de la phase prédite à partir de la densité d'UIPS prédites autour de l'UIP

Paramètres de l'algorithme sélectionnés en fonction de la meilleure performance – indice de Jaccard.

- Taille de fenêtre : **10**
- Minimum d'uips voisins dont la phase a été prédite : **4**

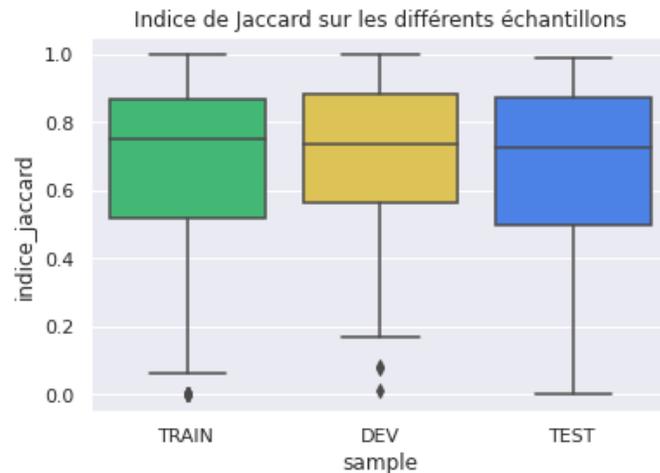


A partir de cet algorithme, on peut calculer les indices de Jaccard

Des performances honorables

Les phases prédites ont été reconstruites, on peut alors regarder la performance par rapport à l'algorithme par règles

INDICE DE JACCARD PAR ECHANTILLON



Sur chaque échantillon, l'indice de Jaccard médian est similaire :

→ autour de **0.73**

$\frac{3}{4}$ des appels ont un indice de Jaccard qui oscillent **entre 0.5 et 0.87**

COMPARAISON RÉSULTATS MODÈLE VS REGLES

On regarde les indices de Jaccard moyens calculés lors de l'étude sur l'algorithme et sur le modèle.

Limites (*) :

- l'indice de Jaccard sur l'algorithme n'a été calculé que sur 70 appels
- L'algorithme de reconstruction des phases du modèle a été construit très naïvement

Indice Jaccard moyen	REGLES	MODÈLE
Sur les phases courtes Moins de 4 min	0.65 (*)	0.63
Sur les phases longues Plus de 4 min	0.84 (*)	0.79



Avec peu de données annotées (700 appels), le modèle donne des performances similaires à celles de l'algorithme

Conclusion

APPRENTISSAGE

- Découpage du problème en 2 tâches :
 - création de score sur chaque uip (scoring)
 - Reconstruction des phases
- Modélisation possible avec peu d'annotations (700 appels)
- Prise en compte d'informations hétérogènes :
 - Données structurées : quantitative (datetime) / qualitative (interlocuteur)
 - Données non structurées : texte (uip et contexte)

COMPARAISON

- Résultats similaires entre l'algorithme et le modèle en terme d'indice de Jaccard sans avoir cherché à optimiser la phase de reconstruction de phase
- Approche par modélisation plus générique, rigoureuse, répliquable et maintenable qu'une conception de règles spécifiques

RÉSULTATS

- Glove a bien fonctionné sur les retranscriptions : la vectorisation donne de bonnes performances en terme de scoring
- Le contexte apporte énormément d'information au modèle, en particulier une fenêtre de 6
- Les UIPS sont moins bien prédites lorsqu'elles sont au bord de la phase

PERSPECTIVES

- Evaluer l'apport de modèles type deep learning (ex : RNN)
- Améliorer les méthodes de reconstruction de phase (kernel smoothing)
- Explorer les approches non supervisées : découper les appels en plusieurs phases sans annotation





Merci

Références

De nombreux icônes sont issus du site <https://www.flaticon.com/>