

Génération automatique de résumés critiques d'articles à but de veille médicale

Implémentation et évaluation d'approches par fonction sous-modulaires
pour le résumé par extraction

Le projet Bibliovid



Bibliovid

**Projet de veille scientifique d'articles en rapport avec la
COVID-19**



www.bibliovid.org



Le résumé automatique

- Une tâche parmi les plus étudiées du traitement automatique du langage naturel
- Une **tâche complexe** :
 - Comment **abstraire** les principaux thèmes abordés au sein d'un document ?
 - Comment tenir compte de l'**ambiguïté** du langage naturel ?
 - Comment sélectionner les phrases les plus **pertinentes** pour un résumé ?
 - Comment produire un résumé **cohérent** ?
 - Quels objectifs pour le résumé ? (critique, rapports des faits, 4e de couverture...)

Le résumé automatique

Deux approches majoritaires

- Résumé par **abstraction** (seq-2-seq, Pegasus)
- Résumé par **extraction** (MMR, LexRank, approches par apprentissage automatique)

Evaluer des approches de fonctions sous-modulaires dans un contexte de résumé par extraction

Approches par fonctions sous-modulaires

Fonctions sous-modulaires

- Fonctions d'ensembles particulières
- Mathématiquement, soit E un ensemble, et $F : 2^E \rightarrow \mathbb{R}$ une fonction ensembliste. F est une fonction sous-modulaire si et seulement si :

$$\forall A, B, A \subseteq B \subseteq E \text{ et } s \notin B, F(A \cup \{s\}) - F(A) \geq F(B \cup \{s\}) - F(B)$$

- Lors de l'ajout d'un élément A , la variation $F(A \cup \{s\}) - F(A)$ est **conditionnée** par le contexte.
- Sous-modularité : dans le cas de la maximisation de F avec contrainte de budget, un algorithme glouton peut trouver une solution de qualité garantie à 0.632 fois l'optimalité.

Fonction sous-modulaire 2

Exemple : installation de capteurs

- Installation d'un nombre fixé k de capteurs afin de couvrir une surface S
- Etant donné C un nombre de capteurs, $F(C)$ est la surface totale couverte par les capteurs de C

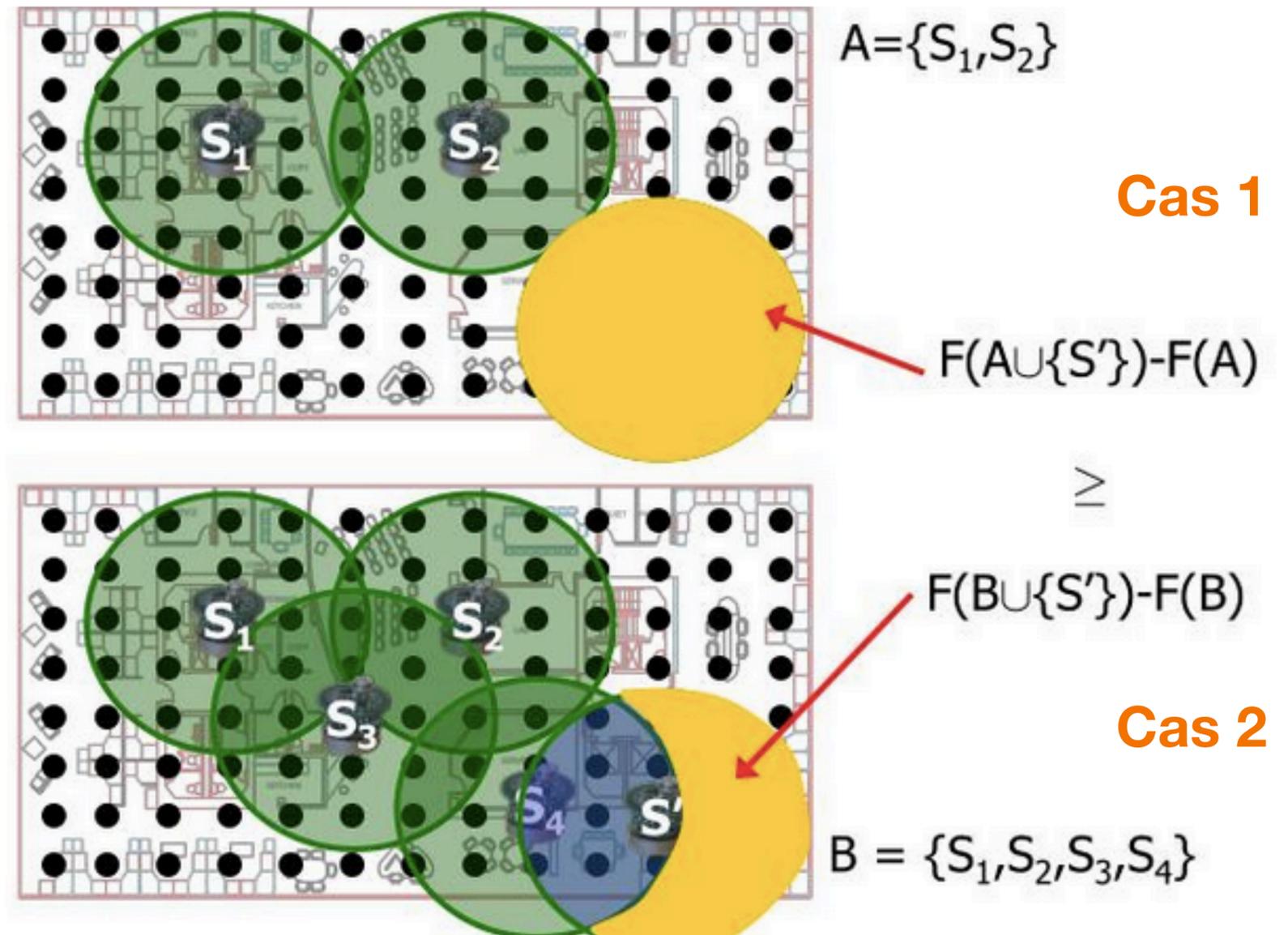


Figure : Illustration du comportement d'une fonction sous-modulaire dans la résolution du problème de dispositions de capteurs

(www.slideserve.com/roman/maximizing-submodular-functions-and-applications-in-machine-learning)

Les fonctions sous-modulaires pour le résumé automatique par extraction (Lin et Bilmes, 2011)

- Construire une fonction permettant d'estimer la couverture et la redondance d'un ensemble de phrases
- En considérant S un ensemble de phrases :

$$F(S) = L(S) + \lambda R(S)$$

- $L(S)$ évalue de la couverture de S par rapport à l'ensemble du document
- $R(S)$ récompense la diversité des phrases sélectionnées dans S

Les fonctions sous-modulaires pour le résumé automatique par extraction (Gillick et Favre, 2009)

- Décomposer l'information au sein des phrases en unités sous-phrastiques : Considérer la redondance de manière indirecte
- Concepts peuvent être beaucoup de choses : entités nommées, n-grammes, relations sémantiques...
- En considérant S un ensemble de phrases :

$$F(S) = \sum_{i \in S} w_i c_i$$

Où c_i est un indicateur de la présence du concept i dans S

w_i est un poids associé à ce concept i

- Solution optimale en programmation nombre entiers, mais un coût plus élevé.

Algorithme glouton pour le résumé automatique

Maximisation de la fonction sous-modulaire (Lin et Bilmes, 2011)

Algorithm 1 Modified greedy algorithm

```
1:  $G \leftarrow \emptyset$ 
2:  $U \leftarrow V$ 
3: while  $U \neq \emptyset$  do
4:    $k \leftarrow \arg \max_{\ell \in U} \frac{f(G \cup \{\ell\}) - f(G)}{(c_\ell)^r}$ 
5:    $G \leftarrow G \cup \{k\}$  if  $\sum_{i \in G} c_i + c_k \leq B$  and
      $f(G \cup \{k\}) - f(G) \geq 0$ 
6:    $U \leftarrow U \setminus \{k\}$ 
7: end while
8:  $v^* \leftarrow \arg \max_{v \in V, c_v \leq B} f(\{v\})$ 
9: return  $G_f = \arg \max_{S \in \{\{v^*\}, G\}} f(S)$ 
```

Figure: Description de l'algorithme glouton pour la maximisation de fonctions sous-modulaires (Lin et Bilmes, 2011)

Recall Oriented Understudy for Gisting Evaluation (Lin, 2004)

- Utilisation de la métrique ROUGE (Lin, 2004) pour évaluer le résumé généré.
- La variante ROUGE-N est basée sur le nombre de n-grammes en commun entre chacune des références et le résumé produit :

$$\text{rappel} = \frac{\sum_{S \in \text{references}} \text{nombre de } n\text{-grammes en commun avec } H}{\sum_{S \in \text{references}} \text{nombre de } n\text{-grammes dans } S}$$

- Utilisation de la variante ROUGE-1 (nombre d'uni-grammes communs)
- Sensibilité de ROUGE sur la variation de la taille des résumés.

Expérimentations

Les baselines testées

- *Fonction de Lin et Bilmes (2011)* : implémentation personnelle de l'algorithme et comparaison avec les résultats présentés dans le papier.
- *Icsisumm (Gillick et Favre, (2009))* : implémentation personnelle de l'algorithme et comparaison avec les résultats présentés dans le papier.
- *HeadExtraction* : extraction des premières phrases des articles, à partir de l'introduction.
- *TailExtraction* : Extraction des dernières phrases des articles, jusqu'à la fin de la conclusion.
- *AbstractExtraction* : les résumés générés correspondent aux abstracts des articles, en respectant la contrainte de longueur.

Les Document Understanding Conferences

- Conférences et compétitions annuelles axées sur les tâches de résumé automatique
- Pour chaque compétition, cinq tâches présentées:
 - Résumé mono-document
 - Résumé multi-documents
 - Résumé multi-documents à partir d'un point-de-vue
 - Résumé multi-documents en réponse à une question
- Développement sur le jeu de données de 2003, test sur le jeu de 2004

Résultats d'évaluation DUC04

- Métrique utilisée : ROUGE-1 (F-score), exprimé en pourcentage.

	Fonction de Lin & Bilmes (2011) (Implémentation personnelle)	Fonction de Lin & Bilmes (2011)	Icsisumm (Implémentation Personnelle)	Icsisumm (Ani Nenkova)	HeadExtraction	TailExtraction
DUC-04	27.23	38.90	27.11	38.41	32	26.62

Figure : Résultats d'évaluation des divers approches sur le jeu de données DUC04.

Le dataset Bibliovid

- Textes complets et exploitables de 252 articles médicaux, issus de Cord-19.
- Les articles ont été traduits du français vers l'anglais de manière automatique.
- Génération des résumés critiques de ces articles basés sur différentes sections: *objectif, résultats, synthèse, niveau et pertinence des preuves, méthodes*.
- Le budget (en nombre de caractères) est donné par la taille du résumé de référence

Résultats Bibliovid

- Métrique utilisée : ROUGE-1 (F-score), exprimé en pourcentage.

	Fonction de Lin & Bilmes (implémentation personnelle)	Icsisumm (Implémentation personnelle)	Abstract	HeadExtraction	TailExtraction
Bibliovid	29.44	26.41	18.78	26.42	26.98

Figure : Résultats d'évaluation des approches sur le jeu de données Bibliovid.

- Remarque : pour chaque article, la longueur du résumé à générer est la longueur du résumé de référence.

Conclusions et perspectives