


Recommandation de documents pour moteur de recherche Coveo

Philippe Blais
Philippe Blouin-Leclerc
William Bourget
Stéphane Caron
Samuel Lévesque


21 décembre 2018

Résumé

Faire ressortir les points saillants de l'article en un seul paragraphe question de titiller le lecteur.

L'ensemble du code et des documents qui ont servi à la résolution de cette problématique et à l'écriture de l'article se trouvent dans le répertoire  du projet.

1 Présentation du problème et état de l'art

Lors de ce projet proposé par Coveo , nous devons utiliser un historique de requêtes faites par des utilisateurs afin de développer un modèle de recommandation de document. Le but du modèle est de proposer une série de 5 documents d'intérêt en fonction de la recherche qui est faite par l'utilisateur et de certaines autres caractéristiques.

Toutefois, dans la plupart des approches les plus populaires, (METTRE DES RÉFÉRENCES DE PAPIERS QUI PARLENT DE TECHNIQUES) le modèle commence par extraire de l'information des documents cible et peut par la suite se définir une mesure de distance entre une requête et chacun des documents pour déterminer quel serait la meilleure correspondance requête-document. Malheureusement, pour ce projet, nous n'avons pas accès au contenu des documents que l'on souhaite prédire, mais bien à un jeu restreint de caractéristiques telles la source du document, son auteur et son titre.

2 Approche proposée

Ici, on présente les grandes lignes conceptuelles qui ont basé notre travail. On souhaite entre autres présenter et référencer les modèles les plus importants qui ont été utilisés dans notre modèle.

On discute aussi des concepts du document recommandation et pourquoi on s'intéresse surtout à certaines variables (pourquoi beaucoup de travail sur les queries, utilisation de techniques du traitement de la langue naturelle, etc.)

3 Méthodologie expérimentale

Ici, on parle de la manière dont on applique les grandes lignes décrites auparavant.

Points à traiter : - Méthodologie de validation (Séparation du jeu de données) - Mesure de score utilisée (Précision sur recommandation de 5 documents car métrique d'évaluation de Coveo) - Utilisation des données sans clicks et à plusieurs clicks - Pipeline et recherche en grille - Paramètres testés

4 Résultats expérimentaux

Présentation des résultats avec tableaux, figures et tests statistiques. On n'analyse rien ici, on ne fait que montrer ce que nous avons obtenue avec l'approche décrite plus haut.

5 Analyse des résultats

Faire du gros blabla sale sur les résultats. Pourquoi notre score n'est pas si élevé que ça, comment on aurait pu améliorer l'efficacité des embeddings. Techniques qui fonctionnent le mieux et avantages/inconvénients des différentes techniques en production (temps d'entraînement, mémoire, etc.)

6 Conclusion

Ouverture philosophique, constats du projet et apprentissages