

GOOGLE MUM, UNE AVANCÉE RÉELLE DANS LE DOMAINE DU SEARCH

Posted on 15 juillet 2021

Hummingbird, Rankbrain, BERT et maintenant MUM : Google, depuis de nombreuses années, multiplie les innovations dans ses algorithmes pour mieux comprendre les intentions de recherche de ses utilisateurs, et leur faire correspondre au mieux la réponse la plus pertinente. MUM (pour *Multitask Unified Model*) est donc la dernière pierre à l'édifice du moteur de recherche que la firme de Mountain View met en place dans ce but. Mais que contient exactement ce nouvel algorithme ?



Après quelques mois un peu morne, Google est reparti dans une frénésie de mise à jour et d'annonces diverses. Ce mois-ci, nous allons évoquer une des plus grosses annonces de l'année qui a été faite officiellement via le blog de Google, par Pandu Nayak, VP search (référence [1]). Il s'agit de la présentation de Google MUM.

Depuis, l'algorithme, ou plutôt la technologie MUM, a été très commenté. Mais on a pu lire tout et son contraire, sur le sujet. Voyons ce que l'on peut en dire à l'heure actuelle.

Présentation simple de MUM

Tout d'abord MUM est un acronyme astucieux, qui signifie *Multitask Unified Model*. Plus tard dans cet article, nous allons parler de la technologie équivalente chez Facebook, qui a clairement moins le génie des noms marketing puisqu'ils ont appelé leur techno "UniT" pour *Unified Transformer*. Pour le définir de manière simple, MUM est une évolution assez impressionnante des modèles de la langue basés sur les transformers (un certain type de chaîne de traitement basée sur des réseaux de neurones). On est donc dans une prolongation de ce que Google a initié il y a quelques temps avec BERT (et que les SEO essaient encore de démystifier et comprendre à l'heure actuelle).

L'objectif de Google avec MUM est encore et toujours le même : mieux comprendre l'internaute qui pose ses questions au moteur, pour mieux y répondre. Là encore on est dans une continuité puisque dès 2012, Amit Singhal annonçait qu'il rêvait d'un futur où le moteur comprendrait le monde et saurait déterminer l'intention des utilisateurs (voir la référence [2]). Les exemples proposés par Pandu Nayak et les divers googlers impliqués sont assez impressionnants, avec des questions très contextuelles, ou encore des questions impliquant texte et image.

Avant de rentrer dans la technique, voyons donc d'abord les principaux points différenciants de MUM par rapport à ce qui se faisait jusqu'ici.

1. **MUM est nativement multilingue.**

Le premier aspect de MUM est que le modèle a été entraîné pour plusieurs langues (75 au total selon la communication officielle). Très concrètement, cela permet bien entendu de réaliser des tâches de traduction, mais cela permet également de répondre à des requêtes qui sont dans une langue en fournissant une réponse qui est une traduction d'une autre langue. Pour cela, le modèle utilise le fait que toutes les langues "existent" au sein du modèle simultanément, et qu'il y a donc une correspondance algébrique dans le modèle entre les vecteurs pour les mots et phrases de langues différentes mais de signification identique. Au final quand vous poserez une question dans une langue, elle sera posée dans toutes les langues et les réponses prises dans toutes les langues, et renvoyées dans la langue qui vous correspond le mieux.

2. **MUM est multimodal.**

C'est pour moi la vraie nouveauté de MUM, et je la trouve fascinante, d'autant plus que j'ai personnellement travaillé à ce que l'on appelle plutôt la cross-modalité que la multimodalité. Le modèle possède des encodages vectoriels pour des informations de plusieurs types. À l'heure actuelle, Google communique sur le fait que le modèle embarque textes et images, mais que dans le futur, l'audio et la vidéo seront aussi pris en compte. L'intérêt d'un modèle multimodal est assez énorme. On peut typiquement poser des requêtes textuelles qui vont renvoyer des images qui correspondent réellement au besoin informationnel de l'utilisateur ("je veux voir un chat noir dormant sur une couverture rouge et blanche, le tout sur un canapé jaune"). Ce n'est pas une nouveauté, dans un article précédent de Réacteur, j'avais évoqué notre travail sur ce sujet. Mais MUM a la capacité de faire cela à très grande échelle, pas au niveau d'un prototype "académique". Par ailleurs, on peut avoir des requêtes images, et aussi des requêtes mixtes ("j'ai cet outil [on colle l'image de l'outil], comment fais-je pour couper ma haie avec ?"). Très concrètement, cela ouvre le champ à des cas d'utilisation nouveaux puisque nous nous promenons tous avec un appareil qui permet de capter des images de notre environnement immédiat. Pour contextualiser nos recherches, ce sera très efficace (imaginez prendre une photo de votre télévision pour demander la taille ou l'âge de l'acteur que vous voyez à l'écran par exemple, ou encore prendre en photo la devanture d'un restaurant pour avoir les avis associés plus rapidement).

3. **MUM est conçu pour plusieurs tâches nativement.**

C'était déjà le cas de BERT qui était entraîné pour faire au moins 2 tâches efficacement (modèle masqué et la compatibilité de phrases qui se succèdent), mais avec MUM, Google passe la seconde avec plus de tâches effectuées par le modèle directement. Dans l'article scientifique qui sert de référence sur le sujet MUM, on voit 5 tâches pour lesquelles le modèle a des performances au-delà de tous ses compétiteurs : la capacité à répondre à des questions, le résumé automatique, l'acceptabilité linguistique, le calcul de similarité et l'analyse de sentiments.

En tant que SEO, vous pouvez imaginer très facilement ce qu'un moteur va pouvoir faire avec une telle technologie...

Maintenant qu'on a vu ce qu'il faut retenir, voyons voir plus en détails ce qui se cachent derrière la technologie MUM.

Sous le capot de MUM...

Concernant MUM, il y a quelque chose d'assez étonnant qui se passe dans la communauté SEO. On trouve des articles qui spéculent sur les algorithmes qui seraient au cœur de cette techno, en se basant sur le nom lui-même. C'est assez étonnant, vu que la communication de Google est pour le coup limpide. Tout le monde s'est focalisé sur une phrase qui dit que MUM est 1 000 fois plus puissant que BERT. Cette information est totalement inutile, mais la phrase qui la

contient est très exactement celle-ci : "MUM uses the T5 text-to-text framework and is 1,000 times more powerful than BERT."

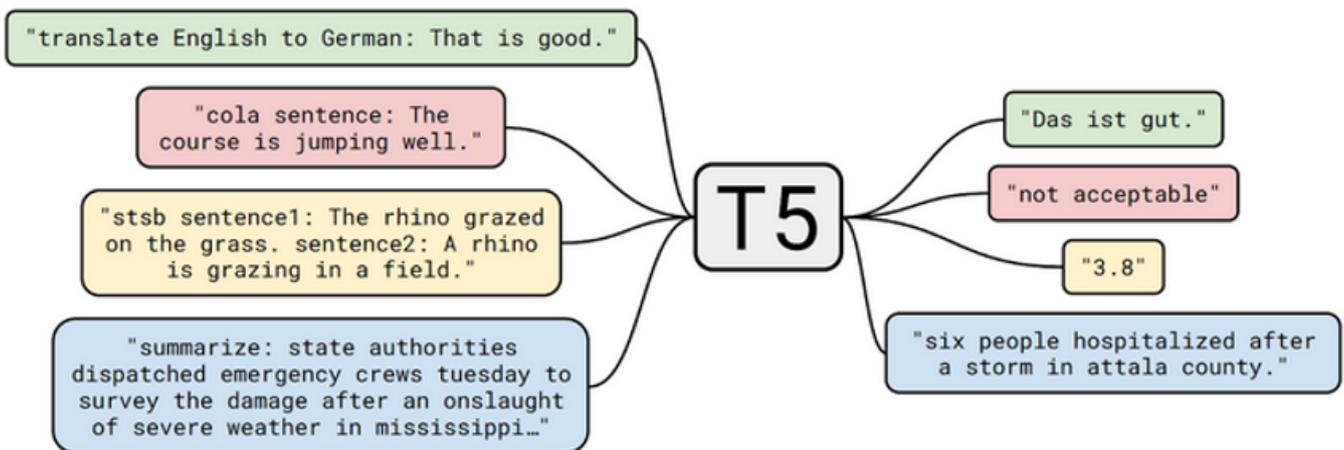
C'est en fait le début de la phrase qui est important : la technologie MUM est basée sur T5, au moins pour la partie texte. Alors la vraie question est donc de savoir ce qu'est T5 !

T5 est le petit nom du *Text-To-Text Transfer Transformer Model*, vous aurez remarqué qu'il y a 5 T dans le nom ;) T5 c'est le produit de l'utilisation de méthodes d'apprentissage par transfert (*transfer learning* en anglais) pour créer des nouveaux modèles de la langue plus performants que les précédents (référence [3]).

L'apprentissage par transfert, c'est l'idée de transférer des connaissances acquises sur une tâche spécifique vers d'autres tâches. L'idée est très intuitive, il s'agit de capitaliser sur ce que l'on sait faire dans un contexte et qui se retrouve en partie dans un autre contexte. Pour l'illustrer simplement, imaginez que vous ayez appris à planter des clous avec un marteau et qu'on vous demande demain d'enfoncer des piquets avec une masse. Ce n'est pas tout à fait la même tâche, mais une partie des gestes que vous aurez appris seront utiles. Vous pouvez transférer une partie de votre expertise à cette nouvelle tâche, et juste affiner votre geste pour prendre en compte le fait qu'une masse c'est plus lourd, que le manche est plus long, etc.

L'autre idée forte de T5 est de tout voir comme un problème text-to-text, c'est-à-dire que l'entrée est un texte, et la sortie aussi. Même dans un problème de calcul de similarité, la sortie qui est en fait une distance, sera vue comme un texte uniquement. C'est une approche qui assez étonnement donne de très bons résultats.

Dans la figure ci-dessous, tirée de l'article [4], on voit l'illustration de 4 tâches sur lesquelles T5 est très bon.



T5 sait tout faire !

Tout d'abord, on voit la traduction automatique. Je ne vais pas trop en dire sur le sujet puisque tous les lecteurs de cet article savent très bien de quoi il s'agit.

La deuxième tâche est l'acceptabilité linguistique (le terme CoLA pour *Corpus of Linguistic Acceptability* est le nom du dataset de test, voir la référence [7]). Il s'agit de dire si la phrase donnée en entrée est acceptable ou non d'un point de

vue d'écriture (est-elle grammaticalement correcte ?).

La troisième tâche est la similarité. On est en terrain connu pour les SEOs, il s'agit de la similarité sémantique : est-ce que les deux phrases ont le même sens ?

Enfin, il est illustré une quatrième tâche, la création de résumé automatiquement. Bien entendu T5 est capable de résoudre bien d'autres problèmes (je vous renvoie aux références [4] et [5] si vous voulez tout savoir sur le sujet).

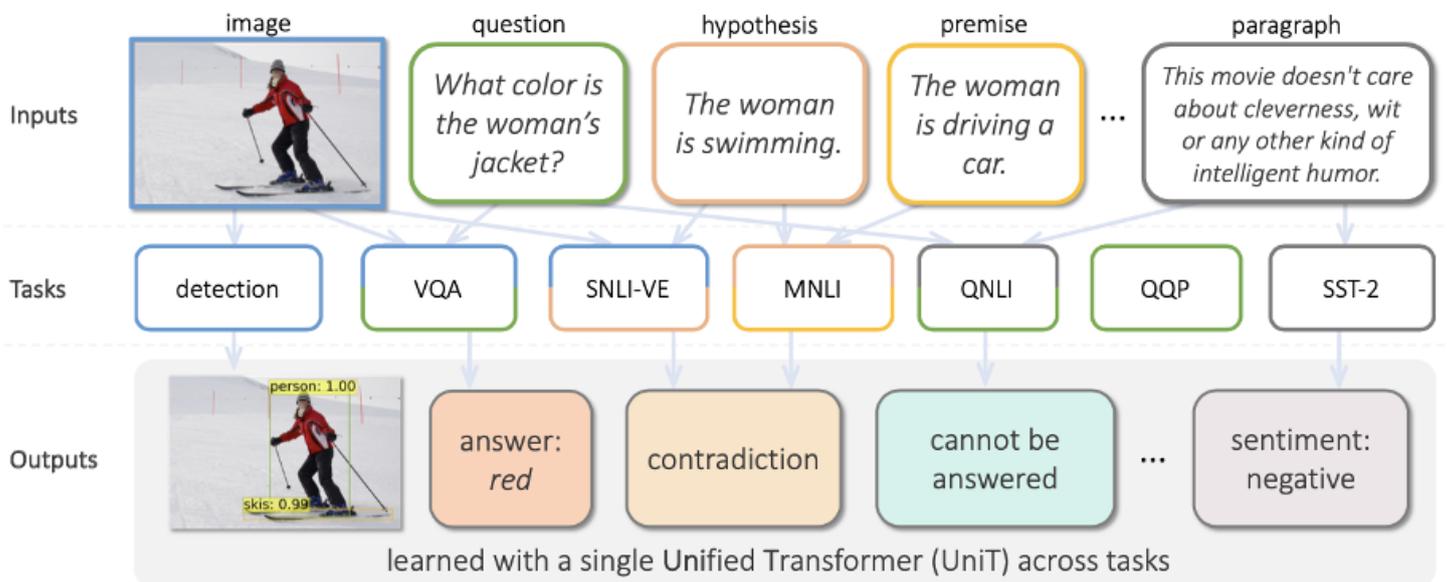
A noter que si vous êtes très motivé, tout est disponible depuis le code source jusqu'au dataset qui a servi à pré-entraîner le modèle. Ce dataset s'appelle C4 et est une version colossale et nettoyée de *Common Crawl* (*Colossal, Clean, Common Crawl* -> 4 C donc C4...). L'avantage de ce dataset c'est qu'il est plus conséquent que ceux utilisés précédemment, et surtout dans un langage réellement naturel (comparé par exemple au dataset wikipedia).

Google est-il devant les autres ?

Avec BERT, tout le monde a cru que Google était largement devant tout le monde, il s'est avéré que la réalité est bien plus complexe, et que si BERT a été industrialisé aux petits oignons comme toujours lorsqu'il s'agit d'ingénierie et de Google, d'un point de vue théorique et même qualité brute des résultats, plusieurs autres opérateurs avaient aussi des modèles équivalents voire supérieurs pour certaines tâches.

Pour MUM dans sa globalité, c'est difficile à dire, mais si on regarde du côté de FAIR (*Facebook AI Research*), on trouve un modèle multimodal sous le nom de "MMF" (*MultiModal Framework*) dès juin 2020, et plus récemment un modèle multimodal et multi tâches appelé "UniT" (voir la référence [8]).

La figure ci-dessous tirée de la référence [8] est édifiante : on voit bien qu'on est sur un modèle qui fait des choses très similaires à ce que propose T5/MUM.



UniT chez Facebook

Compréhension de l'image, requête portant sur le contenu visuel, analyse de sentiments, réponse à une question, tout y est.

Sur la partie "image", UniT est même très impressionnant, comme le montre les exemples ci-dessous, toujours tirés de la référence [8].

visual question answering (VQA_{v2})

question: How are the zebras related?
answer: mother and child



question: Which food contains the most potassium?
answer: banana



Réponse à des questions relatives à des images avec UniT

Le plus intéressant est que la partie texte de UniT est faite avec BERT, qui est donc plus ancien que T5, mais que par contre l'approche utilise un modèle générique avec de la spécialisation qui ne fait pas perdre ce qui est appris par la partie générique. On est donc dans une forme de *transfer learning* qui ne dit pas son nom.

En tous cas, on voit que les deux géants sont proches l'un de l'autre encore une fois, et on peut parier que les plus discrets (Amazon, Apple et OpenAI) ont sans doute aussi quelques algorithmes en réserve également de leur côté.

Et le SEO dans tout ça ?

Lorsque (et si) MUM sera généralisé au niveau du moteur, il y aura un impact sur le SEO, du même type que l'impact de BERT. BERT est un bel exemple de pragmatisme pour Google : comme les SEOs sont toujours à la lutte, ce qu'il faut c'est faire en sorte d'avoir des algorithmes qui font coïncider l'optimisation SEO et l'amélioration du service à l'utilisateur.

MUM pousse le curseur encore plus loin, avec une compréhension plus fine encore du texte, un spectre bien plus large pour les réponses avec la capacité à répondre avec des contenus de toutes les langues, et des images.

Pour le SEO, ça veut dire qu'au-delà de la mise à l'équerre des sites, tout le reste passe par le service à l'utilisateur, et

donc par l'analyse de ses attentes et ses comportements. L'approche qu'on qualifie de data SEO, avec la sémantique en tête, va donc encore gagner en popularité, et les stratégies basées sur les contenus vont probablement devenir de plus en plus efficaces.

Références

- [1] <https://blog.google/products/search/introducing-mum/>
- [2] <https://blog.google/products/search/building-search-engine-of-future-one/>
- [3] <https://ai.googleblog.com/2020/02/exploring-transfer-learning-with-t5.html>
- [4] Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, Peter J. Liu.
<https://arxiv.org/abs/1910.10683>
- [5] <https://github.com/google-research/text-to-text-transfer-transformer>
- [6] <https://www.tensorflow.org/datasets/catalog/c4>
- [7] <https://paperswithcode.com/dataset/cola>
- [8] UniT: Multimodal Multitask Learning with a Unified Transformer. Ronghang Hu, Amanpreet Singh.
<https://arxiv.org/abs/2102.10772>



Sylvain Peyronnet, concepteur de l'outil d'analyse de backlinks [Babbar](#).