

Des ordinateurs apprennent le maltais

Des chercheurs ont entraîné un modèle linguistique sur des données textuelles maltaises et lui ont appris à identifier les sentiments et noms de balises dans un texte.



© graphicwithart/Shutterstock.com

Les technologies linguistiques fondées sur l'IA ouvrent de nouvelles voies pour la communication numérique dans toutes les langues européennes. Cependant, les outils et les ressources en matière de technologies linguistiques font défaut pour le maltais. Le projet LT-BRIDGE, financé par l'UE, s'efforce de combler cette lacune depuis son lancement en 2021.

Dans le traitement du langage naturel, des modèles linguistiques sont entraînés pour associer des mots à d'autres dans un contexte particulier à l'aide d'approches de réseaux neuronaux. Des chercheurs de l'université de Malte (UM), coordinatrice du projet LT-BRIDGE, ont entraîné un tel modèle, le modèle BERTu, sur des données textuelles maltaises.


Remplir les blancs

Mais qu'est-ce qu'un modèle linguistique? Kurt Micallef, doctorant à l'UM, le décrit dans un récent [article](#) publié sur le site web «Times of Malta»: «Les modèles linguistiques sont une compréhension abstraite d'une langue. Vous pouvez considérer cela comme une "intuition" de ce qu'est une langue. Par exemple, si vous deviez remplir le blanc de la phrase "Jien _____ il-gazzetta" (je _____ le journal), vous pourriez écrire "qrajt" (lire) ou "xtrajt" (acheter), mais il est moins probable que vous suggériez "kilt" (manger) ou "karozza" (voiture).»

L'utilisation de la modélisation du langage masqué est un moyen d'entraîner de tels modèles linguistiques. Les mots d'un texte sont aléatoirement masqués, ou camouflés, et le modèle doit prédire le mot manquant. «Ainsi, dans l'exemple ci-

dessus, le modèle devrait idéalement prédire “qrajt”», explique Kurt Micallef. Ce processus est reproduit pour de nombreuses phrases afin que le modèle linguistique puisse apprendre le maltais. Le réseau neuronal est mis à jour à chaque phrase à l’aide d’algorithmes d’apprentissage automatique, et une probabilité est attribuée aux mots qui peuvent être ajoutés dans la phrase.

Autres tâches

Deux autres tâches sur lesquelles BERTu a été entraîné sont l’analyse des sentiments et la reconnaissance d’entités nommées. «L’analyse des sentiments est le processus d’identification du sentiment d’un texte donné», précise le chercheur dans [un autre article publié sur «Times of Malta»](#) . «La forme la plus simple consiste à classer un texte selon qu’il véhicule un sentiment positif ou négatif par rapport à un sujet ou un concept. Par exemple, au vu des annonces budgétaires de Malte, ce commentaire est-il favorable ou défavorable aux annonces faites? Ce type de tâche est appelé un problème de classification, car pour le texte entrant que nous recevons, nous produisons une balise de classification (positive ou négative dans cet exemple).»

Kurt Micallef décrit ensuite la deuxième tâche: «La reconnaissance d’entités nommées est une tâche de balisage, où nous produisons une balise pour chaque mot du texte entrant. Selon le texte d’entrée donné, la tâche consiste à classer les balises qui font référence à des entités nommées et le type d’entité qu’elles représentent. Comparée à l’analyse des sentiments, cette tâche est de très bas niveau et est généralement utilisée pour compléter d’autres systèmes linguistiques. Par exemple, nous pourrions utiliser les données de classification pour identifier les noms de personnes et les rendre anonymes, afin de respecter les lois sur la protection des données.»

L’équipe de recherche a affiné le modèle pré-entraîné de BERTu sur ces tâches en rajoutant une autre couche sur le modèle pour chaque tâche, puis en exécutant des algorithmes d’apprentissage automatique sur l’ensemble des données pour apprendre les paramètres de la couche supplémentaire. BERTu s’est avéré plus performant que d’autres modèles linguistiques, parfois de plus de 20 %. Il permet désormais d’explorer des tâches plus complexes de compréhension du langage en maltais. Le projet LT-BRIDGE («Bridging the technology gap: Integrating Malta into European Research and Innovation efforts for AI-based language technologies») s’achève en décembre 2023.

Pour plus d’informations, veuillez consulter:

[site web du projet LT-BRIDGE](#) 

Mots-clés

Projets connexes



LT-BRIDGE

“Bridging the technology gap: Integrating Malta into European Research and Innovation efforts for AI-based language technologies”

31 Mars 2025

PROJET

Articles connexes



PROGRÈS SCIENTIFIQUES

Ouvrir la voie à une IA fiable



2 Novembre 2022

ACTUALITÉS



Une application d'IA personnalisée pour l'enseignement des langues étrangères par la conversation



20 Mars 2020

RÉSULTATS EN BREF



Comment les machines interprètent le langage humain



18 Mars 2022



Une plateforme d'IA vocale et conversationnelle innovante dope la productivité des entreprises



18 Mars 2019

Dernière mise à jour: 2 Decembre 2022

Permalink: <https://cordis.europa.eu/article/id/442665-computers-learn-maltese/fr>

European Union, 2025