COMMUNIQUE DE PRESSE

Martigny, le 27 juillet 2023

POUR DIFFUSION IMMEDIATE

Vers une intelligence artificielle moins gourmande en énergie

Alors que les technologies d'intelligence artificielle atteignent chaque jour de nouveaux sommets, leur coût énergétique augmente également. A une époque où l'énergie est de plus en plus précieuse, les chercheurs de l'Idiap proposent une nouvelle approche.

Les technologies d'intelligence artificielle comprennent de mieux en mieux le sens du langage humain. Ce progrès représente un élément crucial pour les applications du monde réel utilisant l'analyse de texte et les outils de reconnaissance vocale. Aujourd'hui, les meilleures technologies sont basées sur des modèles appelés "transformers", qui sont très exigeants en termes de ressources. Par conséquent, le nombre d'opérations mathématiques requises pour traiter l'information augmente très rapidement et le temps de calcul nécessaire à l'analyse de textes et de discours plus longs devient vite excessif, même avec plus de puissance de calcul. Conscients de cette écueil, les chercheurs d'Idiap ont élaboré une stratégie pour réduire les ressources informatiques et énergétiques nécessaires au fonctionnement de ces technologies.

Un nouveau modèle

« Lorsqu'ils travaillent avec des algorithmes gourmands en ressources, les chercheurs doivent souvent ajuster leurs données pour obtenir des résultats dans un délai raisonnable. Réduire les coûts de calcul est crucial pour la recherche et ses applications », explique Florian Mai, premier auteur de l'article et assistant de recherche dans le groupe Natural language understanding. Pour réduire ces coûts de calcul, les chercheurs ont décidé de revisiter un modèle vieux de plusieurs décennies, appelé perceptrons multicouches. Un modèle généralement considéré comme inadapté au traitement du langage en raison de son incapacité à gérer des entrées de longueurs variables. Cependant, les chercheurs de l'Idiap ont découvert qu'en passant d'un modèle de traitement statique à un modèle dynamique, les données liées au langage peuvent être traitées efficacement. Ils ont appelé ce modèle HyperMixer.

Une IA plus efficiente

En plus de leurs améliorations du modèle, les chercheurs ont également pu démontrer empiriquement que ce dernier est plus performant ou équivalent aux alternatives traditionnelles. Par rapport aux meilleurs modèles actuels, HyperMixer atteint des résultats comparables à des coûts informatiques nettement inférieurs en termes de temps de traitement, de données d'entraînement et d'ajustement des paramètres.

Au-delà de cette avancée scientifique, HyperMixer fait un pas important dans la direction de la diminution de l'impact environnemental des technologies d'IA en démontrant des performances similaires mais avec une consommation d'énergie bien plus faible. À l'heure où les prix de l'énergie grimpent en flèche et où les ressources se raréfient, la recherche doit jouer son rôle. « Le slogan de l'Idiap "L'intelligence artificielle au service de la société" doit aussi se refléter dans les algorithmes », conclut James Henderson, responsable du groupe Natural language understanding.

Références scientifiques

- “HyperMixer: An MLP-based Low Cost Alternative to Transformers” Florian Mai, Arnaud Pannatier, Fabio Fehr, Haolin Chen, François Marelli, François Fleuret, James Henderson: https://publications.idiap.ch/publications/show/5073

- “HyperConformer: Multi-head HyperMixer for Efficient Speech Recognition” Florian Mai, Juan, Zuluaga-Gomez, Titouan Parcollet, Petr Motlicek: https://publications.idiap.ch/publications/show/5054

- Code informatique des deux modèles : https://github.com/idiap/hypermixing

L’**Institut de recherche Idiap** est un des spécialistes mondiaux de l’intelligence artificielle depuis plus de 30 ans. Reconnaissance vocale et visuelle, interactions homme-machine, robotique, ou encore analyse du langage sont quelques-uns des champs de compétence de l’Institut. Basé à Martigny en Valais, l’institut est impliqué dans des projets locaux, nationaux et internationaux. La Fondation à but non lucratif Idiap a été créée en 1991 par la Ville de Martigny, l’Etat du Valais, l’Ecole polytechnique fédérale de Lausanne, l’Université de Genève et Swisscom.

Contact

- Arnaud Pannatier, co-auteur de la publication, +41 77 439 30 16, arnaud.pannatier@idiap.ch
- Nicolas Filippov, responsable communication, +41 79 139 92 65, nicolas.filippov@idiap.ch