Définition automatique du niveau de la difficulté textuelle de documents

Molka Tounsi Dhouib*, Ekaterina Kostrykina**
Catherine Faron*,**

**Université Côte d'Azur, Polytech Nice Sophia, France
nom.prenom@univ-cotedazur

*Université Côte d'Azur, CNRS, Inria, I3S, Sophia-Antipolis, France
nom@i3s.unice.fr

Résumé. La lisibilité ou la difficulté textuelle représentent une information importante pour de nombreuses applications éducatives telles que les systèmes de recommandation. Aujourd'hui, les meilleurs résultats pour cette tâche sont obtenus en utilisant des techniques de traitement automatique de la langue (TAL) et d'apprentissage profond. Dans ce travail, nous proposons une approche multilingue qui se base sur la combinaison des modèles de réseaux de neuronaux avec des caractéristiques linguistiques extraites du texte afin d'améliorer la qualité de l'évaluation. Nous avons testé notre approche sur deux benchmarks, et les résultats montrent que cette combinaison améliore les performances du système.

1 Introduction

De nos jours, de plus en plus de ressources électroniques pour l'auto-éducation sont disponibles. Cependant, la manière dont l'apprenant absorbe le matériel dépend en grande partie de son expérience dans le domaine d'étude et de la lisibilité d'une ressource électronique particulière. La lisibilité textuelle est une détermination de la facilité ou de la difficulté à comprendre le texte lu. La lisibilité est un aspect important de l'apprentissage, car il est nécessaire d'estimer si le texte transmet l'information d'une manière appropriée pour les apprenants. En outre, la rapidité avec laquelle une personne comprend un texte écrit détermine sa capacité à acquérir de nouvelles connaissances et à rester motivé pour cela. C'est pourquoi l'automatisation de la tâche d'évaluation de la lisibilité d'un texte est de plus en plus populaire aujourd'hui. Cette tâche peut être considérée comme un problème de classification de textes selon leur degré de complexité (Mirończuk et Protasiewicz, 2018). L'objectif du travail présenté dans cet article est d'établir et d'évaluer une approche multilingue permettant de définir automatiquement la difficulté textuelle des documents. Les questions de recherche que nous aborderons sont les suivantes : (i) Quel modèle pouvons-nous utiliser pour évaluer la difficulté textuelle ? (ii) Quelles caractéristiques linguistiques pouvons nous utiliser pour cette tâche ? (iii) Est ce que ces caractéristiques peuvent être facilement adaptées à d'autres langues ? Le reste de cet article est structuré comme suit : La section 2 donne un aperçu de l'état de l'art sur l'évaluation