Apprentissage multimodal basé sur des modèles d'attention pour la classification de documents dans un contexte déséquilibré

Ibrahim Souleiman Mahamoud*, Joris voerman*, Mickaël Coustaty**
Aurélie Joseph*, Vincent Poulain d'Andecy*, Jean-Marc Ogier**

*1 Rue Fleming, 17000 La Rochelle, France
Email:
{ibrahim.souleimanmahamoud,aurelie.joseph,vincent.poulaindandecy}@getyooz.com
**La Rochelle Université, L3i
Avenue Michel Crépeau, 17042 La Rochelle, France
Email: {joris.voerman,mickael.coustaty,jean-marc.ogier}@univ-lr.fr

Résumé. Les documents administratifs ont la particularité d'être identifiables par leur contenu textuel (contenu sémantique) ou par leur mise en page (contenu visuel) et pourtant la classification de ces documents ne se fait généralement qu'à partir d'une de ces informations. Chacune d'entre elles constitue pourtant une part essentielle du document qui peut rendre impossible la distinction entre certaines classes. Les méthodes multimodales de l'état de l'art nécessitent une large base étiquetée pour l'ensemble des classes alors que dans la vie réelle les données sont généralement déséquilibrées. Nous proposons ici un modèle adapté à cette contrainte composé d'un RNN texte et d'un CNN visuel. Leur combinaison permet d'obtenir une description multimodale. Un modèle d'attention est également proposé pour chaque modalité afin de classifier plus efficacement une large variété de documents administratifs. Cette combinaison offre un gain de performance de 1% sur notre base de données privée et 3% sur la base de données publique RVL-CDIP.

1 Introduction

Les entreprises ont besoin de gérer chaque jour une grande quantité de documents. Ces documents représentent le cycle de vie de l'entreprise et sont très variés en termes de classes et d'origine. Ils sont généralement liés à la partie administrative et comptable de l'entreprise (factures, lettres, reçus, ...) ou directement associés au coeur de ses activés. De nombreuses entreprises font appel à des systèmes de "Digital Mailroom" (Schuster et al. (2013)) pour automatiser la gestion des documents. L'entrée de ces systèmes se modélise sous la forme d'un flux de documents définissant plusieurs contraintes : un fort déséquilibre de représentation entre les classes au sein de l'ensemble d'entrainement, un temps de traitements qui doit être très faible pour traiter de grands volumes de documents, ou encore la nécessité de limiter les erreurs qui