

RECONNAISSANCE D'ENTITES NOMMEES

POUR DES BESOINS EN ANALYSE CRIMINELLE

Lenfant Frédéric, frederic.lenfant@intelfe.com, IntelFe, Toulouse

Arrotis Grégory, gregory.arrotis@kpconsulting.fr, KPC, Bordeaux

Gabriel Julien, gabriel.julien@kpconsulting.fr, KPC, Bordeaux

Thématique : Intelligence Artificielle

Résumé :

IntelFe, société représentée par Frédéric Lenfant et spécialisée dans l'analyse criminelle, l'investigation numérique et la formation aux méthodes d'investigation, effectue entre autres la mise en évidence méthodique de relations entre éléments par des techniques d'analyse criminelle. Cette mise en relations est réalisée à partir de différentes entités : des noms de personnes physiques, de personnes morales, des localisations.

Ainsi, KPC au travers de son Data Lab propose une approche basée sur le traitement du langage naturel pouvant extraire de manière automatique les entités citées ci-dessus.

Mots clés : *NLP, Analyse criminelle, Anacrim, Analyse Relationnelle Visuelle, Intelligence artificielle*

1. Introduction

Le Data Lab de KPC ainsi que IntelFe ont entamé un projet en R&D en février 2022 pour l'aide à la reconnaissance d'entités nommées dans des sources textuelles (articles de presse, fichier pdf...) afin de pouvoir par la suite établir des relations entre ces différentes entités.

L'objectif premier est donc d'analyser une source textuelle afin d'en ressortir des entités : noms de personnes physiques, nom de personnes morales, localisations, mais également d'autres informations telles que les nationalités.

Enfin, il faut permettre l'accessibilité de l'outil et pour se faire l'interface métier qu'utilisera IntelFe prend la forme d'une application web. Cette application permet de charger le texte que l'expert souhaite analyser mais également d'apporter des modifications aux éventuelles erreurs que l'algorithme a commis et ainsi lui permettre peut-être d'apprendre ses erreurs. Cette interface comporte également des éléments sur l'algorithme utilisé : la version et la performance de celui-ci au travers de métriques adaptées.

2. Méthodologie

Une étape importante de ce projet était la réalisation d'un algorithme permettant la reconnaissance d'entités nommées. Nous avons ainsi choisi les réseaux de neurones. En effet, la littérature est unanime

sur l'utilisation et la performance de tels algorithmes. Dans notre solution, nous avons fait le choix d'utiliser un réseau de neurones récurrent avec une couche Long Short-Term Memory (LSTM) ainsi qu'une couche de Conditionnal Random Field.

Une autre étape importante a été la collecte de données afin de pouvoir apprendre les algorithmes mais aussi mesurer leur performance et les comparer. En effet, nous avons trouvé des données open source qui correspondaient avec notre problématique. Les données sont des textes provenant de médias par exemple. Nous avons créé un dictionnaire (un mot = un nombre entier), afin que les phrases deviennent des listes de nombres entiers, qui est un format qui peut-être traité par l'algorithme.

Enfin, une interface a été conçue pour permettre aux utilisateurs d'utiliser l'algorithme, de pouvoir apporter des modifications et consulter la version et les performances de l'algorithme utilisé. Ces résultats sont ensuite téléchargés au format csv pour qu'ils puissent être traités par IntelFe.

A la suite du travail réalisé avec cet outil, l'expert en analyse criminelle utilise les entités pour créer un graphe relationnel permettant de mettre en évidence de manière visuelle ces entités.

Il est important de garder en mémoire que cet outil constitue une base de travail pour faciliter le travail. Toutefois, en aucun cas cet outil est destiné à remplacer l'humain.

3. Originalité / perspective

L'utilisation de l'intelligence artificielle dans le domaine de l'analyse criminelle, bien qu'étonnante, se révèle être intéressante. En effet, dans un premier temps, la reconnaissance d'entités nommées permet un gain de temps pour la saisie des données par l'expert et lui permet de faciliter l'analyse des documents textuels pour réaliser des recoupements.

De plus, le choix de l'algorithme utilisé ici a une importance en termes d'efficacité. En effet, il faut que la solution allie efficacité et rapidité. La solution est un outil pour permettre à l'expert de gagner du temps sur la partie d'extraction des éléments.

Pour aller plus loin, l'extraction de relation entre les entités est la prochaine étape de l'intelligence artificielle dans l'analyse criminelle. Cela aurait pour conséquence de pouvoir traiter un grand nombre de sources textuelles dans un temps assez court et de permettre à l'expert d'effectuer un travail de vérification de ces éléments.

Références

LSTM :

- Hochreiter S. et Schmidhuber J., *Long Short-term Memory*, 1997
https://www.researchgate.net/publication/13853244_Long_Short-term_Memory

CRF :

- Charles Sutton, Andrew McCallum, *An Introduction to Conditional Random Fields*, University of Edinburgh and University of Massachusetts Amherst, 2010
<https://arxiv.org/pdf/1011.4088.pdf>
- Daphne Koller, *Champs aléatoires conditionnels*, Université de Stanford, 2017
https://www.youtube.com/watch?v=rc3YDj5GiVM&t=943s&ab_channel=MachineLearningTV

Analyse criminelle :

- Frédéric LENFANT, L'analyse criminelle : de quoi s'agit-il ?
<https://www.intelfe.com/analyse-criminelle/>
- Frédéric LENFANT, Méthode MARVADRISK
<https://www.intelfe.com/formation-analyse-criminelle/>