Title

Deep Learning Insights to handle NLP (Re)insurance Domain Issues: A Claims Analysis Perspective

Speaker

Aurélien Couloumy, Reacfin

Abstract

Unstructured data such as text remain quite untapped nowadays in the (re)insurance industry. The first basic reason of this probably comes from the unawareness of the way to handle well these texts. Natural language processing (NLP) held propose various techniques to address text analysis tasks such as extraction or classification. However, insurance industry suffers from incovenient domain specifities that often limit the setting up of such methodologies. The presentation aims at presenting through a claims analysis and data extraction case study, some of these issues while presenting and applying deep learning techniques to illustrate how to overcome them.

Data collection will firstly be discussed to highlight insurance lack of data as regards of a common NLP projects. Data augmentation techniques such has synonym replacement, random insertion, etc. will be presented and applied. Then data preparation will be introduced to illustrate insurance vocabulary specicities. Pretrained word embeddings (such as Word2Vec, GloVe or ConceptNet) will be compared to custom word embeddings approaches. Data quality aspects and impact on data extraction tasks such as regular expression will be described. Named Entity Recognition (NER) will be explained (Character embedding and Bi-LSTM CRF) and applied to demonstrate the effectiveness of deep learning systems. Annotation part will be also discussed and will precise cost issues. Active learning models (CNN least confidence, DO-BALD and BB-BALD) will be experimented. Results contextualization will be finally presented through different modelling scenarios including sequence representation techniques such as RNN, CNN, HAN and benchmarked, to show domain customization complexity. Methodological perspectives, results regarding claims analysis and usability aspects will be discussed to conclude.

Biography

Aurelien Couloumy is in charge of the Data Science Department at Reacfin. He is a qualified actuary of the French Institute of Actuaries (IA) and the Institute of Actuaries in Belgium (IA|BE). He is also a Lecturer at the Institut de Science Financière et d'Assurances (ISFA – Université Lyon 1) in France and a member of the SAF Laboratory where he works on various R&D topics mixing Actuarial Science, Insurance and Natural Language Processing.