From opinion mining to document search: an overview of applicative use-cases along with practical implementation issues and open research questions

Géraldine Damnati[1,†]

[1]*Orange Innovation, Lannion, France*

Abstract

Recent advances in Natural Language Processing and Information Retrieval, along with easier access to very powerful Language Models and libraries to manipulate them, have led to a notable expansion of the NLP application field. Applicative domains that were traditionally exploiting language technologies like bio-medicine, journalism, social media, customer relationship or digital humanities are progressively addressing more complex tasks, from more complex and heterogeneous data. In the meantime, research communities are structuring themselves around new applicative domains like finance, e-commerce, justice, legal or administrative organizations. From an operational point of view, the prevalence of Data Scientist positions is growing in many companies and public institutions and language technologies skills has become key in many applicative domains. However, despite the huge acceleration of progress in NLP and IR in the deep learning era, state-of-the-art models still represent practical implementation and cost issues in order for them to be deployed in operational services. Additionally, the mismatch between public datasets and business data remains a limitation to bridge the gap between the research community and operational applications.

In this talk I will present several applicative use-cases developed at Orange Innovation in order to feed operational services towards customers and employees. I will discuss the current bottlenecks for the deployment of state-of-the art models in operational services. From a broader perspective, I will present how tasks that are studied in research can yield new opportunities and, conversely, how feedbacks from operational projects can yield new research directions. For instance, as an international group, multilinguality plays a central part in our research activities. I will also highlight how we manage to bridge the gap between academic and collaborative research and business use cases through the construction of specific datasets that can be of interest for the community and transpose to operational contexts.

Keywords

Information Retrieval, Natural Language Processing, Information Extraction, Opinion mining, Technological transfer