

Chair Data Science and Artificial Intelligence for Digitalized Industry and Services

Internship project

Subject

Modeling the graph of citations in case law

Possibility to continue as a PhD candidate

YES (Funding to be confirmed)

About the chair

The Chair Data Science and Artificial Intelligence for Digitalized Industry and Services (DSAIDIS), lead by Florence d'Alché-Buc, a Professor in the department Image, Data, Signal of Telecom Paris, unites five industrial partners: Airbus Defence & Space, Engie, Idemia, Safran et Valeo. It's general objective is to develop, in collaboration with the partners, teaching and research of the international level.

Its four principal research directions are:

1. Building predictive analytics on time series and data streams.
2. Exploiting large scale, heterogeneous, partially labeled data.
3. Machine Learning for trusted and robust decision.
4. Learning through interactions with environment.

Description of the internship

Supervision

Nils Holzenberger (<https://perso.telecom-paristech.fr/holzenberger/>)

Location and dates of the internship

Address : Télécom Paris, 19 Place Marguerite Perey, 91120 Palaiseau

Date of the beginning of the internship : 2024

Team where the thesis will be written

Department INFRES, Team Data, Intelligence and Graphs (DIG)

Keywords

Natural Language Processing, graph analysis, law

Detailed subject

Courts conclude legal cases by writing up a decision. This is an authoritative document that explains the outcome of the case, the decision that the judge(s) came to, and arguments that support the conclusion. As part of its argumentation, the decision cites to other legal cases and to laws. This creates connections between documents, and thus creates a graph. This is comparable to scientific papers citing one another.

Evans and Rzhetsky (2010) show how novel scientific discoveries can be produced by exploiting the graph of citations among science papers. This method has been further extended by Sourati and Evans

(2021). The graph of legal citations has been studied in the context of US case law (Adusumilli et al., 2022), but only for descriptive purposes.

This project is in 2 steps. First, the graph of citations needs to be extracted from open collections of case law (case.law, Judilibre, ECHR) and stored explicitly. Second, this graph will be the basis to apply graph and NLP tools, to pursue any subset of the following goals:

- Predict missing citations
 - Similarities between cases
 - Applicable laws
- Predict (unexpected) interactions between existing laws
- Detect incompatibilities between existing legal decisions
- Suggest novel thematic groupings for cases and laws
 - Re-number laws (Blair-Stanek and Van Durme, 2021)
 - Retrieve case law
- Improve case retrieval

You are free to pick whichever topic in that list interests you. Success in any of these tasks would have direct applicability in a legal context, and could lead to a project in collaboration with e.g. the *Cour de cassation*.

Candidate profile

- M2 student, interested in research
- Coursework in statistical machine learning, probabilities
- Good level of programming in Python
- Good command of English
- Optional: familiarity with deep learning and NLP libraries

Application

To send to nils.holzenberger@telecom-paris.fr:

- Curriculum Vitae
- Personalized motivation letter that explains interest of the candidate in the subject (can be directly in the body of the email)
- Grade reports for recent years
- Contact of a person willing to give recommendation

Incomplete applications will not be considered.

References

Adusumilli, K., B. Brown, J. Harrison, M. Koehler, J. Kutarnia, S. Michel, M. Olivier, C. Pfeifer, Z. Slater, W. Thompson, et al., 2022: The structure and dynamics of modern united states federal case law. *Frontiers in Physics*, 9, 755.

Blair-Stanek, A. and B. Van Durme, 2021: Ai for tax analogies and code renumbering.

Evans, J. and A. Rzhetsky, 2010: Machine science. *Science*, 329(5990), 399–400.

Sourati, J. and J. Evans, 2021: Accelerating science with human versus alien artificial intelligences. *CoRR*, abs/2104.05188.

