



Paul NULTY

Technology centre: Center for Applied Data Analytics (CeADAR), University College Dublin (UCD)

Academic Mentor: Dr. David Lillis

Commercial Partner: Corlytics

Commercial Mentor: Ray O'Donnell

Paul received his bachelor's degree in University College Dublin (UCD), Ireland, focusing on A Neural Network Eyespace Detector for the Game of Go. In 2005 he started his doctoral degree in School of Computer Science and Informatics, University College Dublin doing research on Lexical Expressions of Semantic Relations between Nouns. Among other achievements Paul developed a pipeline for applying natural language processing and visualisation to conduct research using a large collection of digitised historical texts, successfully applied as an individual researcher to the Amazon Cloud Credits for Research program and helped to organise the Machine Reading the Archive programme at Cambridge University.

Dr. David Lillis

Dr. David Lillis is an Assistant Professor in the UCD School of Computer Science. He is a Principal Investigator in the CeADAR centre for applied data analytics, and an SFI Funded Investigator in the Crop Optimisation through Sensing, Understanding and Visualisation (CONSUS) project. Dr. Lillis also maintains research collaborations with groups in the University of New Haven (UNH), USA, and Beijing University of Technology (BJUT), China. His collaboration with UNH came about through being awarded a Fulbright Scholarship in 2017. He has published over 50 peer-reviewed research papers in a variety of subject areas including Information Retrieval, Multi Agent Systems and Digital Forensics.

Ray O'Donnell

Ray O'Donnell has responsibility within Corlytics for Product Engineering, Technical Architecture, and Technical Strategy. Prior to working with Corlytics, Ray was Chief Technical Architect for NetReveal that focused on the detection and prevention of financial crime, providing realtime, batch, and offline mission-critical analytical products to almost 200 financial institutions worldwide. Ray has over 35 years of experience in architecting and delivering high-performance analytical software products to global financial institutions. Ray has a particularly strong background in educating clients and colleagues on technology subjects, and in guiding teams in technology-based projects to a successful conclusion. Over the course of his career, Ray has engaged in deep-level mentoring and coaching.

Center for Applied Data Analytics (CeADAR)

CeADAR is the National Centre for Applied Data Analytics and Machine Intelligence. CeADAR is a market-focused technology centre that drives the accelerated development and deployment of data analytics and machine intelligence technology innovation. The Centre's work focuses on developing tools, techniques and technologies that enable more people, organisations and industries to use analytics and machine intelligence for better decision making and competitive advantage. CeADAR is the bridge between the worlds of applied research in data analytics and machine intelligence and their commercial application.

Corlytics

Corlytics are leaders in regulatory risk analysis. To conduct this analysis, the company sources, extracts and analyses unstructured text from multiple regulatory sources. The primary challenge the company seeks to solve for its customers is to serve them up with relevant, clear and concise information from 10,000's of documents. The company is seeking to apply sophisticated text analysis approaches to a) help them more efficiently scale their analysis and b) continually include a broader set of content.

Paul's project

“Natural Language Processing for Classifying and Understanding Regulatory Risk”

The project aims to improve the state-of-the-art in natural language processing (NLP) techniques for analysing and classifying complex legal and financial texts, with a particular application to regulatory risk in the financial services sector. Extracting and representing information from unstructured text is an increasingly fruitful method for many industry applications, since so much of modern social and commercial communication is conducted through digital text.

Discovering, classifying, and representing concepts using the statistical distribution of words in natural language is also an important basic research question in computational linguistics. The proposed research will approach this scientific question within the specific context of documents generated by regulatory bodies and regulatory enforcement, with the aim of using discoveries in classification and data analytics methodology to improve the performance of automated systems that assess regulatory risk in order to aid companies in achieving compliance and reducing cost in an increasingly complex regulatory environment.

Regulatory risk analysis is a particularly appropriate testing ground for machine learning methods using linguistic features, with an emphasis on model interpretation and transparency. The proposed company partner, Corlytics, currently assesses the risk levels indicated in regulatory documents by manually categorising them according to the type of financial product or service and category of regulatory control they refer to, using this to quantify risk.

Current methods for automatic document classification often rely on low-level features combined with complex learning algorithms, targeting performance on standard datasets to the detriment of feature interpretation and robustness. Corlytics emphasises ‘white-box’ methods so that classification can be constantly examined and revised with reference to a fastchanging regulatory system and the priorities of clients. By using NLP to extract higher level features such as named entities and semantic roles from documents, the proposed project aims to produce robust and accurate predictions using interpretable statistical learning methods.
