KPMG: Task complexity, not data availability, drives the choice of machine learning method

Summary

NLP and text mining approaches often rely on Deep Learning for Named Entity Recognition. These approaches enable meaning to be extracted (i.e., from sentences and passages) and are often able to perform well on complex tasks. However, KPMG's experience with document search and classification illustrates that when a particular sub-task is simple enough, traditional machine learning may be the best approach.

Opportunity: Better substantiation for R&D tax credit claims

Tax incentives for Research and Development (R&D) in the US can be significant. They can be up to 11%-15.8% of eligible incremental R&D expenditures and the benefit can further increase with many states also offering a state R&D credit. For smaller research organizations in particular, these tax credits can be decisive in making an R&D project commercially viable, or in securing investor funding to initiate projects in the first place. KPMG is hired by businesses of all sizes to document the R&D they've done and to help ensure that the business gets the maximum tax credit for that R&D.

Challenge: Lots of manual effort!

In the US, the Internal Revenue Service (IRS) assesses the merit of R&D tax credit claims using a 4-part test. The test checks whether the activities being claimed for: 1. Involve the creation of a new business component, or improvement of an existing one; 2. Are technological in nature (i.e. based on hard sciences or engineering); 3. Discover new information that eliminates uncertainty related to the methodology, capability or design of the business component; 4. Involve a process of experimentation through e.g., simulation, modelling or testing. Clearly there is subjectivity in this assessment, so providing strong evidence is key to a good outcome.

The evidence is typically collected from an organization's documentation and can take many forms. It could include presentations, emails, meeting minutes, lab reports, test records, and engineering drawings. Content is often unstructured, of intractable volume, stored in various repositories, or in some cases (e.g., in Agile or Continuous Improvement / Continuous Delivery environments) it may be very limited. In any case, regulations do not specify what qualifies as "sufficient" evidence, so it's important to review as much information as is available and, present as much of it as possible that is relevant and high quality.

KPMG supports clients in their audit readiness and is experienced in managing the discovery process. Traditionally this has involved a top-down approach, starting with a list of projects, trawling through document repositories related to those projects, manual searching through the documents (e.g., using keywords), reading the documents and tagging specific sections that satisfy each of the 4 tests. This is a significant manual effort, and necessitates some prioritization, which risks excluding valuable evidence and takes up valuable time from client scientists and engineers to support the effort. KPMG wondered if machine learning could help them do better.

Solution: Rules-based approaches outperformed machine learning

KPMG initiated an internal hackathon with four teams to compete on solving a subset of the problem (document chunking), using alternative methods. The teams were given 1000 documents with labelled sections and asked to present a confidence score for each of the documents' relevance to each of the 4 tests. Documents were chunked into sections by tokenizing words and sentences, and the teams tried a range of approaches including statistical learning (e.g., regular expressions, Support Vector Machines, Decision Trees and Random Forest), Deep Learning (DL) for Named Entity Recognition, and rules-based approaches. They found that accuracy ranged from 55% using out-of-the box document discovery software (about as effective as a manual keyword search), to above 70% for DL. However, the best approaches were rule-based, with accuracy exceeding 85%. This was likely due to a relatively high degree of standardization between document formats, making the document chunking task relatively straightforward.

Outcomes: Better evidence provides increased ability to claim credits

The system is now in operation with several KPMG clients to great effect. Each month it processes upwards of 5,000 documents, and critically the search is transformed from a selective, top-down approach to a bottom-up, exhaustive approach. Tax law is relatively static over time, and so the system requires minimal maintenance and improvement, while providing greater leverage for human expertise. There is also anecdotal evidence to suggest significant impact; for instance, one KPMG client was able to secure an additional 40% tax credit for their R&D tax credit study as a result of utilizing machine learning to review R&D project documentation to determine eligibility for the tax credit.

It is worth reflecting on the relative performance of Deep Learning to other methods. The results reinforce the assertion that even if there is a large enough dataset, DL tends to be superior only when the data or the problems are extremely complex. In this case simple rules and keywords were sufficient to identify the relevant information for each test, while also providing greater explainability.