On the Radar: Cortical.io Contract Intelligence v2.4 extracts key information from contracts

Semantic folding-based AI solution for semantic fingerprinting of legal documents

Summary

Catalyst

One of the challenges in the legal world is that the vocabulary used in legal documents is rather narrow, and small differences in how these words are expressed can lead to significant contextual or semantic differences. This makes the task for a human reading such documents doubly challenging. Cortical.io Contract Intelligence (CCI) helps organizations speed up a reader's absorption of legal print by distilling key concepts and information buried in legalese through the aid of an artificial intelligence (AI) solution.

Cortical.io has created technology it calls semantic folding, and this owes its origins to a strand of AI that has over the years been running in parallel with connectionist AI models such as the currently highly popular deep neural networks (DNNs). This strand is rooted in neuroscience, and its most prominent champion is the Numenta research organization that Jeff Hawkins and team created. Cortical.io is a partner of Numenta and has taken the memory model used in Numenta's hierarchical temporal memory (HTM) theory and adapted it to natural language and real-world use cases. Unlike the statistics-based connectionist AI models such as DNNs, semantic folding features fast learning from a small set of examples and has high granularity in the concept details it can hold.

Key messages

- CCI extracts key information from thousands of complex contracts that use disparate and diverse language.
- CCI can work with existing contract management software by populating it with key information extracted from legal documents.
- The scope of text documents can apply to a wide-ranging business domain and can include amendments, certificates, approval notes, and letters.
- The semantic fingerprinting is finely detailed so that language ambiguity can be resolved.
- Languages currently supported are English, German, and Spanish, but other languages can be added on request because the solution is language independent.

Ovum view

Cortical.io recognized the potential of the memory model, known as sparse distributed representation (SDR), used in Numenta's HTM and built an AI system on top of SDR that can be used to solve real-world problems. The system Cortical.io created, semantic folding, maps a business domain into its internal memory by ingesting a mass of available reference material from within that domain: related Wikipedia pages, business books, and other documents. Cortical.io uses a 128×128 node SDR matrix; since each node holds a binary value, this allows 2 to the power 16,384 possible patterns to be stored, or more than 10 to the power 4,932. With 2% of nodes active for each memory pattern, this reduces to 10 to the power 98. The estimated number of atoms in the universe is 10 to the power 80. In conclusion, there is no shortage of space in the system.

CCI has been trained on the legal domain and can be enhanced by continued training on the customer's specific documents. In production use, CCI is presented with a legal document and will

extract meaning to a report output. CCI will process all types of legal documents: contracts, credit agreements, lease agreements, ISDA master agreements and annexes, bond indentures, amendments, certificates, approval notes, letters, etc.

Cortical.io has produced one of the few SDR-based AI systems currently available on the market and is a pioneer in demonstrating the power of this technology.

Recommendations for enterprises

Why put Cortical.io on your radar?

There are many solutions on the market that aim to process unstructured and structured corporate data and extract meaning: this has been a key application for business intelligence systems and, with progress in the machine learning (ML) branch of AI, has also become a hot application for that technology. Cortical.io has taken an original approach, starting with technology from Numenta as a basis and building its own technology on top to produce a different type of AI assistant. In the legal domain, which has a strong need for help in digesting complex documents, CCI has advantageous properties such as high semantic precision, fast training, and fast production processing. Ovum believes that Cortical.io's solution deserves evaluation.

Highlights

CCI is a natural language processing (NLP) assistant that combines unsupervised ML algorithms with Numenta's SDR to produce semantic folding. Semantic folding has properties that are similar to the human neocortex:

- It is continuously learning.
- Its training is unsupervised and performed by exposure to real-world data.
- The AI system is its memory (there is no separation between memory and processing); learning is accomplished by adaptation of the neural topology.
- Temporal sequences can be stored.

The SDR is stored in the retina engine (see Figure 1). A bit pattern in the SDR represents a semantic feature. The first step is to train CCI on the application domain, where reference documents are exposed to CCI using an encoding process and stored in the retina engine as a semantic word fingerprint dictionary – this is level 1: word fingerprints.

In production use, a document is input into CCI and semantic text fingerprints are created based on the previously trained word fingerprints – this is level 2: text fingerprints. Semantic text fingerprints can then be compared to evaluate proximity of concepts across text pieces. Similar meanings will have a high degree of proximity; for example, "done deal" and "signed contract" text fingerprints will overlap by as much as 30%.

CCI will quickly generate consistent and comparable summary abstracts and spreadsheets as output, and using the API the user can fill any existing contract management software with automatically extracted data.

The Cortical.io Contract Intelligence Engine analyzes the meaning not just of keywords but of whole sentences, paragraphs, and long text so that the problems of language ambiguity and vocabulary mismatch within and across documents are overcome

The Cortical.io AI system is independent of language, although for its first products in the market the company has launched English, German, and Spanish solutions. The extractable document types include .pdf, .doc, and .txt.

CCI is accessed via a simple user interface. The solution has a simple user manual and no training is required, other than a brief explanation of the main features and capabilities. It is designed to be used by the business and does not require any AI knowledge – the technology is fully automated. A half-day of training is included in the initial professional services package.

Figure 1: Architectural diagram of CCI
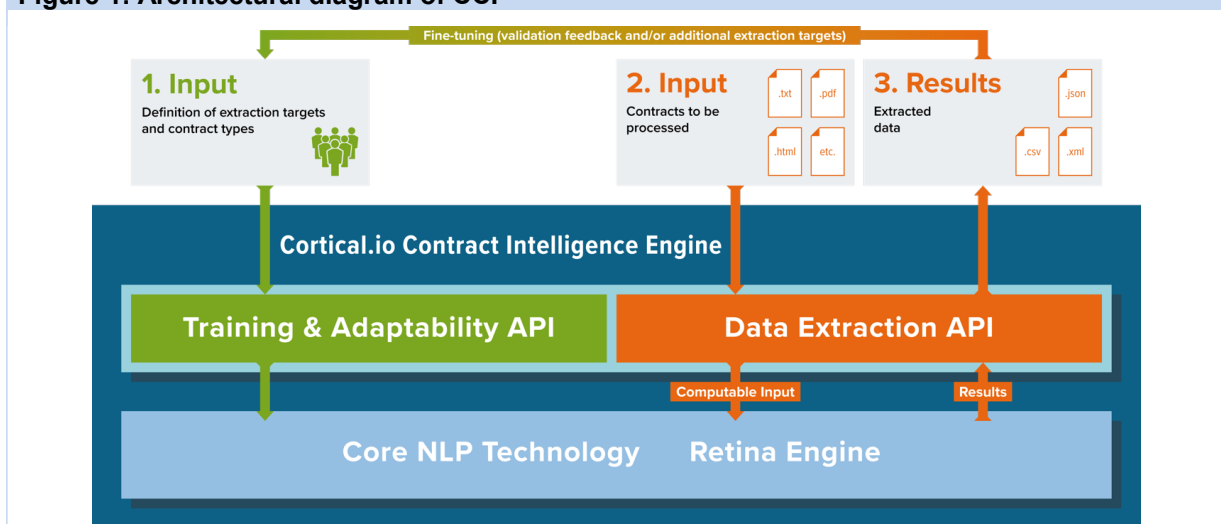
Source: Cortical.io

Background

Cortical.io was founded in Vienna, Austria in 2011 to test the semantic folding theory developed by CEO Francisco Webber. Francisco had been following the research done by Jeff Hawkins at Numenta and had the intuition that his theory of SDRs could be applied to NLP. In 2012, the idea was tested by building a prototype, which was financed with Austrian governmental research funds. The prototype of the retina engine confirmed that converting text into SDRs dramatically improved the processing of natural language. Private investors joined the company in 2013, enabling the development of the technology stack. An API was launched in 2014, followed by an enterprise version in 2016. In 2015, a strategic partnership agreement was signed with Numenta and an office opened in Silicon Valley. In 2017, the first enterprise license was signed and an office opened in New York. In 2018, Cortical.io entered a joint business relationship with PwC Germany.

Current position

Cortical.io has 22 employees (15 in Austria, seven in the US). Depending on the next financing round, the company expects to grow to 50 employees by the end of 2019. The company expects to increase its revenue by over 70% over the next year. The company is at an early stage of growth, with two customers in production and many in pilot phase.

On the roadmap

Cortical.io has product features and applications that are currently in progress:

- Semantic Enterprise Search for large-scale data collections.
- Applying semantic folding to provide intelligence support in high-technology product manufacturing.
- Semantic data feeds such as news, SEC documents, and job profiles.

Data sheet

Key facts

Table 1: Data sheet: Cortical.io

Product name	Contract Intelligence	Product classification	NLP
Version number	2.4	Release date	November 2018
Industries covered	All	Geographies covered	Global
Relevant company sizes	Large	Licensing options	Term: annual subscription per use case. SaaS: volume based.
URL	www.cortical.io	Routes to market	Direct sales Partner sales
Company headquarters	Vienna, Austria	Number of employees	22

Source: Cortical.io

Appendix

On the Radar

On the Radar is a series of research notes about vendors bringing innovative ideas, products, or business models to their markets. Although On the Radar vendors may not be ready for prime time, they bear watching for their potential impact on markets and could be suitable for certain enterprise and public sector IT organizations.

Authors

Michael Azoff, Distinguished Analyst, Information Management

michael.azoff@ovum.com

Ovum Consulting

We hope that this analysis will help you make informed and imaginative business decisions. If you have further requirements, Ovum's consulting team may be able to help you. For more information about Ovum's consulting capabilities, please contact us directly at consulting@ovum.com.

Copyright notice and disclaimer