Artificial Intelligence in Action:

Programme Advisers and Recommendation Systems

Case Study: AgResearch

May 2019

TOWARDS OUR INTELLIGENT FUTURE TE ARA MŌ TĀTOU ATAMAI O ĀPŌPŌ aiforum.org.nz

forum NEW ZEALAND
Te Kāhui Atamai Iahiko o Aotearoa

IDC
ANALYZE THE FUTURE

Special Study

Artificial Intelligence in Action: Programme Advisers and Recommendation Systems

Monica Collier

May 2019

Within academia, the scientific community and the commercial world there exists a vast amount of published literature. The pace at which scientists publish research is increasing. Various estimates suggest researchers publish between 1.8 million and 2.5 million new papers each year. The International Association of Scientific, Technical and Medical Publishers 2015 report says the volume increases by 4-5% per year.

Commercial companies and academia have access to digitised data; for example, repositories of datasets, reports, books, images, and photographs. These sources could be internal or external to the company.

The problem is that there is too much literature for humans to consume. Companies and academia can't leverage the optimal value of datasets and literature. This slows down the pace of innovation.

Artificial Intelligence (AI) speeds up the process of extracting and classifying content. AI can use this data to provide advice or recommendations. For example, recommending a reading list, finding testable hypotheses, or suggesting a product or service for a customer.

PROGRAMME ADVISERS AND RECOMMENDATION SYSTEMS

AI based recommendations and advice provide users with assistance in a wide range of use cases. AI helps businesses capture and extract detailed information and insight from enterprise documents. For example, processing mortgage application needs and requirements. This includes automatic classification and understanding of the document. It includes integration with recommendation systems. These are intelligent filtering engines that narrow the decision-making process.

Practical applications include:

- Banking systems. Banks are maturing in their ability to extract insight from enterprise documents. They pair this with recommendation systems to match products and services.
- Shopping advisors and product recommendations learn about prospective customers and make shopping recommendations. For example, Amazon book recommendations.
- The securities and investment industry. IDC expects AI to play an increasing role in calibrating trading decisions in real time.

MINING ACADEMIC LITERATURE FOR SCIENTIFIC HYPOTHESES

The following case study is a niche use case in the scientific research sector. It is difficult for scientists to keep up with the amount of literature available today. This AI use case assists a scientist to pinpoint related data or facts within large amounts of literature. The system then forms a testable hypothesis. Using AI speeds up the process of finding solutions to scientific problems. This is particularly valuable for translational research.

Translational research applies knowledge from one field of science to a different field to solve a problem. It takes results of lab research and applies it to solve real world problems. Scientists use translational research to discover innovations; new therapies, medical procedures or diagnostics.

AGRESEARCH'S COGNITIVE TEXT MINING FOR NITRATE LEACHING SOLUTIONS

Research is a New Zealand Crown Research Institute. Its science and technology intend to benefit New Zealand agriculture and the wider economy and society. The institute conducts research and develops solutions across all sectors of the pastoral agriculture including pasture seeds, pest control, high value food production, and farming systems. The faster scientists can find solutions to problems the agricultural sector faces the more advantage New Zealand gains on the world markets.

Often solutions to problems already exist but scientists need to find linkages and solutions that apply in their domain. For example, a scientist discovers a statement that says, "a affects b". In another source a statement says, "b affects c". The scientist can infer a relationship between 'a' and 'c'. These statements are often in different pieces of research literature. As mentioned before, it's getting harder for humans to conduct literature reviews when there is a vast number of new scientific publications.

AgResearch recognises how AI has the potential to change the way it does science. It wants to utilise AI powered literature related discovery as a way of finding out new hypotheses to test. The company is particularly interested in discovery within the field of nitrogen leaching. There are a number of expert scientists in this field in New Zealand. The subject will be a good test of the system to see if it can find hypotheses scientists have not yet found. AgResearch says the issue is a pressing challenge for the New Zealand agricultural sector.

AgResearch is taking a 'learning by doing' approach to understand what AI can do. In part, this learning will help the institute make a later decision around whether to 'buy' or 'build' a production solution.

The Solution

AgResearch is in the early stages of solution development to help its scientists form testable hypotheses. The company intends to have a web-based application with a mechanism to ask the system to find statements of hypotheses. It is likely to have a graphical display where users can click their way through to the results that interest them.

The solution is grounded in AI. Natural Language Processing (NLP) is a core enabler. Recent advances in NLP are a fundamental reason why the adoption of AI is accelerating. It is becoming possible to work with natural language in sophisticated ways.

There are three main processes the system will undertake:

Store and Manage documents

The solution will tap into streams of published data using research publishing houses' APIs, for example. These include Dutch publishing giant Elsevier and international science publisher Springer Verlag. With the API connection, the system can interrogate the publishers' web-based databases to obtain papers and data that are open source or to which AgResearch subscribes. The system will store and manage the documents it downloads.

Examine and extract content

The system ingests the inbound documents and extracts its content in a structured manner into a large textual database.

Find relevant and contextual linkages between statements.

The system will apply NLP across the textual database. This is to find the "a affects b' type of statements within the texts. This is a Natural Language Processing method known as "information extraction". A key part of this is that the NLP can understand words within their context. NLP deconstructs the syntax of sentences to ensure it captures the intended meaning in the database. For example, the word "water" can be both a noun (a body of water or a chemical compound) or a verb (to water the garden). Next, the system relates the contextual mention of "water" to a universal, unambiguous, definition. Now the system can compare statements about "water" between different documents.

OVERCOMING CHALLENGES: AGRESEARCH'S LEARNINGS

The specific challenges that AgResearch is facing on its journey include:

- A lack of skill sets. Finding people with the knowledge and skills to develop an end to end solution can be difficult.
- Speaking a common language. The team working on the project includes social scientists, literature librarians, statisticians and a technology consultant. A challenge is getting everyone to speak a common language to develop a base understanding of each other's expertise.
- Understanding the literature and how to access it. Many existing databases for science literature are locked down. AgResearch says it was difficult to gain access. In the end, the solution was to use the base data from the literature publishers supplemented with public databases. For example, WikiData for universal definitions and the Directory of Open Access Journals.
- b Another challenge was the switch in mindset required of the system users, in this case, soil scientists. The system aims to solve the problem of "how do we export and analyse all the possible data to seek a solution". The scientist's mindset traditionally was "How do we organise this data and filter it down to what we already think might be the solution". AgResearch is taking a 'learning by doing' approach to engage end users in exploring different methods of using literature to solve problems.

THE BENEFITS

The two largest benefits of using an AI driven literature discovery, according to AgResearch, are:

- Time saved discovering novel solutions. It is a scientist's job to come up with new solutions. Having AI discover testable hypotheses saves considerable time.
- Help new, emerging researchers come up to speed in a field. There is potential for the system to map out a large domain of science and identify key papers for researchers to consume.

THE RISKS

AgResearch recognises there is a risk the project might not end up producing something usable. To mitigate this, the project intends to produce a prototype system, even if that is not the end-stage production system. For example, success could be automating the aggregation of literature into a database.

Data management and validation is also a risk. An example here is in how the system identifies the key concepts in a research paper. Often this is reliant on author specified key words which may not fit within a standard categorisation system. If the metadata for each paper is not correct or is inconsistent, this will reduce the output quality. The system could miss finding valid linkages.

There is a risk that a solution already exists that would meet AgResearch's needs. The team has found it difficult to assess what alternatives are available and suitable. However, the more the project team is "learn by doing" the more expertise they will gain. This will help them assess alternatives in the future.

WHAT ALTERNATIVE SOLUTIONS EXIST?

There are several systems for data mining and hypothesis seeking that have been developed specifically for research scientists and academia. However, many of these are not specifically built upon Artificial Intelligence, and the majority are designed for medical science. These translational research platforms include: TranSMART, SIDRAiTrip, BRISK, caTRIP, iCOD, iDASH. See *Table 1.*

TABLE 1

Translational Research Platforms

SOLUTION NAME	DESCRIPTION
TranSMART	Open source platform for storing, analysing, and testing hypotheses. Can contain large volumes of clinical and translational data.
SIDRAiTrip	The Sidra Integrated Translational Research Informatics Platform
BRISK	An open source data management tool developed by the University of British Columbia, Vancouver
caTRIP	Cancer Translational Research Informatics Platform
iCOD	Integrated Clinical Omics Database
iDASH	This collaborative platform focuses on algorithms and tools to share data in a way that preserves privacy.

Source: IDC, 2019

In the commercial world a number of platforms are available for literature review. Many of these contain AI components. These tools extract information from content, usually using NLP. Systems such as iris.AI then create a visual map of relationships to help the user drill down into the results that are more relevant to them. See *Table 2.*

TABLE 2

Commercial Platforms

SOLUTION NAME	DESCRIPTION
Iris.ai	Research discovery with artificial intelligence. Researchers can enter the URL of a research document or enter a research question to search millions of documents. Clickable visual results lets a user drill down to a precise reading list. Try iris.ai
Dimensions.ai	Similar to iris.ai, Dimensions.ai lets users explore the connections between different research publications. Try dimensions.ai
Euretos AI	Contains over 200 omics databases and links together literature with clinical and experimental data. Researchers use this data for target and biomarker discovery and validation. View.Euretos AI
Semantic Scholar	Academic search engine. Try semantic scholar
Dataset Search	A google search engine for searching for data sets online. Try Dataset Search

Source: IDC, 2019

WHAT THE FUTURE HOLDS

AgResearch is open to either a "buy" or "build" solution should it decide to productionise. The company aspires to apply the system across all its research. It could also provide the system to other Crown Research Institutes. With expansion, the underlying principles remain the same. The boundaries around the literature to bring into the database will increase.

In the commercial world there are opportunities to improve productivity and value and make 'next best activity' more intelligent particularly in the banking, finance and insurance domains by using AI based programme advisers and recommendation systems.

IDC projects that spending on Programme advisors and recommendation systems worldwide will grow with a CAGR of 38.9% between 2017 and 2022.

IDC'S GUIDANCE

Rule-based chatbots tend to be easier to implement while conversational AI-powered platforms can scale and self- The AgResearch case study provides some good 'best practice' for any business looking to deploy AI. In particular, the lessons to apply are:

- Learn by doing. The team at AgResearch had little understanding of AI before embarking on the project. Now the team is building capacity by creating expertise in house and through partnering with external experts, without committing to a full-scale investment.
- Cross functional project team. The AgResearch team includes technical people, scientists, statisticians, and librarians. Each brings in a different perspective and different expertise.
- Involve your end users from the start. The project team has involved its end users, the soil scientists from the start. It has shown them the concepts and some initial analyses. In this way, the project team is building engagement while discovering barriers to adoption early.
- Take your time to decide between "buy" vs "build". There are pros and cons to either approach. By taking the time to build in house knowledge and expertise, the company can put together better arguments for or against each case.
- Set smaller goals. AgResearch realises that its "learning by doing" approach may not result in it building a full production system. It has a goal to ensure it delivers a usable output of some sort, even if that is a subset of the full system.

International Data Corporation (IDC) is the premier global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets. IDC helps IT professionals, business executives, and the investment community make fact-based decisions on technology purchases and business strategy. More than 1,100 IDC analysts provide global, regional, and local expertise on technology and industry opportunities and trends in over 110 countries worldwide. For 50 years, IDC has provided strategic insights to help our clients achieve their key business objectives. IDC is a subsidiary of IDG, the world's leading technology media, research, and events company.

IDC New Zealand, Level 11, 203 Queen Street, Auckland, New Zealand. Ph +64 9 377 0370
Twitter: @IDC | idc-community.com | www.idc.com

The Artificial Intelligence Forum of New Zealand is a non-government association with a mission to harness the potential of Artificial Intelligence (AI) to help bring about a prosperous and inclusive Future New Zealand.

The rapid development of AI technologies presents major opportunities and challenges for our country: from creating world leading AI businesses, nurturing a pool of talented AI engineers, applying AI technologies to our agriculture, government, manufacturing and service industries to holding a meaningful national debate on the broader implications for society, New Zealand needs to actively engage with AI now in order to secure our future prosperity.

The Forum brings together citizens, business, academia and the government connecting, promoting and advancing the AI ecosystem to help ensure a prosperous New Zealand.

AI Forum of New Zealand
L1 Building C, 14-22 Triton Drive, Auckland 0632, New Zealand
Ph +64 9 394 7693
www.aiforum.org.nz