Machine learning advances intelligence analysis

[Content preview – Subscribe to **Jane's Intelligence Review** for full article]

Assumptions about technology that does not yet exist underpin many analyses about artificial intelligence's likely impact on classified and open-source intelligence work. *HI Sutton* examines how artificial intelligence is currently used by open-source analysts

Key Points

- Artificial intelligence has changed the way that intelligence analysis is conducted, with tools now freely available that were once development priorities for state intelligence agencies, yet truly cognitive artificial intelligence remains elusive.
- The increasing adoption of machine learning in analytical workflows will require greater integration between data scientists and intelligence analysts, access to large volumes of genuine data, and an awareness of the challenges posed by the inscrutability of these processes.
- Intelligence analysts will have to make greater use of software that can visualise the processes or outputs of machine-learning systems to help them act as 'data translators' for non-technical intelligence customers.

Advances in artificial intelligence (AI) will transform intelligence collection and analysis, including open-source intelligence (OSINT). However, with AI being aggressively marketed by private companies and within governments and militaries around the world, it is necessary to look beyond the hype and consider what is currently possible with this technology and how these capabilities are likely to change the intelligence landscape in the immediate future.

Many of the fundamental problems that AI can address were laid out by the Japanese government's Fifth Generation Computer System (FGCS) programme, which steered global AI research throughout the 1980s. At that time, the focus of AI research was a method known as 'expert systems', whereby the computer is programmed to answer specific problems by applying human-defined rules.

In practice, it was very difficult to define rules that cover enough eventualities to create a program that humans would regard as truly intelligent. AI research petered out in the late 1980s, but a revival in research took place in the 2000s when improvements in computer power enabled an alternative approach to AI known as machine learning (ML).

Unlike the expert systems approach, in ML the computer 'learns' by itself. Moreover, at a time when human analysts are increasingly overwhelmed by the volume of data accessible in open sources, ML systems perform better in cases in which there is more data to analyse. Although the core logic for ML dates to the 1950s, it requires high levels of processing power to be effective, and hence has only become a realistic proposition for commercial uses since the mid-2000s.

© 2018 IHS. No portion of this report may be reproduced, reused, or otherwise distributed in any form without prior written consent, with the exception of any internal client distribution as may be permitted in the license agreement between client and IHS. Content reproduced or redistributed with IHS permission must display IHS legal notices and attributions of authorship. The information contained herein is from sources considered reliable but its accuracy and completeness are not warranted, nor are the opinions and analyses which are based upon it, and to the extent permitted by law, IHS shall not be liable for any errors or omissions or any loss, damage or expense incurred by reliance on information or any statement contained herein.

A man uses a Google vocal translate app on his smartphone during a Google promotion event at the City of Fashion and Design in Paris on 4 November 2014. With machine translation, multiple sources in different languages can all be translated into a common language and then analysed as a single body of evidence. (Thomas Samson/AFP/Getty Images)

1710283

Notwithstanding this shift in approach, the fundamental problems that AI research seeks to tackle largely remain those set out in the FGCS programme. This laid out goals including natural language processing (NLP), automatic voice recognition, intelligent image analysis, and machine translation.

Although the motivation for the researchers involved was to improve the human-machine interface, it is unsurprising that these technologies had applications for intelligence agencies seeking to process the vast amounts of data generated through technical intelligence collection. The extent to which these applications of AI have become available to the public is remarkable, to the point where many users no longer think of these as artificial intelligence.

Existing AI capabilities

NLP is the ability of a computer to interpret text communication that is written in a natural, free-flowing way. This includes the ability for a computer to respond to complex sentences including slang, poor or incorrect grammar, or missing words – all characteristics of normal written language. NLP has advanced to the point where it is now widespread. The 'chatbots' that help visitors to airline websites to book flights are an everyday example of NLP.

In the intelligence world, the same capability could enable a computer to process large numbers of social media posts and identify ones that are significant for military or law enforcement purposes. This capability goes beyond what can be achieved with key word analysis (searching for specific

© 2018 IHS. No portion of this report may be reproduced, reused, or otherwise distributed in any form without prior written consent, with the exception of any internal client distribution as may be permitted in the license agreement between client and IHS. Content reproduced or redistributed with IHS permission must display IHS legal notices and attributions of authorship. The information contained herein is from sources considered reliable but its accuracy and completeness are not warranted, nor are the opinions and analyses which are based upon it, and to the extent permitted by law, IHS shall not be liable for any errors or omissions or any loss, damage or expense incurred by reliance on information or any statement contained herein.

words or short phrases) because the whole text is analysed, thereby reducing false positives and allowing refined filtering among large numbers of positive results.

Automatic voice recognition can be viewed as a subset of NLP and means the ability to understand commands given audibly. On top of the problems inherent in NLP, this requires the ability to understand accents, speech errors, and differences between written and spoken expression. Another less obvious challenge is that the audio data may be of poor quality, contain background noise, or not cover the entire audio spectrum (traditional 'narrowband' telephones trim the highest and lowest frequency sounds from human speech to save bandwidth).

Intelligence applications for this technology could include the transcription of audio or videos so that they can be analysed together with text comments, or the identification of an individual's voice to subsequently locate that same person in other recordings. On 19 January 2018, online news publication The Intercept reported that by 2006 the US National Security Agency (NSA) had developed the ability to "identif[y] people by the sound of their voices". The report cited leaked classified internal NSA memoranda describing a process for developing 'voiceprints' of individuals from intercepted communications.

Intelligent image recognition means the review and categorisation of an image. Use of this technology is now an everyday occurrence. Google's reverse image search tool can be used to find other examples of an image on other websites, to locate the most likely original online source of the image, and to suggest the location if it includes a landmark.

Intelligence uses would include filtering videos posted online to find those that are likely to show footage of a particular event. These could then be reviewed to build up the 'ground truth' of the event. Another use is facial recognition, which can be used to identify the same person in multiple but seemingly unrelated images.

The last of the FGCS 'problems' was machine translation. Translating between languages has many of the pitfalls of NLP. However, with AI it is possible to translate texts in one language into another and produce a readable result. Services such as Google Translate and Bing Translator can automatically translate text or entire websites with a level of accuracy that was not possible, at least in open sources, at the start of the 2010s.

Traditionally, intelligence analysts required strong language skills to interpret texts or audio in a foreign language. Machine translation has not obviated the need for language skills. However, it is no longer the case that an analyst needs a high level of fluency in a language to have a basic level of access to the information contained in a document. Machine translation also means that multiple sources in different languages can all be translated into a common language and then analysed as a single body of evidence.

The cognitive myth

These tools are all dependent to an extent on ML. Unlike earlier applications of AI, with ML the computer effectively learns by doing.

For example, to teach the computer to identify photographs of people it could be provided with two sets of data, one of images containing people and one of images which do not. From this the program would work out which variables to use to differentiate the two groups.

© 2018 IHS. No portion of this report may be reproduced, reused, or otherwise distributed in any form without prior written consent, with the exception of any internal client distribution as may be permitted in the license agreement between client and IHS. Content reproduced or redistributed with IHS permission must display IHS legal notices and attributions of authorship. The information contained herein is from sources considered reliable but its accuracy and completeness are not warranted, nor are the opinions and analyses which are based upon it, and to the extent permitted by law, IHS shall not be liable for any errors or omissions or any loss, damage or expense incurred by reliance on information or any statement contained herein.

[Continued in full version…]

(1035 of 3379 words)

For the full version and more content:

Jane's Military & Security Assessments Intelligence Centre

This analysis is taken from Jane's Military & Security Assessments Intelligence Centre, which delivers comprehensive and reliable country risk and military capabilities information, analysis and daily insight.

*IHS country risk and military capabilities news and analysis is also available within **Jane's Intelligence Review**. To learn more and to subscribe to **Jane's Intelligence Review** online, offline or print visit: http://magazines.ihs.com/*

For advertising solutions visit Jane's Advertising

© 2018 IHS. No portion of this report may be reproduced, reused, or otherwise distributed in any form without prior written consent, with the exception of any internal client distribution as may be permitted in the license agreement between client and IHS. Content reproduced or redistributed with IHS permission must display IHS legal notices and attributions of authorship. The information contained herein is from sources considered reliable but its accuracy and completeness are not warranted, nor are the opinions and analyses which are based upon it, and to the extent permitted by law, IHS shall not be liable for any errors or omissions or any loss, damage or expense incurred by reliance on information or any statement contained herein.