Alex Wright

Algorithmic Authors

Natural Language Generation software begins delivering on the promise of automated prose.

SOON AFTER THE Sonoma County Little League's Falcons faced off against their rivals the Mustangs on April 28, 2015, the league's website ran this game recap:

Anthony T got it done on the bump on the way to a win. He allowed two runs over 2-1/3 innings. He struck out four, walked two, and surrendered no hits.

Anders Mathison ended up on the wrong side of the pitching decision, charged with the loss. He lasted just two innings, walked two, struck out one, and allowed four runs.

That story—and millions more like it—was written not by a cub reporter, but by Quill, a software program developed by Chicago-based Narrative Science. By analyzing game statistics entered using a mobile app and uploaded via the Internet, Quill identifies patterns in the data to produce original stories intended to mimic the voice of a flesh-and-blood scribe.

Along with companies like ARRIA, Yseop, and Automated Insights, Narrative Science is helping to launch a new market for editorial content created using the principles of Natural Language Generation (NLG).

In contrast to the better-known field of Natural Language Processing (NLP)—in which software transforms raw text into structured datasets—NLG addresses the opposite problem: turning structured data into prose.

There is nothing terribly new about NLG per se; the foundational research goes back decades, and it has long been a staple of Artificial Intelligence programs. The field has enjoyed a resurgence in recent years, thanks largely to the rise of big data.

"With all the investment in big data, the world has a tremendous amount to tell us," says Kris Hammond, professor

of computer science at Northwestern University and chief scientist for Narrative Science. "The need for this kind of technology has exploded."

Hammond began exploring the possibilities of computer-generated journalism along with his research partner Larry Birnbaum at Northwestern University, where he supervised a student research project called StatsMonkey, an experiment in auto-generating college sports stories in collaboration with the school's journalism department.

Since co-founding Narrative Science with former Doubleclick executive Stuart Frankel in 2010, Hammond has seen the company grow beyond its roots in the sports world to serve not just beleaguered Little League coaches, but professional purveyors of prose like Forbes.com, and several Fortune 500 firms who use Quill to produce management reports from internal business data.

Automated Insights has produced more than a billion stories to date ranging from quarterly earnings reports for the Associated Press to car descriptions for Edmunds.com, digital marketing material for Comcast, and game recaps for Yahoo! Fantasy Football.

With its reliance on structured data like scores, player stats, and win/loss records, sports journalism is a natural fit for NLG applications. Similarly, financial and other business data lends itself well to automated report writing. The potential applications extend into any number of domains where human beings are trying to make sense of data.

Both Quill and Automated Insights' Wordsmith software perform similar functions: gathering data from third-party sources, performing data analysis to glean a set of insights, and then applying NLG techniques to translate those insights into readable stories.

Wordsmith can ingest data in a wide range of formats (like XML, CSV, spreadsheets, and various APIs), then analyze the material to identify trends, changes, and other pertinent patterns. The software can then produce stories in any number of narrative formats, from traditional inverted-pyramid-style news stories to bullet-point summaries, tweets, and headlines. The software exports the resulting content in any number of formats, including XML, JSON, HTML, Twitter, or email.

For Hammond, the deeper challenge lies not in generating copy, but in finding the most pertinent meaning in a given dataset. "It's not just about converting numbers to language."

While recapping baseball games or churning out company earnings reports might seem like reasonably straightforward operations, the vagaries of human language and cognition can pose thorny computational challenges in creating credibly human-sounding stories.

At minimum, machine-generated content must meet basic standards of journalistic accuracy. "NLG should always strive to be well written, but the text must always support the data," says Automated Insights' James Kotecki. "In some cases, changing one word, adding an adjective, or even switching the placement of words can create unintended factual errors."

As any professional editor will tell you, a good story has to do more than just get the facts straight; it also has to engage the reader's attention. This is where the real AI challenges come into play.

"Truly human-sounding content must adhere to the contextual and tonal expectations of its intended audience," says Kotecki. To bring a story to life, the software must imbue the data with one of the essential virtues of human language: variation. This means not just mixing up synonyms, but determining the appropriate tone and tenor of the story, given the patterns in the data.

"If I'm telling a story about a game that was a blow-out, where one team took the lead and held on and pounded the other team into the ground, that's

a different story than if it was a squeaker," says Hammond.

For Hammond, the deeper challenge lies not in generating copy, but in finding the most pertinent meaning in a given dataset. "It's not just about converting numbers to language," he says. "Those numbers need context."

That context comes not from data analysis alone, but from understanding the larger domain and, in some cases, historical patterns. The software applies core AI principles like inference and reasoning to analyze the data in context, and to deduce the most relevant and interesting set of potential meanings. "It goes beyond simple NLG," says Hammond. "How do you introduce ideas, expand on them, and explain them?"

Quill uses its own proprietary case-frame language, which Hammond characterizes as an "old-school AI" strategy of capturing the entities and languages within a given domain and encoding them in an ontology.

For example, the entities in a base-ball game might include pitchers, hitters, balls, strikes, and runs; the system understands the semantic relationships between those entities, and can produce a set of variable terms to describe all the possible interactions between them. "For every relationship/object/predicate, you need variability built into the system," says Hammond.

Can this programmatic combination of data analysis and AI techniques really produce credible journalism? Will automated stories ever pass the journalistic equivalent of a Turing test?

In 2014, Swedish researcher Christer Clerwall published a study entitled "Enter the Robot Journalist" that gauged reader responses to a news article produced by Automated Insights's Wordsmith program with a story on the same topic penned by a flesh-and-blood scribe from the *Los Angeles Times*.

Participants in the study found the software-generated story to be more informative and trustworthy than the one written by the journalist, but they also characterized it as "descriptive, boring, and objective."

In some cases—like quarterly earnings reports—boringness might ac-tually seem like a virtue, but in other domains, this lack of narrative sparkle may preclude machine-generated stories from making it to the front page anytime soon.

For now, these companies are focused primarily on producing the kinds of content that are unlikely to displace any workaday writers. Hammond likes to talk about producing stories that serve "an audience of one." There are not too many full-time reporters out there working the Little League beat, after all.

In a similar vein, Automated Insights is developing plans to make its software more accessible to the general public, so individual consumers can upload their own data and generate automated stories on topics of their choosing.

Future applications for NLG may extend well beyond the realm of the traditional long-form narrative. At Aalto University in Finland, Ph.D. student Eric Malmi has created a software program called DopeLearning that is capable of generating its own rap lyrics. Other promising NLG projects include companies like x.ai, which offers an artificial intelligence-driven calendaring service that provides users with a virtual assistant named "Amy Ingram," who can arrange meetings for a user by sending email in a computer-generated writing style that is difficult to distinguish from that of a human assistant.

As NLG applications continue to evolve, developers may set their sights on more complex topics than sports and corporate earnings reports. Hammond and others have high hopes for the potential of NLG techniques to turn out high-quality, original journalism.

Moving beyond these kinds of just-the-facts news stories will require another conceptual leap in NLG application development, taking computer-generated writing beyond the realm of interpreting structured data sources and into the far messier world of interpreting free-form written material, much of it currently generated by human journalists. To bridge that gap, NLG software may need to start incorporating some of the text-parsing methods of its kindred discipline Natural Language Processing (NLP).

Already, Narrative Science is start-ing to incorporate data drawn from other news stories' headlines. For example, the software can analyze another news outlet's headline about an executive hiring or firing, then relate that data to subsequent movements in the stock price. However, scanning for keywords in a headline is a far cry from understanding the complexity and nuance of more interpretive journalism.

Generating information from unstructured text sources also poses much thornier problems of fact-checking, since it is one thing to double-check a stock price, but quite another to assess the reliability and motivations of a human subject. Seasoned editors and reporters rely on their years of journalistic experience to help vet their stories and screen out the implausible, which may prove difficult to replicate algorithmically.

While there is no shortage of computational challenges on the horizon, the sheer demand for written material across the Internet seems to point toward continued innovation in this nascent field. Could there be a robot Hemingway on the horizon? Unlikely as that may seem, Hammond has already gone on record as predicting a computer program will one day win the Pulitzer Prize.

Whether or not that ever happens, Hammond is convinced NLG applications can serve a basic human need: to help us make sense of the world around us not as a set of data, but as a web of deeply interconnected narratives. "We're all wired for storytelling." ▣

Further Reading

Christer Clerwall.
Enter the Robot Journalist. *Journalism Practice.* 8, 5 (2014), 519-531.
http://bit.ly/1HDkz8C

Reiter, E., and Dale, R., 1997.
Building applied natural language generation systems. *Nat. Lang. Eng.* 3, 1 (March 1997), 57-87.
http://bit.ly/1CGSsJ8

Malmi, E., Takala, P., Toivonen, H., Raiko, T., and Gionis, A.,
DopeLearning: A Computational Approach to Rap Lyrics Generation (May 18, 2015).
http://bit.ly/1CGSvVn

Alex Wright is a writer and researcher based in Brooklyn, NY.