When the best librarian is an AI

As the editorial page in issue 1/2020 announced, one of the goals we intend to work towards is the reinforcement of the journal's international attitude. The goal will be pursued via regular contacts with the professional community and by building on the correspondence with members of the International Advisory Board. David Weinberger, to our great pleasure, recently joined the Board. We asked him to offer the readers of *AIB studi* a reflection on a topical issue for the library profession. The theme is the role AI (Artificial Intelligence) and machine learning may play in the 'bibliographic recommendation' services, both internal and external to libraries.

Weinberger offers the point of view he gained both during his experience as co-director of the Harvard Library Innovation Lab and from the activity he carried out with Google's People + AI Research (PAIR). The latter was a multidisciplinary team exploring the human side of AI, starting from the idea that, in order to realize its full potential, machine learning must be participatory and involve the affected communities. Weinberger addresses the community of librarians and encourages them to take on a pro-active role in the use of AI when serving their communities.

AI – as any technology – is neutral with respect to the values and purposes for its use: for this reason, Weinberger deems that libraries should strive to counterbalance its purely business-oriented usage, and infuse into machine learning the democratic and humanitarian values that libraries uphold, and that, at the present time, demand renewed attention. In order to do this, librarians should not surrender in front of complexity; on the contrary, they should aim at increasing the interoperability of the tools they constructed over time and promoting the competencies they acquired. In this sense, it is increasingly critical that they reason from a system perspective and prevail over fragmentation which is probably the weak spot of libraries when confronting commercial entities.

A fascinating challenge which we hope also Italian librarians – in the present moment – will be willing to take on, for the future of our libraries and of the communities we serve.

Chiara Faggiolani, Anna Galluzzi

****

So far we've largely been largely able to shrug off the challenge AI poses to libraries by insisting that humans make better, more nuanced decisions about people and culture than any soulless machine can.

We have narrowed that claim some, for AI is showing itself to be helpful with many tasks where statistics are an important guide for action: generating lists of items that are candidates for removal from the shelves, perhaps predicting items that may see a surge in popularity, and definitely coming up with maximally efficient layouts for archive storage facilities.

Translated by Matilde Fontanin

But, when it comes to the sort of decisions that librarians make based on their hard-earned experience, wisdom, and their commitment to advancing their communities, some continue to insist there's a magic circle protecting librarians – a 'No machines!' zone.

That's understandable but risky. It seems that every day there is another announcement of the machines' success at tasks we used to consider to be purely the realm of human intellect, creativity, and emotion. We don't yet know how wide the magic circle will end up but it's likely to continue to shrink.

Some areas seem more secure from the spread of AI. For example, while the role of physical space is changing, libraries are likely to continue as irreplaceable safe, useful, and peaceful community places.

Similarly, libraries are unlikely to be challenged by AI as a source of free access to cultural materials. Challenges may come from other directions – for instance, copyright law might someday catch up with the way the internet has liberated culture – but that seems independent of AI.

Then there is the simple fact that for most communities, a library building stands as a sign and symbol of the community's commitment to culture, learning, and equal opportunity.

But AI is already posing a challenge to one of the most important virtues of libraries: a library taken as a system knows a huge amount about the items it tends for us and the communities it serves; a recent article in *Code4Lib* by Greg Sohanchyk and Dan Briem details ways of expanding that community knowledge. That's why users turn to libraries and librarians for recommendations of the next thing they should read, view, or listen to.

I'm going to guess that we're a long time away from being willing to chat with an AI as if it were a human librarian, enjoying the back and forth and feeling like we're getting to know someone. But when it comes to the transactional part of the conversation – the actual recommendation – AI already has significant advantages over humans. To begin with AI can know far more about books and individual users than could fit in any human being's head.

For example, Amazon knows the details about millions of books. It knows about the patterns of choices users make about each of those books. It knows about the intersecting networks of users' real and potential social connections. It knows the buying and clicking and reviewing behavior of its users, and not just when it comes to books. It knows which books users have downloaded a sample of onto their Kindles and which of those samples have led to a sale. It may well know which Kindle books users have started but not finished. In terms of the pure quantity of data, no librarian has ever known even a fraction of what Amazon knows about books and readers.

But, I can hear you objecting that Amazon is a ruthlessly capitalist enterprise that is interested only in selling its users something. Exactly right. That gives libraries their opportunity. And if libraries don't seize that opportunity, AI in the hand of commercial entities is the greatest threat to libraries since the physics of fire, for the usefulness of the AI provided by profit-seeking companies will continue to grow, leaving libraries in the dust.

Libraries have two things on their side.

First, the overall populace generally believes librarians are wiser about books than any machine can be. But, I'm afraid, that will not be enough for long. Not only can a machine learning system know about millions of books and other resources – including the entire long tail of works rarely read or recommended – commercial entities have strong incentives for training machine learning systems to match people's interests ever more accurately.

The second reason is more robust: libraries are on the side of their users and communities. Amazon doesn't care which book we buy so long as we buy one, so its algorithms make 'recommendations of least resistance'. If Amazon sees that we have been buying books about leftwing or rightwing politicians, about climate change or climate change denial, about one religion and not others, it will offer us more books as much like the ones we have been buying as Amazon can find.

That's because machine learning finds patterns, but librarians like to disrupt patterns. If you've just completed a series of romance novels set in 19th century England, your local librarian will let you know about a similar series, or other works by the same author. But the librarian might also mention a novel by Jane Austen or perhaps a nonfiction biography of a woman of that era. Amazon just wants you to buy a book, but librarians use their judgment to nudge open your interests. Being on the side of the user often means expanding the user's interests, opening up their world a bit.

That also means that libraries are on the side of open, tolerant, communities. Amazon, on the other hand, knows that it's easier to get us to buy a book that confirms our beliefs than one that challenges them.

So how can libraries help blunt a cultural environment in which the next cultural item you'll engage with is likely to be highly successfully targeted at you by AI employed by commercial entities that, unlike libraries, don't care about stretching people's horizons or creating a better informed, more open-minded, more sympathetic, more tolerant, and more compassionate community?

We could change our library infrastructure.

That sounds big and expensive, which taken to its maximum it would be. But it could also be done incrementally, with solid benefits for local libraries along the way.

Let me paint one maximalist picture, and then talk about some of the smaller steps along the way.

The most obvious solution to the problems posed by commercial entities is to build a non-commercial alternative. For the moment, let's call it the 'open library graph'; a graph is a densely linked representation of huge amounts of disparate data. Imagine a node for Dante's Inferno that has connections to all the books that refer to the Inferno, to scholarly books about Hell and Heaven, to information about the white guelphs, the black guelphs, the ghibellines, to Bosch's *Garden of earthly delights* and from there to modern surrealism, to the Christian Bible, the Jewish Bible, the Koran, the fate of Heavenworthy pre-Christians such as Socrates, then out from there to works of Western philosophy, Eastern philosophy – a linked concatenation of cultural works and references, ever expanding and ever growing richer. This would be a global resource of incalculable value. And it would draw upon global data already available: Europeana, Digital Public Library of America, many nation's national library systems, university bibliographic data, etc.

Libraries should build and 'own' this open resource.

And it should learn from what libraries know about their communities, while of course preserving the anonymity of individual users.

Here's an example. When I was co-directing the Harvard Library Innovation Lab, we computed a 'stackscore' for every work in the university's catalogue by apply a weighted formula that considered aggregated, anonymized data such as how often the work was checked out, how often it was put on reserve for coursework, and other such measures. The result was a number between 1 and 100 for every book that served as a crude but useful measure of the Harvard community's valuing of that resource (even these computations need some extra steps to ensure anonymity). Then we wrote a catalogue browser that ranked results by their stackscores, harnessing the community's use of materials as a guide to useful works.

One of the good things about stackscores is that every library could come up with its own formula for computing its community's values. Perhaps one library would count more heavily how often a book is requested, while another library would count the number of times a work is renewed, or weight reference librarian recommendations more heavily. Thus the measures of value can be local, while the stackscore – a number between 1 and 100 – allows for comparisons across libraries.

In fact, it's plausible to use machine learning to compute more accurate and useful stackscores. The analysis could be more detailed, and, importantly, it might be possible to compute predictive stackscores for items so deep in the long tail that they haven't been read in decades, thus getting more value from the library's collection.

Stackscore is just one idea. But this sort of idea has special value when compared with commercial rankings, for a few reasons.

First, it counters personal preferences with community preferences, treating local communities as especially valuable in their local-ness.

Second, any recommendation system based on data about the prior usage of materials is likely to create a feedback loop: if a book is recommended because of high usage, each recommendation will increase that usage. But if we had a shared infrastructure for library data, a community could routinely use other communities' stackscores to nudge people to works they might otherwise not have seen. For example: 'Here are the ten books on the topic of evolution that our community has most used and liked. But here are the top ten from a local university's biology department, and from its theology department, and here are the top choices from a community culturally quite different from ours'.

Computing and sharing stackscores is relatively easy. But all aspects of communities learning from one another, and ultimately perhaps contributing to something like an 'open library graph', would become much simpler if the software used to manage libraries provided easy ways for libraries to share information. The low end way to do this is to bring pressure to bear on the providers of Integrated Library Systems to make that sort of information available in standard formats and protocols. The high end way to do this is to at long last provide free or very low cost open source cloud-based ILS services that make the aggregated information about collections and behaviors available world-wide – while of course carefully preserving user privacy. That would be an amazing source of data for libraries and others to use for machine learning applications that support users and their local communities.

All of these specific suggestions are just examples. They may not be feasible or even desirable. But one way or another, libraries need to open themselves up to the ways that machine learning can support their users, and communities. Leaving this to commercial entities puts culture itself at risk. It is up to libraries large and small to humanize machine learning to ensure it serves the deeply democratic and humanitarian values of even the most humble local library.

David Weinberger

Ai fini della citazione, utilizzare esclusivamente il testo in lingua italiana, che presenta il DOI, la paginazione, l'abstract e gli altri dati ufficiali.

When citing, please always refer to the Italian translation only, complete with DOI, page numbers, abstract and other data.

[David Weinberger, *Quando il bibliotecario migliore è un' IA*.
AIB studi, vol. 60 n. 2 (maggio/agosto 2020), p. 213-217. DOI 10.2426/aibstudi-12478]