

All the News That's Fit for You

Personalized news promises to make daily journalism profitable again, but technical and cultural obstacles have slowed the industry's adoption of automated personalization.

IN 2006, a pair of entrepreneurs started a company called Outbrain with a promising premise: sifting through the massive volume of news available online to find what's most interesting to each reader. As Outbrain CEO Yaron Galai puts it, "You might have just read something that was a waste of time, and others feel the same way, so we could get some collaborative filtering going. From a reader's perspective, it's very tempting to solve this problem." But a year into product development, the Outbrain founders decided to change course because of a factor they hadn't appreciated: the sheer velocity of online news. "News goes away too quickly," says Galai, "and there's not enough time to collect deep-enough data to see who likes what and how it relates to other readers."

Delivering personalized news poses much harder problems than delivering personalized recommendations of books and movies as Amazon and Netflix do. Yet, despite the difficulties, personalized news seems all the rage these days. In February alone, *The New York Times*, *The Washington Post*, and Yahoo! all announced some form of automatic personalization, and Google is quietly running its own experiments in personalized news delivery.

These companies, along with several startups, see the revenue potential of giving each user more of what he or she wants. "I don't think it's rocket science to say that people who read the product more frequently and more thoroughly are more likely to subscribe to it than those who do not," says Marc Frons, the chief technology officer at *The New York Times*, which earlier this year started limiting the number of articles non-subscribers can read for free.

Joshua Benton, who directs Har-

The screenshot shows the New York Times homepage with a personalized news feed. The main headline is "The Times Introduces Digital Subscriptions." Below it, there are several article teasers, including "OSAMA BIN LADEN: Friedman: Will Bin Ladenism end?" and "Methods of Questioning Debated Again." A "MARKETS" section shows the Dow Jones and Nasdaq indices. A "RECOMMENDED FOR YOU" section lists five articles, including "Net Income Declines 8% at Chevron as Costs for Rare Metals Rise" and "Tensions Rise as U.S. Officials Press Pakistan for Answers." The page also features a "WHAT'S POPULAR NOW" section and a "Log in to see what your friends are sharing on mytimes.com" prompt.

The New York Times and other media companies are implementing personalized news delivery.

vard University's Nieman Journalism Lab (mission: "to help journalism figure out its future in an Internet age"), agrees that personalization offers enormous business potential. "*The New York Times* has well over a decade of data about what stories I've read, how many seconds I've spent on each story, and what sections I've read, so you would think they would be able to tailor my experience in a way that would be more pleasing to me," Benton says. "As a result, the page becomes a more valuable piece of property to an advertiser." Regardless of how the *Times'* payroll pans out, more advertising revenue would be particularly welcome in an industry whose sharply declining print circulations have led to decreases in ad sales and, in many cases, the death of entire newspapers.

But despite the promise of algorithmic personalization, the idea is far

simpler in theory than in practice, and newspapers have struggled to figure out how to do it without giving up their traditional role as arbiters of news.

"Computer scientists may think it's nirvana to get what you want to get," says Penelope Abernathy, professor of digital media economics at The University of North Carolina at Chapel Hill. "But a newspaper person will say, 'My role is to bring you the world, and it may be news you didn't know you needed to know.'"

Print journalists don't usually view themselves as mere news retailers but as crucial players in democratic society—and news personalization threatens to erode that function. As in the balkanization of culture in general—from niche books and movies to ideologically focused television news channels—personalized news risks doing the opposite of what newspapers have traditionally done: Instead of uniting people through a common discourse, personalization may make readers ever more insular.

A Hybrid Approach

The solution at *The New York Times* has been a hybrid approach. The site is supplementing its home page, with its standard mix of editor-selected content, with its recently introduced Recommendations page, which shows a ranked list of stories each logged-in user might find interesting based on his or her reading history.

Industry analyst Ken Doctor, a veteran journalist and the author of *News-economics*, says most newspapers' inertia stems in part from a lack of expertise. While editors focus on bringing readers the world, and the business side sells ad space around the articles, no one is left to tackle the considerable problem of delivering customized content.

"Newspaper companies aren't very tech-savvy," says Doctor, noting that

even their upper management usually has an operations background.

The New York Times, with its deeper resources, is showing itself to be an exception; but even so, its foray into automated personalization is still rudimentary. “We spend a lot of time and effort tagging content,” says Frons, referring to the system of semantic tags that tell the recommendations engine what each story is about. The site keeps track of your reading habits at the rather coarse level of these tags—noting, for example, that my most-read topics include “Demonstrations, Protests, and Riots” as well as “Television” and “Search Engines”—and makes recommendations accordingly. Apart from ignoring stories you’ve clicked on inadvertently, the *Times* recommendations engine doesn’t consider how long you’ve spent on each story or how much you liked a piece.

Furthermore, since the *Times* does not use collaborative filtering, it has no way of discovering your latent interests, since these aren’t revealed in the semantic tags. What’s more, because the engine looks only at articles, and not at your search history, it can’t suggest stories on topics you care about that the *Times* hasn’t previously covered. However, Frons says the challenge is less about the underlying technology and more about delivering the kind of user experience that readers expect from the paper of record. “I think *The New York Times* is probably unique among many publishers and content providers online in that people who come to us do so not only for the individual articles but also for our news judgment—for what the *Times* thinks are the most important stories of the day.”

The *Times* is not entirely special in this regard, though. Neil Thurman, senior lecturer in the graduate school of journalism at City University London, heard a version of Frons’ sentiment from many of the editors he surveyed for his recent study about the use of personalization at 11 national news sites in the U.S. and the U.K. “The editors said to me that they felt users had, to an extent, personalized their news experience by choosing what newspaper to buy or which site to visit,” he says. Automated recommendations can, depending upon how they are

implemented, sometimes break down the news organization’s editorial voice, as exemplified by an incident on the right-wing *Telegraph* Web site, when the site linked out from its pages on the environment to several liberal blogs and Web sites. “Those decisions about linking were run algorithmically, so you got some surprising links being generated,” Thurman says.

The most established companies have the most to fear from unwelcome surprises—and even those with state-of-the-art technology, like Yahoo!, are proceeding cautiously. “The footprint of what we do is very broad” [more 100 million daily visitors on the Yahoo! home page alone], “so we want to take careful steps,” says Raghu Ramakrishnan, the company’s chief scientist for search and cloud computing.

The technical obstacles are monumental—from the scalability challenges of combing through terabytes of daily click logs on thousands of servers worldwide to the difficulty of learning from nearly real-time feedback. The Yahoo! CORE algorithm, which is at the heart of the matchmaking process between content and users’ interests, starts with the system’s prior knowledge about users, but then proceeds to learn more about how each user will respond to a particular article by using explore-and-exploit strategies on a slice of the user base.

“What to exploit is easy,” says Ramakrishnan, “but what to explore is where the secret sauce comes in.”

The CORE algorithm must simultaneously optimize for a complex mix of variables, figuring out the right trade-off between click-through rate and engagement, between getting a broad-enough sample of users and not wasting these users’ attentional bandwidth, and between different types of ads. Thanks to CORE, says Ramakrishnan, the click-through rate on articles is already up more than 160% compared to what it was based on human editorial judgment alone. Despite this feat of algorithmic personalization, Ramakrishnan is mindful of the need for human judgment, which can help avert machine-learning problems like overfitting. “If I keep showing you what I think you want to see,” explains Ramakrishnan, “there’s a danger that I might get very narrow in what I show you.”

Yahoo! gives editors “dashboards,” decision-support tools to help them make sense of the ever-changing analytics and select particular stories for particular groups of users. As for individual personalization, Yahoo! has yet to begin testing it on either its home page or its news page, starting instead with lower-profile properties like its sports page. “We’re being cautious because we don’t want to compromise our users’ trust,” says Ramakrishnan. Google, too, is still at the experiment stage, showing some users news customized to their clicking history.

It’s too soon to tell if personalization will save the journalism business, but it will certainly provide a reality check for journalists. Sree Sreenivasan, digital media professor at Columbia University’s Graduate School of Journalism, recalls a time not long ago when, as a reporter, you could say that a million people read your story because your newspaper had a million circulation. “You fooled yourself,” Sreenivasan says. ■

Further Reading

Agarwal, D., Chen, B.-C., and Elango, P. Explore/exploit schemes for Web content optimization, *Proceedings of the 2009 9th IEEE International Conference on Data Mining*, Washington, D.C., Oct. 6–9, 2009.

Kamba, T., Bharat, K., and Albers, M.C. The Krakatoa chronicle—an interactive, personalized newspaper on the Web, *Proceedings of the 4th International World Wide Web Conference*, New York, NY, Dec. 11–14, 1995.

Li, L., Chu, W., Langford, J., and Schapire, R.E. A contextual-bandit approach to personalized news article recommendation, *Proceedings of the 19th International Conference on World Wide Web*, New York, NY, April 26–30, 2010.

Liu, J., Dolan, P., and Rønby Pedersen, E. Personalized news recommendation based on click behavior, *Proceeding of the 14th International Conference on Intelligent User Interfaces*, New York, NY, Feb. 7–10, 2010.

Thurman, N.J. Making “the daily me”: Technology, economics and habit in the mainstream assimilation of personalized news, *Journalism: Theory, Practice & Criticism* 12, 4, May 2011.

Based in San Francisco, **Marina Krakovsky** is the co-author of *Secrets of the Moneylab: How Behavioral Economics Can Improve Your Business*.

© 2011 ACM 0001-0782/11/06 \$10.00