

May 1, 2003

## **Making Intelligence a Bit Less Artificial**

By LISA GUERNSEY

WHEN Mike Kellogg of Chicago shops at Amazon.com , he can view a list of recommended products that match his taste in music with uncanny precision. The store knows he loves Wilco, a folksy rock band, and offers him CD's with a similar sound.

When Anne Heilemann of Iowa City visits Amazon, she is immediately reminded how clueless the recommendations can be. A few months ago she bought a book about flower girls for a friend's 7-year-old daughter, who will be in her wedding this month. Now Amazon offers her more flower-girl books (one is more than enough, Ms. Heilemann said), and assumes that she must also need a book about bargains on baby products.

"I would freak if I needed a baby-bargains book right now," she said.

Mr. Kellogg's experience is a dream come true for developers of recommender systems - software that analyzes patterns in a customer's choices to predict what else that person might want or need. In addition to Amazon's system, the better-known examples include TiVo , the digital television recorder, and NetFlix, the online DVD rental service.

But Ms. Heilemann's experience may ring truer to most people. Only 7.4 percent of online consumers who noticed these systems said they often purchased recommended products, according to a report issued in February by Forrester Research . About 22 percent said they found the recommendations valuable, and about 42 percent said the products listed were not of interest.

To improve the recommendations, many software developers are doing an about-face from the mid-1990's, when they put their energy into getting computers to do all the work. Today they say that automated programs that look for patterns in customer data are not smart enough to detect a gaffe. Something more sophisticated is required: the human mind.

People are becoming a critical component: analysts who understand why a particular type of music appeals to some people, categorization experts who know how to cross-reference material, retail executives who tweak the system to improve the bottom line and reviewers who check for nonsensical or offensive results.

"The holy grail is to be able to capture all the customer's interactions in detail and get smarter about what not to recommend," said Usama Fayyad, chief executive and president of digiMine, the software company behind the online recommendations of J. Crew and Barnes & Noble.

"We can recommend very well. Knowing when not to bother someone is much harder."

Odd pitches and poor matches have led to an outpouring of anecdotes. A discussion on a

bulletin board at Salon.com this year titled "When Customer Profiling Goes Wrong" described people's befuddlement upon receiving off-the-wall recommendations from Amazon. Someone named Molly wrote that she bought "a single trashy romance novel" and is now "branded for life."

"The best results are achieved from powerful technology and human intervention," said Matt Turck, president of TripleHop Technologies, a company that has built recommendation engines for USA Today's online travel section and SkiMatcher, which advises travelers on ski resorts.

All this talk of human intervention sounds very different from the hype of the dot-com boom, when startup companies spun visions of computer programs that could help people discover their yet-to-be-revealed tastes in books and music. Imaginative online software was more coveted than gold, and recommendation systems looked like a step toward the creation of artificial intelligence. It was hard not to be intrigued by the idea that with the right data and the right mathematical formula, a computer might be able to grasp a person's preferences better than friends and family, suggesting books or movies that the consumer would not have discovered otherwise.

"There is a sense that with people you are perceived stereotypically, but that the system might give you a totally different chance," said Rashmi Sinha, the founder of Uzanto Consulting, a company that focuses on end-user experiences with technology. She is also a cognitive psychologist who has conducted studies of how people respond to recommender systems.

In the mid-1990's, much attention was drawn to collaborative filtering, a technique that matches a user to a group of others who have purchased or praised similar products, then analyzes the group's data to predict what else the user might like. Patti Maes, an expert in the field and a professor at the Media Lab of the Massachusetts Institute of Technology, called it "automating word-of-mouth." Firefly, a company she helped start, became the symbol of the technology's promise. The New York Times Magazine ran a 4,200-word article in 1997 about Firefly, which was then valued at \$100 million, Microsoft bought the company the following year.

It turned out that Microsoft hadn't bought Firefly for its collaborative filter. It wanted the software that kept track of user profiles, which Firefly had called "passport," for what became Microsoft's own Passport software for the quick transfer of personal data.

"They weren't so much interested in the recommendation engine," Dr. Maes said in an interview last month. "It wasn't because they didn't believe in it, but it wasn't as good a match for their strategy."

Collaborative filtering is only a piece of today's recommendation technology. "There was this great expectation that it was going to be this killer app, and it didn't meet people's expectations," said Jack Aaronson, who helped to design the technology for a recommendation company called Open Sesame and who now runs the Aaronson Group, a consulting practice.

Some of the problems with collaborative filtering are common enough that they have earned nicknames, like the cold start problem and the popularity effect.

To make interesting matches, a company needs a large number of people who have rated or purchased a large number of products. The cold start phenomenon arises when a Web site opens but neither of those criteria has been met. Joe Smith might buy the same book as Jane Doe, but without more data, it would be a stretch to predict that the next book Joe buys is one that Jane might want.

To counter that, some companies employ human editors to make the first connections between products and likely purchase patterns. At Barnesandnoble.com, an editorial staff makes recommendations. Choicestream, the company that creates the MyBestBets technology used by America Online, has brought in analysts to distill the defining attributes of television programs ("thought-provoking" is one example) and uses the computer to match them with other programs that have been similarly categorized.

"If it is not vetted and monitored by humans and not complemented by actual hand-selling, as we say in the book industry, it doesn't feel like there is anybody there," said Daniel Blackman, vice president for books, video and music for Barnesandnoble .com.

The popularity effect is at work when results delivered by the computer are boring and obvious. MediaUnbound, the company that develops the recommendation engine for the MP3 service PressPlay, has been analyzing the more than four million MP3 collections that are open for browsing through Napster, KaZaA and other file-sharing services. Michael S. Papish, the company's chief executive, said that if he were to run that data through a collaborative filter to predict musical taste, one band would be at the top of the list for every single person: the Beatles.

Because of the Beatles' name recognition and popularity, they are likely to be in anyone's collection, regardless of their taste. But if MediaUnbound were to put the Beatles at the top of every recommendation list, their service would seem stale and uninspired.

To address that problem, the company built some new rules into the software. But it also hired music analysts to scout the music scene for new bands, seed the databases with interesting acts and build genre maps to show how musical tastes are connected.

"The computer is good at averaging things; it rounds it out, sands it down," Mr. Papish said. "And then we use the humans to bring back the exciting rough edges."

Colin Wambsgans, an analyst at MediaUnbound, said he was working last week on a recommendation list generated for consumers who like the Flatlanders, a 1970's band from Texas that recently released a new album. He moved Squirrel Bait, a rock band with a harder sound, down a few notches and pushed another band, the Palace Brothers, up. "The style of music wasn't quite the same," he said. "The Palace Brothers have more of a folksy sound."

Even Mr. Kellogg's experience with helpful taste-matching at Amazon was a case of a human's coming to the rescue. To make sure that he gets recommendations that follow his tastes, Mr. Kellogg, 29, a product manager in Chicago, has spent a good deal of time editing his online profile. He wants Amazon to know that he loves Bob Dylan, so he has given high ratings to those CD's and dozens of other favorites. When he makes a purchase for someone else - like a

karaoke CD for his 7-year-old cousin - he checks a box that tells Amazon to ignore it.

"I actively work at my recommendations," Mr. Kellogg said.

Jason Kilar, vice president of worldwide software operations for Amazon, would not comment on how many people use the editing feature; Christopher M. Kelley, an analyst at Forrester, said he figured that most people never take the time to use it. But it has become an important tool for taking into account the complexity of human interactions. Mr. Kilar said that customers need to be able to communicate, for example, "Please exclude this one because I was temporarily out of my mind and I thought I liked Sade."

Business sense is the latest layer to be added to recommendation technology. Some companies want to be able to weight the results that appear on a recommendation list so that products they want to clear out appear at the top and out-of-stock items are suppressed. "The technology needs to be able to support this," said Dr. Fayyad, whose digiMine software offers such options. But Dr. Maes and some software developers warn that if companies allow the bottom line to dictate their recommendations, shoppers may distrust their systems.

Amazon says it is holding out against such tweaking. "We let the recommendation engine do its magic," Mr. Kilar said. "We are extremely pure."

That is giving Ms. Heilemann, a 29-year-old development director for a university in Iowa, a good laugh. "Am I missing something in my life?" she asked. "Maybe I really do need this booster bundle thing for my computer or this Wiggles 'Yummy Yummy' DVD," a video for the toddler set.

"But they do think that I need a total body yoga workout," she added, "and they might be right about that."

HEN Mike Kellogg of Chicago shops at Amazon.com, he can view a list of recommended products that match his taste in music with uncanny precision. The store knows he loves Wilco, a folksy rock band, and offers him CD's with a similar sound.

When Anne Heilemann of Iowa City visits Amazon, she is immediately reminded how clueless the recommendations can be. A few months ago she bought a book about flower girls for a friend's 7-year-old daughter, who will be in her wedding this month. Now Amazon offers her more flower-girl books (one is more than enough, Ms. Heilemann said), and assumes that she must also need a book about bargains on baby products.

"I would freak if I needed a baby-bargains book right now," she said.

Mr. Kellogg's experience is a dream come true for developers of recommender systems - software that analyzes patterns in a customer's choices to predict what else that person might want or need. In addition to Amazon's system, the better-known examples include TiVo, the digital television recorder, and NetFlix, the online DVD rental service.

But Ms. Heilemann's experience may ring truer to most people. Only 7.4 percent of online consumers who noticed these systems said they often purchased recommended products,

according to a report issued in February by Forrester Research . About 22 percent said they found the recommendations valuable, and about 42 percent said the products listed were not of interest.

To improve the recommendations, many software developers are doing an about-face from the mid-1990's, when they put their energy into getting computers to do all the work. Today they say that automated programs that look for patterns in customer data are not smart enough to detect a gaffe. Something more sophisticated is required: the human mind.

People are becoming a critical component: analysts who understand why a particular type of music appeals to some people, categorization experts who know how to cross-reference material, retail executives who tweak the system to improve the bottom line and reviewers who check for nonsensical or offensive results.

"The holy grail is to be able to capture all the customer's interactions in detail and get smarter about what not to recommend," said Usama Fayyad, chief executive and president of digiMine, the software company behind the online recommendations of J. Crew and Barnes & Noble.

"We can recommend very well. Knowing when not to bother someone is much harder."

Odd pitches and poor matches have led to an outpouring of anecdotes. A discussion on a bulletin board at Salon.com this year titled "When Customer Profiling Goes Wrong" described people's befuddlement upon receiving off-the-wall recommendations from Amazon. Someone named Molly wrote that she bought "a single trashy romance novel" and is now "branded for life."

"The best results are achieved from powerful technology and human intervention," said Matt Turck, president of TripleHop Technologies, a company that has built recommendation engines for USA Today's online travel section and SkiMatcher, which advises travelers on ski resorts.

All this talk of human intervention sounds very different from the hype of the dot-com boom, when startup companies spun visions of computer programs that could help people discover their yet-to-be-revealed tastes in books and music. Imaginative online software was more coveted than gold, and recommendation systems looked like a step toward the creation of artificial intelligence. It was hard not to be intrigued by the idea that with the right data and the right mathematical formula, a computer might be able to grasp a person's preferences better than friends and family, suggesting books or movies that the consumer would not have discovered otherwise.

"There is a sense that with people you are perceived stereotypically, but that the system might give you a totally different chance," said Rashmi Sinha, the founder of Uzanto Consulting, a company that focuses on end-user experiences with technology. She is also a cognitive psychologist who has conducted studies of how people respond to recommender systems.

In the mid-1990's, much attention was drawn to collaborative filtering, a technique that matches a user to a group of others who have purchased or praised similar products, then analyzes the group's data to predict what else the user might like. Patti Maes, an expert in the field and a professor at the Media Lab of the Massachusetts Institute of Technology, called it

"automating word-of-mouth." Firefly, a company she helped start, became the symbol of the technology's promise. The New York Times Magazine ran a 4,200-word article in 1997 about Firefly, which was then valued at \$100 million, Microsoft bought the company the following year.

It turned out that Microsoft hadn't bought Firefly for its collaborative filter. It wanted the software that kept track of user profiles, which Firefly had called "passport," for what became Microsoft's own Passport software for the quick transfer of personal data.

"They weren't so much interested in the recommendation engine," Dr. Maes said in an interview last month. "It wasn't because they didn't believe in it, but it wasn't as good a match for their strategy."

Collaborative filtering is only a piece of today's recommendation technology. "There was this great expectation that it was going to be this killer app, and it didn't meet people's expectations," said Jack Aaronson, who helped to design the technology for a recommendation company called Open Sesame and who now runs the Aaronson Group, a consulting practice.

Some of the problems with collaborative filtering are common enough that they have earned nicknames, like the cold start problem and the popularity effect.

To make interesting matches, a company needs a large number of people who have rated or purchased a large number of products. The cold start phenomenon arises when a Web site opens but neither of those criteria has been met. Joe Smith might buy the same book as Jane Doe, but without more data, it would be a stretch to predict that the next book Joe buys is one that Jane might want.

To counter that, some companies employ human editors to make the first connections between products and likely purchase patterns. At Barnesandnoble.com, an editorial staff makes recommendations. Choicestream, the company that creates the MyBestBets technology used by America Online, has brought in analysts to distill the defining attributes of television programs ("thought-provoking" is one example) and uses the computer to match them with other programs that have been similarly categorized.

"If it is not vetted and monitored by humans and not complemented by actual hand-selling, as we say in the book industry, it doesn't feel like there is anybody there," said Daniel Blackman, vice president for books, video and music for Barnesandnoble .com.

The popularity effect is at work when results delivered by the computer are boring and obvious. MediaUnbound, the company that develops the recommendation engine for the MP3 service PressPlay, has been analyzing the more than four million MP3 collections that are open for browsing through Napster, KaZaA and other file-sharing services. Michael S. Papish, the company's chief executive, said that if he were to run that data through a collaborative filter to predict musical taste, one band would be at the top of the list for every single person: the Beatles.

Because of the Beatles' name recognition and popularity, they are likely to be in anyone's

collection, regardless of their taste. But if MediaUnbound were to put the Beatles at the top of every recommendation list, their service would seem stale and uninspired.

To address that problem, the company built some new rules into the software. But it also hired music analysts to scout the music scene for new bands, seed the databases with interesting acts and build genre maps to show how musical tastes are connected.

"The computer is good at averaging things; it rounds it out, sands it down," Mr. Papish said. "And then we use the humans to bring back the exciting rough edges."

Colin Wambsgans, an analyst at MediaUnbound, said he was working last week on a recommendation list generated for consumers who like the Flatlanders, a 1970's band from Texas that recently released a new album. He moved Squirrel Bait, a rock band with a harder sound, down a few notches and pushed another band, the Palace Brothers, up. "The style of music wasn't quite the same," he said. "The Palace Brothers have more of a folksy sound."

Even Mr. Kellogg's experience with helpful taste-matching at Amazon was a case of a human's coming to the rescue. To make sure that he gets recommendations that follow his tastes, Mr. Kellogg, 29, a product manager in Chicago, has spent a good deal of time editing his online profile. He wants Amazon to know that he loves Bob Dylan, so he has given high ratings to those CD's and dozens of other favorites. When he makes a purchase for someone else - like a karaoke CD for his 7-year-old cousin - he checks a box that tells Amazon to ignore it.

"I actively work at my recommendations," Mr. Kellogg said.

Jason Kilar, vice president of worldwide software operations for Amazon, would not comment on how many people use the editing feature; Christopher M. Kelley, an analyst at Forrester, said he figured that most people never take the time to use it. But it has become an important tool for taking into account the complexity of human interactions. Mr. Kilar said that customers need to be able to communicate, for example, "Please exclude this one because I was temporarily out of my mind and I thought I liked Sade."

Business sense is the latest layer to be added to recommendation technology. Some companies want to be able to weight the results that appear on a recommendation list so that products they want to clear out appear at the top and out-of-stock items are suppressed. "The technology needs to be able to support this," said Dr. Fayyad, whose digiMine software offers such options. But Dr. Maes and some software developers warn that if companies allow the bottom line to dictate their recommendations, shoppers may distrust their systems.

Amazon says it is holding out against such tweaking. "We let the recommendation engine do its magic," Mr. Kilar said. "We are extremely pure."

That is giving Ms. Heilemann, a 29-year-old development director for a university in Iowa, a good laugh. "Am I missing something in my life?" she asked. "Maybe I really do need this booster bundle thing for my computer or this Wiggles 'Yummy Yummy' DVD," a video for the toddler set.

"But they do think that I need a total body yoga workout," she added, "and they might be right about that."



