SPONSORED BY:

MOBILE

BIGDATA REPUBLIC | Transform Your Business With Data

ABOUT US REGISTER LOGIN

HOME BLOGGERS MESSAGES POLLS WEBINARS RESOURCES VIDEOS | STRATEGY ANALYTICS TECHNOLOGY FB TW LI G+ RSS

FINANCIAL SERVICES HEALTHCARE GOVERNMENT EDUCATION RETAIL

BIG DATA STRATEGY

Case Study: Amazon Mechanical Turk

Robert Plant, Associate Professor, School of Business Administration, University of Miami
9/11/2013
Comment
10 comments

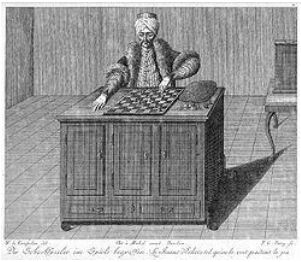
Login
50% 50%

Amazon's reach is vast and continues to grow. Since the company's beginnings in book sales, it has proliferated to offer seemingly endless categories of consumer goods.

Behind the scenes, it has developed a complex ecosystem of technologies to manage these activities: Initiatives like Amazon Web Service's Elastic Cloud Compute environment; Kiva Systems (the robotic warehouse management solution); and Liquavita, Samsung's color display unit acquired in May 2013 for potential deployment in its e-readers to power its business model.

What data isn't showing us

With all of this technology, big data, and computing resources, it's interesting to see that Jeff Bezos and his team recognize that there remains a boundary between data, systems, and human knowledge. This acknowledgement can be found in the Amazon Mechanical Turk project, or MTurk for short.



An etching of the Mechanical Turk, via Wikipedia.

The Mechanical Turk name originates from an 18th century chess playing automaton, in essence, a desk behind which sat a mannequin, dressed in traditional Turkish attire, and which played skilled chess games against a human adversary drawn from a theater audience. The "Turk" was created by Hungarian inventor Wolfgang von Kempelen in 1769 and toured with him in Europe until his death in 1804, at which point it was acquired by Johann Nepomuk Mälzel, who toured the United States. The machine was in fact controlled remotely using rods by a chess master. This hoax was revealed in several essays at the time, including one by Edgar Allan Poe, in what could be considered an early Turing Test.

The modern Mechanical Turk

The aim of MTurk is to provide an online, crowdsourcing marketplace to enable users, known as "requesters," to obtain answers to a wide variety of queries. These queries are ones that Amazon customers would be unable to perform on the Amazon site; would normally require some expertise in AI, machine learning, big data, and programming; or are just so open-ended and unstructured as to be impossible to solve without conjecture and contextualization.

These questions are posed on MTurk as "Human Intelligence Tasks," and human responders, known as "Providers," give answers, usually for a small amount of compensation. Example requests include:

- Given a Facebook user, enter their profile information and find the corresponding Twitter handle. Reward $0.05, Time allotment: 5 Mins.
- Verify address, phone, website details for restaurants through online research. Reward $0.15 Time Allotment: 10 Mins.
- Transcribe a short audio. Reward $0.05 Time Allotment: 3 Mins.
- Compare two results of a product search side by side and judge which one is better. Reward $0.00 Time Allotment: 10 Mins.

Some of the questions require that the Provider is qualified in certain areas. For example:

- "Locate and format data from the web"
- "Identify music mood" (Korean)
- "What is the best way to complete these three paragraphs?"

But why not a knowledge-based system?

MTurk raises several interesting questions pertaining to the system itself, big data, and AI. First, why can't simple questions such as "provide the best choice between three sentences" be answered on Amazon through a knowledge-based system, or a neural network?

MORE BLOGS FROM ROBERT PLANT

A Review: Big Data in 2013
39 comments
Data scientists, NSA, C-suite buy in... Big data changed a lot in the last 12 months.

What Would Seymour Cray Think of Big Data?
13 comments
How would the tech legend respond to the 4 Vs?

The 360 Degree View of a Customer
56 comments
Big data is helping companies like Netflix see what you really want (and even when you go to sleep).

Lessons for Big Data From President Obama's Healthcare Implementation
24 comments
Three big data lessons for CEOs from HealthCare.gov's bumpy start.

Is Big Data Ruining Sports?
29 comments
From sailing to the NBA -- big data could increase the gap between sporting haves and have-nots (while ruining the magic in the process).

More from Robert Plant

FLASH POLL

All Polls

BDR IN YOUR INBOX

Enter email SUBMIT

FEATURED VIDEO

Big Data Explained: What Is Supervised Learning? 4

A gentle video intro to supervised learning.
Watch This Video

More Video Blogs

FOLLOW US ON TWITTER

Computational linguistics could be used to scan through Amazon's literary collection and perhaps relate the sentences to the frequency of their use and the rankings of the books in which the sentences are contained, or if no exact match was made, their nearest neighbor. A similar outcome may be possible for the data surrounding the Korean mood music -- the genre could potentially be discerned from a system performing pattern recognition against a music library in which artists have provided meta-data for their own music.

More sophisticated methods are used by firms such as Pandora, in their Music Genome Project, or Shazam, which uses a dataset of 9 billion tags on 30 million tracks in 33 languages to identify tunes.

Clearly, for many questions posed, the person designing the survey is, in fact, not using any scientific experimental design principals, but, if they are asked with just the intent to obtain quick prototype feedback, then it doesn't demerit them from their utility.

However, I'm still puzzled at the site's very existence. Helping people make about a dollar an hour is noble, but somewhat circumspect. Perhaps the true reason is that this is an outward facing lens that Amazon uses to look upon the world, and the truth is that even with their huge amounts of internal big data, Amazon can't answer the bigger questions Jeff Bezos and the team may ask -- for example, "what don't we know?" and "what is missing from our data?"

As such, Amazon may actually use this site to look for new ideas, product offerings, and innovation it can then leverage. After all, as they used to say on the TV show the *X-Files*, "the truth is out there," you just got to know the right questions to ask.

Related posts:

- Case Study: Examining Airbnb
- Case Study: Fresh & Not So Easy
- Big Data Case Study: Tesco

Email This Print Comment

COMMENTS

Newest First | Oldest First | Threaded View

Pradeepta Mishra, User Rank: Exabyte Executive
9/29/2013 | 1:10:21 PM

Re: Not sure about this one.
@Daniel, Amazon's product recommendation engine is based on the algorithm called market basket analysis. It is association of a product with other similar products based on the past purchase behavior of the customers. On top of that they have done some profiling based on the demographic features of a visitor and then based on the algorithm the engine suggests the products of which a customer might be of interest. Sometimes it is very relevant and sometimes it is not even half correct. It is very subjective and sometimes based on the person's mood.

Login
50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
9/16/2013 | 4:29:22 AM

Re: Not sure about this one.
Plenty of room for improvement there @Daniel - I think there's something they can do, studying the way we look up objects (basically a behavioural approach) which can determine the way we shop for ourselves, or research a product for an article (a particular problem for me!) or look for a one off gift which can determine what is suggested in future.

Something along the lines of financial institutions noticing that someone logging into their account and immediatly withdrawing a truck load of money is a highly fraudulent way of behaving, whereas real customers tend to wander at leisure, checking balances, looking at previous transactions before making any concrete decisions.

Login
50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Daniel Gutierrez, User Rank: Blogger
9/13/2013 | 6:48:04 PM

Re: Not sure about this one.
@Saul, yes that's correct. Taking the Amazon recommender system as an example, when you login to your account and you are presented with books Amazon "thinks" you might like it is based on a probabilistic model that tries to predict what you might like based on what other shoppers who have similar interests as you have bought in the past. It is not an exact science. In fact, when I see these recommendation in my Amazon account, they're wrong about 1/2 the time. But it beats what was done years ago, just arbitrarily display products that have no basis in your preferences.

Login
50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
9/13/2013 | 4:30:16 AM

Re: Not sure about this one.
So is there something to be said here for getting a consensus on 85% of personal opinion? Especially in Amazon's retail case, where they might want to sell more into the personal tastes of people?

Login
50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Daniel Gutierrez, User Rank: Blogger
9/12/2013 | 12:57:30 PM

Re: Not sure about this one.
@Saul, yes you're right, adding elements to the classification equation can only help make accurate predictions. That being said, we're talking about probabilistic results here, so a 85% likelihood is pretty good.

Login
50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
9/12/2013 | 6:04:37 AM

Re: Mturk to train classifiers
Eeep, I didn't see this comment first @Daniel, fascinating use. That combination of experience/opinion and raw data - so many possibilities!

Login

Big Data Republic@BigDataRepublic 15 Apr
AllJoyn: A Common Language For Internet Of Things - Open-source lingua franca aims to let IoT devices and services... ow.ly/2FHbY7
Retweeted by Yakamoto
Expand

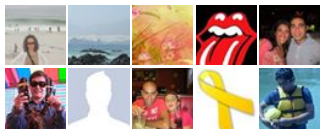
Hajime Ozaki@hajozaki 22h
BDR Top 20 #BigData David Smith David Smith List of top tweeters on Big Data from @BigDataRepublic wp.me/p4Aavi-2
Show Summary

Marcus Borba@marcusborba 28 Apr
#BigData #Analytics #BusinessIntelligence #FS @informationweek @AllAnalytics @BigDataRepublic @BigDataNetwork @BigDataClub

LIKE US ON FACEBOOK

Big Data Republic
Like

4,168 people like Big Data Republic.



Facebook social plugin

ACCOLADES

Media Pioneer Award Winner

50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Saul Sherry, User Rank: Blogger
9/12/2013 | 6:03:19 AM

Re: Not sure about this one.
@Daniel, useful as in could combine this relatively subjective data with a wider source and create a semi-supervised learning project?

Login

50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Daniel Gutierrez, User Rank: Blogger
9/11/2013 | 7:50:15 PM

Re: Not sure about this one.
Actually, the way it works with Mechanical Turk is to ask multiple human participants the same question to obtain a consensus on what the answer is, but often the response is very subjective like which of these looks better or is more functional, etc. But if you can determine that 85% of the respondents feel the same way, that knowledge is useful.

Login

50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

Daniel Gutierrez, User Rank: Blogger
9/11/2013 | 7:43:47 PM

Mturk to train classifiers
In my opinion, speaking from the point of view of a data scientist, Mechanical Turk's most useful function is to help train classifier algorithms using supervised machine learning techniques. For example, once I developed a classifier using the logistic regression algorithm. The feature-set had to do with the fashion industry and the algorithm was supposed to predict whether a specific style would do well during a fashion market week. We used Mturk to process a training set by having human classifiers make the assignments. Used in this way, Mturk is a fantastic resource for big data.

Login

50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

legalcio, User Rank: Exabyte Executive
9/11/2013 | 7:28:42 PM

Not sure about this one.
Amazon's assumption is that someone out there in crowdsouce land actually does know the answer to a specific query. And the difference between Amazon MT and Yelp, or Orbitz or God knows how many other blog sites, social networking sites, etc., is.....? Big Data may not have all the answers, but a lot of Big Data analysis and value is based on the query. Ask the right one and the answer very likely is out there.

Login

50% 50%

Reply | Post Message | Edit/Delete | Messages List | Start a Board

BIG DATA REPUBLIC

ABOUT US | CONTACT US | HELP | REGISTER
TWITTER | FACEBOOK | LINKED IN | GOOGLE+ | RSS

in partnership with InformationWeek UBM Tech powered by UBM DEUSM

bigdatarepublic : /bigdatarepublic/section/2635 : /bigdatarepublic/section/2635/267601

UBM Tech

UBM TECH

OUR MARKETS: **Business Technology** | **Electronics** | **Game & App Development**

Working With Us: Advertising Contacts | Event Calendar | Tech Marketing Solutions | Corporate Site | Contact Us / Feedback

Terms of Service | Privacy Statement | Copyright © 2014 UBM Tech, All rights reserved