Inside IBM Research

Kush Varshney I Volunteers and algorithms deliver funds

**TOPIC:** How mobile technology is helping make cash transfers to Kenya and Uganda’s poor.

**INTERVIEWEE:** Kush Varshney, Research Staff Member, Mathematical Sciences and Analytics Department, IBM Research, and Data Ambassador, DataKind

**INTRODUCTION:**

Long-distance targeting: The words conjure up images of missiles and drones. But a collaboration involving IBM Research, a data science community called DataKind™, an NGO called GiveDirectly and a “data journalist” funded by the Knight Foundation got together in New York to deliver cash directly to extremely poor villagers in Kenya and Uganda. Armed only with computers, Google satellite images and an algorithm, the team ultimately facilitated the distribution of cash payments via mobile phone — and gave people a way to better their lives.

In this episode of Inside IBM Research, IBM’s Kush Varshney talks about the work he did with data journalist Brian Abelson, a crew of social innovators and twenty idealistic volunteers to target a problem half-way around the world.

**PRESENTATION:**

So, in this part of the world, most people are very poor to begin with. But the people who are a bit more poor are the ones who tend to live under roofs that are made out of thatched materials. In comparison, the other roofing material is metal, tin sort of things. This is exactly the same criteria that GiveDirectly uses for their unconditional cash transfers. Within a village, once it’s been selected, they’ll send a staff member there, and it’s only people who live under thatched roofs that receive these unconditional cash transfers.

The challenge was put forth to us, “How can a data scientist or an electrical engineer help a charitable organization do this village selection problem?” And then our own brainstorming led to a bunch of different things that we could put together into one unified solution.

The first step was locating where the houses were, and the roofs, regardless of the type. So, we were using satellite imagery, and if you were to give a satellite image to a human, they would be able to fairly easily tell the difference between a thatched roof and a metal roof. So, we trained the machine to do basically what a human would do as well. We used a method called template matching, which comes up in image processing. You start off with a small piece of an image that you think looks like a roof, and you scan it over the entire image. And wherever there’s a strong match to what’s underneath from what you thought the template looked like, there’ll be a strong
response. So, that will allow us to identify where these roofs are.

And then the second stage is classifying the roofs being either thatched or metal. And the color distribution of the pixels is a very strong indicator of that. It turns out that the thatched roofs are very green and brown, whereas the metal roofs are white. So, it’s fairly easy to distinguish them.

So, those were the two sets of features that we had within an image patch. Then we had to train this learning algorithm. So, in order to obtain training data, we built a crowdsourcing platform. And we got volunteers to label houses for us, who would click on the location of the houses as well as say whether it was thatched or iron. And then we built a random forest regression model to take the features from the template matching and the color histograms into account and determine the proportion of thatched roofs in a particular image.

The DataKind folks advertised this, and within an hour or so, like twenty volunteers signed up and wanted to do this clicking on images to label houses.

It was Brian who built this interactive web tool, which was very, very intuitive and easy to use. So, it would serve up an image to a person on a web page. And they would just have to click. So, Brian set things up so that multiple people could log in at the same time and could click on houses and so on. And he let them go for a weekend. And when he returned after that weekend was done, he saw that his system had crashed. And he was concerned, like, what had happened. And the one thing that he didn’t include in his code was a check for if all of the images had been labeled. It turned out that all these volunteers were so enthusiastic that they had finished labeling the entire set of images over a weekend.

This kind of illustrates in a data science project that there are all sorts of technologies that you have to put together, starting from the basic data acquisition, data cleaning things, all the way to the final machine learning sort of algorithms.

In February of this year [2014], GiveDirectly was running their largest campaign to date, and so they were dispersing around $4 million. For that campaign, the fifty villages they went into were based on our ranking from the algorithm. So, yeah, I mean, $4 million to a large number of people was the result of our algorithm.

The whole point of this GiveDirectly movement is that people should be allowed to do whatever is in their best interests with that money. And that’s actually been proven through randomized control trials that it helps alleviate poverty in a better way than if a particular choice is already made for the person receiving the money.

In terms of the algorithms or the methods that we used, they were nothing ground-breaking in any sense from the research perspective. I wouldn’t have thought that you could even use satellite images to characterize poverty in a particular region of the world, and then use it to support an unconditional cash transfer program. But putting everything together into one unified system that solves a very real-world problem and really does help people come out of poverty is a very interesting and novel thing. So,
yeah, it’s been a really great experience.

OUTRO:
We weren’t the only ones to find Kush’s work with DataKind and GiveDirectly interesting and novel. In August 2014, his research paper — “Targeting Direct Cash Transfers to the Extremely Poor” — received the Best Social Good Paper Award at the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. You can find a link to the paper on Kush’s IBM Research personal page.

You’ve been listening to Inside IBM Research. I’m your host, Barbara Finkelstein. Our music is “Happy Alley” by Kevin MacLeod and is licensed under Creative Commons Attribution 3.0. Share this episode with colleagues and friends — and keep an ear out for our next episode.”

Listen to “Volunteers and algorithms deliver funds” on Soundcloud:

https://soundcloud.com/ibmresearch/volunteers-and-algorithms-deliver-funds