

Module 1 - An Overview of Algorithmic News Media

Hello and welcome to week 1 module 1. As a reminder before I get into today's content, please be sure to watch the other videos this week as well as to read the three articles that I posted on the syllabus.

Those include an introductory chapter called "The era of news algorithms" as well as another piece on what machine learning can do and a piece about computational story discovery.

Please also see the discussion questions I've posted. Those relate to content that we'll be covering in the videos and finally after you've had a look at all that, don't forget to take the quiz this week.

All right! So, today I want to give you a broad overview of how algorithmic approaches are being used throughout journalism. I'll talk about what algorithms are good at, what people are good at (with respect to algorithms) and I'll talk about how to blend algorithms and people together in productive workflows.

So, here I want to give you a sort of broad overview of some different scenarios where algorithms automation and A.I. are used in news production. I'll show a lot of specific examples of these scenarios later this week and in the following weeks, I just want to give you an overview right now.

One of the more promising approaches is to use algorithms for data mining to discover new and valuable knowledge from data and documents sets. Data Mining can be used to enhance business operations like optimizing subscribers to a website as well as in a number of editorial scenarios.

So things like story discovery, monitoring social media, predicting things like elections which you may have seen on the New York Times at 538, as well as helping to support activities like fact checking or content moderation. Algorithms can also be used to automatically produce content or automatically generate content like writing text articles or producing videos or data visualizations.

There are algorithms that can help summarize content, take a large article and crunch it down into a shorter representation and algorithms can be used to drive news bots and automatically produce content that then gets shared on social media channels.

Algorithms can also be used to aid in news distribution, so things like headline testing help to optimize for a particular headline which increases the click through rate of an article and certainly algorithms can help to personalize and recommend content to enhance the relevance and engagement of content on a website.

All right! So just to step back for a minute, I want to talk again a bit about what an algorithm is. An algorithm is a series of steps that's undertaken in order to solve a particular problem or to accomplish a defined outcome. This recipe here is an algorithm although it's an algorithm that's usually done by a person rather than a computer. Mostly we're interested here in algorithms that run on digital computers because that allows them to scale up and sometimes to automate entire processes.

So, what is it exactly that news algorithms do? There are five main ways that algorithms can be used in information processing tasks related to news. These include things like calculation, prioritization, classification, association and filtering, and sometimes these operations are chained together in order to accomplish more complex things.

For instance summarizing an article may involve filtering away some unimportant sentences and then prioritizing the remaining ones for some kind of output summary. I want to talk a little bit more in detail about each of these five things.

So, in terms of calculation some information tasks involve calculations of more or less non-controversial mathematical equations. These are things with demonstrably correct answers like let's say scanning in a document and converting it into a machine readable text. This process is called "Optical Character Recognition" or "OCR" and you can see an example of that in the slide here.

Many tasks though don't have an easily demonstrably correct answer, but instead involve a lot of subjective judgment and I want to talk about some of those more subjective things that algorithms do next. So algorithms can prioritize information and prioritization decisions are really important in news because of the limits of human attention and information overload. So it's really helpful to be able to find the most important thing, so that you can show it to someone you know at the top of the screen and draw their attention to it. We're all familiar with things like you know search algorithms which prioritize results with respect to some query that we've put.

Classification decisions are about deciding which category something belongs to. This can be really useful in content production because it helps in finding information that's classified in a certain way that you might be looking for. Like let's say you're looking for a photo that contains a picture of a particular politician. If it's categorized with the name of the politician that would make searching for it very easy. This example is from a New York Times project called "who the hell" and what it does is it recognizes the faces of Congress people and images, and then it labels those faces with the name of the politician as well as how certain the algorithm is and if that's the face of that particular politician.

Association decisions are about creating connections between different entities. So, for instance in an investigation like the Panama Papers you could potentially use an algorithm to find associations or relationships between different entities like officers of shell companies and you can imagine how that would be useful for developing and finding new threads in an investigation.

Finally algorithms can be used for filtering information. Filtering again is very important because of information overload issues. Deciding whether to show or hide something can be very powerful. The Facebook news feed is a perfect example of a filtering algorithm. It determines of all the posts that it could show you, it determines which of those it's actually going to show you when you log in on the website or in the app.

So again news algorithms they calculate, they prioritize, they classify, they associate, they filter. That's what algorithms are good at. What are humans good at? And what do they do better than algorithms?

Well, I would say there's three main things. People are very good complex communicators. Complex communication is about listening, it's about negotiating, explaining, interpreting. It's also about adapting content to different channels, putting information in context, it's about social intelligence. Reporting collecting information from different sources can involve undertaking really difficult interviews with sources that are unmotivated to share information or maybe sources that are deceptive or antagonistic in their interactions. We still need humans doing reporting in order to get information out of other human sources.

Expert thinking is another area where humans still excel with respect to algorithms. Expert thinking is about in-depth problem solving using domain knowledge and you can imagine in investigative journalism often involves this type of expert thinking. An investigation might include analyzing documents, data and other sources information for relationships and associations that might not even be known ahead of time whether or not those relationships are significant or not, and the significance of those relationships might only become clear through the interpretation of an expert with very deep domain knowledge in a particular area like say finance.

Expert thinking also released to this idea of many cognition. So, knowing when to step back and knowing when a particular strategy might need to be switched up. Algorithms don't have any of that sort of ability to take a step back and consider the bigger picture, but people really excel at that.

And then finally I think humans still have an edge in terms of their ability to be dynamic in the world. So algorithms are really good at repetitive work and routine work, but they're basically not able to flex to a changing world and the world is always changing.

Algorithms will need to adapt if they're if they're ever going to keep up with the world, but people are good are good at being flexible for these types of non routine events and I would also say that people have an edge in terms of creativity, originality and being able to develop novel interpretations of what's going on in the world.

Two themes that I'd like you to come away with today are the importance of human values and algorithms and also this idea of blending or hybridization between algorithmic and human effort.

So this idea of human values relates to all of the different subjective decisions that algorithms make. So prioritizing, classifying, associating, filtering, there's all kinds of subjectivity is embedded in those operations, and that means that there are human subjectivities that get built into algorithms automation and A.I. by the people that design and build these systems.

That means that editorial decisions, subjective editorial decisions are sort of everywhere in the code that constitutes these systems. I think that raises an interesting question which is: Whose values do you want running your news organization? Whose values do you want embedded in your news algorithms?

The second theme relates to this idea of blending or hybridization of effort between algorithms and people. Basically the issue is that algorithms are good at some things, people are good at a different set of things, and in order to get the best of both worlds we need to figure out ways to blend algorithms and people together so that they complement each other, so that we get more than the sum of their parts.

This raises an interesting question I think, in terms of how journalistic practices will change or will need to change as our rhythms and A.I. are blended into news routines, and I hope some of you taking this course will help invent these new blended news routines.

You'll see these two themes of human values and hybridization again and again in this course. They'll come up in slightly different ways when we talk about computational story discovery, content automation and news distribution.

That's it for now. In the next video I'll get into more detail about computational story discovery tools and data mining and then I'll talk about our computational thinking and the final video this week. Thanks.