

How do we solve the problem of Fake News?

Dr. Bhavani Thuraisingham

January 3, 2017

Welcome to my first motivational article for 2017 and what better topic to write about than Fake News. We have been plagued with fake news stories especially during the latter half of 2016 relating to the US 2016 Elections as well as other events. The “Pizzagate” scandal brought the fake news topic into the forefront of the news stories. As someone who has conducted research and education work in cyber security and data management/analytics, I would like to discuss how we can possibly address aspects of fake news.

First of all fake news is nothing new. Story telling (real or make believe) has been around for centuries. Fairy tales are among the best fake news. We have read stories of the Handsome and Rich Prince falling in love with the Beautiful Princess; they get married and live happily ever after. While we feel better after reading such wonderful stories they are not good for our long term health as we expect our marriages to be “happily ever after” and real life is far from perfect. So this is an example of fake news that is detrimental to our happiness and well-being.

Many of us have come across colleagues or even friends that make up stories about their lives. These are people likely with low self-esteem who have to essentially lie to make them feel good. Some of us who listen to these fake stories get envious and that will have a bad effect on us.

Another example is social media such as LinkedIn. Do we really trust what everyone says about themselves? I know of people who have likely been laid off from their companies (or are no longer with their companies) but give the impression they are still working there. If you complain to LinkedIn, my understanding is that LinkedIn will inform the person posting false information but might also send the email address of the complainer. The guilty person could always say, I forgot to change my employment history and then possibly sue the complainer. Furthermore, recruiters may select the person who has entered false information instead of an honest person.

Then of course there are the fake news web sites that maliciously spread blatant lies about others. This information will get propagated by Facebook and Twitter and within seconds the entire world knows about the fake stories which seem too good to be fake. By the time the fake news sites are detected (and in many cases they are undetected), it is too late. Malicious stories have been spread about the person such as “Hillary Clinton is a child sex trafficker”.

One could ask the question, so what is the difference between such fake news web sites and Tabloid newspapers that often publish fake news. I believe the offense is the same in both cases, except that not all of us read Tabloid newspapers and we read the headlines in supermarket stores (e.g., the Queen is dying and the throne will go to Prince William and not to Prince Charles”) and quite often we don’t believe such stories. They have a reputation of publishing false information. Now, if the New York Times published such stories we will of course believe them. The problem is, there is no way today to distinguish between good and bad sites.

So, now the question is what are the solutions? Several techniques have to work together to provide viable solutions. I discuss a few of them next.

Recommender Systems: One of the key technologies is recommender systems. That is, we need the capability to have recommender systems that will go through each and every web site and post

recommendations about these sites and assign a score. For example, a score of 10 could mean site is extremely reliable and 0 means is completely unreliable. A score of 5 could mean one can trust the site 50% of the time. The next question is how does the recommender system know whether the web sites are reliable or not. They use data management and analytics/mining techniques. Here are a few of such techniques.

Data Provenance Management: For each piece of information on the web, we need techniques to determine where did the information come from? Suppose John posts some news on Twitter. He got this information from web site A which got its information from web site B. Now web site B's recommender score is 2. Therefore John's news is likely not reliable and could be fake. Therefore, we need reasoning techniques to manage the provenance of the data.

Data Integrity Management: Integrity checks have to be placed at all levels. These checks may use provenance management techniques. In a Database Management 101 course we learn about database integrity. That is, we learn techniques on how integrity constraints may be enforced. Similarly we need to develop ways to check the integrity of the data posted on the web.

Data Analytics/Mining: Regardless of whether we are developing recommender systems or data provenance management systems, data analytics plays a major role. We need techniques to analyze the vast amounts of data on the web and be able to classify the news as to whether they are say great, good, moderately trustworthy, or bad. Data analytics techniques will include classification, making associations, and detecting anomalies.

Data Security Management: CIA (Confidentiality, Integrity, Availability) are the three tenets of cyber security. Integrity means the data is accurate, trustworthy and has not been maliciously modified. We need techniques to check whether the data has been maliciously corrupted. That is, has the web site been attacked? If so, can we attribute the attack to a specific group of hackers?

Data Governance: We need techniques to understand the risks as well as enforce appropriate policies. That is, every person who creates a web site, or posts something on Facebook or Twitter has to be governed by a set of policies. He or she must certify that the information he/she posts is accurate to the best of his/her knowledge and understands the consequences of posting false information.

The Law: I wish for once we do not have to deal with lawyers, but unfortunately we have no choice. That is, we need appropriate laws to ensure that those to spread fake news receive some form of punishment depending on the crime. For example, if a web site publishes fake news that could drive a person to commit suicide (which has happened in many cyber bullying situations), and then the punishment could be several years in jail. If someone has posted false information on say LinkedIn, then perhaps a fine or public shaming for lying should suffice. Regardless of what the offense is, the person has to be accountable for what he or she does.

In conclusion, this article has stressed that spreading fake news is not a laughing matter. It could have serious and often unintended consequences. Therefore, we need not only technologies to detect fake news, but also laws to ensure that the offenders are punished depending on the severity of the crime. Essentially what we need is a collection of Big Data Management, Analytics and Security techniques combined with legal aspects to address the problem of fake news. Whether we can solve the problem of fake news remains to be seen. But at least we should make every effort to limit the damage caused by the problem.