Junk Information Detector

Donghao Lin(lindo), Qibang Liu(liuqib)

January 15, 2018

1 Problem

With the increasing demand of internet, more and more junk information are sent by many bad or boring people that are impacting our life annoyingly. Developing a App that can detect whether the information is junk becomes an important problem.

2 Evidence

In the article Fake Information On The Internet. It All Goes Down To How We Perceive Them, the author Hafiz Rahman S has stated When they spread, they can influence a chain reaction that will dominate the headlines for weeks. Not just fake information, fabricated information that went viral can also put trouble in the growing trend of media contents.Those fake information simply bring darkness to an already dark hole. It's dimming the lights to create a much wider problem. While the truth is out there, less people are seeking for them. Why? Because they're aren't as interesting as the fake ones. Top fake stories can receive more engagement than the "true" top news stories. This is human behavior. And on the internet that has become more social, this behavior can be easily manipulated. Junk Information can sometime be critical and serious to certain user that are not very familiar with computer and internet scam. Like for my sometime I browse a gaming website it has those little window flowing in the web page advertising some game that looks fantasy but the real purpose of that window is to lure user to click into a scam website and those people that post the window there can plant some virus on the victim's computer or maybe something worse.

3   Difference

However, it is also very important to detect junk information correctly, because sometime the computer could get it wrong and make some trouble. Like gmail, sometime my classmate send me a link or a invitation to get collaborate in a project that mail usually gets into spam, at first when I dont know how this work it took me sometime and tries to get it right, so making it work correctly is also important. Those detector the companies use now like the one gmail is using just block keywords or some vicious accounts by using filters that they collect data from users. But it has low success rate that detect junk information right, as a result, some useful information is block by mistake and some junk information is not blocked. So, we have a smarter approach that is to use filter that not only collect data from user, but also use cloud computing to analyse the content and accounts. Some keywords are actually not junk in some sentences, therefore, we need to analyse the word in a sentence but not just detect the separate word and block all the content with that word. Then we create a database that have billions of word combinations and will add more combination that are not junk during running, which means it has ability of self-learning.

4   Challenge

The challenge in developing this junk information detector is to create a huge database and let it learn new word combinations itself. And some users may not share their datas to the server, so it is hard to collect data. To minimize the risk, we can use some data exist now, and add combinations as most as we found to the database.

5   Refference

Fake Information On The Internet. It All Goes Down To How We Perceive Them, Hafiz Rahman S, https://www.eyerys.com/articles/people/1359179061/opinions/fake-information-internet-it-all-goes-down-how-we-perceive-them