The Evolution of Security in Personal Computing

Technological advancements are happening at record speed and have been for the past several decades. According to UNCTAD’s 2018 report on technology and innovation this is in part due to the fact that more technological advancements bring … you guessed it; more technological advancements[[1]](#footnote-1). At the same time, Moor’s law[[2]](#footnote-2) states that “computing power” doubles roughly every second year, culminating in a positive feedback loop, an ever-increasing upward avalanche of bits and bytes.

The personal computer was introduced to the masses in the early 1990’s[[3]](#footnote-3) and the share of personal data we put online has only continued to grow since then (I don’t actually know this, but I have a pretty good hunch). This brings up the issue of how to best protect our data and what safety rules to follow when we “travel online”. And while some still think their birthday is a perfectly fine password for anything from social media accounts to web banking, many are beginning to see the necessity of properly protecting their data.

## Case: Security Now Podcast

The podcast “Security Now” (SN) has been running continuously since 2005. Its two hosts, Steve Gibson and Leo Laporte, in their own words “*spend somewhat shy of two hours each week to discuss important issues of personal computer security. Sometimes we'll discuss something that just happened. Sometimes we'll talk about long-standing problems, concerns, or solutions. Either way, every week we endeavor to produce something interesting and important for every personal computer user.*”

## Data

Full transcripts of ~736 SN-episodes are available for download as .txt files through the website <https://www.grc.com/securitynow.htm>. I have started creating a python script that can download them automatically, which is going well, but still needs a little more work. (If all else fails I could always put on a few episodes of FRIENDS and click-download each file manually (which would probably be faster in the end)). (But I won’t though) (…but Netflix…) (But no seriously, I’ll make the script work) (for reals).

## Analysis

Since the landscape of personal computing has so drastically changed in the decade since the SN podcast started, my plan is to perform **topic modelling** on the transcripts to infer which topic changes have occurred over time. Jacobi, van Atteveldt and Welbers did roughly this in 2015, when they analyzed the coverage of nuclear technology in The New York Times (Jacobi et al., 2015).

It could also be interesting to look at the timing of specific events/introduction of new terms (one might expect these two “experts” to be ahead of the curve on novel topics compared to mainstream media).

I have previous experience with using the “stm” package for topic modelling in R which uses Latent Dirichlet Allocation (LDA) to generate topics based on the co-occurrence of words in a corpus. This time, in order to learn something new, I would like to use python (e.g. **word2vec**). LDA is not the only method for topic modelling, though. Subramani, Sridhar and Shetty (“it’s spelled with an e!”) proposed a neural network approach, which is apparently more effective when you have larger datasets (Subramani et al., 2018). The *lda2vec* model, similarly, looks intriguing … (Moody, 2016).

First task would be to look more into these different methods to determine which is more appropriate for the dataset.

## References

Jacobi, C., Van Atteveldt, W., & Welbers, K. (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, *4*(1), 89-106.

Moody, C. E. (2016). Mixing dirichlet topic models and word embeddings to make lda2vec. *arXiv preprint arXiv:1605.02019*.

Subramani, S., Sridhar, V., & Shetty, K. (2018, November). A Novel Approach of Neural Topic Modelling for Document Clustering. In *2018 IEEE Symposium Series on Computational Intelligence (SSCI)* (pp. 2169-2173). IEEE.

1. <https://twitter.com/unctad/status/996450139261857792> [↑](#footnote-ref-1)
2. <https://en.wikipedia.org/wiki/Moore%27s_law> [↑](#footnote-ref-2)
3. <https://en.wikipedia.org/wiki/Personal_computer> [↑](#footnote-ref-3)