

# Machine learning for distinguishing fake news

In recent years there has been a dramatic increase of fake news in the media. This article takes a machine learning approach, trying to tackle the problem.

A main problem for social media networks trying to deal with fake news is the pure amount of articles being published and shared every day. It is practically impossible for social networks like Facebook or Twitter to manually go through every article and sort out the fake ones. Thankfully recent years advances in artificial intelligence and machine learning makes it possible to automate the process of determining if a news article is trustworthy or not.

The process of automatically classifying a text into two or more classes is called text classification. In this case the classes will be “fake” and “real”. In this study a method based around “hierarchical attention networks” is used. The idea behind a “hierarchical attention network” is that the information in the article is contained not only in the words but in the context. So instead of counting the occurrences of words like simpler methods do, each word is instead put into context of the surrounding words. The same is done for sentences. Each sentence is treated in the context of its surrounding sentences. In this way the network learns what information is important at each level. That means that it can distinguish what words are important for a sentence and what sentences are important for the text.

For example, fake news would usually be about politics and rarely about catastrophes. So the word “earthquake” might be important in determining if the article is fake or not. In the same way a sentence describing the current weather conditions of a city might not be relevant on its own as it could be part of a fake article about how a political candidate uses taxpayer’s money to travel but could just as likely be part of a travel guide.

In addition to applying this method to fake news the study has examined different ways of “training” the model. Training is the process of feeding the model with examples, allowing it to improve how well it performs. The best result that could be reached on the “test set” (a set of data, not part of the training process) was 96% accuracy. A result this good could of course decrease the amount of manual work that would be required for a company to sort out the fake news in its feeds. At the very least it could be used to make a rough filter that could decrease the amount of time required to rout out fake news from our news feed.