**Building the New York Times recommendation engine**

The purpose of this article is how we can incorporate new recommendation techniques to give suggestions users about a topic or genre. Their previous algorithm recommends articles like those that have already been read, which was naïve and content-based filtering. Their next algorithms talk about collaborative filtering which states that if one reader’s preferences are very similar to another reader’s, articles that the first reader reads might interest the second and vice-versa. However, the engineers at New York times then incorporated a more robust technique by collaborating the 2 algorithms in which they model each reader based on their topic preference and then recommend them articles based on how closely their topics match with the other person. The engineers basically divided the challenge into 3 parts: 1) How to model an article based on its text: for this they used a Latent Dirichlet Allocation (LDA) which is a content modelling algorithm and looks at the body of each article and from it learns a mixture of topics. LDA was quick and accurate for their purposes as the topics in LDA tend to be very broad and allows them to relate pieces from different viewpoints. 2) How to update the model based on audience reading patterns: LDA uses words as inputs but words are often ambiguous and they corrected this by adding offsets to model topic error, their algorithm also incorporated reading patterns on top of content modelling, they used CTM and Collaborative Poisson Factorization to calculate the offset. 3) How to describe readers based on their reading history: The methods used for adjusting article topics also calculate reader preference, but they can’t scale to all users, so they averaged together the topic of all articles. By modelling article content and reader preference with topics they have reconceptualized their recommendation engine.

**The spread of true and false news online**

The purpose of this paper is analyzing the spread of fake news and rumors on social media. To understand this, it is necessary to examine diffusion after differentiating true and false stories and controlling for the topical and stylistic differences between the categories themselves. The author defines cascades as: if rumor “A” is tweeted by 10 people separately but not retweeted it has 10 cascades but if it is independently tweeted by 2 people and each of those 2 tweets is retweeted 100 times then the rumor consists of 2 cascades of size 100 each. Their results proved that a greater fraction of false rumors experienced between 1 to 1000 cascades whereas greater fraction of true rumors experienced more than 1000 cascades. They also proved that false political news traveled deeper in the network and more broadly and reached more people then legitimate news. The authors used a Latent Dirichlet Allocation topic model with 200 topics and trained on 10 million English-language tweets. This generated a probability distribution over the 200 topics for each tweet in their dataset, then they measured how novel the information in the true false rumors was by comparing the topic distribution of rumor tweets with the tweets the user was exposed to in the last 60 days. Their study resulted that these false rumors were significantly more novel than the truth across all novelty metrics. The good thing about this is that the authors not only found out the true false news based on topic modelling but developed a metric of how exposed the users were to false news. They verified their results by 3 students and averaged a 90% accuracy. The author states that research needs to be done on behavioral explanation of differences in the diffusion of true and false news.