Team Name: Social Corps
Team Members: Akshay Jagadeesh, Dior Xiong, Jee Su Byun, Zepeng Huo

Analyzing Corporate Social Responsibility Reports with Modeling Public Mood and Emotion Phenomena

What is the need? Who wants or benefits?

Companies are increasingly ramping up their focus on social responsibility, whether it is women's rights, protecting the environment. Embracing socially responsible policies is essential to a company's long-term success. Periodically, companies in the U.S release a CSR (corporate social responsibility) report on how well they do with regards to sustainability and environmental, or rather social good in general.

In our project, we will analyze CSR reports of different companies to quantize each company's contribution to society. This will let all the potential customers for these companies aware of how they are integrating the principles of sustainable development into their every day operations. It will also benefit companies themselves that are indeed found to be socially responsible, as customers will gladly pay a premium for goods, knowing that part of the profits will be channeled towards social causes near and dear to them.

What data (or datasets)?

CSR(Corporate Social Responsibility) reports

We examine the extent to which environmental sustainability has become embedded in corporate policy and core business discourse by analyzing CSR Reports of a sample of Fortune 500 companies. We will first attempt to find preprocessed dataset of CSR reports online and then either crawl or download complete CSR reports for the remaining companies in PDF format from the corresponding official corporate websites. We will also attempt to create a good balance of different types of industries so that the algorithm can have a more balanced approach.

Twitter

On the basis of a large corpus of public Twitter posts we also look specifically at the interplay between corporate social activities and the public's attitude toward the companies.

What is your "data mining" toolkit? You should list specific methods you will implement.

We will use latent Dirichlet allocation (LDA), a generative statistical model to find the topics in the corpus we gather, and will categorize them into sub-groups. Each group will represent a broadly-defined topic, so that we can make semantically meaningful topic labels to each sub-group and see how each company's CSR report match the topics.

We then extract the social media's public reactions towards those company's reports from Twitter. We perform a sentiment analysis of tweets filtered by the keywords '#company_name'

and ‘#social_keyword’ in the same time period as the years CSR reports (explained below) are published. We then extract 6 emotional mood states (tension, depression, anger, vigor, fatigue, confusion) [2] to compute a 6-dimensional mood vector representing each company.

Preliminary sketch of what you hope to find.

We hope to see if these companies are actually doing what they say are doing. A good way to measure this would be to look at the perceptions of the general public, which will be represented by Twitter. We will look at the overall Twitter sentiment value toward a company and see if it is positive or negative. If it is positive, we will say that a company is upholding its ethical image. However, if it is negative, we would assume the opposite.

Another thing we hope to find is how these CSR reports change over time. There is generally, one CSR report released a year, so we want to look at how a company’s CSR report is different from the previous years. For example, we would look at McDonald’s CSR report from 2008 to 2018 and see how their values have changed over time.

Reference

[1] Tremblay, M. C., Parra, C., & Castellanos, A. (2015). Analyzing Corporate Social Responsibility Reports Using Unsupervised and Supervised Text Data Mining. New Horizons in Design Science: Broadening the Research Agenda Lecture Notes in Computer Science, 439–446. doi: 10.1007/978-3-319-18714-3_36

[2] Bollen, Johan, Huina Mao, and Alberto Pepe. "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena." Fifth International AAAI Conference on Weblogs and Social Media. 2011.