

Trend Analysis Using Probabilistic Topic Modeling

Evaluating ESG Trends in Earnings Call Data

Anton Strähle*

June 2021

Abstract

This thesis examines how topic modeling, specifically probabilistic topic modeling, can be used in order to gauge trends in textual data. We introduce two probabilistic topic models, LDA and DMM, and present two types of approximative inference for both models as well as methods of hyperparameter estimation. Furthermore we also introduce methods of guiding the models in the direction of uncovering topics relating to the trends that we are interested in.

The two models are applied to different corpora consisting of questions from the Q & A sessions of financial earnings calls. We focus on how ESG related trends can be examined using these methods, hence we guide the models towards uncovering such topics. We find that DMM models are not applicable to the data examined as the assumptions made are not fulfilled to such an extent as initially thought. LDA models on the other hand, when applied to specific subsets of the data, yield quite promising results.

When further nudging the LDA models in the direction of uncovering specific ESG topics we find that we are able to uncover topics where we have some prior knowledge of the topics existence within the corpus. The ability to uncover topics in such a manner indicates that the framework established in the thesis is usable in practice but that it requires some prior knowledge regarding the topics one wants to uncover.

Using the trained models we also construct two metrics that can be used to gauge the trends of these topics when they are evaluated on a test corpus that spans a longer period of time. We also discuss the drawbacks of the framework introduced as well as some ways in which it can be improved.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden.
E-mail: anton.t.strahle@gmail.com. Supervisor: Taras Bodnar.