

Weighted coherence as topic models' interpretability measure

Friday, 17 May 2024 18:15 (12 minutes)

Topic modeling is very useful for analyzing text data. It can be used to analyze large collection of text data such as articles, reviews, social media, and others. This helps in clusterization documents by topic, extracting keywords, and identifying patterns in the data. There are a lot of automatically calculated criteria of informativeness of thematic models. One of these criteria is coherence. But the problem with coherence is that it does not take into account most of the text in the calculation, which makes evaluating the quality of the topic by this criteria unreliable. The aim is to propose a new method for calculating coherence that takes into account the distribution of the topic throughout the text.

Primary authors: ZHGUTOV, Kirill; VORONTSOV, Konstantin; ALEKSEEV, Vasily

Presenter: ZHGUTOV, Kirill

Session Classification: 17 Computer & Data Science

Track Classification: Computer & Data Science