WE1S's use of MALLET diagnostics to assess the quality of topic models.

Topic modeling is a leading method of machine learning that discovers "topics" in texts by analyzing the statistical co-occurrence of words (see card M-2). It finds out which words tend to come up together in a document set (and in individual texts) when people discuss something or, as in newspapers, many things. Co-occurring words suggest "topics." Topic modeling is an "unsupervised" machine-learning method, meaning that it does not require pre-training of the computer on a gold standard of texts that humans previously analyzed. The specific kind of topic modeling that WE1S uses is Latent Dirichlet Allocation (LDA) as implemented in the MALLET Machine Learning for Language Toolkit.

Topic models are sensitive to a number of parameters, particularly the number of topics the user chooses to generate. They are also sensitive to peculiarities of the data supplied, such as variability in the length of texts and the occurrence of concentrations of rare words in some texts. As a result different models produced from the same texts can be more or less useful for human interpretation. Some topics may consist of "junk" (numbers, abbreviations, or typographical errors), or less meaningful content such as overlay generic words. Models and topics of this nature are considered less interpretable and therefore less useful for understanding the texts from which they are generated. Many attempts have been made to identify statistical methods for identifying "quality" models and topics. MALLET produces a "diagnostics" file in XML format, which compiles many

statistical metrics that can be useful in assessing the quality of models and topics, and for identifying significant and aberrant patterns in them. WE1S has adapted MALLET's diagnostic visualization tool into its own MALLET Diagnostics Visualization Tool to make examining MALLET's diagnostic data easier.

Research has shown that statistical measures of topic quality in particular tend to differ from human intuitions (Chang *et al*.). To address this, WE1S has developed an interpretation protocol (see M-3) that places human assessment of topic models at the center of its methodology. The statistical metrics generated by MALLET's diagnostics can nevertheless be useful as means of looking at the model from different perspectives that privilege particular criteria such as topic coherence or document length. Used in conjunction with the WE1S interpretation protocol, MALLET diagnostics can help us gain a fuller understanding of topic models.

Resources

Interpretation & Interpretability research: WE1S bibliography

Recommended article:
Chang, Jonathan, et al. "Reading Tea Leaves: How Humans Interpret Topic Models" (2009), doi: 10.5555/2984093.2984126.

Tools:
MALLET Diagnostics Tool