Evaluation Metrics

Perplexity:

In information theory, perplexity is a measurement of how well a probability distribution or probability model predicts a sample. It may be used to compare probability models. A low perplexity indicates the probability distribution is good at predicting the sample.

Perplexity on 200 articles:

LDA: 1703.8

PAM: 1509.42

Coherence:

Coherence Measures

1. ***C_umass*** is based on document co-occurrence counts, a one-preceding segmentation and a logarithmic conditional probability as confirmation measure

The UMass metric defines the score to be based on document co-occurrence:

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)}$$

where $D(x, y)$ counts the number of documents containing words x and y and $D(x)$ counts the number of documents containing x. Significantly, the UMass metric computes these counts over the *original corpus* used to train the topic models, rather than an external corpus. This metric is more intrinsic in nature. It attempts to confirm that the models learned data known to be in the corpus.

Model	Topics	U_{MASS}
Good Model	('system', 'user', 'eps', 'human', 'interface') ('graph', 'trees', 'minors', 'survey', 'time')	-14.7
Bad Model	('computer', 'system', 'user', 'trees', 'graph') ('system', 'graph', 'trees', 'user', 'eps')	-14.7

4.5. Choosing the Best Coherence Score

There is no one way to determine whether the coherence score is good or bad. The score and its value depend on the data that it's calculated from. For instance, in one case, the score of 0.5 might be good enough but in another case not acceptable. The only rule is that we want to maximize this score.

Usually, the coherence score will increase with the increase in the number of topics. This increase will become smaller as the number of topics gets higher. The trade-off between the number of topics and coherence score can be achieved using the so-called elbow technique. The method implies plotting coherence score as a function of the number of topics. We use the elbow of the curve to select the number of topics.

-8.5744 for u_mass.

High is better for "c_v" score and low is better for "u_mass". That is it. – Eduardo Freitas Apr 30 at 19:09