|  |  |  |
| --- | --- | --- |
| **Method** | **Characteristics** | **Limitations** |
| Correlated Topic Model(CTM) | Uses a normal logistic distribution to create relations among topicsAllows the occurrences of words in other topics and topic graphs | Requires lots of calculationResults in lots of general words inside the topics |
| Dirichlet Multinomial Regression (DMR) | Uses Gibbs samplingProvides inferences about hidden variables | Tendency to underestimate abundant features and overestimate marginal featuresResults in a larger number of variables to sample and a more complicated sampling distribution (low efficiency) |
| Hierarchical LDA (HLDA) | Discovers topics within a corpus in a hierarchical wayPuts abstract terms at the base of the hierarchyLocates detailed and specific terms near the leaves of the hierarchy | Ignores the lexical co-occurrencePoor consideration for word dependenciesThe more the hierarchy levels, the slower the performance (long execution time) |
| Latent Dirichlet Allocation (LDA) | Requires manual removal of stopwords Previous studies have found that representation of the relationships among topics is out of LDA's scope | Inability to model relations among topicsThe number of topics (K) must be knownFailure in the face of a large number of vocabularies |
| Pseudo-document based Topic Model (PTM) | Analyses topics without using auxiliary contextual informationAssumes each short text relates to only a single pseudo documentAvoids overfitting when the training corpus is in a relative shortage | Cannot be applied directly to raw input data (needs to use some heuristic methods to enrich the input data)Lacks a quantitative relationship among wordsCan generate some high-frequency but topic-irrelevant wordsCannot deal with extremely sparse and noisy data |
| Supervised Latent Dirichlet Allocation (sLDA) | Assigns a label on each training document (in distinction from the LDA model)Offers improved predictions over regressions on words aloneApplicable, besides text, on social networks image classification | Requires marking documents with a response variableCannot be used for multi-class classification problemsLabour-intensive and expensive to apply it on a large dataset due to the labelling process |