**Evaluation of Topic Modelling by Topic coherence**

* Topic modelling provides us with methods to organize, understand and summarize large collections of textual information.
* There are many techniques that are used to obtain topic models - Latent Dirichlet Allocation
* LDA is a widely used topic modelling technique to extract topic from the textual data.
* Topic models learn topics—typically represented as sets of important words—automatically from unlabelled documents in an unsupervised way

**Disadvantage of topic models:**

* **No guaranty on the interpretability of their output.**
* Therefore, coherence measures have been proposed to distinguish between good and bad topics.

**Example:**

Let’s start learning with a simple example and then we move to a technical part of topic coherence.

*Imagine you are a lead quality analyst sitting at location X at a logistics company and you want to check the quality of your dispatch product at 4 different locations: A, B, C, D. One way is to collect the reviews from various people – for example- “whether they receive product in good condition”, Did they receive on time”.* You may need to improve your process if most people give you bad reviews. So basically, you are evaluating on the qualitative approach, as there is no quantitative measure involved, which can tell you how much worse your dispatch product quality at A is compared to dispatch quality at B. To arrive at the quantitative measure, your central lab at X set up 4 different quality lab Kiosk at A, B, C and D to check the dispatch product quality (let’s say quality defined by % of conformance as per some predefined standards). Now, while sitting at the central lab, you can get the quality values from 4 Kiosks and can compute your overall quality. You don’t need to rely on people reviews, as you have a good quantitative measure of quality.

Here the analogy comes in:

The dispatch product here is the topics from some topic modelling algorithm such as LDA. The qualitative approach is to test the topics on their human interpretability by presenting them to humans and taking their input on them. The quality lab setup is the topic coherence framework, which is grouped into 4 following dimensions:

* Segmentation: A lot of dispatch product divided into different sub-lot sizes, such that each sub-lot product is different.
* Probability Estimation: Quantitative Measurement of sub lot quality.
* Confirmation Measure: Determine quality as per some predefined standard (say % conformance) and assign some number to qualify. For example, 75% of products are good quality as per XXX standard.
* Aggregation: It’s the central lab where you combine all the quality numbers and derive a single number for overall quality.

From a technical point of view, Coherence framework is represented as a composition of parts that can be combined. The parts are grouped into dimensions that span the configuration space of coherence measures. Each dimension is characterized by a set of exchangeable components. First, the word set t is segmented into a set of pairs of word subsets S. Second, word probabilities P are computed based on a given reference corpus. Both, the set of word subsets S as well as the computed probabilities P are consumed by the confirmation measure to calculate the agreements ϕ of pairs of S. Last, those values are aggregated to a single coherence value c.

There are 2 measures in Topic coherence:

**Intrinsic Measure**

It is represented as UMass. It measures to compare a word only to the preceding and succeeding words respectively, so need ordered word set. It uses as pairwise score function which is the empirical conditional log-probability with smoothing count to avoid calculating the logarithm of zero.

**Extrinsic Measure**

It is represented as UCI. In UCI measure, every single word is paired with every other single word. The UCI coherence uses point wise mutual information (PMI).  
Both Intrinsic and Extrinsic measure compute the coherence score c (sum of pairwise scores on the words w1…wn used to describe the topic).

If you are interested to learn in more detail, refer this paper: - [Exploring the Space of Topic Coherence Measures](http://svn.aksw.org/papers/2015/WSDM_Topic_Evaluation/public.pdf)