Determining the novelty of patents using topic models

The novelty measure builds on work by Kaplan & Vakili (2013), who use topic models to find 'breakthrough technologies'.

The novelty measure (λ) for each patent p in each time period y is determined by the sum of the novelty score (γ) in that time period for each topic t over the cutoff score c. This is found by a simple algorithm:

1. For each topic-period, find the sum of patents with a topic proportion over the threshold c (where β_{pt} is the proportion of topic t in the distribution of topics over patent p):

$$\theta_{ty} = \sum_{i=1}^{p} x_i \text{ where } x_i = \begin{cases} 1 & \text{if} \quad \beta_{pt} \geq c \\ 0 & \text{if} \quad \beta_{pt} < c \end{cases}$$

2. To find the novelty score for each topic-period (γ_{ty}), find the period of the first period where $\theta_{ty} \geq 1$ (y_{init}) and set γ_{ty} to 1, find the period of full diffusion ($y_{\max[\theta_t]}$), then set each intervening period to one minus the ratio of the cumulative patent count to the maximum:

$$\gamma_{ty} = \begin{cases} 1 & \text{if} \quad y = y_{init} \\ \sum_{i=1}^{y} \theta_t \left(1 - \frac{\sum_{i=y_{init}}^{y_i} \theta_{ti}}{\sum_{i=y_{init}}^{y_{\max[\theta_t]}} \theta_{ti}} \right) & \text{if} \quad y_{init} < y < y_{\max[\theta_t]} \\ 0 & \text{if} \quad y \geq y_{\max[\theta_t]} \end{cases}$$

3. Calculate the novelty measure:

$$\lambda_p = \sum \gamma_{ty} \forall \{ t_p : \beta_{pt} \geq c \}$$

This provides us with a scalar novelty measure (λ_p) for each patent p.

References