PMI normalized between 0 and 1

Paolo Dragone

July 21, 2016

$$p(x) := \frac{\text{\# of recipes containing } x}{\text{\# of recipes}}$$

$$p(x,y) := \frac{\text{\# of recipes containing } x \text{ and } y}{\text{\# of recipes}}$$

$$\operatorname{pmi}(x,y) := \log \frac{p(x,y)}{p(x)p(y)} = \log p(x,y) - \log(p(x)p(y))$$

$$\operatorname{npmi}(x,y) := \frac{pmi(x,y)}{-\log p(x,y)}$$

The normalized pmi is between -1 and 1, where -1 is complete inoccurrence, 0 is independence and 1 is complete cooccurrence. To shift this measure between 0 and 1, sum 1 and divide by 2:

$$\begin{aligned}
\operatorname{npmi}_{0\text{–}1}(x,y) &= \frac{\frac{pmi(x,y)}{-\log p(x,y)} + 1}{2} && (1)\\
&= \frac{\log p(x,y) - \log(p(x)p(y)) - \log p(x,y)}{-2\log p(x,y)} && (2)\\
&= \frac{\log(p(x)p(y))}{2\log p(x,y)} && (3)
\end{aligned}$$

If x and y never co-occur together, then

$$\begin{aligned}
p(x,y) &= 0 && (4)\\
\log p(x,y) &= -\infty && (5)
\end{aligned}$$

which being at the denominator makes $\operatorname{npmi}_{0\text{–}1}(x,y) = 0$ in the limit.

If instead x and y always co-occur, then

$$\begin{aligned}
p(x) &= p(y) = p(x,y) && (6)\\
p(x)p(y) &= p(x,y)^2 && (7)\\
\operatorname{npmi}_{0\text{–}1}(x,y) &= \frac{\log(p(x)p(y))}{\log p(x,y)^2} = 1 && (8)
\end{aligned}$$