**1. Bootstrap**

seeds/seed tuple

图片包含 文字, 白板

描述已自动生成

缺点：

规则的准确性低；人力成本；

存在一些修饰语（解决：近似匹配）；

Error accumulation；

Bootstrap：1.生成规则；2.生成tuple；如此循环

Snowball：1.生成规则；2.生成tuple；3.评估规则；4.评估tuple；循环

**2. snowball**

**图片包含 屏幕截图, 电子产品

描述已自动生成**

**图片包含 屏幕截图

描述已自动生成**

为什么可以用内积表示相似度：

因为在把规则表示成向量的时候，已经做了归一化。所以不用除以模长。

2.1 step1生成模版pattern

2.2 step2生成tuple