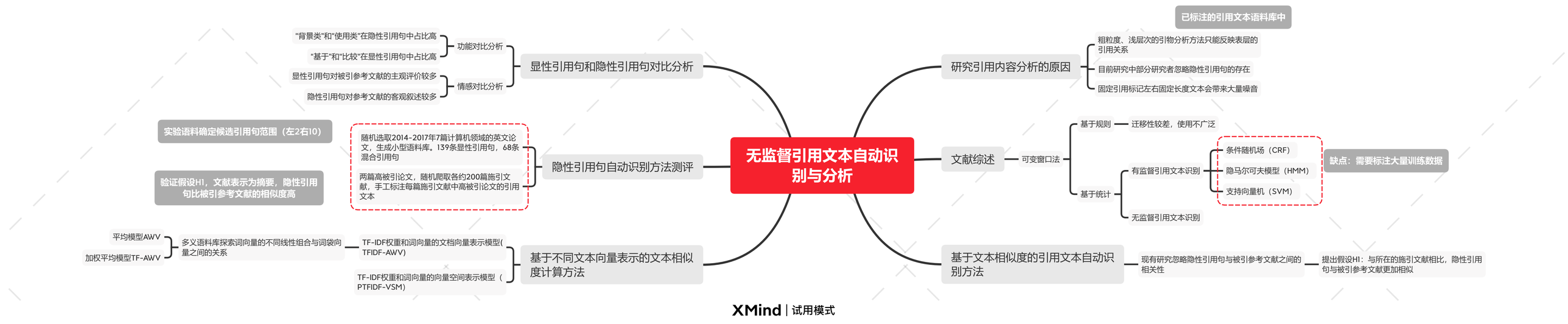
无监督引用文本自动识别与分析
显性引用句和隐性引用句对比分析
功能对比分析
"背景类"和"使用类"在隐性引用句中占比高
"基于"和"比较"在显性引用句中占比高
情感对比分析
显性引用句对被引参考文献的主观评价较多
隐性引用句对参考文献的客观叙述较多
隐性引用句自动识别方法测评
实验语料确定候选引用句范围（左2右10）
随机选取2014-2017年7篇计算机领域的英文论文，生成小型语料库。139条显性引用句，68条混合引用句
验证假设H1，文献表示为摘要，隐性引用句比被引参考文献的相似度高
两篇高被引论文，随机爬取各约200篇施引文献，手工标注每篇施引文献中高被引论文的引用文本
基于不同文本向量表示的文本相似度计算方法
多义语料库探索词向量的不同线性组合与词袋向量之间的关系
平均模型AWV
加权平均模型TF-AWV
TF-IDF权重和词向量的文档向量表示模型(TFIDF-AWV)
TF-IDF权重和词向量的向量空间表示模型（PTFIDF-VSM）
研究引用内容分析的原因
已标注的引用文本语料库中
粗粒度、浅层次的引物分析方法只能反映表层的引用关系
目前研究中部分研究者忽略隐性引用句的存在
固定引用标记左右固定长度文本会带来大量噪音
文献综述
可变窗口法
基于规则
迁移性较差，使用不广泛
基于统计
有监督引用文本识别
条件随机场（CRF）
隐马尔可夫模型（HMM）
支持向量机（SVM）
缺点：需要标注大量训练数据
无监督引用文本识别
基于文本相似度的引用文本自动识别方法
现有研究忽略隐性引用句与被引参考文献之间的相关性
提出假设H1：与所在的施引文献相比，隐性引用句与被引参考文献更加相似

XMind | 试用模式