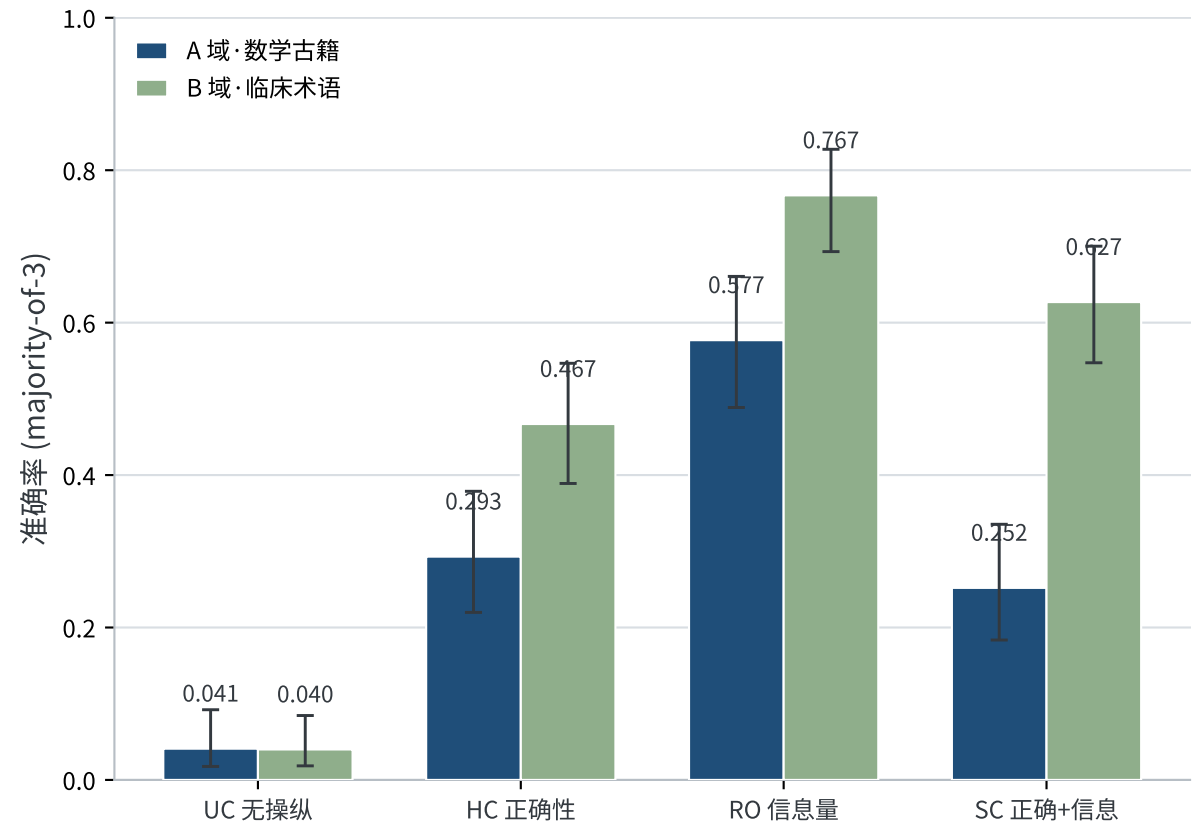
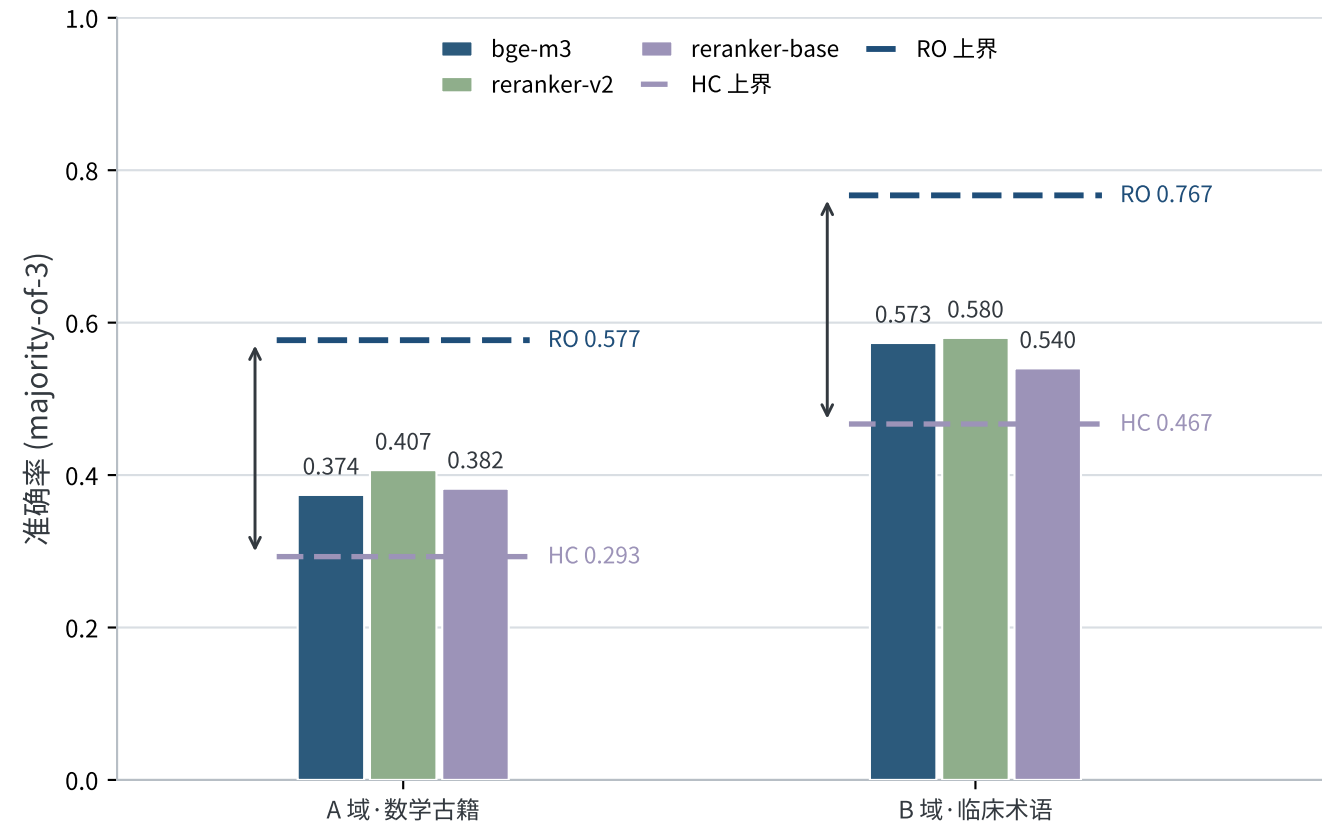


信息量 > 正确性：四条件行为证据与基线因果锁定

(a) 四条件准确率与 95% 置信区间



(b) 基线检索器 vs HC vs RO (三方对比)



检索器 top-1 全面高于 HC，却均无法追平 RO → 增益并非来自“候选含正确答案”，而来自信息量