

CS848 Project Proposal: study promotion effect of paper

Yifan Zhang

November 23, 2023

Abstraction

Promotion effect happens when the subsequent papers citing a certain previous work also cite its references.

metrics

these are the metrics that can be used thorough RQ1-3.

- promotion effect score of a paper P : $\frac{1}{P_r \cdot P_c} \cdot \sum_{j=1}^{P_c} \sum_{i=0}^{cr_j-1} 2^i$, where P_r is the number of references, P_c is the number of citations, and cr_j is the number of P 's references cited by citation j .
- promotion effect distribution (how many citations cited 1 reference, how many cited 2 references, etc)
- promotion that paper(reference) Q received from paper P : $\frac{1}{P_r} \cdot \sum_{j=1}^{P_{cQ}} \sum_{i=0}^{cr_j-1} \frac{2^i}{cr_j}$, where P_{cQ} is the number of citations of paper P that cite paper Q ;

the standard deviation of promotion received by all references of a paper P can be used to characterize the level of dispersion of P 's promotion effect

- how does the citing context match? (for each paper, randomly sample and exam some cases when a citation cites only one reference)

RQ1: does promotion effect differ between high influence and low influence papers

collect 15 high influence paper (around 1000 citations) and 15 low influence paper (50-100 citations), and compare the three metrics

RQ2: does promotion effect decrease after a paper is retracted

collect 5(?) retracted high influence paper, and 5(?) retracted low influence paper, compare the metrics before and after their retraction

RQ3: does author experience affect promotion effect

reusing the dataset from RQ1 and RQ2, calculate the author experience (e.g average author index) and study its relation with the promotion effect metrics

Limitation

semantic scholar API only returns up to 1000 citations, so in order to avoid bias we only focus on (relatively high influence) papers that receive around 1000 citations. The study is also limited by the request limits of the API and our computing devices. With more information and better computing technology, we can look at more influenced papers, such as *Fitting Linear Mixed-Effects Models Using lme4*, which has over 50,000 citations.

This study is correlation, not casual.

There might be inherent difference in the citations of high influence paper vs low influence paper, e.g high influence paper are cited by both senior & junior researchers, while low influence paper only cited by junior researchers, which requires future study.

Also different types of citations: method citation vs background citation, etc

self citation, or similarly citing each other's work within a small group;

compounding factors: e.g multiple papers providing promotion effect for the same reference

implicit promotion effect: paper C reads paper B's related work session and finds paper A, but it then cites only paper A but not paper B