

Credit Allocation Problems

Chunheng Jiang

Credit Allocation

Credit allocation is to solve the problem that how to allocate credits to the coauthors of a joint work. It's a.k.a credit assignment.

- ▶ Collaboration among researchers is a vital component in academia.
- ▶ No general guideline to place authors: contribution, seniority or alphabetical ordering.
- ▶ Number of coauthors is increasing.

There are several common factors that affect the allocation, including the order of the coauthors, the number of coauthors [?] and the reputation in the community [?, ?].

Existing Methods

Shen and Barabási [?] developed a collective credit allocation procedure. It estimates the credits to the coauthors based on the votes from the whole community rather than the coauthors themselves.

Xu et. al. [?] investigated 15 authors credit-assignment schemas. These schemas rely on the order of the coauthors and the total number of coauthors in the publications. These approaches present obvious similarity and strong correlation.

Based on [?], Bao and Zhai [?] proposed to incorporate a reinforcement mechanism and a power-law temporal relaxation function.

Shen's Method

- ▶ Target Paper: p_0
- ▶ Authors: $\mathcal{A} = \{a_1, a_2, \dots, a_m\}$
- ▶ Papers Citing: $\mathcal{D} = \{d_1, d_2, \dots, d_\ell\}$
- ▶ Papers Co-cited: $\mathcal{P} = \{p_0, p_1, p_2, \dots, p_n\}$

The credit shares of a coauthor is determined by whether she continues working with other coauthors on the topic of p_0 . Here, the set \mathcal{D} plays as a committee from the same community, and the citations are their votes. A bipartite network can be constructed.

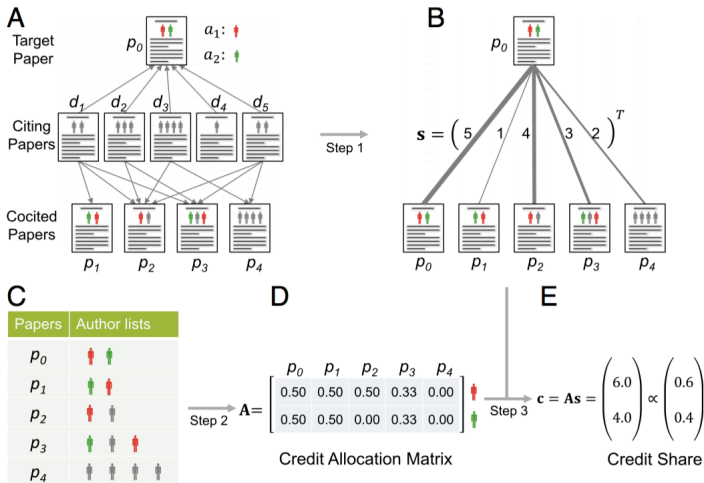
Shen's Method

The credit share of a_i in p_0 is computed with the formula:

$$c_i = \sum_{j=1}^n A_{ij} s_j,$$

where A_{ij} is the credit a_i earns from the cocited paper p_j , and each coauthor gets the same amount of shares from p_j . The relevance level (or cocitation strength) s_j of p_j to p_0 is defined as the number of documents in \mathcal{D} referring both p_0 and p_j .

Shen's Method



Priors

There are various author-list based prior rules to construct the credit allocation matrix A :

Disciplines	Priors				
	Count	Fractional	Harmonic	Axiomatic	Zhang
Physics	76%	76%	52%	44%	68%
Chemistry	71%	83%	67%	67%	88%
Medicine	86%	86%	43%	36%	100%
Overall	76%	81%	56%	51%	82%

Priors

- ▶ Count prior: Each author of paper p_j gets one credit.

$$A_{ij} = I\{a_i \in p_j\}$$

- ▶ Fractional Prior: All authors of paper p_j equally share one credit

$$A_{ij} = I\{a_i \in p_j\}/n_j$$

where n_j is the number of coauthors in p_j

- ▶ Harmonic Prior: All authors of paper p_j share one credit, where the share of the i th author is proportional to the reciprocal of its rank in the author list.

$$A_{ij} = \frac{1}{i} / \sum_{k=1}^{n_j} \frac{1}{k}.$$

Priors

- ▶ Axiomatic Prior: All authors of paper p_j share one credit without exogenous information

$$A_{ij} = \frac{\sum_{k=1}^{n_j} \frac{1}{k}}{n_j}.$$

- ▶ Zhang's Prior: The total credit of all authors is 3. The first author and corresponding author each get one credit. The remaining authors placed at rank i ($1 \leq i \leq n_j - 2$) gets credit

$$A_{ij} = \frac{2(n_j - i)}{(n_j + 1)(n_j - 2)}$$

Modified

Total credits C for all papers, including both the cocited and citing papers.

The qualities of the citing papers are also important, because their importance can affect the cocitation strength vector s . Let w_1, w_2, \dots, w_ℓ be the importance score of the citing papers. We modify Shen's method with the information:

$$s'_i = \sum_{j=1}^{\ell} w_j I\{d_i \text{ cocites } p_0, p_j\}.$$

The importance of a paper is proportional to its total credits earned from all its authors:

$$\begin{cases} C &= \sum_i w_i, \\ w_i &= \sum_j c_{ij}. \end{cases}$$

Datasets

- ▶ American Physical Society (APS)
 - Period: 1893 – 2009
 - Papers and citations from all the 11 journals of APS
 - 463,348 papers, 4,710,547 citations, and 248,738 authors
- ▶ Web of Science (WOS)
 - Period: 1955 – 2012
 - Multidisciplinary
 - 37,553,657 papers, 672,321,250 citations, and 8,724,394 authors

Evaluation

To evaluate the performance, the Nobel-prize winning papers in Physics (1995 - 2013), Chemistry (1998 - 2013), Medicine (2006 - 2013) and Economics (1998 - 2013) are collected to build the citation network. The credit allocation algorithm produces a credit share for each coauthor appeared in the awarding Nobel Prize papers.

The experimental results indicated that Sen's method performs very well, and successfully assigns the Nobel-prize winners the highest credit share in the awarding papers.

Further Studies

- ▶ Integrating both the order of coauthors and the contextual info to design better credit allocation matrix $A = (A_{ij})_{i=m}^{n+1}$
- ▶ Differentiate the values of the citing papers \mathcal{D} because of the reputation differences, produce more reliable relevance estimation of s (link-analysis based analyze over all papers)
- ▶ Authors' institutions and contextual text of the citation (definition, ideas)

