

A Paper Recommendation System

Dah Ming Chiu

December 17, 2008

Part of research is reading other people's publications - *papers*. This activity can take a tremendous amount of time, considering the rate at which papers are produced. Therefore a good researcher must selectively read papers.

There are many ways to select papers to read. Students seek advice from their supervisors. Professors may be able to judge a paper's worthiness more quickly, from experience, from reputation of the authors, or from other community services such as peer review activities. Conferences and journals help select the fittest papers for their topic areas, at different stages of a piece of research work. Word-of-mouth also help propagate the knowledge of good papers. These mechanisms all help the process of research.

Once we scale up the community¹, the effectiveness of the above mechanisms quickly diminish. From a personal perspective (as a researcher in computer networking), I am publishing in a dozen or more conferences/journals, seeing conference, workshop and other publication venues at an estimated rate of a hundred per year. Some conferences are rather large; for example IEEE Infocom publishes 200+ papers each year. The scale has reached a point that I cannot keep up with even keeping track of those papers that directly relate to my research.

Since all the researchers in the community should be in the same shoes, why not create a mechanism to collectively solve this problem? The idea is to build a *paper recommendation system*. Each (experienced) researcher recommends a paper based on the subset of papers visible to him/her. Ideally, the recommendation comes with a digest of what the paper is about and the reasons why it is recommended.

To ensure these recommendations are useful, they should satisfy the following properties:

1. Quality - They should come from experts who have read the paper carefully.
2. Accessibility - They should be categorized appropriately for easy access. For example, each user should be able to customize his/her own view of recommended papers of interest to him/her.
3. Ranking - They should be ranked. The recommendations will have different quality, freshness, and relevance. Like the results returned by a search engine, the ranking of recommendations is very important.

The following are major challenges in building such a system:

1. Business model - This is not necessarily related to making a business out of this service. Rather, it is about how to have an incentive system so that the system can self-sustain; namely how to make sure good recommendations keep coming. Ideally, it can self-sustain as a sizable social network. Contributors contribute mainly because of social reasons: for example, they become thought-leaders, or they get satisfaction from helping others or contributing to a good cause. Another possibility, which I believe is necessary, is to provide some form of financial

¹This can happen for a variety of reasons, such as research funding increase, or internalization of research communities.

rewards to contributors. This can be in the form of book coupons or prestigious stationary items (e.g. nice pens), which might be obtained from sponsors. If a sizable viewership is built up, it might be able to generate advertizing revenues to help self-sustain itself.

2. Reputation system - One way to help generate more good recommendations is to automatically discover good contributors via a reputation system. Ideally everyone should be able to contribute a recommendation, and others can determine the potential usefulness of the recommendation based on the author's reputation. This is similar to the type of reputation system used by eBay or Amazon. We can tie the reputation to real-world reputation by requiring real identity be used by authors; alternatively, identify can be anonymous and reputation can be built up gradually through previous recommendations. A mixture of the two based on contributor's own selection is also possible. The choice of identity has implications to collusion resistance, discussed below.
3. Collusion resistance - The biggest threat to the system is collusion, any form of exchange of favor by contributors which compromise the quality of the recommendations. The use of real identities reduces the risk of collusion, since all recommendations are public and eventually collusion may become self-evident via data mining. The reputation (real or virtual) will decrease, which becomes a deterrent to collusion.
4. Extensibility - The system will need to grow over time. The design must allow it to easily expand and evolve. Initially, the system may only focus on a few topic areas, and new topic get gradually added without major disruption to user experience. Reputation system and various collusion resistance mechanism can be added and improved over time. On top of this platform, various new service should be easily added. For example, search capabilities, viewer personalization, viewing statistics recording, viewership demographics analysis, and many other value-added services.

Beyond the above issues, how to implement a high usable and scalable system is of course a standard engineering challenge.

Finally, a word about *bootstrapping*. Before a reputation system and collusion resistant mechanisms are designed, the system can initially be based on invited contributions from well-known researchers. There can be some rewards for these contributions from a bootstrapping fund. Initially, the system can be announced to only a few selected university communities.

Acknowledgement

This idea came up during a lunch discussion with Minghua Chen, John Lui, Chuan Wu, Jianwei Huang and Angela Zhang. John and Dah Ming have talked about various forms of academic social networks in the past few years.