Causality in collective filtering

Mario Paolucci // Stefano Picascia // Walter Quattrociocchi
LABSS-ISTC-CNR [http://labss.istc.cnr.it] Rome, Italy

We describe a proposal for improving the practice of web-based collective filtering. We aim to build, offer to the community, and promote a simple social tool, based on the concept of causality.

The functioning of democracy relies on accuracy and exhaustivity of information. For a decision to make sense, it must be based on real facts; even more so for decisions that involve the attribution of representative power. We must suppose that electors are correctly informed on the matters on which they are called to vote. Asking people that are uninformed, or even worse mystified, amounts to throwing a dice, or to cover decision taken elsewhere with a false blanket of popular consent. If people need to be correctly informed, then independence and freedom of press become an institution of democracy at least as important as a house of representatives. Freedom alone, however, is not enough; it should be accompanied by a sincere research of truth on the part of the professionals of the press - the journalists. While the first can only exist if protected by law, the second is probably harder to maintain; some kind of ethical code and the passion of the readers for truth can contribute. Nothing new here, but a little repetition won't hurt.

How is the current situation with regards to that? Is the principle of independence of press upheld? Experts - journalists, intellectuals, technicians - used to play the role of information selectors. The functioning of this information selection mechanism works in two phases; a first selection process individuates a number of individuals; these are then attributed a social power, that of selecting relevant and truthful information, and in turn present this to the general public. We could call this mechanism "centralized selection."
Now, with the overwhelming amount of data made available by the diffusion of network connections, another way of selecting information has emerged, not in a top down way, but bottom up: collaborative information filtering.

In web-based collaborative information filtering platforms (think of digg.com or reddit.com) the community selects the issues that will get most attention with a mix of votes and recommenders' reputation. The obvious question here is - which one of the filters works better? The collaborative, bottom up, new paradigm, or the old one, based on processes that would (hopefully) select the best individuals who will in turn select what information will be consumed, and what will be wasted?

The question as stated, however, ignores an important dimension that introduces a new degree of freedom and increases the distance between the two approaches - the dimension of design. While the traditional approach is co-evolved with culture, the new one has a much more relevant design part; the presentation of filtering mechanisms, the assignment of values to votes, the grouping of issues; any choice to show or to hide, to allow or to deny action, is an explicit design choice in an online system. None of these choices can be considered neutral; they always introduce a bias, be the system designer aware of this fact or not.

The consequence on the initial question is that we should not simply ask which system works better, but also ask how can we design collective filtering in order to draw the best results - and what do we mean for the "best" in this case.

Our concern in this contribution is the process by which users attribute sense to news item - and how they connect them.

While complex approaches (Semantic Web and Topic Maps technologies, as cited in the call for abstracts) already exist, they seem way too complex to be of any real help to the general user. What we propose is a ***meaningful linking module*: a semantic layer operating on top of Web 2.0-style social networking/content filtering infrastructures, dedicated to drawing casual inferences between items.**

Take Digg.com or Reddit.com, popular social news sites that represent archetypical bottom-up news and discussion arenas on the Web. Generally, multiple news items about a specific topic or keyword coexist in the platforms' database. What is missing, and in our view would be extremely useful, is an easy way for users to establish links between pieces of content on the basis of causality among them. With the same principle of rate-and-rank that inspires these platforms users should be able to draw, enforce and deny causal links between facts. A visualization module should provide diagrams and comprehensive ways to glance the outcomes of the process and visualize the causal relationships between facts and interpretations.

The purpose of such a tool is twofold - first, obtain information not only on what people deems relevant about a specific issue - but also to induce circumstantiated cause-effect thought in people that volunteers to participate in the filtering process. The cause-effect relations we propose will also allow to specify support material, driving the user's choice to "official" sources of facts - meaning statistical offices tables and the such.

In sum, we propose to implement a system that improves the current state of the art in collaborative information filtering by a) supporting informed, evidence-based, cause-effect oriented discussion and b) employing new and more advanced interaction mechanisms.