Causality in collective filtering

Mario Paolucci, Stefano Picascia, Walter Quattrociocchi

No Institute Given

Abstract. In this paper, we describe a proposal for improving the practice of web-based collective filtering, in particular for what regards discussions and selection of issues about policy, based on the intuitive concept of causality. We give examples of its working and we state suggestions on how to built it.

1 Collective filtering and why it matters

Collective filters and similar platforms are the representative of web2.0 in its somewhat purest form. They implement the principles of rating and ranking to present information in a hierarchical way, ordered according mesures of relevance and community appreciation. They exploit the participatory practices typical of web2.0 to challenge traditional media outlets in the gatekeeping and agenda-setting functions. The basic principles that inform collective filters have been proposed for deliberative platforms, ICT systems aimed at facilitating debate and policy monitoring.

Web 2.0 has introduced new possibilities in the process of information sharing by providing strategies for content classification and presentation. Tagging consists of user generated keywords aimed to describe a resource, or some aspects of a resource, as perceived from the cognitive perspective of the individual. In the collaborative/filtering social environment tags are used to organize resources within a personal information space, but also with the possibility to be shared, by allowing other users to browse and search through tags posted by other users.

Another typical offspring of the Web 2.0 approach are social news sites: here users select and discuss upon "interesting" and valuable news headlines proposed through a process of collaborative content filtering and displayed as a list on the homepage. The process of selection is quite straightforward and typical of the Web 2.0 "smartmobs" approach: users post links to news items discovered on the web; fellow users can comment each post, vote it "UP" it if they consider it interesting, relevant, or generally "worth reading", or down it if not useful/interesting, inconsistent. A basic reputation mechanism [7]is included: users submitting popular (i.e. those that get many "ups") stories or valuable comments gain in 'karma', or 'reputation points'. Posts with a certain amount of positive votes collected over a certain amount of time and submitted by "reputable" users get featured in the homepage. The same applies to the debate that is triggered around single pieces of content: users rate others' comments and the most well received appear on the top of the list, while those falling below a certain threshold become invisible. Sometimes the voice of high-karma users counts more towards the reaching

of the threshold needed for a post to get on the frontpage, or a comment to be displayed.

With several millions of users collectively filtering and discussing news items, sites like Digg.com and Reddit.com started presenting themselves as grassroot challengers to professional mainstream media news- desks in the gate-keeping process. At the same time heavy criticism was raised and the alleged democratic nature of these media was questioned, as it became clear that only a small amount of "power users" was responsible for the most part of the frontpage of digg.com; we consider Digg as archetypal of the whole class of tools discussed here. The "democracy" problems can be seen as a canary value of the huge biases that probably, unquestioned, affect these platforms.

Interestingly, a superficial similarity exists between the Google's PageRank and social moderation methods as implemented on e.g. reddit, digg, and youtube: what counts is the number of references, or of links, that carry some reputation with them. However, the difference is in the semantics of a vote. PageRank is unable to discern between different semantics of a web link. For example, when linking to a fraudulent web page as a warning to other users, PageRank will count the link as a positive vote. Social reputation systems operate on a semantic layer above PageRank, by allowing people to consciously vote for pages, and even to express negative ratings.

1.1 Collective filtering, policy and democracy

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 con-

nections, 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.

In the rest of the paper, we make the point that it could be possible and fruitful, especially for policing matters, to introduce yet one additional semantic layer over the basic reputation algorithms, to embed the principle of causal relationships in the design of the collective filtering platforms.

1.2 From repuatation to causality

Practices aimed to provide a needful shortcut to information in the overloaded peer production web, including epistemic tools and cognitive tools, already exist. Basically, the reputation [8] of an item, that is how others value and rate the item - is the only way we have to extract information about it. To these set of practices, we propose to add one more: (causal) correlation.

In the system that we envisage, a visual enriched layer will allow users to provide new (or remove existing) correlations between pieces of information extracted and deterministically linked together by the system via text mining algorithms. Textmining will serve only as a preliminary inference within context to provide users an already defined, but preliminary diagram, in order to facilitate users' interaction. Thus, we aim on one hand to provide tools for users communication over the web and on the other hand to provide an important set of indicators for policy maker in order to better understand the users' perceived meanings of debated issues, the perceived relevant factors and possible solutions.

2 About causality

Causality, or the relation of cause and effect, is perhaps the most important tool that we use in making sense of the world around us. However, causality is all but an obvious, agreed-upon concept. The literature about causal reasoning is too extensive to be reviewed here. We will just set a few pointers that will help to frame the problem from the point of view of the application we envisage, that is, collective filtering systems. We can roughly divide the literature in three sections: common sense causal reasoning, human causal reasoning as seen from neuroscience, and statistical interpretation of causal reasoning.

In common interpretation, the attribution of causality depends on the graining level of events. Consider a soccer player scoring with a long kick. At the level of the game, there is little doubt that the player kicking the ball caused the score, and not a player that was at that time in another part of the field. However, this simple certainity gets muddled as soon as we change the scale of description.

Consider again the other players. Isn't the player who touched the ball just before the scoring one playing a role in causing the event? Or, for that matter, what about the goalkeeper, who perhaps made a mistake in evaluating the intention of the scoring player, or a mistake in the trajectory of the ball?

This basic consideration, however, can also be read the other way around - given an event which is included in a causal relationship, there is at least one description level in which that causal description feels right - the correct level of description for that causal relation. We base our proposal on this consideration. Deeper reflection on this matter would take us to considerations regarding free will [1], and from there into neuroscience [2,4] or statistics. We will review these quickly before proceeeding further.

2.1 Causality, timing and neurosciences

The first, striking set of experiments aimed at understanding what causality is for humans are the famous moving balls of [6]. In that series of experiments, simple stimuli as the images of two balls were shown to subjects as they collide - but with a twist. The reference experiment showed a plausible ball hit - the moving one stops, setting the second in motion - was correctly interpreted as causal. The experimental conditions, instead, played a trick with the (imaginary) physics of the situation.

A temporal or spatial delay was added, changing the interpretation of the motion from causal to non-causal. Neuroimage analisys was used to supplement self-reported causal interpretation, showing that [5] the visual system itself can produce causal interpretations.

In fact, for the simpler situations (the physically consistent ones) the activation of brain areas dedicated to complex visual analysis seemed to be enough. Instead, for the cases interpreted as not causal, a higher activation of the frontal cortex was evidenced, perhaps signaling a call-at-arms, a search for causality where other mechanisms failed.

Temporal matters can be especially relevant; in fact, for perceptual experience, causality and the experience of time seems to be directly connected. The issue is reviewed in [2], that reports several striking examples.

However, things change when more complex tasks are involved, and "finding a single region of the brain that uniquiely represents causal thinking is likely an unrealistic goal" [4].

2.2 Causality and statistics

Another basic mechanism that induces causal reasoning consists in the observation of event associations. This leads on one hand to a more rigorous definition of causality by statistics, and on the other hand to the interpretation of human causality attribution due to associative learning. This second is shown to be essentially due to surprise [5] and, to a measure, to the innate tendency to attribute meaning through narrative explanation.

All the experiments mentioned above, however, are based on formation and detection of simple inference rules. This is to be expected, as our understanding of causality and of how our brain deals with it is in its beginnings. Here, we use simple as opposed to complex in its technical sense; a system can be called complex, for example, when it is composed of many parts that interact in a non linear way (for more definitions, see [3]).

3 An architecture for causality attribution to news items

Current collective filters present multiple ways to organize content and to retrieve information: rating+ranking systems let the most relevant topics and the most valuable contributions to the debate emerge from individual preferences; tagging systems let items of content relate to each other on the basis of basic topical affinity. We believe that the technologies are ready to add one more semantic layer to this class of systems: causality.

The practice of tagging had a huge diffusion in the past years thanks to its immediateness and "cognitive cheapness" paired with great advantage (handyness) in terms of content classification and ease of retrieval. The same principle of projecting one's implicit mental taxonomy that underlies the practice of tagging could easily be implemented with regard to causal relationships between news items. The main design challenge, though, becomes that of keeping the necessary cognitive investiment at a minimum.

A clarifying example of current practices, that we propose in contrast, is that of a controversial blog entry and its comment. The entry [1] was posted in 2010 by Michael O'Hare, professor of public policy at Berkeley, USA, as a message to his students. In the entry, he tells a story about California educational system, starting from the large public investments that created it:

[1] http://blogs.berkeley.edu/2010/08/24/a-letter-to-my-students/

".. agreed to invest money... into the world's greatest educational system, and into building and operating water systems, roads, parks..", to the squeeze of the '80s where money was pulled back. In addition, the squeeze was, in the interpretation of the author, intentional and caused by generational matters: "Posterity never did anything for me!".

The subject is captivating, but what is even more interesting is the reactions that it had elicited in the comments[2]. Amid the plain emotional and ideological comments, a highly technical fight for the definition of the problem on the base of real data emerged. See a few examples of the issues debated[3]:

"There has been no reduction in taxes paid by legal California residents that has led to the present condition. If he is referencing Proposition 13, he also needs to consider the inflation adjusted price of real estate in 2010 versus 1978" (JH)

"..though it has only about 12% of the total U.S. population, California's share of the welfare caseload has risen from 22% in 2002 to over 30%." (AC); "What is your source for this assertion? .. spending dropped $349 million between 1996-97 and 2009-10, without adjusting for inflation" (RJ, in answer to the above)
"Again, it might help to use real data, real information, and real sources. The State of California has not paid into the pension fund of the University since 1990-1991" (RJ, in reply to a comment about pension overspending)

As it is clear even from these few examples, there is a need to carry on an informed discussion. However, even in web 2.0, most of the debate takes place in natural language and with little use of pointers to data. Our proposal aims to improve on both accounts: a graphical representation of causes and effects would rule out irrelevant comments, and automated link suggestion to official data would help sharing a realistic view of facts.

Of course, this would not solve all of the problems. Incompatible world views would interpret the same data in different causal schemes. Moreover, even the data themselves are subject to interpretation. But the new tool would leverage on the power of graphical representation, and the ease of linking, taking the debate to a new, better level.

3.1 The semantic linking module

Thus, we propose to improve existing collective filtering with a new module, that we call the semantic linking module. It should allow for the effective visualization and manipulation of emerging topics and their interrelations, by introducing a layer over existing folksonomies and algorithm-generated correlation.

[2] retrieved on Sept. 1st 2010.
[3] Comments taken from the blog webpage. Usernames posting the cited comments are Jonathan Hanson (JH), Alan McCann (AC), Rosemary Joyce (RJ).

Here, by folksonomy we mean a set of cumulative evaluations of resources by users in relation to a given context, expressed as tags, aimed at synthetically describe emergent/perceived properties of the resources. It is, in substance, a conceptual structure, created by a community, to classify and link concepts, built upon the practice of tagging: the employ of user generated keywords to describe a resource, or some perceived aspects originating by the cognitive perspective of the individual.

The semantic linking module will allow users to tag News, entries and comments posted to the baseline platform; tags will be elaborated and compared with respect to the classification of resource and its social perception. The module will filter data and create a visual canvas by the means of text mining procedures.

Tags will be categorized and connected in arguments (occurrences of similar tags on the same news, similar social evaluations). In this way, tagging occurring in the same class of news, or similar tags common to different resources will be connected and their perceived meanings and connections will be inferred. The unsupervised data transformation and interpretation will be displayed in a diagram showing correlations between topics.

By providing diagrams describing a semantic correlation within facts derived from users perceptions of contents and the information will be enriched also by the users feedbacks on the social perception (users ideas about the supposed correlation).

In addition, user feedback on the same or similar topic will be shown, in a double presentation specifying what has been machine inferred, and what content manipulation has been done by other users. The user performing the new tagging operation will have the choice of elaborating the automated proposal, or adopting/manipulating someone else's causal view.

Thus, the correlation maps will be validated by users, by changing correlations, by adding content and additional links in order to support and justify their opinions. The system will also automatically feed on a fact engine - an automated link to certified data, provided by official sources as for example national and international statistical institutes.

The proposal we are putting forward, thus, leans on two basic principles.

First, there is a connection between visual thinking and causal thinking. This has been shown extensively in the neurological literature presented above. Thus, we foresee that actually drawing the connections between items will help users to clarify what they actually think and believe, and help them communicate this effectively. Also, that this kind of argumentation would allow to separate emotional (and perhaps even ideological) issues from the real.

Secondly, politcal debate shows, in our opinion, a growing gap between the real issues and the issues being debated. This is shown with striking evidence in the case of Italy, that was the target of a deliberate exaggeration of security issues, also facilitated by media concentration. This is likely to be a world wide tendency; with some banalization, it can simply be considered as an effect of positive feedback of power.

The availability of data would contrast this tendency, and this in a widely accessible and used information access system. In conjuction with the simplified view granted by the visual causal tool, data could improve awareness and participation, allowing collective filtering to replace the experts opinion without falling into the trap of minimal common denominator.

3.2 The semantic linking module in action

To clarify our proposal, we show two example of what the system in action would look like:

Emma gets informed Emma is surfing italian blogs while planning her trip to Italy where she will attend a one year Erasmus programme. One blog links to a video hosted on youtube. It appears to be an extract from an Italian news channel and shows a black woman lying naked on the floor, apparently passed out. The images are quite shocking, Emma learns that the woman is probably a prostitute arrested during a police control who passed out after being interrogated at a police station. She decides to post it on the collective filtering system. Using the bookmarklet the procedure is straightforward, matter of clicking a button. While she is tagging the post, the 'related' tab shows a number of postings with similar content documenting the backdrops of the latest crackdown on crime promoted by the Italian government. The new entry will be featured in the "upcoming" section and, since Emma is a new user, will have to gather a big support from fellow users to reach the frontpage. Emma clicks on a blinking link in the page. Then appears a pop-up page showing both a list containing the similar news,contents with the related comments, and a diagram. The diagram's words are: immigrant, violence, racism, ignorance, crimes. The diagram is the last of a series of discussions within the members of the community an it seems to be continuous updating because the direction of the arrow connecting immigrant and crimes is changing every 20 seconds. Thus, Emma understand that there is a controversy on the attribution of causes in that community.

Audun and the traffic It is early morning in Rome, Audun, a visiting researcher, is buying his lunch at the supermarket. He hears people talking about the topic of the day, which all Italian t.v. news dedicated at least 3 reports: the double park. The mayor of Naples asks citizens to support the police against the abusive parking. The subject seems important for the two men, raising a heated discussion whose points seems to be that it is important to control and facilitate the traffic, and that the transgressors must be punished. Audun wonders if what he heard is just a coincidence, of if the problem of parking is really important for people, this impression reinforced by other bits of overheard conversation.

At work, Audun accesses the collective filtering web site, looking for information about the problem of parking. He learns that the most perceived problem is that of transit: people feel frustrated wasting a lot of time in the traffic, and now seems that the cause has been found in the ones who, carelessly, make

double parks causing discomforts to other citizens. The causal diagram of the perceived meanings pops up, clearly showing a connection within hit and run drivers, parking, traffic and carelessness.

Audun doesn't agree with the correlations, and wonders if the matter is just being pushed by the media. After some research, he discovers that the real problem is that the street network has been really improved in the 1970s, then no more works have been conduced, while the population has growth exponentially during years. Official reports show that the number of "legal" parking lot is not adequate to the real number of cars and citizens circulating. Audun bookmarks the correlations with a negative rate and provides an additional link to the official reports to provide hints to better inform the debate and the opinion formation with his contributions.

4 Conclusions

In this paper, we have drafted the features of a new tool that we consider important and necessary to improve the quality of collective filtering platforms. The problem is not only technological but social; the very 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.

From everyday experience, however, it seems that this is not enough. Lacking the experts, filtering tends to promote oversimplification and emotional responses. To improve on this state of things, 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.

To assess our hypothesis on the functioning of a causal module, we should test the effectiveness of these improvements trying to discern whether and when they contribute to harmonization and integration of the points of view expressed, and identify, by simulation, the configurations in which the mechanism would instead promote atomization and conflict. It is well known that users have private information that can differ from the public information due to the reputation effects, social rewards and trust.

We hope that our idea could initiate community mechanisms aimed at reducing the bias between private information and infomation that users communicate, also introducing features to allow users to go beyond the level of debate where they defend their usually public position. These features will facilitate new ideas to emerge from and within the group.

Acknowledgments

We acknowledge the help of Rosaria Conte and of all partners that helped us arrange a previous version of this idea for a project propsal. These include Frédéric

Amblard, Davide Rossi, Jorn Kohlhammer, Audun Josang, Dino Karaberg, and Chris Snijders. Mistakes and naive assumptions are, however, ours only.

References

1. Daniel C. Dennett. *Freedom Evolves*. Penguin, February 2003.
2. David M. Eagleman and Alex O. Holcombe. Causality and the perception of time. *Trends in Cognitive Sciences*, 6(8):323–325, August 2002.
3. B. Edmonds. What is complexity? - the philosophy of complexity per se with application to some examples in evolution. pages 1–16, 1999.
4. J. A. Fugelsang, V. A. Thompson, and K. N. Dunbar. Examining the representation of causal knowledge. *Thinking & Reasoning*, 12(1):1–30, 2006.
5. Jonathan Fugelsang and Kevin N. Dunbar. Brain-based mechanisms underlying causal reasoning. In Eduard Kraft, BalÃ¡zs GulyÃ¡s, and Ernst PÃ¶ppel, editors, *Neural Correlates of Thinking*, volume 1 of *On Thinking*, chapter 16, pages 269–279–279. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
6. A. Michotte and T. R. Miles. *The perception of causality*. Methuen London, 1963.
7. Mario Paolucci, Tina Balke, Rosaria Conte, Torsten Eymann, and Samuele Marmo. Review of internet user-oriented reputation applications and application layer networks. *Social Science Research Network Working Paper Series*, September 2009.
8. Mario Paolucci and Rosaria Conte. *Reputation: Social Transmission for Partner Selection*, pages 243–260. Hershey: IGI Publishing., 2009.