

Will collaborative filtering 2.0 please stand up?

By Garrett Engle

Website competition is tough. Any edge a website can offer could mean the difference between success and failure – on-line as well as off – and the latest technology domain to harness that edge is collaborative filtering (the science of product recommendations). Think about it; for the most part the Internet has always been about consuming information we know we want – like a huge on-line database. Collaborative filtering, on the other hand, is really the first major internet technology to bring information to us that we wouldn't otherwise be aware of; it offers the promise of new experiences, experiences that we are guaranteed to like! So why is it that the best recommendations I get are still from my good friends? How is it, then, that the current state of collaborative filtering, which I call collaborative filtering 1.0, falls drastically short of its promise or, should I say, its potential? Currently, the recommendations I get from collaborative filtering 1.0 are hardly convincing enough for me to blindly invest my money towards; definitely not the same as a friend's recommendation. And that, I believe, is what we want, isn't it? A collaborative filtering system that delivers a product recommendation with the same reliability and confidence as if it were a good friend. We want collaborative filtering 2.0.

Like internet search technology before Google came along no one has really figured out collaborative filtering; yet, at the same time, everyone knows it's the future. Take Netflix, for example. They are an industry leader and if they knew all the answers would they have resorted to a public contest (which is still going on) to gain a mere 10% improvement in accuracy? This is interesting because it means a fully-realized collaborative filtering system hasn't materialized yet and we have to imagine what this would look like. In other words, we have to ask ourselves, what would it mean to have a system give us a recommendation that was as good as if it came from a friend? If this were the case, clearly, we wouldn't be bombarded by the slew of recommendations we get today driven by the need for more sales. If your friend recommended to you 5 new products every single day how long would it take for you to stop caring? How sincere could they be? Likewise, collaborative filtering 2.0 will focus on quality and not quantity. It will offer fewer recommendations but each one will be taken *seriously*. Perhaps there would even be a period of time when no recommendations were made at all (heaven forbid!); but, when one was made, you would be sure it was *good*. Can you imagine a service like that? Who wouldn't want it?! This is, in essence, collaborative filtering 2.0 and it will raise the bar in what we expect out of a web service. Like Google's search technology, once the threshold has been crossed there is no going back.

The collective-limit or collaborative filtering 1.0

So what's the hold-up? What is so hard about collaborative filtering that has kept it back in version 1.0? Let's take a look at the current state and see what we've got.

In essence there are two types of collaborative filtering: user-centric and item-centric. In a user-centric system you look for a group of users that have similar tastes and recommend a product based upon that user-group. In an item-centric system relationships between a group of items are derived and used to make a recommendation based upon the item-group (Slope One being a popular algorithm). In either case there is no getting around the fact that a recommendation is being made based upon how others rated products – i.e. the collective's opinion.

It doesn't matter what kind of methodology is used they all derive their recommendation from the collective. However, is this even the right approach? Let's take a step back for a moment... While, sure, you'd be able to get product recommendations from a group, this mentality will never be able to consistently help you because, as an individual, you have many characteristics and preferences that differ from the group – *this is by definition*. If you shared every thought and feeling like the group then you would cease to be an individual, you would become the group. But, of course, we know everyone is an individual. So is it a surprise when we see less-than-perfect product recommendations when we base all of our methodologies around the collective's opinion; the lowest common denominator? When we consciously ignore the very aspect of our users that makes them unique? By inferring that the reasons a user will like a product must be the same as the collective's completely ignores the unique aspects of a

user. With any flavor of this paradigm collaborative filtering will never fulfill its potential. While there is an obvious overlap in many preferences between an individual and a group, the recommendation process will hit a hard limit of accuracy as long as it implicitly infers the reason why a user likes a product based upon the collective's opinion. The delta between the individual and the group (and again, by definition, there will always be a delta) will never be crossed – how could it? This barrier of accuracy, which I call the collective-limit, can not be crossed by collaborative filtering 1.0 simply due to the fact that it ignores the individual's uniqueness and, thus, *implicitly* infers why a user will like a product from the group instead of *explicitly* deriving why from the user.

Clearly, until a collaborative filtering paradigm emerges that can utilize the collective's input while, at the same time, capturing the individual's tastes and preferences we will never get beyond the inherently limited version 1.0.

New collaborative filtering approaches have emerged but they, too, miss the mark. For example, Pandora.com, a popular music recommendation site, seems to have made some head-room by creating a list of characteristics for every song so when a user likes a song they also have a given list of characteristics that the user likes as well. These extra characteristics are used to more accurately recommend another song and the response has been (more) positive. The only problem is that someone had to create this list of characteristics *per song* prior to implementing the methodology. What if you wanted to rate movies this way; who is going to catalogue the characteristics of every movie? It's like the human-genome project for products – it could take years before a catalogue is ready. Also, who is to say that the person doing the cataloguing has even got it right? How can we guarantee his/her opinion is accurate relative to the collective's opinion with which he/she is trying to match? Finally, none of this really addresses the problem of explicitly understanding why the user likes a song. Even though Pandora has many characteristics associated with a song they still must implicitly infer which characteristics matter to the user which, ultimately, must come from others who have rated the song – the collective-limit is still intact.

My good friend, collaborative filtering 2.0

It seems to me collaborative filtering 2.0 must achieve a few things if we are to ever reap its benefits:

- It must capture me as an individual. While pulling the collective's opinion to offer me a recommendation is crucial it can't do it in a vacuum. There has to be some understanding of my unique tastes and preferences to bust through the collective-limit and consistently offer me genuinely, good recommendations.
- In the pursuit of the above point it must not barrage me with questions, force me to maintain a preference-profile or compromise my privacy in any way. I'm busy; I don't want to be forced to maintain pages (or even one page) across the Net in the hopes of getting a better recommendation. A real Internet service should service me; I shouldn't have to service it.
- It must also evolve as I evolve. If my opinions and/or preferences change collaborative filtering 2.0 needs to change with me to continually offer me fresh recommendations based upon who I am today, not who I was last year.
- It must be easy to deploy and flexible enough to be applied to any product.
- It must be easy to use! If I have to write an essay for every product rating to get a good recommendation then forget it.

Again, the best way to visualize collaborative filtering 2.0 is as if it were your good friend recommending you a product. When your friend says to you, "Remember that conversation we had last week? Well, I found this great movie that is right up your alley; I think you'll really enjoy it!" Then you can't help but get excited about the potential of the new experience. Your friend knows you, knows what you like and knows which areas you are willing to explore. Shouldn't collaborative filtering 2.0 treat us the same way? For any website it's the long-overdue extra edge that I highly recommend.

*To see how to overcome the limitations of collaborative filtering 1.0 explore Attribute Profiling at
<http://www.egretendeavors.com/ap/ap.htm>.*