

# **MS&E 235, Internet Commerce**

Stanford University, Winter 2007-08

Instructor: Prof. Ashish Goel, Notes Scribed by Shrikrishna Shrin

## **Lecture 6: Reputation Systems**

### **Reputation Systems in Everyday Life**

A reputation system can be thought of as an algorithm which attempts to determine ratings for a collection of entities, given a collection of opinions that those entities hold about each other. This is similar to a recommendation system, but with the purpose of entities recommending each other, rather than some external set of entities (such as books, movies, or music). (Some parts taken from Wikipedia.org)

We use reputation systems in many ways in our everyday life. Sometimes, it is obvious that we are interacting with a reputation system. For example, your friends might recommend a particular movie among various choices. At other times, it is not obvious that we are interacting with a reputation system. For example, an e-mail program might automatically classify certain messages as spam and moves them to a spam folder.

Google search is another example of a reputation system as they rank the results in a certain order with the most relevant result at the top. Other examples include Amazon's recommendation system where they recommend products that you might be interested in or the reviews system at epinions.com where you can get reviews about products that you would like to buy. Any system where you can search for items and get an ordered list based on your search is a reputation system.

### **Difficulties of Reputation Systems**

Note: In the following discussion, the phrase "online world" refers to the internet world and the phrase "real world" refers to real life.

#### **Objective measures of quality? relevance?**

In order to generate an ordered list of items, it is important that we are able to measure or estimate the quality of the item and its relevance in an objective fashion. In the real world, it becomes difficult to build reputation systems around purely objective metrics as subjectivity also becomes important. Furthermore, in the real world, there is dearth of data to use in objective analysis.

It is easier to obtain data in the online world as compared to the real world and this makes objective analysis possible on the internet. However, while data might be available, any objective analysis is complicated by factors such as bias in the data and imperfect information. For example, only people who like the tripadvisor.com website will use it to plan their trips. Therefore, data

obtained from tripadvisor.com may be biased to represent the views of just people who like tripadvisor.com. Furthermore, suppose it is found that people leave a shopping website quickly without making a purchase, it may be that they are not interested in buying the products on sale. However, it is also possible that the customers did not buy anything because the website took too long to load and therefore they became impatient and left. In this case if the only data we had was the average time a user spent on the website, we would not be able to accurately determine the root cause for poor sales from the shopping website. Therefore, while data is available in the online world, objective analysis is not trivial.

### **Personalization**

Personalization is easier in the online world than it is in the real world. For example, it is very difficult to make sure that a store is decorated the way a customer wants when he/she walks in and the decoration keeps changing based on who is in the store. In the online world, however, customers can add their own images, create a profile, change the background of the web page etc. very easily. Also, data is easily obtained in the online world to help personalize each individual user's experience. For example, Amazon knows which products a particular customer has bought and is able to recommend products that he/she might be interested in. However, even if Walmart or another store is able to determine which products the customers who are currently in the store have bought, it is very difficult to recommend different products to each one of these customers individually. In the real world only your friends who know your tastes can provide personalized recommendations.

### **Presentation Bias**

Presentation Bias is the concept of something or someone gaining reputation/fame/notoriety just because it/they are already reputed/famous/notorious. It can be thought of as the concept of "rich get richer" while it is hard to find gems in the trough.

For example, the first video on youtube.com is likely to get a huge number of clicks thus further increasing its total number of views and popularity. On the other hand, in real life, it is difficult for something to spread as virally as it can in the online world. Therefore in terms of virality, the online world makes dealing with presentation bias harder. On the other hand, however, suppose there is a top news article about the notoriety of a celebrity whose notoriety is the reason for the article in the first place, then, the newspaper publisher can eliminate presentation bias in the online world by displaying a different top story every time a visitor visits their website. In such a case, real world print media cannot eliminate presentation bias in the said manner. What appears as headlines, appears as headlines for everyone who reads the paper. Another example of a presentation bias would be the top ranked results for hotels on websites such as tripadvisor.com. The hotels at the top of the search results are likely to get more customers and to a certain extent, their reviews (if favorable)

ensure that this hotel remains at the top of the results.

Whether presentation bias is a bigger problem in the online or real world depends on the situation at hand. However, if handled correctly, presentation bias can be mitigated to a certain extent in the online world.

### **Problem of cheap pseudonyms**

In the online world, it is easy to obtain many fake identities whereas in real life it is much harder to do so. Therefore, if we met a random stranger on the road, we are likely to consider him/her more reputed than the worst possible person we might know. Unfortunately, on the internet, any stranger has to be treated like the worst possible person we might know due to the problem of cheap pseudonyms.

This problem is one of the primary issues that is a much bigger problem in the online world than it is in the real world. This is because it is easy and cheap to obtain fake identities on the internet whereas the same is not true in the real world.

### **Problem of scale**

Before the internet, the amount of content created in a distributed fashion was not much and it was possible for a centralized arbiter to decide what reputation to assign to each content. For example, every year centralized committees decided which people are reputed enough to appear in compilation such as the "Top 1000 most influential people in the world". In the online world, every time you search for a name, say "George", Google or any other search engine has to decide what reputation to assign to the millions of people who have the same name George as the search engine has to return an ordered list of results. This problem is once again that of the long tail. As more and more content is created in a distributed fashion (thanks to the internet), the amount of content in the long tail is enormous. It is impractical to manually assign reputation to each web page on the internet and therefore, the process of assigning reputation has to be automated.

The problem of scale is a significant issue in the online world but it is not such a big problem in the real world.

### **Misalignment of incentives**

Misalignment of incentives is another challenge that reputation systems need to address in the online world. A user of a reputation system has to be sure that the results the reputation system returns are those that are the best for him/her rather than those that are the best from the reputation system's point of view. For example, if the reputation system could make money by re-ordering the results, it might have an incentive to not return the search results in the order that is best for the customer requesting information. Search engines such as Google have successfully addressed this issue by displaying organic search results

on the left and sponsored links on the right. The organic search results are those that are the best (most relevant) to the person searching for information whereas the sponsored links are optimized to help Google make more money. This generates trust in the search engine's results as it is apparent to the user that there is no misalignment of incentives between him and the search engine.

Suppose, a website where users flocked to read reviews about various products also happened to be selling those products. The website is better off if someone buys a product that is costly as opposed to one that is cheap. There is therefore a misalignment of interest as the user is looking for an unbiased recommendation of which product to buy while the website has an incentive to make the user buy a product that is costlier than another which would also serve the users' needs just fine. In such cases, it becomes important to ensure that through some mechanism, it is made obvious that there is no misalignment of interest.

### **Quid pro quo**

Sometimes, when at a restaurant and you notice something undesirable in your food, the restaurant might offer a free meal to indicate that they regret their mistake. Similarly, in the online world, a restaurant might offer a free meal to someone who writes a bad review (on websites such as yelp.com) to make up for the customer's distress and the customer might decide to remove his review. Therefore, the data available to reputation systems is in a certain sense "adulterated". This is not necessarily bad as the restaurant owned up to its mistake and offered the free meal as a nice gesture to an unhappy customer. What is even worse is as another case (discussed next) where the restaurant or business creates fake identities to write favorable reviews or pays customers to write favorable reviews.

### **Ballot Stuffing**

Ballot stuffing is the act of influencing the opinions of users/customers through deceit. Businesses sometimes create fake identities and write favorable reviews while posing as customers. Sometimes, they also pay customers who write a favorable review. The problem of ballot stuffing in the online world is a challenge for reputation systems to overcome. This problem is made worse by the problem of presentation bias and cheap pseudonyms. This is much harder in real life as it is difficult to create fake identities.