Reputation System for Marketplaces

Business Analysis

Introduction

The rise of internet marketplaces has resulted in major changes in consumer shopping behavior. No longer constrained by the availability provided by local stores, consumers are now free to purchase items online from around the globe to be delivered directly to their doorsteps. Yet with this vastly increased choice in product and supplier selection, comes an array of bewildering questions. Most notable is the question of which product, supplied by which supplier, at what price, with what delivery option, best fulfills a particular buyer's needs? To address this question, marketplaces often use ratings systems coupled with recommendation engines to aid consumers in finding products. But can we trust these ratings systems and recommendation systems? Recently a number of scam campaigns involving financial incentives for the creation of fraudulent reviews have come to light, affecting up to an estimated 61% of reviews within certain product categories on Amazon. In this blogpost, we analyse the current state of a variety of different types of ratings systems, and look at how well our current reputation system distinguishes honest versus fake reviews.

Gaming of rating and reputation systems in e-commerce is a large and challenging issue. To improve their reputations, suppliers sometimes resort to using fraud and manipulation. Some suppliers offer low quality products and/or services and enhance their reputations by using groups of fake consumers to artificially inflate their reputations. We refer to this kind of behavior as a scam.

Online reviews are an important tool for customers when making buying decisions. As Chae et al (2016)[1] and Floyd et al (2014)[2] discovered, they have a large spillover effect, substantially increasing final product sales in online marketplaces. A local consumer review survey in 2014 by BrightLocal[3] reported that an estimated 88% of customers trust online reviews as much as personal recommendations; that 31% of customers are likely to spend more on businesses with "excellent" reviews; and that 72% of respondents said that positive reviews increase their trust in a local business. A study by Öğüt & Onur

[1] Chae, I., Stephen, A.T., Bart, Y. and Yao, D., 2016. Spillover effects in seeded word-of-mouth marketing campaigns. Marketing Science, 36(1), pp.89-104.
[2] Floyd, K., Freling, R., Alhoqail, S., Cho, H.Y. and Freling, T., 2014. How online product reviews affect retail sales: A meta-analysis. Journal of Retailing, 90(2), pp.217-232.
[3] https://searchengineland.com/88-consumers-trust-online-reviews-much-personal-recommendations-195803

Taș (2012)[4] of hotels in Paris and London revealed that a 1% increase in online customer rating increased sales per room up to 2.68% in Paris and up to 2.62% in London."

Since reviews can make such a huge impact on buying decisions, there is ample incentive for suppliers to create fake reviews to spur sales. Given that fake reviews need to be viewed as genuine, it is difficult to accurately estimate what percentage of reviews is fake. Amazon filed lawsuits against over 1000 fake reviewers in 2015[5]. Amazon also estimates that only around 1% of reviews are fake[6]. Independent researchers, however place this number much higher. ReviewMeta, a website with an algorithm that helps identify fake reviews, estimates that between 9-11% of Amazon reviews are fake[7,8]. Some of the most fraudulent product categories on Amazon have (according to Greg Sterling, MarketingLand)[9] up to 61% of all reviews being faked.

Other marketplaces are not exempt from this problem. Mayzlin et al (2014)[10] estimated that between 5 - 15% of all reviews on TripAdvisor are fake. TripAdvisor's fake review problem was also investigated by Schuckert et al (2016)[11] who estimated 20% of its reviews were fake. Similarly Luca et al (2016)[12] estimated that 16% of reviews on Yelp are fake and Anderson et al (2014) made an analysis for and unnamed private apparel company and found out that at least 5% of reviews there are fake.

In the internet era, trust has become a concept that can be easily manipulated due to anonymity. Consumers and users in general want assurance concerning the information they receive. To help fulfill this desire, algorithms are necessary to help users filter out content which is likely to be fake or irrelevant.

What we Did

Given the widespread use of fake ratings and reviews, it is essential for a reputation algorithm to sufficiently separate suppliers who earn reputations through fake ratings from those who earn their reputations honestly. In order to do this, we designed a reputation algorithm to sufficiently separate fake ratings from real ones. We tested our reputation algorithm in a simulated marketplace environment, taking into account as much detail about customer behavior as our data supplied.

Our reputation system accounts for numerous factors. These factors include time liquidity (in which reputation transfers throughout time and reputation scores from earlier times is influence future

---

[4] Öğüt, H. and Onur Taș, B.K., 2012. The influence of internet customer reviews on the online sales and prices in hotel industry. The Service Industries Journal, 32(2), pp.197-214.
[5] https://www.nbcnews.com/tech/internet/amazon-files-suit-against-1-000-people-fake-reviews-n447101
[6]
https://finance.yahoo.com/news/rise-fake-amazon-reviews-spot-175430368.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAL_t73ZJqHa4ReDJZpbD_uEzSARlBJ4fUEeFY4NOw0JV8IOAUXHdzxJcrNzbSXPxbOjjJM7VUf1-3CSs89sB5EqaehmWTEcgYmFQST7GHdh4rllyywndfbM1Dhv5JVuo5MWpAnzrJ_uJogBKgNWA3I2n4KzSc8jdrBf6qCoMYado
[7] https://thehustle.co/amazon-fake-reviews
[8] https://www.buzzfeednews.com/article/nicolenguyen/amazon-fake-review-problem
[9] https://marketingland.com/study-finds-61-percent-of-electronics-reviews-on-amazon-are-fake-254055
[10] Mayzlin, D., Dover, Y. and Chevalier, J., 2014. Promotional reviews: An empirical investigation of online review manipulation. American Economic Review, 104(8), pp.2421-55.
[11] Schuckert, M., Liu, X. and Law, R., 2016. Insights into suspicious online ratings: direct evidence from TripAdvisor. Asia Pacific Journal of Tourism Research, 21(3),
[12] M. Luca, G. Zervas, Fake it till you make it: reputation, competition, and yelp review fraud, Manag. Sci. 62 (2017) 2016

scores); reputation flows from raters (buyers) to ratees (suppliers) (so that raters influence the reputation of products as well); time on the market (how long the supplier or consumer is staying on the market); and rates of spending flow. The algorithm can also be configured to suit different marketplaces since some have different market conditions and dynamics than others, and in order to optimally assign the reputation to participants. Because of very liquid and flowing manner in which our reputation system assigns and transfers reputation we call it the **weighted liquid rank reputation algorithm**.

While our reputation looks good in theory, we can not just blindly trust it. We tested our algorithm in large scale marketplace simulations in which buyers and suppliers exchange goods. We relied on market studies to provide realistic parameter values, focused on individual buyers and suppliers and scaled the simulation to a large number of market participants.

In creating our simulation, we included many real world factors based upon detailed studies of real-world marketplaces such as:

- Price distributions of different marketplaces. For example, there are more inexpensive items than expensive ones in an average marketplace. As prices increase, the number of products decreases roughly exponentially;
- The fraction of customers/clients that leave reviews online reviews depends greatly upon marketplace design;
- The rate at which customers rate and/or review items;
- The rate of change with time of the number of transactions and users (transaction growth);
- How the number of transactions and users is usually changing with time (user growth);
- The average customer/client purchase frequency;
- The number of customers who leave reviews and ratings;
- The factors affect influencing the number of customers leaving reviews;
- The number of customers do suppliers receive per day (the distribution of different suppliers and their relative success in marketplace).

We also conducted research on marketplace manipulation and scam, specifically the following;

- Percentages of scam reviews and transactions;
- Relationships between product price and scam;
- Number of reviews (scammers usually have more reviews, though they often come from different accounts;)
- Rating entropy (scam reviews are more likely to be extreme;)
- Review timing (usually fake reviews and ratings are paid and come shortly after the product is listed or after a certain period of time;)
- Account lifetimes (scam reviews tend to come from short-lived accounts;)
- Ratios of scam buyers and “real” or fair buyers;
- Ratios of fair and scam suppliers;
- Frequency of repeat purchases (users have memory, that is if a buyer *a* bought something from supplier *b*, and the quality has turned to be low, they will not repeat a purchase form the same supplier since they were disappointed with the product.)

Since we have considerable information about each agent, we can analyze the success of particular market simulations. We are attempting to indirectly detect manipulation and scam in order to provide accurate reputation scores. For this purpose, we introduce the following two financial performance metrics that can be easily retrieved from our simulations:

Loss to scam (LTS): there we sum up the volume of transactions by honest buyers to dishonest sellers and divide that to the spend of all honest buyer transactions. This metric shows what proportion of money spent by honest buyers is spent on the products offered by dishonest sellers.

Profit from scam (PFS): we sum up the volume of transactions by honest buyers to dishonest sellers and divide this by spendings of dishonest buyers. This means the return to the money spent by dishonest buyers on their own transactions. It shows how profitable it is to run fraudulent products in our marketplace. This metric should only be understood as relative value, such as what is the improvement of PFS over different systems, since one could argue that dishonest sellers do not need to spend their whole product price to get a fake rating.

Our goal is to minimize both the financial loss to buyers due to scam (LTS), as well as the amount that scammers profit from scam (PFS.)

In our paper[13] we also introduced additional "societal" metrics for measuring utility and inequity.

We also define different reputation system set-up scenarios as follows and used multiple runs on each scenario.

No reputation system (our control): participants are making decisions relying only on their own memories and not referring to any reputation system.

Regular reputation system (standard version of our reputation system): Does not take into account any factors other than values of ratings that consumers make to suppliers.

Weighted reputation system: When considering ratings as regular reputation system does, accounts to financial values of transactions between participants so that rating values are weighted by costs of transactions that are rated.

TOM-based reputation system: In addition to weighting ratings with financial values per-transaction, this setting weights the ratings based on the rater's time on the market (TOM) as a "proof-of-time". That is, the raters (buyers) are implicitly rated based on how long they have been on the market. So, rating by buyer with a longer history influences reputation of a seller more than the one made by rater with shorter history.

SOM-based reputation system: In addition to weighting ratings with financial values per-transaction, weights the ratings based on rater's spendings on the market (SOM) as a "proof-of-burn" value. That is, the raters (buyers) are implicitly rated based on how much they spend on this market. So, rating by buyer with a lot of spendings influences reputation more than the one made by rater with smaller spendings.

[13] Kolonin, Goertzel, Pennachin, Duong, Argentieri, Znidar (2019): A Reputation System for Marketplaces - Viability Assessment, arXiv e-prints

Our Results

We made simulations for different populations of 100, 1000 and 10000 suppliers and consumers being active during one quarter, one half year and one year. In the simulations 90% of participants were consumers and 10% were suppliers. In all simulations 5% of all suppliers were considered dishonest, charging money for fake services, and 5% of consumers were dishonest, providing fake positive ratings to dishonest suppliers. The behavior of the remaining 95% of consumers and suppliers was honest. Honest consumers encountering fake services served by dishonest suppliers rated the dishonest suppliers negatively and never returned back to such suppliers. All configurations resulted in very similar results to those we present below.

Dishonest suppliers and consumers conducted campaigns with durations of length we called the “scam period.” Once the scam period ends, dishonest suppliers and consumers changed their previous accounts (identities), so that the scammers’ identities, learned by honest consumers, periodically became inactive and the old scammers reappeared with new identities. In this manner, multiple dishonest suppliers and consumers appeared at different times on the marketplace as aliases of real scammers.

We tried four different scam periods: 10, 30, 92 and 182 days. In this way we compared how the effectiveness of the reputation system changed depending on how long scammers stayed on the market.

We show results for 1000 participants active for a half-year in Table 1 below. The leftmost columns “LTS Relative Decrease” and “PFS Relative Decrease” illustrate the performance of different reputation systems under different scam periods. “LTS Relative Decrease” is the relative increase of loss to scam in relation to having no reputation system. Similarly, “PFS Relative Decrease” is the relative increase of profit from scam in relation to having no reputation system.

We can see that for any scam period, regular reputation system does not provide any improvement, but rather makes things worse, making loss to scam and profit from scam greater. In turn, weighted reputation system provides stable improvement, decreasing loss to scam ad profit from scam. Finally, in cases of TOM-based (assessing raters’ reputation with time on the market) and SOM-based (assessing raters’ reputation with spendings on market) reputation system with shorter scam periods we have the best improvements.

Outcomes:

Scam Period	Reputation System	Loss to Scam (LTS)	Profit from Scam (PFS)	LTS Relative Decrease	PFS Relative Decrease
182	No	2.4%	44%		
182	Regular	2.7%	49%	-13%	-13%
182	Weighted	2.3%	42%	2%	3%
182	TOM-based	1.4%	30%	41%	31%
182	SOM-based	2.2%	40%	8%	7%
92	No	3.0%	54%		
92	Regular	3.5%	65%	-19%	-20%
92	Weighted	2.8%	52%	5%	4%
92	TOM-based	1.7%	36%	43%	33%
92	SOM-based	2.6%	47%	13%	12%
30	No	3.9%	73%		
30	Regular	4.7%	86%	-19%	-18%
30	Weighted	3.3%	59%	17%	19%
30	TOM-based	1.5%	31%	63%	58%
30	SOM-based	1.5%	27%	63%	63%
10	No	4.4%	81%		
10	Regular	4.7%	88%	-7%	-8%
10	Weighted	3.0%	54%	33%	33%
10	TOM-based	0.2%	3%	**96%**	**96%**
10	SOM-based	0.3%	6%	**93%**	**93%**

Table 1: If we look at different scam periods, it is clearly seen that if no reputation system is used then losses of those agents to scam and profits of scammers increase when the scam period is shortened. The regular reputation system appears rather ineffective for any scam period. Weighted reputation system on the other hand, always shows improvement, increasing with shorter scam periods. The performance of TOM-based and SOM-based systems for longer scam periods of 182 and 92 days exceeds that of the weighted reputation system. We also note that these systems significantly improve with decreases in the scam period and provide the best improvement across all scenarios with the shortest scam period of 10 days, making losses of honest consumers and profits of scamming suppliers nearly 10 times smaller.

Table 1 above shows that by using our TOM- and SOM-based **weighted liquid rank algorithm** reputation system, online marketplaces can allocate products much better and prevent scams well. The biggest winner here are the buyers of the products. They will still lose a bit of money on scamming sellers, however it will be significantly less than they would in case there were no reputation system. In a decentralized marketplace with no centralized service for scam prevention, it is still necessary for buyers to spend some money on scam sellers, because at the beginning there is no way of knowing someone is selling good or bad product – only after some feedback from buyers and transactions can we take appropriate action in order to allocate recommendations better.

The other impact of the reputation system used by marketplace is making the marketplace prohibitively inefficient for scammers from a financial standpoint, so that the decrease in PFS makes scams much less profitable.

The capability of the reputation system to discriminate honest suppliers providing generally real services from dishonest suppliers charging money for fake services is illustrated with the following figure where the vertical axis represents reputation rank in range from 0 to 100 while the horizontal axis represents market suppliers, dishonest in red on the left side and then honest in blue on the right side. Multiple dishonest suppliers are shown including all aliases being active across all scam campaign periods. Every alias of dishonest supplier agents is represented separately. The reputations of known honest suppliers on the right half is significantly higher than reputations of known dishonest suppliers on the left half - based on simulation with 1000 market participants acting over 6 months, with dishonest suppliers having a scam period of 10.

Conclusions

We have designed and evaluated a reputation system based on **weighted liquid rank algorithm**, which aims on giving better recommendations to products on online marketplaces. Our reputation system is particularly well equipped to tackle online reputation manipulation, which happens all the time in modern online marketplaces. The introduced **weighted liquid rank algorithm** is equipped with measures to weight reputation of the raters based on spending on the market (SOM-based) and on time on the market (TOM-based). The simulations of the marketplace with different scam patterns let us conclude the following.

- The regular reputation system “does not work”. It fails in comparison to versions of our **weighted liquid rank algorithm**. The regular system also makes profits from scam higher, so that the scam becomes more attractive. This provides incentives for scammers to provide fake ratings and engage manipulative behavior. Our experiments confirm what is happening on real world markets on different online marketplaces - scam is widespread because it is profitable.
- The weighted reputation system makes it possible to save up to 1/3 of losses to scam and makes scam less profitable. This in turn reduces incentives for scammers to engage in fraudulent behavior.
- Shorter periods of scam activity increase loss to scam and profit from scam. The presence of the non-traditional (not regular) reputation system decreases losses and scam profitability.
- With shorter periods of scam activity, the TOM-based and SOM-based reputation systems provide decrease loss to scam and profit from scam by up to 96% of the original volume.

Using the **weighted liquid rank algorithm** can therefore help customers choose optimal product configurations in a better way on current online marketplaces, thus increasing customer satisfaction and the popularity of those online marketplaces using such a system. The system is, moreover, very scalable.