November 2021

Reliable Normalized Customer Reviews

Alexander Ventura

Andy Lavery

Reliable Normalized Customer Reviews

ABSTRACT

Online reviews suffer from certain reliability problems such as lack of reviewer objectivity, reviewer bias, lack of domain familiarity on the part of the reviewer, etc. This disclosure describes techniques to improve the reliability of review scores obtained from online review facilities. A set of meta-parameters, referred to as review reliability parameters (RRP), is defined and associated with the underlying numerical rating of a review. RRPs are inputs to a machine learning model that derives a weighted review. The review of a reviewer with good RRPs is weighted more than the review of a reviewer with poor RRPs. RRPs can be based on factors such as the reviewer’s domain-specific expertise, history, demographic, meta-reviews, reviewing experience, etc. The techniques enable the creation of reliable, normalized reviews.

KEYWORDS

- Online reviews
- Review objectivity
- Review bias
- Review reliability
- Meta-review
- Bot detection
- Weighted review
- Star rating scale

BACKGROUND

Online reviews suffer from certain reliability problems that affect the individual reviews that constitute the final (aggregate) score. Some example factors that can affect an individual review include:

- Lack of domain familiarity: The reviewer has poor familiarity with the pertinent domain and is unable to provide meaningful feedback.

- Reviewer is biased: The reviewer might have a business or personal incentive to tarnish or enhance the reputation of the reviewed entity.
- Reviewer lacks objectivity: The reviewer injects subjective emotion into the review and is unable to be nuanced about their feedback. This often manifests as a reviewer who tends to rate at the extremes of a scale (e.g., a reviewer that only gives 1/5, or 5/5 star reviews).

Most review systems utilize an equal-weights approach where all reviews are averaged to compute the final score for the reviewed entity. As explained above, not all reviews provide the same meaningful information; highly meaningful and pertinent feedback may be obscured by a large number of meaningless reviews. This high noise-to-information ratio has led the consumers of online reviews to treat the process with some skepticism. Wide availability of the review process, while democratizing it, can drive online reviews to be treated by prospective customers as low trust and low information density.

Some online review services have created an elite class of reviewers who purportedly provide frequent and high-quality reviews and photos. Such review services also enable users to look either at all reviewers, or just the elite class of reviewers. Yet, even the elite class of reviewers is not free from bias, and there is no attempt to normalize reviews authored by them. Additionally, some businesses may have no reviews from elite reviewers, which doesn’t help users trying to assess the business.

Some review services, e.g., for streaming services, use the actual number of services purchased (e.g., shows watched) by customers. If the review service can assess exactly how much time was spent watching a show to gauge interest from the customer, then it is less dependent on star-type ratings, which, as explained above, can be subject to bias, subjectivity, domain unfamiliarity, etc. Such a review service will neither receive rogue one-star reviews from disgruntled but misinformed customers, nor will it get rogue five-star reviews from insiders. However, not every business is amenable to the precise measurement of time spent or interest displayed by customers as is possible for a streaming service.

Some review services calculate a meta-review, e.g., enable a review to be upvoted or downvoted, effectively enabling the reviews themselves to be reviewed. However, upvoting or downvoting doesn't change the aggregate score of the product or business. Additionally, the upvote/downvote facility can enable a rogue user to upvote or downvote many reviews to try to bias the votes.

DESCRIPTION

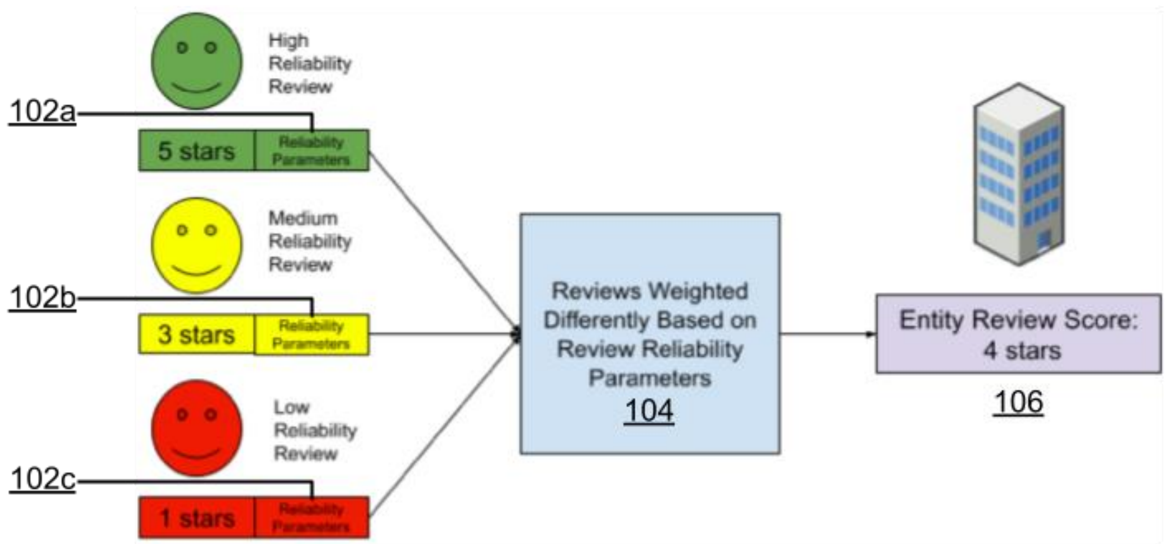

Fig. 1: Use of review reliability parameters to generate reliable normalized reviews

This disclosure describes techniques to improve the reliability of review scores obtained from online review facilities. Per the techniques, illustrated in Fig. 1, a set of meta-parameters, known as review reliability parameters (RRP, 102a-c) are defined and associated with the underlying numerical rating of a review. RRPs operate as inputs to a machine learning model (or other computational engine, 104) to derive a weighted review (106).

For example, in Fig. 1, the review (number of stars) provided by the green reviewer, who has good reliability parameters, is weighted more than the review of the yellow reviewer, who has relatively poor reliability parameters. The review of the yellow reviewer is in turn weighted more than the review of the red reviewer, whose reliability parameters are poorer than those of the yellow reviewer. In this manner, review reliability parameters assigned to reviewers enable the creation of normalized reviews with higher reliability.

Generation of the RRPs

The described RRP framework can be applied to any domain of any business (e.g., restaurants, hotels, vineyards, theme parks, etc.). RRPs can be based on several factors such as:

- Domain-specific expertise: For example, a wine reviewer who has experienced higher quality brands can have their review weighted more heavily than a new wine drinker. However, the same individual, who is a seasoned wine drinker, may not have much expertise in assessing hotels. Therefore, hotel reviews left by that individual are given relatively less weight. In this manner, a given individual can be associated with a vector of weights, one weight for each domain.
- Reviewer history: A reviewer with a lengthier history of published reviews can be weighted more heavily than a less experienced reviewer.
- Demographic parameters: Population-segmentation parameters, e.g., age, gender, income, etc., can be used as factors that contribute to an RRP as appropriate.
- Reviewer chain: Sometimes a set of fake reviewers will act in a coordinated fashion to contribute biased (positive or negative) reviews to a group of products. Reviewers found

to belong to such a reviewer chain can be ignored, e.g., given very low or zero weightage. Such reviewer chains can be detected by analyzing reviewers to see if they are all contributing to the same reviews to increase (or decrease) the score artificially.

- Meta-reviews: Upvotes or downvotes of a given review can be used to calculate a weighted aggregate score of the product or business. This can also tie in with reviewer-chain analysis to detect fraudulent upvotes/downvotes.

Grouping of RRPs

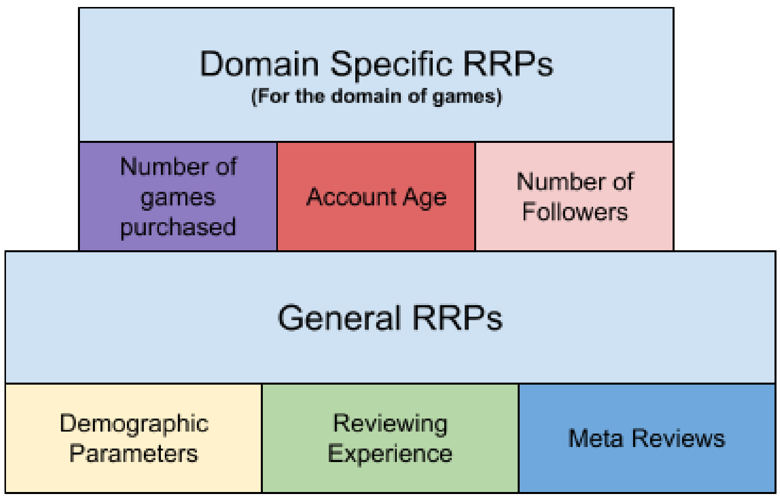

Fig. 2: Domain-specific and general RRPs

RRPs can be grouped into different categories, e.g., domain-specific RRPs, general RRPs, etc. Domain-specific RRPs are any product or business-specific parameters believed to be pertinent to the overall reliability of a review for a specific domain. These parameters can be derived from user-specific metrics not typically shared by separate product lines or businesses. An airline-specific RRP might be the number of miles a reviewer travels in a year, whereas a game-specific RRP might be the number of games a reviewer has purchased. Domain-specific RRPs are metrics that measure the competence of a user to leave a meaningful review for a given service. General RRPs are parameters applicable across domains and RRP implementations, e.g., the demographic (age, gender, etc.) of the user, etc. As illustrated in Fig. 2, the overall RRP comprises domain-specific RRPs layered on top of a base of general RRPs.

Examples of general RRPs

- Demographic parameters: standard demographic parameters, e.g., age, gender, education, etc. Such parameters can be filtered or tuned to provide a better weight for the review; for example, education can represent technical qualification, such that a licensed electrician writing a review for an electrician's tool is given a greater weight than a non-electrician.
- Reviewing experience: Even if a reviewer is new to a particular platform, there is some likelihood that the individual has left reviews on other platforms. This collective review history across all review platforms that the reviewer has interacted with provides an aggregate history that parametrizes the reviewer regardless of the current platform.
- Meta-reviews: Upvoting/downvoting of reviews, if enabled by the review platform, can influence the weight given to a reviewer.

Examples of domain-specific RRPs, where the domain is games

- The number of games purchased, which is a proxy for the experience that the reviewer has on the platform.
- Account age, which prevents bots from leaving biased or heavily weighted reviews. (bot-user accounts tend to be new; real humans tend to have lengthy and rich histories.)
- The number of followers, which is a proxy for overall taste-making influence.

Examples of domain-specific RRPs, where the domain is restaurants

- The number of restaurants a reviewer has visited.

- The publications (or platforms) where a reviewer has published, e.g., food critic versus regular customer.
- The number of restaurant reviews published by the reviewer.

Building review profiles for users

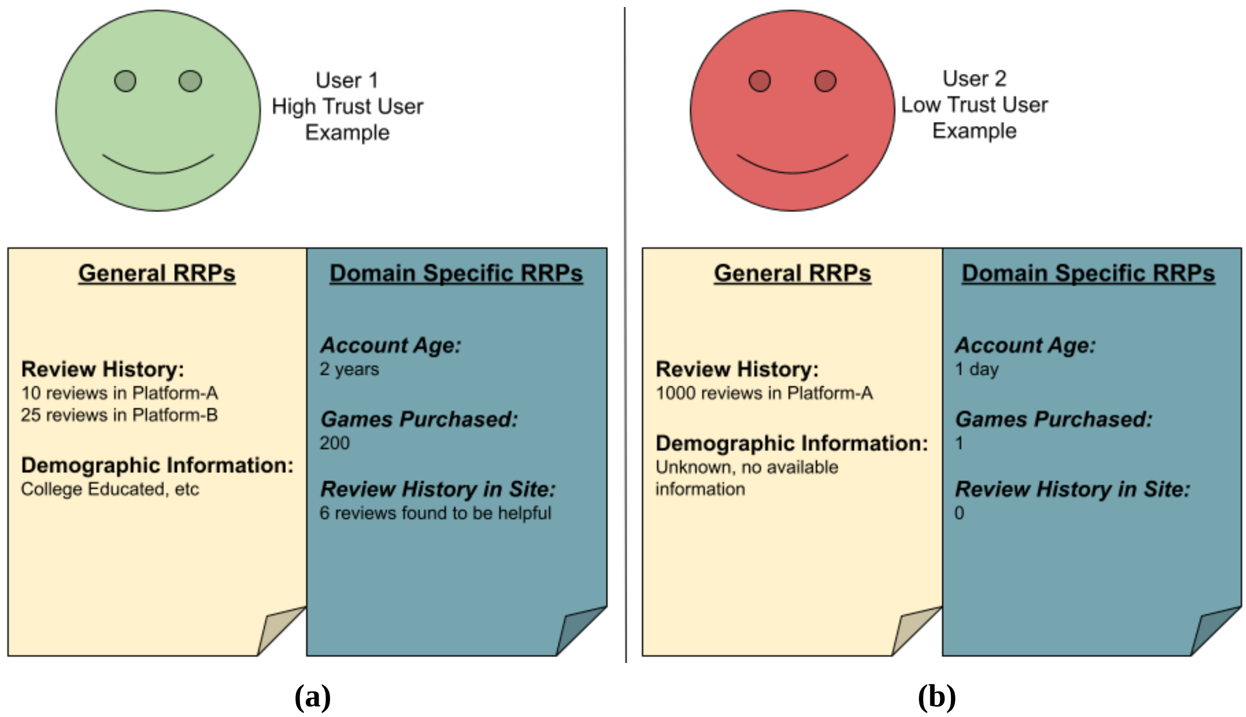

Fig. 3: Building review profiles for users

Fig. 3 illustrates building review profiles for users, using games as an example domain. Fig. 3(a) illustrates a user whose general RRPs reveal multiple reviews on two platforms (platform A and platform B), with a reputable demographic profile. Further, the reviewer's account age is not too young, and the reviewer has ample experience, having purchased two hundred games and having multiple (six) reviews that have been found to be helpful. Such a user is marked as a high-trust user, and their reviews are weighted more heavily relative to reviews from other users.

On the other hand, the user of Fig. 3(b) has an unfeasibly large number of reviews (one thousand) in an account that is a mere day old. None of the very large number of reviews have been found helpful. The user has purchased just one game and belongs to an unknown demographic. Indications are that the user is a bot; at any rate, a low-trust user. Accordingly, the reviews of such a user are weighted less heavily relative to reviews from other users.

The machine learning model adjusts the review weights based on the RRPs of each user. The final aggregated score for the business or product (a game in the case of Fig. 3) reflects reviews from high-trust users having a greater weight as compared to low-trust users.

In this manner, the disclosed techniques address weaknesses of current review methodologies by generating weights that favor reviews from users with a high degree of reviewing competence. The techniques also achieve greater review objectivity through the usage of RRPs. The techniques are applicable wherever users provide reviews, e.g., in application stores, in maps, in business/product (restaurant, etc.) review platforms, in user-experience feedback forms, in e-commerce websites or apps, etc. Further, the disclosed techniques enable users to filter point-based (or star-based) ratings, e.g., by enabling queries such as “show reviews by experienced wine tasters,” “show reviews by reviewers who match my age (or gender or other demographic),” “weight reviews based on calculated reliability from review count, review lengths, etc.”

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs, or features described herein may enable the collection of user information (e.g., information about a user’s reviews, demographics, social network, social actions or activities, profession, a user’s preferences, or a user’s current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level) so that a particular location of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques to improve the reliability of review scores obtained from online review facilities. A set of meta-parameters, referred to as review reliability parameters (RRP), is defined and associated with the underlying numerical rating of a review. RRPs are inputs to a machine learning model that derives a weighted review. The review of a reviewer with good RRPs is weighted more than the review of a reviewer with poor RRPs. RRPs can be based on factors such as the reviewer's domain-specific expertise, history, demographic, meta-reviews, reviewing experience, etc. The techniques enable the creation of reliable, normalized reviews.

REFERENCES

[1] Yelp Elite Squad https://www.yelp.com/elite

[2] "Netflix Is Shutting Down User Reviews This Summer" July 6, 2018, available online at https://variety.com/2018/digital/news/netflix-user-reviews-shut-down-1202866623/