

A Taste of Spam in your VOC Program? Tips for Filtering Social Media Data



For many financial services companies, social media is the next frontier in their Voice of the Customer (VOC) program. Capturing insights from social media has many clear benefits. First, it is rich with comments from both customers and potential customers, enabling you to monitor customer experience issues and also focus on acquisition. Second, people talk about your brand, products, services, and competitors on social media, which allows for many different types of analysis. Finally, unlike traditional VOC data, social media data is largely unsolicited. This means that people can talk about whatever they want, so you get an unfiltered picture.

On the other hand, social media data analysis is extremely difficult. Collecting social media data requires a sound infrastructure, and analysis can be frustrating for even the most seasoned VOC analysts. Why? Spam is rampant in social media, and spammers are getting increasingly more sophisticated in disguising their content as genuine brand mentions. While this makes accurate analysis seem impossible, if you understand the different types of spam in social media, you'll have a better chance of weeding it out.



A few flavors of spam

- **Shortened URLs** - One of the most common forms of spam is shortened URLs. Bit.ly and other URL-shortener tools are helpful to many legitimate Twitter users by shortening lengthy hyperlinks. However, spammers use these services to disguise links that lead to advertisements for their products. Often, what looks like a genuine tweet is really an ad, and the only way to tell is by clicking the link.

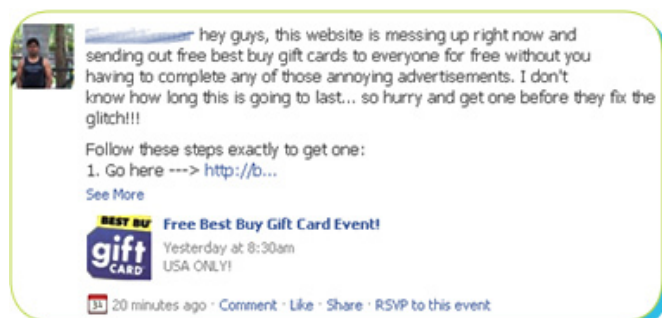
A strong VOC program can filter out some of this spam by unwinding shortened URLs as the data is captured. With the link no longer disguised, you can flag it more easily as an advertisement. Many social media analytics programs have a dynamic list of common terms, usernames, and URLs to flag as spam. The unwound URL can then be checked against this list, and the post can be flagged and filtered so it is not included in the analysis.



This Bitly link leads to a car financing ad.

- **Social media chain reposts** - Another common form of social media spam is Facebook and Twitter chain reposts. These are messages that are not generated by the user, and have no use for analysis, yet their pervasiveness in social media can potentially lead your analyses astray.

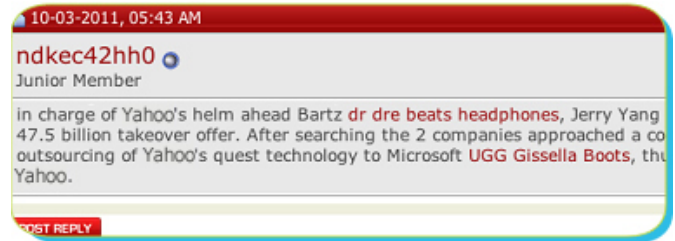
The good thing about these posts is that their content doesn't change. Therefore, once you notice it, you can easily set up a filter based on the content to block all identical posts.





To find these recurring posts, sort all duplicates together and look at the most duplicated. Then read through these and determine whether they are spam or legitimate retweets / reposts. It is essential to manually check your data. If you aren't reading your data, this and other types of spam will certainly skew your results.

- **Spam intertwined with legitimate content** - Finally, one of the most insidious forms of social media spam is largely made up of legitimate looking content, with spam terms interspersed. It looks like a genuine post, but contains links to advertisements.



It seems like it would be easy to filter using common spam terms such as “free credit score,” but when you take into account that there are millions of ways to phrase “free credit score,” that strategy is no longer feasible. So, how can social media analysts manage this type of spam without losing their minds?

A little spam won't kill you

Of course you want your analyses to be accurate so your organization can take appropriate action to improve customer experience, target promotions, etc. Because spam can skew analyses by amplifying certain keywords, you need to balance accuracy with the limitations of machine / human analysis. Check the machine yourself. Use your own understanding of your business, and ask yourself, “Do these results make sense?” If they don't, analyze further to see what's driving them. You may be seeing spam that you can filter out, then re-rerun the analysis.

You're not going to be able to filter out all spam posts- new types of spam pop up every day. Some are inevitably going to slip into your analysis. What you need to keep in mind when looking at your results is that *relative levels and fluctuation are much more important than precise counts*.

For example, if you are a bank looking at the top themes surrounding your brand, the fact that there are 150 ATM related posts is not useful information by itself. It is only useful if you understand what the usual number of posts is, and whether this is a significant departure. If it is, then it is worth diving in and determining what's driving the activity.

Remember to keep the initial goals of your Voice of the Customer program in mind: you're trying to solve business problems or meet key objectives. To do so, the exact number of posts within each category does not need to be 100% accurate. The most important thing is to understand what the data tells you so it can help you make an educated decision for an informed course of action.

Let's Talk

Beyond the Arc, Inc. is based in Berkeley, California, with offices in Seattle and Los Angeles.

Toll-free: [1. 877.676.3743](tel:1.877.676.3743)

Email: info@beyondthearc.net

Website: beyondthearc.com

Blog: beyondthearc.com/blog

Social:    