

Using Data Science to Score Marketing Content

Do you know how well your content is performing

The growth of the Internet in recent years has caused an evolution in on-line advertising. Static advertisements gave way to dynamic pop-up and banner advertisements which now, in turn, have given way to organic, or “native,” marketing content that blends in with the Internet viewing experience. This native marketing content not only adheres to the design of the surrounding content, it may provide a wide variety of informative content rather than abrupt sales appeals. And this native marketing content is presented in a various media (including video, audio, animation, slides, articles, white papers, catalogs, buyer’s guides, etc.) and on various types of web sites and platforms (including video sharing, entertainment, blogs, microblogs, email, professional networking, social networking, etc.)

This evolving and complex marketing environment has produced a demand for equally capable tools for evaluating the effectiveness of the wide array of content. The problem of evaluating content effectiveness falls into two major categories: Evaluating a particular content piece, and evaluating the effectiveness of an overall campaign, consisting of many content pieces.

Evaluating a particular content piece

The problem of evaluating a particular content piece is challenging because, as noted above, content comes in several different media and is hosted on different types of web sites. Therefore even the metrics available to evaluate content varies. These content performance metrics may be associated with an external host platform, if the content is hosted off of the home website. In this case, the hosting and the metrics are “offsite.” These offsite metrics may include number of visits, comments, likes, shares, and so forth.

On the other hand, the content performance metrics may be associated with the home website. In this case, the hosting and the metrics are “onsite.” These onsite metrics may include number of landing page hits and unique visits, forms completed, emails or phone calls generated from the Contact page, and sales.

The challenge in evaluating a particular content piece is two-fold. First, the piece needs to be accurately given credit for the activity that it generates. But what qualifies as “activity”? This brings us to the second challenge. The metrics used to describe the activity should be key performance indicators (KPIs). Offsite metrics such as visits, comments and likes are generally less key than onsite metrics. In fact, these metrics generally do not reflect sentiment. A blog post may generate a high number of visits and comments, but they may be disagreeable. And as for the onsite metrics, hits are less key than unique visits, which in turn are less key than contacts and finally sales. Clearly, some metrics are more key than others.

The problem is that the more key the metric, the more difficult it generally is to associate with any particular content piece. For instance, accurate metrics are usually available from offsite analytics. These give measurements of the visits, comments, likes, shares, and so forth. And these metrics can be unambiguously associated with a particular content piece. But these metrics are less key.

On the other hand, the most important metric is sales, but rarely will a sale result from a customer interacting with a single content piece. Most customers will engage with several brand touchpoints before making a purchase decision.

So the important onsite metrics are difficult to associate with an individual content piece. This is particularly true when the content is hosted offsite. One strategy is to offer the customer a coupon code that they later must supply. This helps to associate offsite content with an onsite metric. Another strategy is to use the onsite analytics which provide referring website data. These referring website data contain a wealth of information but unfortunately they do not easily resolve the association problem. The referring website data provide, for example, the number of hits or pageviews associated with specific referring URLs. They may also provide search keywords used in navigating to the landing page. But the referring website data are incomplete. They include only a fraction of the hits. Also, hits and pageviews are less key than unique visits, forms completed, contacts and sales. Furthermore, some offsite content may not contain live links to the landing page. And finally, many of the URLs supplied cannot be unambiguously associated with a particular content piece. For instance, the most frequent referring website is often google.com or "Direct Request." Nonetheless, for those URLs that can be associated with content pieces, the referring website data do provide helpful insight.

Evaluating the effectiveness of an overall campaign

The evaluation of the effectiveness of an overall native content campaign needs to account for the body of content that is posted. As noted above, a wide variety of media and types of platforms are possible, and the associated content across this spectrum works together to form the overall campaign tapestry. It is possible, for instance, for a particular content piece not to score high according to the evaluation methods discussed above, yet it could nonetheless play an important role in the campaign tapestry because it is the only content piece the campaign has to offer in a particular medium or platform.

Therefore evaluating the effectiveness of an overall campaign is more than merely tallying up and aggregating the individual content piece scores. Such an evaluation must account for three important attributes: quality, quantity and diversity. The quality of a campaign can be determined from the single-piece scores discussed above. The quantity, on the other hand, looks at the volume of pieces produced and posted. And finally, the diversity accounts for the breadth of the campaign, in terms of how many of the different off-site platforms, including email, are used.

Instant Content Scoring Algorithm

The scoring of content, both at the content piece level and at the campaign level has several challenges. First, it is a complicated problem involving a substantial baseline of historical data. Second, even given

such a baseline of data, how can the effectiveness of a particular offsite content piece be estimated using the important onsite KPIs? Third, how can campaign content scores be normalized while also allowing for substantial expansion, improvement, and maturity in a client's online content? And finally, it is inconvenient for clients to wait for several weeks or months, to accumulate the historical data. For many clients, there is a need for a rapidly-generated score. A good approach to solve all these challenges is by creating a heuristic scoring algorithm, described below.

At the point of initial registration, clients need to be able to view their current content scores. Since historical data will not generally be available to compute this initial score, a data scientist could design a scoring algorithm based on heuristic rules. These rules could entail best methods and practices that have been learned in the online content marketing industry. This heuristic scoring algorithm can be used at any and all times. In other words, its application is not limited to a first-use, or merely at the point of registration. Therefore this heuristic scoring algorithm can be used as part of a longer term solution, to help address the challenges in computing the long-term content score.

This heuristic scoring algorithm could compute component-level scores for each type of content. Here is a candidate list of the type of content, or categories, the heuristic algorithm will account for:

- Video
- Blogs
- Microblogs
- Newsletters
- Product pages (specifications, User Guides, etc.)
- Email
- Social networking

This list merely illustrates the possibilities. For some clients there may be types of content that are not relevant. For instance, a service provider will not have product pages. It will be important for this heuristic algorithm to know the relevancy of each content type.

Therefore, in the data collection process, any content types that are not relevant would need to be noted as such. The data collection process will likely consist of a survey, or series of questions, answered by the client.

Once the data are collected, the instant content scoring algorithm takes the data as input. The instant content scoring algorithm has three basic components:

1. Data formatting. The input data are prepared for use in the second component.
2. Content type models. The prepared data are input to each model, for each content type, and a score is output for that particular content type. If the input data indicate that the content type is not relevant, then a "Not applicable" score is output.

3. Aggregate model. The scores for each content type, output from the second component, are used together in a model that computes an overall, campaign, score.

Therefore, the output from this heuristic algorithm is both (i) a single, normalized, aggregate score and (ii) the individual content type scores. The individual content type scores could be accessed, for example, by hovering or clicking on the aggregate score.

Content Prescription Algorithm

The next logical step for users, after learning their content scores, is to learn what is their best strategy to improve and maintain their score. The content score breakdown (i.e., the scores for each specific content type) gives users ideas for improving their score. But the content score breakdown, itself, can only provide limited guidance. This is because these scores do not indicate the relative importance, or the relative cost, of each of the different content types.

A data scientist could construct algorithms that take as input (i) the answers to these questions, and (ii) the content scoring outputs (such as from the instant scoring content algorithms). The algorithms then compute a suggested strategy that efficiently uses the client's resources to meet the goals, and improve the content scores.

Importantly, a content prescription algorithm could support a "what-if" capability, allowing the user to input a hypothetical strategy. The algorithm would take this hypothetical strategy as an input and produce as output the results of implementing that strategy.

Long term content scoring algorithm

This algorithm is more challenging than the previous two because it incorporates data filtering, modeling, and tracking components. In addition to being more challenging, this algorithm requires a substantial baseline of data describing the content and describing the on-line activity (such as from Google Analytics). The data into two main categories: independent and dependent data or, more generally, input and output data.

The input data consist of the descriptions of the content pieces that have been posted to the Internet. For each content piece, these descriptions include three key parameters:

1. The type of content piece (e.g., blog post, microblog post, email, social network post, video, web page document, newsletter, etc).
2. The date of the posting.
3. A metric of the relative quality or length of the content (1-10).

The output data consist of the results of the content postings. These fall into offsite and onsite metrics, as described in the Introduction above. The offsite metrics may include number of visits, comments, likes, shares, and so forth. The onsite metrics include data typically associated with the home website

and may include number of landing page hits and unique visits, forms completed, emails or phone calls generated from the Contact page, and sales.

The following is a real-world campaign promoting a new educational service. In this simple case study, the campaign consisted of four content pieces posted to the Internet in June of 2014:

1. A posting on a relevant blog promoting the new service.
2. An advertisement on a relevant website promoting the new service.
3. An advertisement on a different, relevant, website promoting the new service.
4. A podcast discussing the details of the new service.

In this example, a data scientist could use the upload dates of these four content pieces as our input data. For our output data, one would use the number of unique landing page visits, as provided by Google Analytics. Figure 1 shows the Google Analytics data time history and the posting dates of the four content pieces.

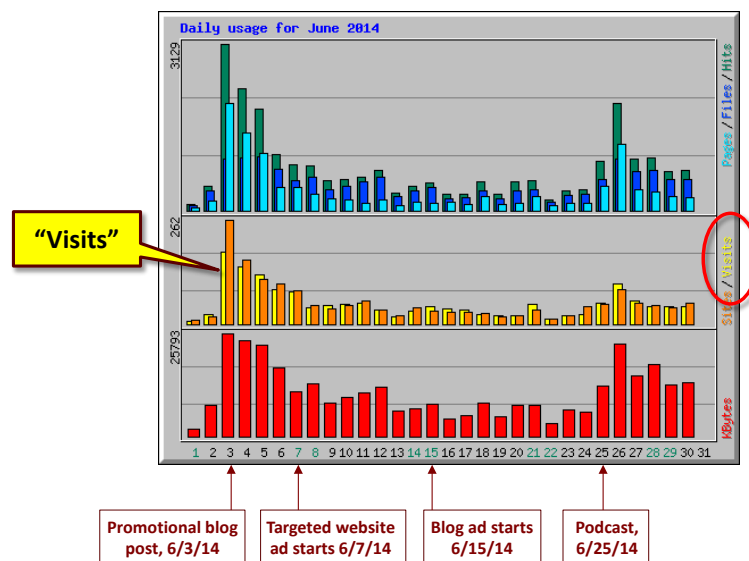


Figure 1 Example campaign.

You begin with a simple two-parameter model to describe the effectiveness of each content piece. The effectiveness is measured as the number of unique landing page visits that the content piece generates. Each content piece is assumed to produce an initial number of visits per day, followed by a decay rate as the content piece becomes stale. Figure 2 illustrates this two-parameter model.

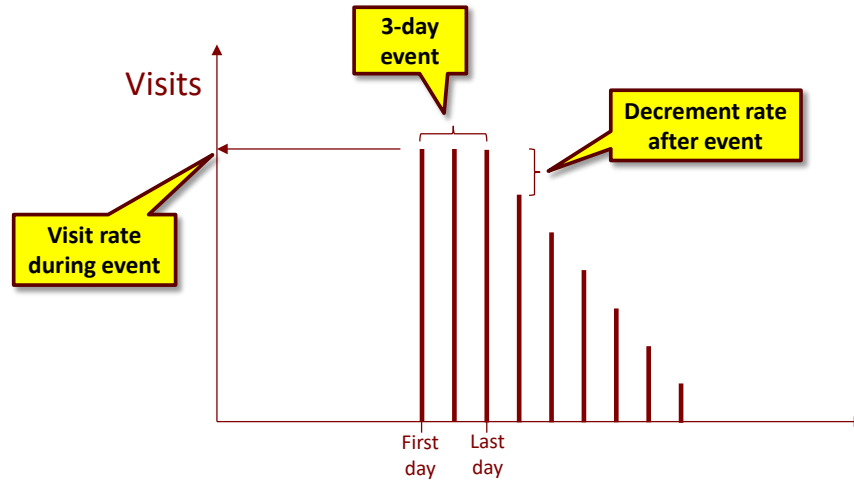


Figure 2 Two-parameter content piece effectiveness model.

As Figure 2 shows, the first parameter in our content piece effectiveness model is the number of landing page visits per day, during the content piece initial introduction. This could be a three-day period, as shown. The second parameter is the reduction, per day, in the landing page visits generated by the content piece. As Figure 2 illustrates, the model uses these two parameters to generate a predicted time history of landing page visits generated by the content piece.

The next step is to combine the four models of our four content pieces. That is, you sum the landing page visit time histories, from our four different content piece models, to obtain an overall, combined, landing page time history. You then compare this modeled time history with the Google Analytics, actual, time history. You can use a non linear batch least squares mathematical technique to fit our modeled time history to the observed time history. That is, you can adjust the two parameters in the models of our four content pieces (a total of eight parameters in all) to minimize the difference between our modeled time history and the observed time history. Figure 2 shows this comparison.

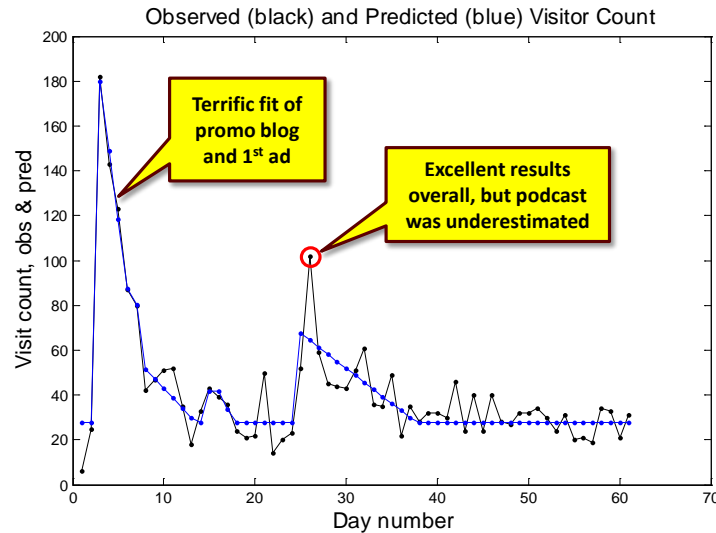


Figure 3 The two-parameter models successfully fit the observed landing page data.

As Figure 3 shows, the two-parameter models successfully fit the observed landing page data with excellent accuracy. By combining the modeled landing page visits from the four models, you could fit the Google Analytics observed visits accurately, and captured the dynamics quite well. The only deviation was the initial spike from the podcast, which was not completely captured by the model. In order to model this spike more accurately, you could use a more complicated model, such as a three-parameter model. A good way to measure and visualize the model accuracy is to plot the residuals, which are the observed visits (from Google Analytics) minus the modeled visits. Figure 4 shows the residuals.

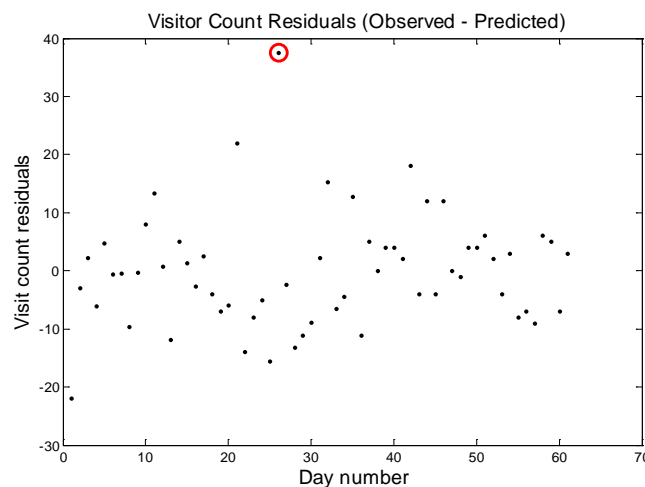


Figure 4 The fit residuals show excellent agreement between the observed and modeled landing page visits.

As Figure 4 shows, the two-parameter model produces well-behaved, unbiased residuals. The lone exception is the initial podcast spike, as discussed above. Now that you have accurate models for the four content pieces, you can compare their effectiveness. Table 1 lists the two parameter values for each of the four content pieces, the total number of visits generated by each piece (i.e., the “Raw score”), and a normalized version of the raw score (i.e., the “Score”).

Table 1 Comparing the effectiveness of the four content pieces.

Event	Initial Visits/day	Decrement/day	Raw score	Score
Promo blog	151.8	30.7	451.5	90.3
Targeted ad	23.7	4.4	100.5	20.1
Blog ad	13.6	8.1	32.8	6.6
Podcast	39.6	3.1	270.6	54.1

The results show that the promotional blog was the most effective of the four content pieces, with a very strong initial production of landing page visits. The podcast was the second most effective. As discussed above, this score is a slight underestimate of the actual podcast production. Nonetheless, it remains a distant second to the promotional blog. Next the targeted advertisement is third most effective, and the blog advertisement is last. Figure 5 shows these content piece scores graphically:

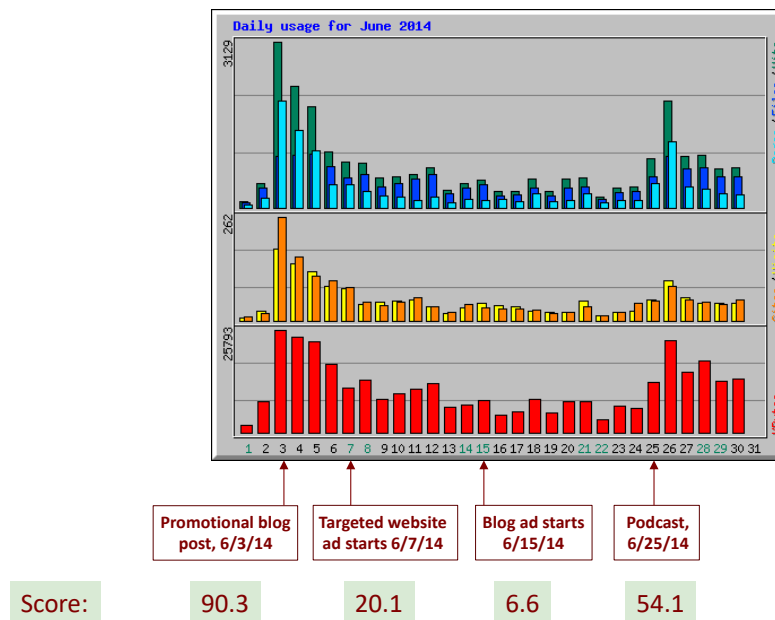


Figure 5 Display the content effectiveness scores graphically.

The Figure 5 results show the own-site content effectiveness scores. As discussed above, the content piece score can also include an off-site score, accounting for comments, shares, likes, and so forth. You can combine the own-site and off-site scores, to obtain an aggregate, or composite, content piece score, as illustrated in Figure 6.

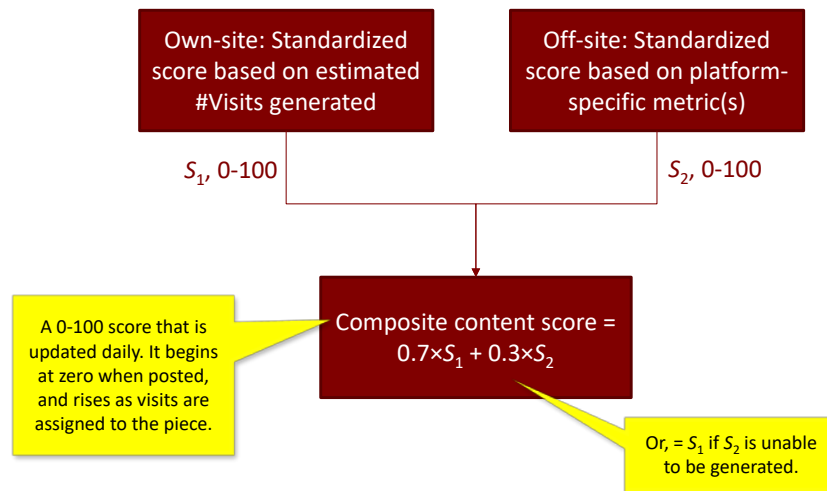


Figure 6 Combine the own-site and off-site content effectiveness scores to obtain a composite score.

The Figure 6 score is for a particular content piece. You can combine the scores for all the content pieces to obtain an overall content quality score. You can then combine that score with a quantity score, and a diversity score, to obtain a composite campaign score, as Figure 7 shows.

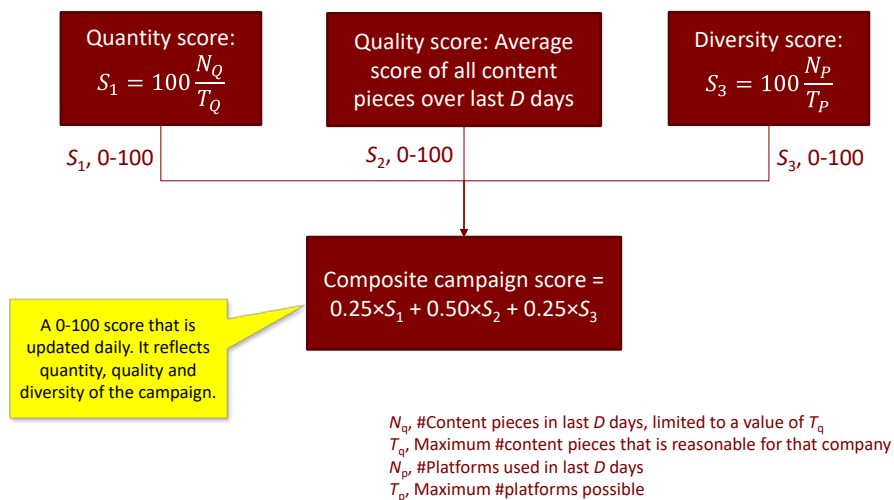


Figure 7 Combine content quantity, quality and diversity scores to obtain a composite campaign score.



Using Data Science to Score Marketing Content

A Mosaic Data Science Whitepaper

As Figures 6 and 7 show, this algorithm is contingent on having available several metrics, including the content piece data (e.g., type, posting date), own-site KPIs (e.g., the landing page visit time history from Google Analytics), off-site performance metrics, and normalizing parameters.

Conclusion

Being able to effectively and efficiently measure marketing content can optimize the way marketers reach their target audience. If you know a certain piece of content works on a certain media channel, you can repeat that process, increasing revenue from your content strategy.