Quantifying Commercial Opportunities in Twitter: Methodology

While the belief about the opportunity in Twitter is widely shared, there have been few reported studies that attempt to characterize and quantify the opportunity with some precision. This is, of course, a hard problem given the huge volume and diversity in the Twitter firehose (~250 million tweets a day across all subjects). A lot of tweeting, while perhaps important to the individual or their followers, is not actionable. The users would not want responses and marketers would not have a meaningful entry point. We call the actionable tweets, Tweets of Commercial Interest, or TCI, and finding them is a true needle in a haystack problem. We have to attack this problem in steps.

In order to develop our estimates we developed 3 filters:

Filter 1: Vertical domain keywords/phrases for selected verticals.

Filter 2: Intention

Filter 3: Actionability

<u>Filter 1</u>

We focused on three simple verticals: auto, consumer technology and food. We generated a small set of keywords/phrases representing a handful of major brands and product types representative of these respective industries. These were applied to the firehose in order to restrict the analysis to these areas. The keyword/phrase filters were applied using our system that is capable of both direct sieve matches as well as approximate matches over word sets (similar to a search engine). The latter is a type of theme matching that is not typically available in social monitoring tools.

<u>Filter 2</u>

We used the Solariat intention classification engine to classify tweets in terms of intention.

In our framework, intention is signaled by what are called 'speech acts', which do one of the following:

- *States a Need*
- *Asks for Something*
- *States a Problem*
- *States a Positive*

The classification system's precision/recall characteristics have been extensively validated and the performance over Twitter data is extremely high (this is an interesting point that bears a separate discussion). This gives us confidence that we have the right tweets to analyze after passing the thematically identified tweets through this filter.

Filter 3

Finally, we developed a general definition of actionability that could be handed off to a crowdsourcing system for labeling to find the actionable tweets from within the set of tweets with intention. We believe that actionability can be characterized in terms of the following features:

- The ‘speech act’ has some kind of intent, in other words the tweet indicates some expectation of a response.

- The ‘speech act’ is topically relevant to the brand, the object of the intent is a commercially relevant concept like a product or service.

- The ‘speech act’ is something that a marketer would likely have some kind of response or content for.

So, a tweet like “hate my laptop, need a new one” is certainly actionable. Likewise, a tweet like “I would love to be able to take great sports photos” would also be actionable. But a tweet like “Good morning Tweeters! Beautiful weather!” is almost certainly not.

Solariat has developed a dedicated crowdsourcing system that performs work similar to generally available crowdsourcing systems like Amazon’s Mechanical Turk, but for more sophisticated linguistic analysis tasks. We built this framework (along with recruiting and training the workers) because the nature of the tasks is typically too complicated for these more generic services. The actionability labeling task is a good example. It requires serious training of the labelers and a more involved workflow.

To make sure that we were accurately identifying actionable tweets, we insisted on unanimous agreement over the labeler group (n=6). The labeling task was tricky because we it involved labeling a tweet as actionable for some non-specific marketer. With customers, the concept of actionability is concrete and specific to the customer’s brand/product and marketing goals. Given potential errors of interpretation in the general case, we elected to apply the most rigorous standard and require unanimity. Note that the labelers worked completely independently. Undoubtedly, we generated some number (possibly large) of false negatives, but we want high confidence that our numbers represented a true lower bound.

We applied the above filters to a sample of the firehose (approximately 40 million tweets) extracted over a period of 3 weeks to avoid:

1. Time biasing in the samples

2. Sets too large to analyze in detail manually

Multiplying the number of tweets produced by the application of the 3 filters by a sampling proportion factor produced the volume estimates over the firehose.

In the operational Solariat system, the actionability label is actually produced automatically by what amounts to an actionability classifier. But for our purposes here, we decided to separate out the speech act and actionability features separately and do the latter in a detailed manual fashion.

For details please contact us at info@solariat.com