Social Media & Credit Decisioning
Analyses by a leading trade credit services provider

By Saleem Khan - Global Leader – Content Innovation and
Rich Ferrera, CCE, Senior Leader of Global Data Assets

Background

Credit managers have traditionally relied on company financials and trade payment experience data when making their trade credit decisions. Social media data presents a potential new source of untapped information that can complement traditional factors in the credit decision process. If used in a pragmatic way, social media data, in conjunction with company financials and trade experience data, may provide a more nuanced view of an entity's credit health.

When available, social media data provides a real-time view into the pulse of an entity. Unlike company financials, which are released quarterly or monthly, social media data may allow credit managers to receive a more real-time assessment of a company's ability to pay its trade credit obligations.

We are in the early phases of testing to see whether social media data can be used as a leading indicator of credit conditions at an entity level. This article will describe approaches being taken and certain key considerations when using social media as a source of content. We continue to experiment with this novel new source and believe there is a potential signal in the noise that can increase the aperture for credit managers.

A Framework for Social Media Sources

In order to incorporate social media sources into a proprietary model a few key factors need to be considered. The approach taken is a deductive one that starts with all potential sources then whittles down to candidate sources and individual variables from those sources.

Cohorts: Not all social media sites are alike; a cohort based filter needs to be applied before deciding where to source your content. Interested in determining credit terms for retail locations, nightlife and restaurants? Your best bet may be Foursquare[1] or Facebook. Interested in consumer goods? Twitter and Pinterest may be your best sources. It's been said that, "Facebook is like your living room, LinkedIn is like your office," etc. This sort of cohort based filter can be a great starting point to determine where to go when analyzing a certain type of business entity.

Quantitative vs. Qualitative: The next filter that needs to be applied is the type of data you are looking to source. Check-in data from users of social media apps can produce a great source of quantitative data. Information about trends in foot traffic, the amount of time spent in a location, and the distance travelled to visit a location can all be used in models that may have correlation with trends in cash flow for a business.

Qualitative data is not to be discounted by any means. A host of tools exist that provide inexpensive entity extraction capabilities and natural language processing in order to extract meaningful sentiment, entity information, and metadata, to name a few. Tools like IBM's Alchemy API[2] can be fed the body of a tweet or a post on Foursquare to extract relevant content as well as sentiment.

Verified Users: When looking to use posts by named users it becomes important to make a determination about the poster's identity. Many social media sites have verified user features which confirm the name and address of a poster before providing a certificate that the user can display (e.g. verified user on Amazon.) Some sites, by virtue of their cohort, would create pressure for users to use their true identity (e.g. LinkedIn account needs to be accurate if you're a job seeker.) Verified user data provides an extra layer of confirmation that there isn't too much noise in the data you are analyzing.

Influencer Scores: The network effect of social media sites also needs to be taken into account when verified user information is sparse or not present. Influencer scores provide a sort of rating for a user based on a multitude of factors, ranging from the number of followers a user has to how many reposts an original poster gets. An example would be the influencer score that Warren Buffet has compared to the obscure author of this article. A post from Warren Buffet - who currently has 1.1[3] million followers on Twitter and whose posts are retweeted by thousands - about an entity would carry much more weight, which may incent the poster to post credible tweets.

Sentiment Analysis: Another important derived factor from social media data is the sentiment around a specific topic or entity. This sort of data is created in a proprietary manner using language processors after content has been downloaded and ingested into a central relational database (e.g. Oracle) or a map/reduce database (e.g. Cloudera). This type of data is obviously a probabilistic determination but can provide important trending signals for an entity.

An Illustrative Example: A credit manager at a wholesale food distributor needs to determine which restaurants in her portfolio may have trouble repaying on time. Using the cohort filtering method described previously, the credit manager determines that her portfolio of Utah-based restaurants are active on a social media site. The sample size is large enough for her to pull quantitative information about check-ins over time at these locations. She runs a simple regression analysis to determine that, historically, there is a high correlation between a 20% decrease in check-ins to payment delays. She further determines that there is a high

correlation to the 20% downturn in check-ins to a negative sentiment score from posters of 65% or more. Armed with this information, which indicates reduced sales and cash available to pay trade creditors, our credit manager can now expedite her collection activities on customers that meet this criteria.

Disadvantages: Social media data is often unstructured, uncleansed and unmatched which makes it a very tough dataset to work with. The use of colloquialisms, abbreviations and jargon often make sentiment and entity extraction difficult, if not impossible. In addition to the data quality issues, the issues around the proliferation of bots3 (automated posts from unverified users) create a tremendous amount of noise.

Conclusion: There are strong advantages to using social media content in your credit decision framework. Credit managers cannot overlook this dataset as a proxy for determining real-time credit conditions. This content and the technology used to make credit assessments is still in an early phase but will soon be another tool in the credit manager's arsenal.

1. FourSquare Data: https://developer.foursquare.com/
2. Alchemy API (Sentiment/Entity Extraction): http://www.alchemyapi.com/
3. http://www.bloomberg.com/features/2015-click-fraud/

About the Authors:

Saleem Khan *is the Global Leader of Content Innovation at Dun & Bradstreet. He is responsible for identifying new content and technology capabilities that will enhance and expand D&B's product offerings. He has worked for startups and large multinationals, and is a proponent of rapid prototyping of data products using the Lean Startup methodology.*

Rich Ferrera, CCE*, is the Senior Leader of Global Data Assets at Dun & Bradstreet. He has been with D&B for 38 years and spent many years leading D&B's data collection teams including Trade and Public Records. Starting in 2005, Rich was involved in the development, customer education and implementation of DNBi, which is D&B's flagship credit risk product. Rich has also led teams that are responsible for implementation of D&B's risk, sales and marketing, and supplier management products, and currently leads D&B's global data assets and data provisioning for products.*