

Business Social Media Analytics: Definition, Benefits, and Challenges

Completed Research Paper

Clyde Holsapple
University of Kentucky
cwhols@uky.edu

Shih-Hui Hsiao
University of Kentucky
shihhui.hsiao@uky.edu

Ram Pakath
University of Kentucky
ram.pakath@uky.edu

Abstract

Based on a review of 27 research papers related to social media analytics (SMA), we develop an integrated, unifying definition of Business SMA, providing a nuanced starting point for future Business SMA research. Our definition goes beyond being entirely customer-focused, encompasses both the external and internal organizational environs, and goes beyond intelligence gathering to also accommodate activities such as sense making, insight generation, problem/opportunity detection and solution/exploitation, and decision making. Further, we identify several benefits of Business SMA, and elaborate on some of them, while presenting recent empirical evidence in support of our observations. The paper also describes several challenges facing Business SMA today, along with supporting evidence from the literature some of which also offer mitigating solutions in particular contexts. Accordingly, this research study helps further an understanding of Business SMA and its many aspects, grounded in recent empirical work, and is a basis for further research and development.

Keywords

Social media, business analytics, social media analytics

Introduction

In recent years, Social Media Analytics (SMA) has emerged as a key focus area within the broad field of Analytics (Kurniawati et al. 2013). Broadly speaking, SMA applies appropriate analytics capabilities to user-generated social media content to achieve a specific goal(s). Such content is held in a variety of repositories (Sinha et al. 2012) including, blogs/micro-blogs (e.g., Blogger/Twitter), social networking sites (LinkedIn), wikis (Wikipedia), social bookmarking sites (Delicious), social news sites (Digg), review sites (Yelp), and multimedia sharing sites (YouTube). There are several techniques that SMA could exploit – e.g., Sentiment Analysis/Opinion Mining, Insight Mining, Trend Analysis, Topic Modeling, Social Network Analysis/Influence Analysis, and Visual Analytics. In a manner of speaking, SMA takes the approach of “listening” to available user-generated social media content, rather than actively “asking” for user input, and acting upon it. A primary reason for burgeoning interest in SMA is the depth and reach of social media in terms of user-generated content volume and content diffusion speed. As an extreme example, a June 2006 video showing a Coke bottle exploding when Mentos (a kind of scotch mint) was placed in it became a top hit on YouTube and, subsequently, on late night news shows (Kaplan and Haenlein 2010).

Despite the growing interest in SMA, the fact that user-generated content is usually ad hoc, free-form, and contains both relevant and irrelevant material from the perspective of specific analytic goals renders SMA an arduous task in many practical settings. Even so, the perceived value in analyzing social media data attracts both academics and practitioners to increasingly devote time, money, and effort in pursuing SMA-related endeavors. For example, Sterne (2010) observes that analyzing relevant social media data can benefit a business measure the volume of user-generated buzz about a product or service, buzz

diffusion over time, buzz topic trends, the resultant impacts on sales, etc. This information can help the firm improve its marketing strategy.

Given its relative infancy, there is a dearth of research on a comprehensive view of what Business SMA is, its defining characteristics, its benefits and limitations, its deployment strategies, and the challenges one may face in deploying SMA solutions. Formal study and understanding of such factors are essential for systematic development of the field. Our research contributes to filling this void by (a) analyzing existing definitions of SMA and arriving at a business-domain-specific definition, (b) documenting, with recent empirical evidence, some of the business benefits of SMA, and (c) articulating, again with empirical support, some of the challenges encountered in applying SMA in business (and other) domains and solutions recommended in particular contexts. Business SMA limitations and deployment strategies, while also important topics for study and research, are beyond the present scope.

This paper is structured as follows. In the next section, we briefly describe our approach to literature search, filtration, and categorization. The results of these efforts provide a justification for undertaking the present, conceptual study. In the following section, we first examine various definitions of SMA found in extant literature and present a definition that is attuned to business contexts. We follow this with an examination of some business benefits of SMA and an articulation of the challenges one faces in applying SMA in business and other contexts. The final section summarizes our work and offers concluding remarks.

Research Methodology

We used Google Scholar for our literature search using all of the words in “social media analytics” and “social media intelligence” and focusing on the time period of 2008 through 2013. This resulted in unearthing 292,900 papers containing the stipulated keywords anywhere in the paper. We then asked that only papers with these keywords appearing in the title be located. This search resulted in 117 hits, a vastly more manageable number. We next eliminated hits that were duplicates, lecture notes, just abstracts, or unrelated to business. This resulted in a reduced set of 21 papers. Upon reviewing the 21 papers, we found that we only had a very small number describing empirical work. A key goal of ours is to review interesting empirical studies as part of this research endeavor. Therefore, to help build up a sufficient number and diverse set of recent empirical studies, we conducted another search using “user-generated content” as the key words and stipulating that these could appear anywhere in the paper. We further targeted specific IS journals (MISQ, JMIS, ISR, DSS, JAIS, IJEC, EM, and ECRA) and the 2010-2013 time period. This yielded 747 papers. We examined these papers to handpick 24 SMA-related papers representing a diverse pool of business SMA applications. For the purposes of this space-limited conference paper, we draw upon 6 of these 24 papers. Thus, our analysis here is based on a total of $21 + 6 = 27$ papers in all. We categorize this set of papers into four classes as shown in Table 1.

| Category | Description | Number of Studies | Percentage |
|------------------------------|---|-------------------|------------|
| Empirical Research | Use existing SMA methodologies on social media data to help answer specific research questions. | 18 | 66.67% |
| Algorithm/Methodology Design | Develop analysis procedures for SMA that seek to improve upon extant methods. | 13 | 48.15% |
| Conceptual Framework | Develop a framework to help better understand the SMA phenomenon. | 8 | 29.63% |
| Case Study | Describe specific scenarios for SMA application. | 2 | 7.41% |

Table 1. Categorization of reviewed literature

We note that the total number of papers across the four categories exceeds 27 because some papers fit more than one category (e.g., a methodology design paper coupled with empirical research). As evidenced in Table 1, the proportion of studies devoted to nuanced SMA applications (including empirical studies,

case studies, and algorithm/methodology design) far outweighs more comprehensive, conceptual studies that could help one become more insightful about the SMA phenomenon as a whole. This paper makes a contribution to the conceptual niche, where a dearth of sufficient research exists.

Business SMA Definition, Benefits, and Challenges

Business SMA Definition

Table 2 lists various definitions of SMA that we discern in the literature reviewed. An examination of these definitions reveals the following:

- (a) Among the definitions in Table 2, SMA has been defined in terms of types of activities pursued during an SMA life-cycle –
- Pre-analytics processing activities – e.g., searching/scanning/monitoring, finding/identifying, collecting, and filtering social media data;
 - Analytics processing activities – e.g., assimilating, summarizing, visualizing, analyzing, mining, and generating insights from the data; and,
 - Post-analytics processing activities – e.g., interpreting, reporting, dash boarding/alerting, and otherwise utilizing the results of the analytics endeavor.

As may be obvious, these activities are not necessarily conducted independently and linearly. Often, analysts must repeatedly cycle through and re-iterate prior activities during the life cycle to arrive at useful analytics outcomes.

- (b) SMA has also been defined by some (Grubmüller, Götsch, et al. 2013; Grubmüller, Krieger, et al. 2013; Yang et al. 2011) as including the collection of tools, systems, and/or frameworks that facilitate the above activity types.
- (c) Some (Yang et al. 2011; Zeng et al. 2010) even regard the development and evaluation of such tools, systems, and frameworks as falling within the purview of an SMA definition (and we agree that some SMA applications will require such development and evaluation as part of the analysis attempt).
- (d) Some definitions elaborate on the nature of what is being analyzed -- data on conversations, engagement, sentiment, influence (Sinha et al. 2012; Yang et al. 2011); postings, comments, conversations (Grubmüller, Götsch, et al. 2013; or, semi-structured and unstructured data (Kurniawati et al. 2013).

| SMA Definition | Source |
|---|---------------------|
| “... developing and evaluating informatics tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data, usually driven by specific requirements from a target application.” | Zeng et al. (2010) |
| “... developing and evaluating informatics tools and frameworks to measure the activities within social media networks from around the web. Data on conversations, engagement, sentiment, influence, and other specific attributes can then be collected, monitored, analyzed, summarized, and visualized.” | Yang et al. (2011) |
| “... scanning social media to identify and analyze information about a firm’s external environment in order to assimilate and utilize the acquired external intelligence for business purposes.” | Mayeh et al. (2012) |
| “... measure behavior, conversation, engagement, sentiment, influence, ...;” “monitor exchange of information on social | Sinha et al. (2012) |

| | |
|---|------------------------------------|
| networking sites.” | |
| “... social listening and measurements ... based on user-generated public content (such as postings, comments, conversations in online forums, etc.)” [using SMA tools] “with different features like reporting, dash boarding, visualization, search, event-driven alerting, and text mining.” | Grubmüller, Götsch, et al. (2013) |
| “Software systems that automatically find filter and analyze user-generated contents produced on social media.” | Grubmüller, Krieger, et al. (2013) |
| “... the use of analytics-based capabilities to analyze and interpret vast amounts of semi-structured and unstructured data from online sources.” “... provides ... insights into ...customer values, opinions, sentiments, and perspectives” | Kurniawati et al. (2013) |

Table 2. Social media analytics definitions

Drawing on Table 2, we advance a business-domain-specific definition of SMA (Business SMA) to both facilitate discussions that follow and to provide a useful starting point for those engaged in Business SMA research:

“Business SMA refers to all activities related to gathering relevant social media data, analyzing the gathered data, and disseminating findings as appropriate to support business activities such as intelligence gathering, insight generation, sense making, problem recognition/opportunity detection, problem solution/opportunity exploitation, and/or decision making undertaken in response to sensed business needs.”

Our Business SMA definition not only incorporates the ideas embodied in the prior definitions in Table 2, but also lends “purpose” as to why a business entity might choose to engage in SMA. Also inherent in the definition is support for “evidence-based” problem solving/decision making as advocated by (Grubmüller, Götsch, et al. 2013; Ribarsky et al. 2013). Whereas some authors (e.g., Kurniawati et al. 2013; Mayeh et al. 2012) view Business SMA as being synonymous with “customer-centric” SMA, our definition does not preclude the inclusion and analysis of social media data from other business-related entities such as employees, suppliers, retailers, competitors, regulatory bodies, and so forth. This view is shared by Mayeh et al. (2012) with the distinction that they regard a business firm’s SMA attempts as being applied only to its external environment. We, however, contend that a firm’s internal environment is also susceptible to SMA, as with monitoring employee/employer-generated internal social media content; our definition accommodates SMA in both the external and internal environs. Finally, as we note, the purpose of Business SMA is not merely intelligence gathering as Mayeh et al. (2012) contend – it goes beyond intelligence gathering to supporting such activities as insight generation, sense making, problem recognition and solution, opportunity detection and exploitation, and decision making.

Business SMA Benefits

Business SMA has the potential to help a firm realize several benefits, as our literature review reveals. Kurniawati et al. (2013) note the following benefits based on a review of 40 SMA “success stories” from IBM, SAS, and SAP: improved marketing strategies (75% of the cases), better customer engagement (65%), better customer service (35%), better reputation management and brand awareness (30%), product innovation (30%), business process improvement (25%), and discerning new business opportunities (20%).

On the other hand, Sinha et al. (2012) discuss the benefits of using “behavioral informatics” and “human resources analytics” for recruiting, training, internal communications, employee engagement, talent management, employee/employer branding, and employee life-cycle management. As evident, Sinha et al. (2012) focus on SMA application largely within the boundaries of a firm’s internal environment whereas the vendor success stories favor external applications.

Based on our definition and the above observations, one may extend the notions of “engagement” and “service” to apply not only to “customers” but also to “employees,” “business partners,” and in the case of socially-conscious businesses, all of the “society” that such a business impacts. Business SMA and its benefits could, in principle, apply to any and all such entities or sectors.

In the limited space available, we briefly describe some of the benefits along with illustrative, recent empirical research evidence. (We looked for independent empirical support, in part, as a means to mitigate “vendor bias” criticisms that may be levelled against Kurniawati et al.'s (2013) survey of vendor “success stories.”)

1. *Improve Marketing Strategy*: Customer-generated content usually contains valuable information about customers’ experiences with a product or service. Such information is available both in review websites such as eopin.com, Amazon, and Yelp and also in personal social networking sites such as Facebook, personal weblogs, and online forums. SMA can provide useful insights for developing and/or refining marketing/sales strategies.

As Kurniawati et al. (2013) note, most (75%) of the vendor success stories reported concern market strategy improvement. There is ample empirical work related to studying this benefit as well. For example, Hu et al. (2013) evaluate the relative impacts of text sentiment and star ratings on book sales at Amazon and determine that textual reviews (specifically, the two most accessible reviews – most helpful and most recent), unlike ratings, directly and significantly impact sales. Dellarocas et al. (2010) find that moviegoers show a propensity to review very obscure movies aside from very popular ones. They generalize that user review volume for lesser-known products may be increased by deliberately obfuscating the true volume of prior reviews.

2. *Better Customer Engagement*: SMA can be used to identify and target customer values and preferred customer channels for two-way, not just B2C, communication. Abrahams et al. (2013) evaluate multiple approaches for identifying customer values (e.g., customer requirements elicitation) with regard to automotive components by mining threads in three discussion forums (Honda Tech, Toyota Nation, and Chevrolet Forum). Goh et al (2013) analyze the brand community Facebook page (i.e., fan page) of a clothing retailer and the online community’s purchase information, to show that undirected communication does better with both informative and persuasive C2C communication, but directed does better with persuasive M2C (Marketer to Consumer, a form of B2C) communication.

3. *Better Customer Service*: Improved Customer Service is something that many firms strive for in today’s hyper-competitive settings. As an instance of SMA use to provide better customer service, Hill and Ready-Campbell (2011) describe a genetic algorithm-based opinion mining tool that helps identify stock picking experts from the online Motley Fool CAPS voting site, containing over 2 million picks made by over 770,000 registered users. They show that picking stocks recommended by these experts does better than picking stocks recommended by the S&P 500 or using input from all voters on the CAPS site. The net result is the identification of stock portfolios that provide better returns to the clients of a financial services firm using opinion mining than possible through using the opinions of its internal experts or an index like the S&P 500. By providing a better (more valuable) stock portfolio to its clients, the firm is also providing better client service.

4. *Reputation Management*: SMA may also be utilized to monitor and maintain or enhance a firm’s internal and/or external reputation related to brand, product/service, employer, employee, facility, and so on. Deloitte (2013) notes that a growing number of global financial services firms are instituting CRO (Chief Risk Officer) positions and, apart from financial risk, more CROs are instituting “stress tests” that take into account reputational risk, along with operational and regulatory risk, when assessing ability to withstand future industry downturns.

As an example of employer brand-reputation enhancement through use of social media (APCO Worldwide and Gagen MacDonald, 2011), a survey of 1000 full-time employees, employed for at least one year at firms with at least 500 employees found: 58% of respondents would rather work for a company that uses internal social media (ISM) tools, 51% said their companies use some form of ISM, 63% felt that their employers used ISM “well,” 61% felt that collaboration with colleagues was easier with ISM, and 60% felt that use of ISM was indicative of company innovation. Other interesting employer brand reputation-related findings are shown in Table 3.

| Observation | Proportion of Respondents At Companies Doing ... | |
|--|--|---|
| | “Well” with ISM | “Fairly with ISM or “Poorly” with ISM |
| Will likely continue as employee for the foreseeable future | 91% | 74% |
| Would likely encourage others to consider employment at company | 86% | 51% |
| Would recommend company’s products or services | 89% | 64% |
| Would give company benefit of the doubt when it’s facing litigation/crises | 88% | 55% |
| Would purchase company stock | 75% | 45% |

Table 3. Results of 3rd Annual Employee Engagement Survey (APCO Worldwide and Gagen MacDonald, 2011)

As an instance of reputation management-related research, Amblee and Bui (2012) analyzed the role of electronic Word-of-Mouth (eWOM) communication in a closed community of low-priced Amazon Shorts e-book readers. They find that eWOM is effective in conveying product (i.e., e-book) reputation, brand (i.e., author) reputation, and complementary goods (i.e., e-books in same category) reputation.

5. *New Business Opportunities*: There is documented evidence that SMA can also help reveal untapped business opportunities by helping identify new product/service possibilities. For example, Colbaugh and Glass (2011) develop a method to spot emerging “memes” (i.e., distinctive phrases which act as “tracers” for topics) and predict which will propagate wildly and result in significantly discussed new topics and trends. They argue that knowledge about the social network positions of meme originators is helpful in spotting memes that are likely to propagate wildly. The results of such insights could help a firm determine, for instance, where to open a new retail outlet, what feature(s) to include in a new product, or which opinion leaders to further engage.

Business SMA Challenges

Like any technology-enabled “big data” solution, Business SMA is not without its own set of challenges. However, as a nascent developing field, these challenges are also opportunities for further research exploration. In Table 4, we summarize key challenges identified. For ease of comprehension, we divide these challenges into two categories – those related to pre-analytics processing activities and those related to analytics processing activities. The former category pertains to: challenges with the processing of free-form, context-sensitive content; challenges given the use of specialized jargon; data validity concerns pertaining to the use of abbreviations, typos, and questionable credibility; data extraction difficulties due to data size, data/source variety, and the challenges of separating useful from useless information; and the complexities of processing, in real time, the (near) continuously streaming flow of data. The latter category focuses on difficulties stemming from the limited life of usable data and its time-varying nature; and methodological issues relating to developing integrative, multidisciplinary, big data-scalable analysis techniques.

We were able to locate very few references specifically identifying and addressing challenges pertaining to post-analytics processing in the reviewed literature. This may perhaps be an artifact of our particular literature review procedure. We believe, though, that there will very likely be considerable difficulties encountered with suitably packaging and disseminating actionable SMA results, especially with substantial levels of automation and in real-time, given the many pre-processing and processing

challenges noted. As such, research on semi-automated and automated, real-time post-SMA processing, if not already underway, should see growing interest in the years ahead by those engaged in Business Intelligence-related research (of which Business SMA is a part).

| Challenges | Source | Description |
|---|--------------------------|--|
| <i>Related to Pre-Analytics Processing Activities</i> | | |
| Context & Structure: free-form statements; unclear broader context | Best et al. (2012) | "...the brevity of most messages, the frequency of data ingest, and the context sensitivity of each message." |
| | Mosley Jr. (2012) | "One major consideration is that social media data tends to be informal..." ... the challenge becomes connecting the right set of social media data together to be able to understand the broader context of a conversation." |
| | Mayeh et al. (2012) | "The unstructured and distributed nature and volume of this information makes the task of extracting useful and practical information challenging." |
| | Ribarsky et al. (2013) | "Text messages from Twitter, Facebook, and several other social media services have general attributes such as unstructured content..." |
| Language Use: special symbols; slang use | Zeng et al. (2010) | "Social media applications are a prominent example of human-centered computing with their own unique emphasis on social interactions among users." |
| | Mosley Jr. (2012) | "There are certain symbols that actually do have a meaning and therefore extra care needs to be taken in cleansing the text." "...to understand sentiment would require a more thorough investigation into the ways that users communicate sentiment, and then attempting to capture those sentiments within the data in a structured way." |
| | Fan and Gordon (2013) | "Language issues add further complications as businesses begin to monitor and analyze social media conversations around the world." |
| Data Validity: abbreviations, typos, and credibility issues | Asur and Huberman (2010) | "...it was difficult to correctly identify tweets that were relevant to those movies. For instance, for the movie 2012, it was impractical to segregate tweets talking about the movie, from those referring to the year." |
| | Mayeh et al. (2012) | "Social media data is unstructured, distributed and of uncertain credibility." "Social media data includes spam, which are non-sensible or gibberish text. There are some intentional misspellings used to show commenter's sentiment." |
| | Mosley Jr. (2012) | "...issues with misspellings and abbreviations will be a larger challenge...there are no system edits that ensure the social media data that was captured is accurate, and this may result in false information and statements that are driven by pure emotion rather than fact." |

| | | |
|---|---------------------------|---|
| | Ribarsky et al. (2013) | "...intrinsic uncertainty as to the validity of the messages." |
| Data Extraction: diversity, scope of social media; isolating useful input | Melville et al. (2009) | "An important consideration is to avoid crawling, parsing and storing parts of the blog sub-universe that are irrelevant from a marketing perspective." |
| | Colbaugh and Glass (2011) | "...most memes receiving relatively little attention and a few attracting considerable interest." |
| | Chae et al. (2012) | "The relevant messages for situational awareness are usually buried by a majority of irrelevant data." |
| Streaming Nature: continuously flowing data | Melville et al. (2009) | "However, the set of domains to monitor may change often, requiring classifiers to adapt rapidly with the minimum of supervision." |
| | Barbieri et al. (2010) | "Stream reasoning moves from this processing model to a continuous model, where tasks are registered and continuously evaluated against flowing data." |
| | Zeng et al. (2010) | "Social media data are dynamic streams, with their volume rapidly increasing. The dynamic nature of such data and their sheer size pose significant challenges to computing in general and to semantic computing in particular." |
| | Best et al. (2012) | "The real-time nature of social media analytics for emergency management poses interesting visualization challenges." |
| Related to Analytics Processing Activities | | |
| Analysis Time Frame: time-varying impacts; limited usable data life | Asur and Huberman (2010) | "For each movie, we define the critical period as the time from the week before it is released, when the promotional campaigns are in full swing, to two weeks after release, when its initial popularity fades and opinions from people have been disseminated." |
| | Barbieri et al. (2010) | "Data streams are unbounded sequences of time-varying data elements that form a continuous flow of information. Recent updates are more relevant because they describe the current state of a dynamic system." |
| | Colbaugh and Glass (2011) | "...which will go on to attract substantial attention, and to do so early in the meme lifecycle...although memes typically propagate for weeks, useful predictions can be made within the first twelve hours after a meme is detected." |
| | Chae et al. (2012) | "There is a need for advanced tools to aid understanding of the extent, severity and consequences of incidents, as well as their time-evolving nature" |
| | Mosley Jr. (2012) | "Trending topics can literally begin in an instant and can become widespread very fast, and if the analysis occurs too long after the topic is trending, it may be too late for the company to do anything useful about it." |
| | Boden et al. (2013) | "A user can adopt the role of "novice user" the first time she registers with a particular online community, and achieve the role of "expert" months after." |

| | | |
|---|---------------------------|--|
| | Ribarsky et al. (2013) | “Events are bursts of activity over a relatively short time period, the time scale depending on the category of the temporal data.” |
| Methodology: integrative, multidisciplinary big data-scalable approaches | Melville et al. (2009) | “Although, clustering and topic modeling techniques can find sets of posts expressing cohesive patterns of discussion, for generating marketing insight we need to identify clusters that are also novel or informative compared to previous streams of discussion.” |
| | Zeng et al. (2010) | “Social media intelligence research calls for highly integrated multidisciplinary research. Although this need has been reiterated often in this growing field, the level of integration in the existing research tends to be low.” |
| | Colbaugh and Glass (2011) | “...in order to identify features of social diffusion which possess predictive power, it is necessary to assess predictability using social and information network models with realistic topologies.” |
| | Boden et al. (2013) | “However, analysts usually face significant problems in scaling existing and novel approaches to match the data volume and size of modern online communities.” |

Table 4. Current challenges of social media analytics

Given the many challenges noted in Table 4, there have been attempts to articulate partial solutions in particular contexts. We provide examples of suggested workarounds with regard to some of the challenges suggested by a few authors:

With regard to Context & Structure-related challenges, Mosley Jr. (2012) describes a step-by-step procedure for cleansing and structuring over 68,000 free-form Allstate insurance-related tweets to a set of 116 keywords associated with the tweets as a prelude to examining keyword associations using both cluster analysis and association rule mining. However, while the procedure for structuring free-form content is relatively easier with (at most) 140 character tweets, the same cannot be assumed with general free-form content (as with, blogs, for example).

Language Use concerns are amongst the most difficult to deal with. Fan and Gordon (2013) allude to the possible use of machine translation to help with the mining of multi-lingual content. However, machine translation is as yet a developing area of research and application and is no panacea. Mosley Jr. (2012) notes that one solution is to strip gathered tweets of all special symbols such as punctuations, quotation marks, parentheses, currency symbols, etc. However, sometimes such symbols so convey useful meaning as with using a smiley. Therefore, care must be taken in pre-determining which symbols may be safely stripped away and which must be retained based on context (e.g., @ and # in the case of Twitter).

White et al. (2012), in reference to Data Validity, note that as much as 50% of Twitter content is estimated to be spam although there is a declining trend. Detecting real content from fake content is a non-trivial task barring some special cases. As part of their botnet detection attempts on Twitter, White et al. observed a large number of identical Tweets from different accounts that were not re-tweets and all dispatched within a few hours of one another. An examination of the associated websites leads them to conclude that these were filler messages intended to fool the anti-spam defenses of Twitter. Whereas it was possible to detect invalid content in this special circumstance, circumventing Data Validity-related challenges continues to be amongst the most difficult of challenges facing SMA.

The diversity of social media each containing voluminous, diverse information results in unique challenges pertaining to Data Extraction. Melville et al. (2009) consider the question of avoiding crawling, parsing, and storing irrelevant parts of a blog sub-universe. They recommend a “focused snowball sampling” procedure where a text classifier first helps determine relevant links in a given blog and a web

crawler adds blogs that these links point to. The process is re-iterated with each new blog identified until some predetermined “degrees of separation” count has been reached.

Best et al. (2012) describe features of a prototype system (the Scalable Reasoning System (SRS)) for real-time visualization of emergency situation management-related Twitter information for use by the City of Seattle. They note that despite the fact that tweets may be continuously flowing, what was important to the city was not continual, but real-time, updates of the visuals to avoid viewer disorientation. Their solution to provide real-time updates was to embed a clock in the user interface to remind users to request timely periodic refreshes. A refresh indicator also displays the number of new tweets accumulated in the interim since the last refresh. The user can decide to ask for a refresh based on either elapsed time, the number of new tweets, or both. We believe that timely updates, rather than continual updates, are likely the preferred choice in most business decision-making contexts as well.

With regard to challenges introduced by the Analysis Time Frame, Ribarsky et al. (2013) discuss the impacts of time-varying behavior of tweet metadata (such as, retweet count, follower count, favorite count, etc.) on an SMA attempt. All metadata values are only accurate as of the moment of tweet collection. Therefore, a tweet collected shortly after generation may show a retweet count value of zero or a follower count of ten, say, whereas if the same tweet were captured later, these values could be substantially higher. If one were to collect tweets only after a sufficient time period has elapsed, then the time-varying behavior of the metadata will be lost as only the most recent values would be available. Finally, if one were to repeatedly capture the same tweets at various epochs to get around the difficulties just mentioned, then one could end up with vastly larger data set and the tweet-gathering process could take many months. The authors then devise a mitigating solution for long-running tweet-collection settings that takes advantage of the fact that an original tweet are embedded within retweets of the same tweet. Thus, one could focus on gathering only popular/influential tweets that tend to get re-tweeted repeatedly and the re-tweets capture within them the time-varying metadata.

As Table 4 notes, there often is a need for utilizing integrative, multidisciplinary, big data-scalable methods during an SMA exercise. Zeng et al. (2010) note that although this has been a well-known fact for quite some time, substantial progress in this regard is as yet lacking. The situation has not improved radically since. One instance of successful integration is that utilized by Yang et al (2011) in mining web forums maintained by hate groups for radical opinions. They use both a machine learning approach and a semantic-oriented approach to extract four types of features that characterize radical opinion in posts at such forums. They next apply the best of three classification methods – Support Vector Machine, Naive Bayes, and AdaBoost -- to classify new posts as being either radical or benign.

Concluding Remark

In this paper, we undertake a review of 27 research papers related to social media analytics (SMA). Drawing on and augmenting available SMA definitions, we develop an integrated, unifying definition of Business SMA, with a view toward providing a nuanced starting point for future Business SMA research. Our definition goes beyond being entirely customer-focused, encompasses both the external and internal organizational environs, and goes beyond intelligence gathering to also accommodate such business activities as sense making, insight generation, problem/opportunity detection & solution/exploitation, and decision making.

We also draw on the reviewed literature to identify several benefits of Business SMA. We elaborate on some of these benefits and present recent empirical evidence in support of our observations. We identify and categorize several challenges facing Business SMA today, along with supporting evidence from the literature documenting such challenges and offering mitigating solutions in particular contexts. However, comprehensive, generally-applicable solutions to several of these challenges remain open questions for future investigations. Accordingly, this research study helps further conceptual understanding of Business SMA and its many aspects, grounded in recent empirical work, and as a basis for further research. Whereas issues like limitations and deployment strategies for Business SMA are also important, we do not address them in this space-limited setting.

This study is not without limitations. Paramount among these is our focus only on very recent literature (2008-2013), given the paper word-length constraints. Tables 1, 3 and 4 and related discussions may be

expanded using a larger time window (say, 2004-Present) for literature extraction and review. Our literature search also focuses on “social media analytics/intelligence” as keywords. It may be worthwhile expanding the search to also include papers using “social network analysis,” “sentiment analysis,” “text mining,” and “web mining” as key words. While these keywords may net several irrelevant papers, insofar as our focus is on analyzing only social media content, we may yet avoid overlooking important work. We also considered a few conference proceedings papers and industry white papers for insights not available, as yet, in the form of published academic journal articles. This, perhaps, is something to be expected, given the infancy of academic SMA research. Future research based on this paper will address several of the cited limitations.

REFERENCES

- Abrahams, A.S., Jiao, J., Fan, W., Wang, G.A., and Zhang, Z. 2013. "What's Buzzing in the Blizzard of Buzz? Automatic Component Isolation in Social Media Postings," *Decision Support Systems* (55:4), pp. 871-882.
- Ambler, N., and Bui, T. 2011. "Harnessing the Influence of Social Proof in Online Shopping: The Effect of Electronic Word of Mouth on Sales of Digital Microproducts," *International Journal of Electronic Commerce* (16:2), pp. 91-114.
- APCO Worldwidw and Gagen MacDonald. 2011. "The 3rd Annual Employee Engagement Survey: Unleashing the Power of Social Media within Your Organization," white paper. from <http://www.gagenmacdonald.com/2013/3rd-annual-employee-engagement-research-study/>
- Barbieri, D., Braga, D., Ceri, S., Valle, E.D., Huang, Y., Tresp, V., Rettinger, A., and Wermser, H. 2010. "Deductive and Inductive Stream Reasoning for Semantic Social Media Analytics," *Intelligent Systems, IEEE* (25:6), pp. 32-41.
- Best, D.M., Bruce, J., Dowson, S., Love, O., and McGrath, L. 2012. "Web-Based Visual Analytics for Social Media," in *AAAI ICWSM SocMedVis: Workshop on Social Media Visualization (AAAI Technical Report WS-12-03)*, Dublin, Ireland, pp. 2-5.
- Boden, C., Karnstedt, M., Fernandez, M., and Markl, V. 2013. "Large-Scale Social-Media Analytics on Stratosphere," in *Proceedings of the 22nd international conference on World Wide Web companion, Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee*, pp. 257-260.
- Chae, J., Thom, D., Bosch, H., Jang, Y., Maciejewski, R., Ebert, D.S., and Ertl, T. 2012. "Spatiotemporal Social Media Analytics for Abnormal Event Detection and Examination Using Seasonal-Trend Decomposition," in *IEEE Conference on Visual Analytics Science and Technology (VAST)*, Seattle, WA, USA: IEEE, pp. 143-152.
- Colbaugh, R., and Glass, K. 2011. "Detecting Emerging Topics and Trends via Social Media Analytics," in *Proceedings of the 2011 IADIS International Conference e-Commerce*, Rome, Italy, pp. 51-51.
- Dellarocas, C., Gao, G., and Narayan, R. 2010. "Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?," *Journal of Management Information Systems* (27:2), pp. 127-158.
- Deloitte. 2013 . "Global Risk Management Survey, Eighth Edition: Setting a Higher Bar," white paper, Deloitte Touche Tohmatsu Limited. from http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/us_fsi_aers_global_risk_management_survey_8thed_072913.pdf
- Fan, W., and Gordon, M. D. 2013. "Unveiling the Power of Social Media Analytics," *forthcoming in Communications of the ACM*.
- Goh, K.-Y., Heng, C.-S., and Lin, Z. 2013. "Social Media Brand Community and Consumer Behavior: Quantifying the Relative Impact of User-and Marketer-Generated Content," *Information Systems Research* (24:1), pp. 88-107.
- Grubmüller, V., Götsch, K., and Krieger, B. 2013. "Social Media Analytics for Future Oriented Policy Making," *European Journal of Futures Research* (1:1), pp. 1-9.
- Grubmüller, V., Krieger, B., and Götsch, K. 2013. "Social Media Analytics for Government in the Light of Legal and Ethical Challenges," in *International Conference for E-Democracy and Open Government*, Donau-Universität-Krems, Austria, pp. 185-195.
- Hill, S., and Ready-Campbell, N. 2011. "Expert Stock Picker: The Wisdom of (Experts in) Crowds," *International Journal of Electronic Commerce* (15:3), pp. 73-102.

- Hu, N., Sian, K.N., and Reddy, S.K. 2013. "Ratings Lead You to the Product, Reviews Help You Clinch It? The Mediating Role of Online Review Sentiments on Product Sales," *Decision Support Systems* (57), pp. 42-53.
- Kaplan, A.M., and Haenlein, M. 2010. "Users of the World, Unite! The Challenges and Opportunities of Social Media," *Business Horizons* (53:1), pp. 59-68.
- Kurniawati, K., Shanks, G., and Bekmamedova, N. 2013. "The Business Impact of Social Media Analytics," in *Proceedings of the 21th European Conference on Information System*, Utrecht, The Netherlands, p. 48-61.
- Lau, R. Y. K., Xia, Y., and Li, C. 2012. "Social Media Analytics for Cyber Attack Forensic," *International Journal of Research in Engineering and Technology* (1:4), pp. 217-220.
- Mayeh, M., Scheepers, R., and Valos, M. 2012. "Understanding the Role of Social Media Monitoring in Generating External Intelligence," in *Proceedings of the 23rd Australasian Conference on Information Systems*, Geelong, Australia, pp. 1-10.
- Melville, P., Sindhwani, V., and Lawrence, R. 2009. "Social Media Analytics: Channeling the Power of the Blogosphere for Marketing Insight," in *Proceedings of the 1st Workshop on Information in Networks (WIN 2009)*, Manhattan, NY, USA.
- Mosley Jr., R.C. 2012. "Social Media Analytics: Data Mining Applied to Insurance Twitter Posts," in *2012 Casualty Actuarial Society E-Forum*, Arlington, Virginia, USA, pp. 1-36.
- Ribarsky, W., Xiaoyu Wang, D., and Dou, W. 2013. "Social Media Analytics for Competitive Advantage," *Computers & Graphics* (38), pp. 328-331.
- Sinha, V., Subramanian, K.S., Bhattacharya, S., and Chaudhary, K. 2012. "The Contemporary Framework on Social Media Analytics as an Emerging Tool for Behavior Informatics, HR Analytics and Business Process," *Journal of Contemporary Management Issues* (17:2), pp. 65-84.
- Sterne, J. 2010. *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. Wiley.
- White, J. S., Matthews, J. N., and Stacy, J. L. 2012. "Coalmine: an experience in building a system for social media analytics," in *Proceedings of SPIE*.
- Yang, M., Kiang, M., Ku, Y., Chiu, C., and Li, Y. 2011. "Social Media Analytics for Radical Opinion Mining in Hate Group Web Forums," *Journal of Homeland Security and Emergency Management* (8:1).
- Zeng, D., Chen, H., Lusch, R., and Li, S.-H. 2010. "Social Media Analytics and Intelligence," *Intelligent Systems, IEEE* (25:6), pp. 13-16.