Teaching Social Media Analytics: An Assessment Based on Natural Disaster Postings

Tiong T. Goh
School of Information Management
Victoria University of Wellington
Wellington, New Zealand
tiong.goh@vuw.ac.nz

Pei-Chen Sun
Graduate Institute for Information and Computer Education
National Kaohsiung Normal University
Kaohsiung, Taiwan
sun@nknu.edu.tw

ABSTRACT

Unstructured data in social media is as part of the "big data" spectrum. Unstructured data in Social media can provide useful insights into social phenomena and citizen opinions, both of which are critical to government policy and businesses decisions. Teachers of business intelligence and analytics commonly use quantitative data from sales, marketing, finance and manufacturing to demonstrate various analytics concepts in a business context. However, researchers have seldom used social media data to analyze social behavior and communication. In this study we aim to demonstrate an assessment structure for teaching social media analytics concepts with the goal of analyzing and interpreting social media content. We base this assessment on forum postings regarding two recent events: the Christchurch earthquake in New Zealand, and the Japanese earthquake and tsunami. The aim of the assessment is to discover social insights. We base the assessment structure on Cooper's Analytics Framework to cover such concepts as term frequency (TF), term frequency–inverse document frequency (TFIDF), data visualization, sentiments and opinions analysis, the Nearest Neighbor K-NN classification algorithm, and Information Diffusion theory. We review how the students performed on the assignment that used this assessment, and we make recommendations for future studies.

Keywords: Social media, Data analytics, Text processing, Social behavior

1. INTRODUCTION

Recent interest in "big data" analytics has escalated the demand for data analytics specialists. According to Deloitte (2012) there will be both a strong demand for skilled big data professionals in the US over the next five years and a shortage of 200,000 IT professionals with deep analytics skills (Manyika et al., 2011). Chiang, Goes, and Stohr (2012) recently suggested that business intelligence and analytics education program development provides a unique opportunity for the information systems discipline. They contended that the IS discipline should address the challenges of big data—especially in business school IS programs—to meet the growing demand for graduates who can aggregate, analyze, model and evaluate organizational data. To meet this challenge, schools are expanding business programs to develop IS graduates with the business, analytics, IT, and communications skills required for successful future business analytics leaders (Henschen, 2013; NUS-SOC, 2013). Chiang, Goes, and Stohr (2012) suggested that since the IS field has traditionally focused on quantitative, structured data, there is a need to address the interpretive aspect of analytics.

The increasingly popular social media applications such as Twitter, Facebook, blogs, and online product reviews lie within the "big data" spectrum and contain relevant information for business decision making. A survey by Wixom et al. (2014) noted the increasing interest in text analysis to extract information from semi-structured and non-structured data. Natural language processing and semantic interpretation are increasingly popular analytics solutions (Gorman & Klimberg, 2014). As technology evolves, undergraduate IS Business Intelligence and Analytics (BI&A) curriculum should include social media analytics (Topi et al., 2010).

In light of the above, we aim to demonstrate an assessment structure for teaching social media analytics concepts with the goal of analyzing and interpreting social media content. The proposed assessment supports the sharing of reusable teaching resources from different social media content for teaching social analytics in the IS curriculum. We have organized the remainder of this paper as follows. First we review the literature on business analytics teaching and the literature on analytics frameworks for assessment design. We then discuss our methodology for data collection and pre-processing, after which we present the learning enquiries and results. Next we discuss our findings. We conclude with our solutions, the limitations of this study, and suggestions for future studies.

2. LITERATURE REVIEW

2.1 Teaching Business Analytics

While new business analytics programs have multiplied dramatically since 2010, teaching resources such as datasets and case studies remain scarce for professors in this field (Wixom et al., 2014). A Pro-Quest database search of the literature for teaching cases and learning issues relating to business and data analytics in the IS discipline yielded few relevant results. There are not many IS publications on research regarding business analytics teaching and learning. Marchand and Peppard (2013) recommend that teachers put more effort into producing teaching resources and cases that focus on how people create and use information, and how to frame questions that data analytics might answer to increase our knowledge and understanding. Technology is changing rapidly, stressing the need for sharing and reusing innovative business analytics teaching practices and resources to resolve challenges concerning the core body of knowledge and the design and delivery of BI&A programs (Marjanovic, 2013).

In the IS discipline, BI&A includes three evolving categories: BI&A 1.0, 2.0 and 3.0. BI&A 1.0 comprises Database Management System transaction based structured content such as credit card and purchase transaction data. BI&A 2.0 comprises web-based unstructured content such as social media. BI&A 3.0 comprises the most recent developments in mobile and sensor based content, including wearable sensors, nano-sensors and mobile phone sensors (Chen, Chiang, and Storey, 2012). This classification helps guide institutions offering BI&A programs, specifically with respect to resources and technology adoption.

Educators should assess the teaching resources and skill sets needed for each category. The adoption of different technologies will influence the academic alliances programs and the ability to access those BI&A resources. For instance, the IBM Academic Initiative requires the adoption of Cognos, ILOG and SPSS.

IS schools must also balance their BI&A content to meet the diverse demands of the market. An interdisciplinary track covering IS, Statistics, Marketing and Finance is becoming a common approach for business analytics programs, and can potentially attract students from a wide range of backgrounds (Wixom et al., 2014). The applied business research class for MBA students at Utah State University's Jon M. Huntsman School of Business (Chudoba, Hauser, and Olsen, 2010) took a BI&A 1.0 approach using Excel, SQL and SPSS to analyze

structured datasets. Chudoba, Hauser, and Olsen (2010) concluded that students exposed to different tools can benefit from being able to match the appropriate tool to the type of analysis, and are better able to gauge the amount of data needed to make statistically significant, informed decisions. One limitation of this approach is that they did not include unstructured data (e.g., consumer opinions) in the analysis to strengthen the decision making process.

Elizondo, Parzinger, and Welch (2011) employed project based learning in an undergraduate business intelligence class, using SAS Enterprise Miner software for predictive analysis and data mining. The approach helped students understand the importance of data integrity and statistical analysis for decision making in different scenarios.

In an approach closer to BI&A 2.0, Baylor University (Edgington, 2011) has offered a graduate level course for Masters of Science in Information Systems (MSIS) and Masters of Business Administration (MBA) students. Using SPSS Text Analytics software, the course focuses on analysis, validation, and the use of technology. The interdisciplinary nature of the course has gained positive feedback and the undergraduate program has adopted the course.

An example of BI&A 3.0 is the Singapore-MIT Alliance for Research and Technology at the Center for Environmental Sensing and Modeling, which is developing analytics systems that use a network of wireless sensors called WaterWiSe to identify state changes and detect water leaks that occur over time (Whittle, 2014).

2.2 Analytics Framework and Assessment Design

Davenport and Harris (2007) define analytics as "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact based management to drive decisions and actions." Analytics is more than just the analytical methodologies or techniques used in logical analysis. It is a process of transforming data into actions through analysis and insights in the context of organizational decision making and problem solving.

Kaplan and Haenlein (2010) defined social media as a group of Internet based applications that enable the creation and exchange of user generated content. In this context, social media analytics is the technology used to monitor, measure, and analyze activity by users of the Web 2.0 (and beyond) in order to provide information for business decisions (Abbasi, 2012). Teaching such concepts requires a systematic analytics framework to provide the underpinnings for an assessment that demonstrates the concepts of social media analytics. Such a framework allows analysts to explore the underlying features and assumptions, and to transition from one target level to the next (Cooper, 2012).

According to Cooper (2012) a framework that is "descriptive, rather than definitive, in its approach" best accommodates the real-world complexity of analytics. Based on identifying the general characteristics of analytics, Cooper devised an extensible and adaptable framework. The framework consists of different dimensions grouped according to the similarity of their characteristics. To ensure the relevance of the framework to real-world problems, Cooper identified dimensions describing interactions between people, and interactions between people and data,

things or ideas. We adopted Cooper's framework to establish our social media case. Table 1 provides an example of our social media assessment.

Dupin-Bryant and Olsen (2014) used one of several general approaches for developing assessments in analytics and data visualization. Using the 5E instructional model

(a.k.a., the "BSCS 5Es instructional model") they developed a heat map tutorial based on stock market data. The 5E model consists of the following phases: engagement, exploration, explanation, elaboration, and evaluation. The 5E model has strong foundation in educational theory and research.

Part A: Analysis subjects, objects and clients	Analysis subjects: postings from an online newspaper forum Analysis clients: social policy makers, government agencies, disaster management teams Analysis objects: citizen communication behaviors
Part B: Data quality and source	Open data from an online forum; Data captured through XML transformation; Pre-processing at posting level
Part C: Orientation and objectives	Orientation: Present Objective type: description and performance
Part D: Technical approach	Both a descriptive and quantitative approach using text processing, classification and model fitting
Part E: Presentation	Numeric: frequency count; Visual: chart, tables, graphic
Part F: Embedded theories and reality	Social communication, social identity, compassionate and disaster communication and information diffusion theory

Table 1. Framework Mapping for Current Assessment

3. METHODOLOGY

3.1 Background Information

A magnitude 6.3 earthquake struck the Canterbury region in New Zealand's South Island on Tuesday, February 22, 2011, at 12:51 p.m. The earthquake center was two kilometers west of Lyttelton, and ten kilometers southeast of Christchurch. In total, the earthquake killed 185 people, ranking it as the second-deadliest natural disaster recorded in New Zealand.

On March 11, 2011, at 5:46:23 a.m., a magnitude 9.0 earthquake struck off the coast of Honshu, Japan's most populous island, near Sendai, the capital city of Miyagi Prefecture. The earthquake triggered a devastating tsunami that swept over cities and farmlands in northern Japan. The disaster reportedly killed at least 15,690 people (MCEER, 2011). After news outlets reported these disasters, readers posted numerous comments in an online newspaper (www.stuff.co.nz).

3.2 Procedures

We extracted the forum postings related to these two disasters, and they represent Parts A and B of the analytical framework. Since the online paper is a local paper, the participants in the forum are likely to be a sample of the local population. The "Event A" data file represents the Christchurch earthquake, and contains 769 postings extracted from February 22 to March 1. The "Event B" data file represents the Japan earthquake and tsunami, and contains 438 postings from March 11 to March 15. The data structure consists of *Name, Date* and *Postings*.

We stored these data in an Excel format for further processing. We performed pre-processing to remove any duplicate postings. We gave these two data files, together with an assignment description, to 45 IS students enrolled in a one-trimester, advanced database course containing a social media analytics module as part of an assessment. The students had approximately four weeks to complete the assignment. They were required to analyze the data using RapidMiner text processing software to compare the content

and to interpret the results. Based on the mapping of the analytical framework, we formulated the students' main learning enquiries as follows:

(1) What are the topics of interest (themes) in the postings?
(2) Are the postings on the Christchurch earthquake more compassionate and sympathetic than those on the Japanese earthquake and tsunami?
(3) What are the sentiments regarding nuclear power, and the percentage of posts on this topic?
(4) What is the classification performance of the K-NN algorithm if one randomly selects a posting from the two event files?
(5) What is the rate of information diffusion of the two events?

We also wanted to evaluate how well students performed in this assignment in order to identify shortcomings and facilitate future improvements to the assessment itself.

Students loaded the data files into RapidMiner to perform text analytics. RapidMiner is an open-source data mining tool capable of various analyses, including unstructured data processing, modeling and classification (Rapid-i, 2012). The software processed the files with these basic steps: (i) Transform Cases, which transforms all characters in a document to either lower case or upper case, respectively; (ii) Tokenize, which splits the text of a document into a sequence of tokens; (iii) Filter Stop Words, which filters English stop words from a document by removing every token which equals a stop word from the built-in stop word list; (iv) Generate n-Gram, which creates term n-Grams of tokens in a document (a term n-Gram is defined as a series of consecutive tokens of length n); (v) Filter Tokens, which filters tokens based on the number of characters they contain; and (vi) Stem (Porter), which stems English words using the Porter stemming algorithm, applying an iterative, rule-based replacement of word suffixes to reduce the length of the words until a minimum

length is reached. Figure 1 depicts a typical text processing sequence used in RapidMiner. These initial processes represent Parts A, B and C of the analytical framework in Table 1, while the subsequent analysis and results represent Parts D, E and F of the framework.
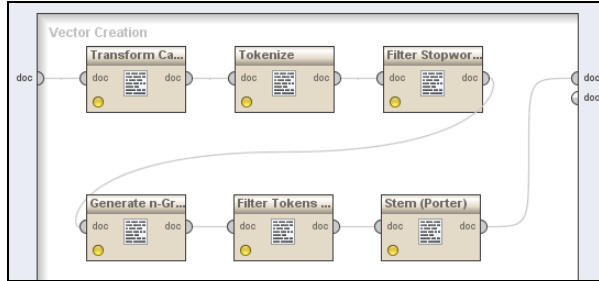

Figure 1. RapidMiner Text Processing Steps

4. LEARNING ENQUIRIES AND RESULTS

4.1 Word Frequency

Determining the themes of the postings is an important aspect of analyzing social media content. Word frequency can provide a quick summary of such themes. Term frequency (TF) is the number of times a term occurs in a document. We obtained the term frequency–inverse document frequency (TFIDF) by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient multiplied by the term frequency. Both the term frequency (TF) and term frequency–inverse document frequency (TFIDF) parameters are available in the RapidMiner text analytics module.

Students extracted the top 20 keywords from each file to compare the two events. Table 2 depicts the term frequency for the two events. The term frequency shows that most of the words in this dataset are related to compassion (Crawford et al., 2011; Thesaurus.com, 2013). Words such as "thoughts," "prayers," "love" and "help" are typical responses to crises (Qu et al., 2011; Wise, 2004). The posters also frequently mentioned disaster locations, so term frequency statistics clearly identified Christchurch (TF:416) and Japan (TF:331) as the crisis locations. Term frequency also showed that the crises clearly involved "people" (A:204; B:155). The posters also frequently mentioned the nature of the disaster, i.e. "earthquake."

Christchurch		Japan	
Term	Term Frequencies	Term	Term Frequencies
christchurch/chch	416	japan	331
thoughts	238	people	155
people	204	quake/earthquake	139
prayers	175	god	83
family	128	thoughts	78
god	112	nuclear	68
love	106	help	67
friends	87	hope	61
hope	86	prayers	61
help	85	time	56
thinking	85	power	54
time	84	world	53
strong	82	nz	52
quake/earthquake	81	christchurch	45
know	77	tsunami	45
city	75	affected	42
heart	70	think	34
safe	69	pray	32
nz	68	disaster	31
go	66	going	31

Table 2. Comparing Term Frequency

4.2 Data Visualization

The goal of visualization is to facilitate the answering of questions, whether anticipated or not, until patterns emerge from the data in the visualization (Stacey, Salvatore, & Jorgensen, 2013). Visualization can sometimes make patterns more easily detectable. It can also provide a more intuitive understanding of the data and compress a large amount of data into a smaller area (Stacey, Salvatore, and Jorgensen, 2013). Visualizing text content is similar to visualizing term frequency.

A word cloud is a popular word visualization typically associated with keywords and text data. Figure 2 shows the word cloud of the two events. Though we did not normalize the term frequency table (Table 2) to the number of posts, the word cloud produced a normalized visualization. The normalized word cloud makes it immediately obvious that the words "thoughts" and "prayers" are relatively more salient in the Christchurch event than in the Japanese event. Other readily available data visualization tools which are suitable for a business oriented IS curriculum include Google, Excel, Tableau and Microsoft Report Builder.

Figure 2. Word Cloud

4.3 Compassion and Sympathy

Knowing what other people think and say may aid significantly in the decision-making process. We can extract sentiments and opinions expressed in social media to help individuals make informed decisions. In disaster situations, postings that relate to compassion and attribution are commonly reported (Wise, 2004). We often associate compassion with times of trouble. The Macmillan dictionary (2013) defines compassion as "feeling sympathy for someone who is in a bad situation because you understand and care about them [sic]". According to Solomon (1998), in compassion, "one suffers with the other, but one need not actually feel his or her pain in any sense." Solomon (1998) also notes that compassion involves a focus on the other, and a desire for the other to overcome adversity, but that one "need not be intimately acquainted with a person or a creature in any sense in order to feel compassion for him."

Since people express compassion through words and actions, we identified the words most frequently associated with compassion and sympathy. For example the posting "*All my thoughts go out to the people of Christchurch. We can only hope that things don't get any worse than it already seems to be =(*" contains the compassionate words "thoughts" and "hope." Comparing the frequency of the compassionate words, we investigated the writers' level of compassion regarding the two events. Table 3 depicts the compassionate words, including stemming. After normalization to the total number of postings, a T-test showed that these compassionate words are more frequently used in Event A than in Event B (p=0.025 t=2.601). Hence, we conclude that writers were more compassionate and sympathetic about Event A. A possible reason is that Event A was geographically closer to the writers, giving it a stronger sense of immediacy than far away Event B. This is also evident from the frequency of the words "family" (event A:158, event B:32) and "friends" (event A:87 event B:23), suggesting that many postings involved someone close to the writer.

Compassionate Words	Event A Frequency	Event B Frequency	Event A Normalized	Event B Normalized
thoughts	238	78	.31	.18
prayers	175	93	.23	.21
love	106	20	.14	.05
thinking	102	49	.13	.11
strong	82	30	.11	.07
heart	94	32	.12	.07
bless	70	29	.09	.07
god	115	83	.15	.19
hope	94	64	.12	.15
sorry	28	9	.04	.02
sad	47	15	.06	.03
xoxoxo	104	8	.14	.02

Table 3. Comparing Compassionate Words

4.4 Emoticons and Special Symbols
Emoticons are strings of symbols widely used in text based communications such as SMS, tweets, e-mail, comments and forums (Ptaszynski et al., 2010). The use of emoticons has become an important part of daily communication (Maness, 2008). For text based communication analysis, these icons and symbols represent the body language and the emotional state of the writers. Table 4 represents three types of symbols used in the two events, mainly "Sad face," "Smiling face" and "Hugs and Kisses." Closer analysis of the use of "Smiling face" shows that writer used it with a compassionate tone rather than a teasing tone, as in "*Our thoughts and prayers are with you!! Keep strong :).*" Hence, similar to the finding of the compassionate word frequency, the compassionate symbol frequency is higher in Event A than in Event B. This further supports our previous finding that writers are more compassionate about Event A than Event B, again, possibly because Event A has a stronger sense of immediacy.

Symbol	Event A	Event B
:(	18	4
=(	3	-
:-(	2	3
:[	-	1
:)	4	1
x (or multiple x)	70	7
xo (or multiple xo)	34	1

Table 4. Emoticons and Symbols

4.5 Sentiments on Nuclear Power
The term frequency of "nuclear" and "power" were rather high in the postings (see Table 2) and visible in the word cloud. This suggests another potential topic of interest within the postings. As there was no nuclear facility in New Zealand, and Event A happened before Event B, Event A contains no postings about nuclear power.

However Event B contains 44 posts (10.05% of the total posts) that mention the topic of nuclear power. Of those, only 5 posts (11.36%) reflect on the topic positively, while the remaining 39 (88.64%) reflect on the topic negatively. These negative postings are mainly opinion related, criticizing the dangers of nuclear power and suggesting reasons for not adopting nuclear power in New Zealand. The opinions expressed reflect the environmental and political culture of this country which has been traditionally opposed to nuclear power.

4.6 Classification of Posts
Classifying postings to determine topic types, emotional states, data sources, relevance, and the positivity or negativity of opinions is an important aspect of content analysis. In this assessment, we explore text classification (TC): the task of automatically detecting one or more predefined categories that are relevant to a specific document (Sebastiani, 2002). For example, newspapers might use an automated system to choose the most suitable category for a given classified ad. Filtering systems use TC to block the delivery of unwanted feeds (Sebastiani, 2002). The goal of the current exercise is to classify a posting as having been generated from Event A vs. Event B. This kind of TC can facilitate search efficiency by first navigating to the node and then restricting the search to a branch of interest (Sebastiani, 2002).

RapidMiner has various types of classification algorithms: decision tree, support vector machine, nearest neighbor, and naive Bayes. We used the K-Nearest Neighbor K-NN algorithm (K=5 with a 2-gram and TFIDF as the feature vector) to determine whether we could identify a randomly selected posting as belonging to Event A vs. Event B. According to expert nominations, citation counts and a community survey, the K-NN is among the top ten most influential data mining algorithms (Wu et al., 2008). Besides performing well in most situations, the K-NN is also one of the simplest algorithms, easily understood by business students. Using the Euclidean distance, the K-NN algorithm searches the pattern space for the k training examples that are closest to the unknown example.

In this exercise the K-NN method achieved an accuracy level of 82.02%. Table 5 depicts the resulting confusion matrix. Interestingly, the software misclassified postings as belonging to Event A far less often than it misclassified postings as belonging to Event B. This may suggest that Event A has strong discriminant words associated with it, and Event B does not.

	True A	True B	Precision
Predict A	705	153	82.17%
Predict B	64	285	81.66%
Recall	91.68%	65.07%	

Table 5. Classification Performance Accuracy: 82.02%

4.7 Information Diffusion
Information diffusion refers to the spread of a piece of information through a network environment. It is analogous to the adoption of an innovation through a network environment. The Bass diffusion model is a common model that describes the diffusion mechanism (Mahajan, Muller, & Bass, 1990). The diffusion process is binary: the consumer either adopts or does not.

In the case of information (e.g., news) a reader may respond to the information by posting a comment or may wait until there is more information (including other postings) to trigger future postings. A forum user who posts a comment after receiving and responding to a piece of information has "adopted" that information. Two factors normally influence the spread of the innovation (information): the innovation effect and the imitation effect.

Figure 3 depicts the cumulative postings for every 10 minute interval, normalized to the total number of postings. The graph clearly shows that the postings for Event A (Christchurch) increase at a faster rate than do those for Event B. Using a nonlinear regression model, we fit the data to the Bass diffusion model below:

$$x(t) = M \times \frac{1 - e^{-(a+b)t}}{1 + \frac{b}{a} e^{-(a+b)t}}$$

where $x(t)$ is the total number of postings in time t, **M** is the maximum postings, a is the coefficient of innovation, and b is the coefficient of imitation.

The coefficient of innovation, sometimes called the coefficient of external influence, represents the influence of mass media, government agencies and advertising efforts on adoption. The coefficient of imitation, also called the coefficient of internal influence, represents the influence of social interactions (e.g., word-of-mouth) on adoption. The resulting coefficients of innovation were 0.031 and 0.008, with R^2 equal to 0.986 and 0.957 for Event A and Event B, respectively. We forced the coefficients of imitation to zero in both cases, assuming no social interaction. The difference in the coefficients of external influence further provides evidence that the Christchurch disaster had greater immediacy and relevancy than the Japanese disaster.
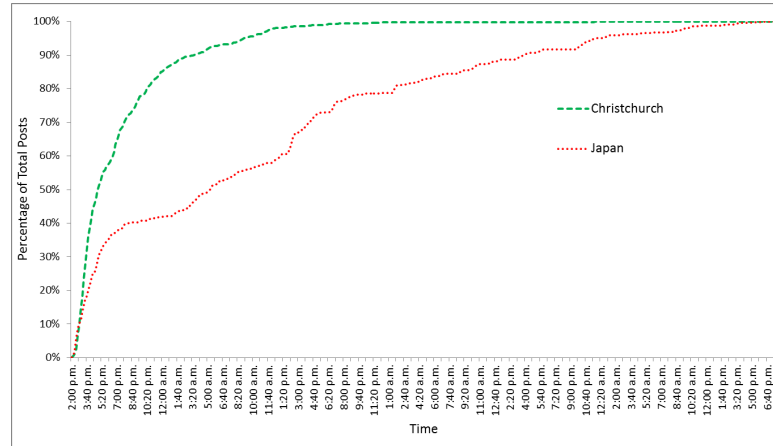

Figure 3. Rate of Information Diffusion

4.8 Student Performance

Table 6 shows the achievement score for each assessment criterion for all 45 students who submitted the assignment. The achievement scores represent the average score attained on a particular assessment criterion. Students had no problem identifying word frequencies in this exercise. On average, they achieved a score of 96.8 for identifying word frequencies. Their sentiment analysis of the posts averaged the next highest score at 84.9. The main difficulty appeared to be their inability to discover new insights from the postings (average score: 26.2). This weakness may be a reflection of both the multi-disciplinary aspect of social analytics and the students' lack of the knowledge required to connect results with theories.

The assessment also revealed a lack of communication skills among the comment posters (e.g., good writing: average score 45.7). Colleges should prepare students with various forms of communication skills including good writing skills. For business analysts, communication skills are highly desirable and valuable, listed among the top five skills desired by employers when making hiring decisions (Wixom et al., 2014).

Assessment Criteria	Student Achievement (maximum 100)
Identify word frequencies	96.8
Perform TF–IDF measure	76.5
Cluster analysis	73.5
Identify entities	70.3
Analysis of compassion and sympathy	55.4
Sentiment analysis	84.9
Analysis of similarities and differences	59.5
Content classification analysis	75.7
Quantify classification performance	50.3
Discovery of new insights	26.2
Poster communication	45.7

Table 6. Student Performance

Despite RapidMiner's user friendly classification and cross-validation interface, students' attempts at quantifying classification performance were rather weak. Quantifying performance requires interpreting the results of the confusion matrix and supporting the acceptability of performance indicators such as recall and precision. According to the assessment evaluation, students could not fully comprehend the concepts or the processes behind classification performance or the confusion matrix (average score: 50.3).

5. DISCUSSION

As one facet of "big data" analytics, social media analytics is becoming increasingly important to organizations which need to transform such data into values with which to make strategic business decisions. Despite the high demand for data analytics professionals, academic courses which teach these relevant skills were neglected in the 2010 IS curriculum (Bell, Mills, and Fadel, 2013; Topi et al., 2010). Though recent trends indicate an increasing number of BI/BA programs, resources for teachers still remain a challenge (Wixom et al., 2014).

In this paper, we introduce an assessment for unstructured data analytics using social media content. The benefits of such an approach are twofold.

First, the abundance of social media content (e.g., tweets, blog posts, Facebook, forums, reviews, etc.) ensures that teaching resources and datasets are renewable and sustainable. Instructors can easily switch to new cases without repeating old ones, and can minimize preparation time for new cases by reusing the same assessment structure. New, relevant and domain-specific cases are available, such as food contamination, crisis communication, disaster postings, political forums, sports comments, consumer products reviews and science forums. Teachers can adopt these cases to provide personalized assessments without changing the assessment structure. Using multiple versions of assessment can also minimize the risk of plagiarism.

Second, by using real-world, up-to-date cases, students will appreciate the relevance of the analysis. Students will also acquire the practical real-world experience and knowledge that hiring managers consider critical (Wixom et al., 2014).

The learning enquiries for the assessment highlight many aspects of social media analytics. For instance, they demonstrate that social media analytics is multidisciplinary. In addition to critical technical skills (e.g., mathematics, statistics, computing), social media analytics also requires communication skills and social domain knowledge. For social analytics, the interpretative and informational aspects are more crucial than the technological aspect (Marchand & Peppard, 2013). For example, to evaluate the resilience of a community, analysts must conceptually understand theories regarding compassionate communication behavior during crises. They must also have the text processing skills to construct a reasonable proposition and build a theory to substantiate and meaningfully support the learning enquiries.

Analysis of the students' performance indicates an inadequacy in their social, behavioral and cognitive skills. For this assessment, instructors must emphasize that only extensive critical analysis can provide a deeper understanding of the degree of community resilience and support revealed by the postings, and that "number crunching" alone is not sufficient (Walsh, 2007). The interpretative aspect of data analytics normally receives less focus in the undergraduate IS curriculum but is necessary in graduate programs, especially IS programs that emphasize social policy research. Our analysis indicates that colleges should strengthen the interpretive aspect at the undergraduate level immediately.

We propose the following solutions. BI&A courses should include behavioral and social cognitive theory to provide an interpretation of analytics findings. We suggest using the *Handbook of Theories of Social Psychology*, a comprehensive text available from Sage Publications (http://www.sagepub.com/home.nav). Schools intend to enhance their business analytics courses with social media analytics must prepare IS undergraduates with adequate levels of multidisciplinary knowledge so that students can make convincing and meaningful interpretations. Teachers can overcome the current weakness in communication skills by creating standard templates with a question format allowing students to fill in the gaps to address problem statements, data collection methods, results, and theories that explain and support findings. Instructors can supply the precise questions or help create them via an instructor-led student brainstorming session.

This study's approach is not without limitations. Cooper's analytics framework includes embedded theories and has known "whose reality" issues. Failure to consider this issue risks compromising the framework's effectiveness in actual implementation. The embedded theories may also be insufficient to explain any newly discovered insights. The current assessment also lacks an engagement plan: there is no task to access prior knowledge and help learners become engaged in a new concept. Lastly, the sample size for the student performance evaluation is small (45), and the evaluation involved only a single, one-trimester course.

6. CONCLUSION

This paper adapts Cooper's Analytics Framework to formulate a social media analytics teaching assessment with the goal of analyzing and interpreting social media content. This assessment can help IS programs to meet the demand for analytics professionals. For instructors, the assessment structure helps them create reusable teaching resources from other social media content. The assessment results showed that our approach is effective for teaching social analytics to undergraduates in an IS program in a business school environment. Using this assessment, students performed adequately. The mean assignment score was 64%, and 85% of the students passed the assessment. Overall, the course received a "very good" rating. According to our survey, students "agree" that the assignment allowed them to think critically and creatively, and they "strongly agree" that the assignment helped them to learn.

While most colleges teach students to generate results using various analytics tools (e.g., Excel, RapidMiner, Microsoft SQL server), the critical thinking skills and relevant knowledge for interpreting those results require further effort. Since RapidMiner contains advanced charts

and statistical analysis such as linear regression, analysis of variance, and R[TM] extension, instructors could include data visualization queries and hypothesis testing in the assignment. RapidMiner's onomastics name recognition extension (NamSor[TM]) can deliver a set of operators for mining proper names in all alphabets and languages. This extension could be part of a segmentation exercise. To enhance students' interpretative abilities and to help them derive conclusive implications, future approaches should include relevant readings in social psychology theory, such as social identity, social influence, social learning, and crisis communication.

In the future, we will adopt the analytics framework to construct other assessments with different orientations (e.g., consumer product reviews, political comments) to further validate both the assessment framework and our pedagogy. We will also include an engagement plan with activities that evoke prior learning experiences, and expose existing conceptual understanding and knowledge. We will explore an approach which blends the current assessment structure with ideas from problem- and puzzle-based learning to elicit problem solving and critical thinking skills to address real-life problems (Presthus & Bygstad, 2012). We anticipate that the assessment framework will also be useful for teaching BI&A 3.0, which involves very big data sets. BI&A 3.0 encompasses data from mobile phones, tablets, healthcare sensors, traffic sensors, RFID sensors, Internet of thing (IoT) sensors, environment sensors, and many more. These data typically fall within the computer science and engineering domain. Future teaching and learning research in BI&A 3.0 must rapidly develop teaching resources that can demonstrate innovative insights. We must identify hidden business values from sensor data via information systems as part of a collaborative, innovative community.

7. ACKNOWLEDGEMENTS

8. REFERENCES

Abbasi, A. (2012). Social Media Analytics. Retrieved from http://misrc.umn.edu/seminars/slides/2012/SocialMedia Analytics_Minnesota.pdf.

Bell, C., Mills, R., & Fadel, K. (2013). An Analysis of Undergraduate Information Systems Curricula: Adoption of the IS 2010 Curriculum Guidelines. *Communications of the Association for Information Systems,* 32(1), 2.

Chen, H. C., Chiang, R. H. L., & Storey, V. C. (2012). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly,* 36(4), 1165-1188.

Chiang, R. H. L., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. *ACM Trans. Manage. Inf. Syst.,* 3(3), 1-13. doi: 10.1145/2361256.2361257.

Chudoba, K., Hauser, K., & Olsen, D. (2010). Teaching Tools for Data Analysis. *International Journal of Management and Information Systems,* 14(3), 15-24.

Cooper, A. (2012). A Framework of Characteristics for Analytics. *CETIS Analytics Series,* 1(7).

Crawford, P., Gilbert, P., Gilbert, J., & Gale, C. (2011). The Language of Compassion. *Taiwan International ESP Journal,* 3(1), 1-15.

Davenport, T. H. & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*: Harvard Business Press.

Deloitte. (2012). Billions and Billions: Big Data Becomes a Big Deal. Retrieved April 1, 2013, from http://www.deloitte.com/view/en_GX/global/industries/t echnology-media-telecommunications/tmt-predictions-2012/technology/70763e14447a4310VgnVCM1000001a 56f00aRCRD.htm.

Dupin-Bryant, P. A. & Olsen, D. H. (2014). Business Intelligence, Analytics and Data Visualization: A Heat Map Project Tutorial. *International Journal of Management & Information Systems,* 18(3), 185-200.

Edgington, T. M. (2011). Introducing Text Analytics as a Graduate Business School Course. *Journal of Information Technology Education: Innovations in Practice,* 10, 28-28.

Elizondo, J., Parzinger, M., & Welch, O. (2011). Using Business Analysis Software in a Business Intelligence Course. *Information Systems Education Journal,* 9(6), 4-10.

Gorman, M. F. & Klimberg, R. K. (2014). Benchmarking Academic Programs in Business Analytics. *Interfaces,* 44(3), 329-341.

Henschen, D. (2013). Big Data Analytics Masters Degrees: 20 Top Programs. Retrieved April 1, 2013, from http://www.informationweek.com/big-data/news/big-data-analytics/big-data-analytics-masters-degrees-20-top-programs/240145673.

Kaplan, A. M. & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons,* 53(1), 59-68. doi: http://dx.doi.org/10.1016/j.bushor.2009.09.003.

Macmillan. (2013). Compassion. Retrieved April 1, 2013, from http://www.macmillandictionary.com/thesaurus/british/c ompassion#compassion_3.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New Product Diffusion Models in Marketing: A Review and Directions for Research. *Journal of Marketing,* 54(1), 1-26. doi: 10.2307/1252170.

Maness, J. M. (2008). A Linguistic Analysis of Chat Reference Conversations with 18–24 Year-Old College Students. *The Journal of Academic Librarianship,* 34(1), 31-38.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity. Retrieved April 1, 2013, from http://www.mckinsey.com/insights/mgi/research/technol ogy_and_innovation/big_data_the_next_frontier_for_inn ovation.

Marchand, D. A. & Peppard, J. (2013). Why IT Fumbles Analytics. *Harvard Business Review,* 91(1-2), 104-112.

Marjanovic, O. (2013). Sharing and Reuse of Innovative Teaching Practices in Emerging Business Analytics Discipline. Paper presented at 46th Hawaii International Conference on System Sciences (HICSS), 2013.

MCEER. (2011). The Great Tohoku, Japan, Earthquake & Tsunami: Facts, Engineering, News & Maps. Retrieved April 1, 2013, from http://mceer.buffalo.edu/infoservice/disasters/Honshu-Japan-Earthquake-Tsunami-2011.asp.

NUS-SOC. (2013). Bachelor of Science (Business Analytics). Retrieved April 1, 2013, from http://www.comp.nus.edu.sg/is/ug-bsc-ba.html.

Presthus, W. & Bygstad, B. (2012). Business Intelligence in College: A Teaching Case with Real Life Puzzles. *Journal of Information Technology Education: Innovations in Practice,* 11(1), 121-137.

Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., & Araki, K. (2010). CAO: Fully Automatic Emoticon Analysis System. Paper presented at the 24th AAAI Conference on Artificial Intelligence (AAAI-10).

Qu, Y., Huang, C., Zhang, P., & Zhang, J. (2011). Microblogging After a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake. Paper presented at the ACM 2011 Conference on Computer Supported Cooperative Work.

Rapid-i. (2012). RapidMiner 5.0. Retrieved April 1, 2012, from http://rapid-i.com/.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization. *ACM computing surveys (CSUR),* 34(1), 1-47.

Solomon, R. C. (1998). The Moral Psychology of Business: Care and Compassion in the Corporation. *Business Ethics Quarterly*, 515-533.

Stacey, M., Salvatore, J., & Jorgensen, A. (2013). *Visual Intelligence: Microsoft Tools and Techniques for Visualizing Data*: John Wiley & Sons.

Thesaurus.com. (2013). Compassion. Retrieved April 1, 2013, from http://thesaurus.com/browse/compassionate.

Topi, H., Valacich, J. S., Wright, R. T., Kaiser, K. M., Nunamaker, J. F., Jr., Sipior, J. C., & de Vreede, G. J. (2010). *Curriculum Guidelines for Undergraduate Degree Programs in Information Systems*. Association for Information Systems.

Walsh, F. (2007). Traumatic Loss and Major Disasters: Strengthening Family and Community Resilience. *Family Process,* 46(2), 207-227. doi: 10.1111/j.1545-5300.2007.00205.x.

Whittle, A. (2014). What Lies Beneath: Sensor Analytics in the Water System. Retrieved from http://newsoffice.mit.edu/2014/what-lies-beneath-sensor-analytics-in-the-water-system.

Wise, K. (2004). Attribution Versus Compassion: The City of Chicago's Response to the E2 Crisis. *Public Relations Review,* 30(3), 347-356. doi: DOI 10.1016/j.pubrev.2004.05.006.

Wixom, B., Ariyachandra, T., Douglas, D., Goul, M., Gupta, B., Iyer, L., & Turetken, O. (2014). The Current State of Business Intelligence in Academia: The Arrival of Big Data. *Communications of the Association for Information Systems,* 34(1).

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., & Philip, S. Y. (2008). Top 10 Algorithms in Data Mining. *Knowledge and Information Systems,* 14(1), 1-37.

AUTHOR BIOGRAPHIES

Tiong-Thye Goh (BSc & MSc Electrical Engineering, Ohio State University; GDipFM, Singapore Institute of Management; MBA (Distinction), Manchester & Wales; PhD Information Systems, Massey University) is a Senior Lecturer with the School of Information Management, Victoria University of Wellington, New Zealand. His research focuses on the relationship between technologies, people and society. In particular, Dr Goh's research involves the understanding of social network, emotion computing, analytics, learning science and user's behavior.

Pei-Chen Sun is currently an Associate Professor and serves as System Manager at National Kaohsiung Normal University, Taiwan. He holds a Ph.D. in Management Information Systems from National Sun Yat-Sen University. His current research interests include e-Learning, electronic commerce, and knowledge management. He published in Journal of Information Management, Computers & Education, Journal of Information Science and Engineering, and International Journal of Innovation and Learning.

STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.