Journal of Psychology & Psychotherapy
ISSN: 2161-0487

Davalos et al., J Psychol Psychother 2015, 5:6
DOI: 10.4172/2161-0487.1000221

Short Communication Open Access

Using Big Data to Study Psychological Constructs: Nostalgia on Facebook

Sergio Davalos*, Altaf Merchant and Greg Rose

Milgard School of Business, University of Washington Tacoma, Tacoma, USA

Introduction

Nostalgia, reflecting on the past, has many aspects. At times, we refer to it as a state of mind - as in, I'm nostalgic. At times, as a feeling – I'm feeling nostalgic. We sometimes use nostalgia as a way to deal with "in the present" emotions or feelings. For instance, we reminisce about past holidays to feel better about the present [1]. Nostalgia can be internally or externally activated. Internally, an individual can evoke feelings of nostalgia. On the other hand, external stimuli can evoke nostalgia, such as a song, an ad, or a brand [2-5].

Previous research has examined nostalgia. However, these studies have utilized small sample, been clinical in nature, and/or employed research techniques such as, surveys, experiments, focus groups and ethnography. For example, nostalgia studies have examined responses to specific stimuli that invoke nostalgia, such as pictures or music. Ideally, it would be expedient to explore nostalgia in as natural a setting as possible. However, until the advent of online social media like Facebook, this was not practical. Now, Facebook users generate posts that provide rich data with all sorts of expression and sentiments. Social media posts encompass the entire range of human expression, including sentiment, information, emotion, encouragement, and opinion. The quantity of posts generated can number into the millions per hour. This provides large data sets of content that can be explored using machine learning techniques.

Machine learning refers to a process where the machine (computer system) learns to perform a task based on past experience, and its performance is measured. For instance, a program can learn to detect deception in email (task) based on examples of deceptive and non-deceptive emails (experience) using the number of emails correctly classified as the performance measure. The techniques can include such methods as decision trees, association rule generations, classifications, cluster analysis, and regression. The key to machine learning is that the machine enhances performance through learning. More formally, the system makes modifications that enhance performance by adjusting parameters based on learning though experience. Two major forms of machine learning are supervised and unsupervised. Supervised learning is when the system learns from exemplars with example inputs and expected outputs. Whereas, unsupervised learning is when the system learns without exemplars. Unsupervised learning is typically used when we do not have a data set that is unclassified and the system is left to find its own structure in the data.

These can be utilized to understand emotional states such as nostalgia by people over long periods of time. Experiments and surveys usually manipulate variables of interest and obtain reactions at specific time states. Obtaining big data from social media sites like Facebook can provide additional insight by studying longitudinal consumer expressions. Figure 1 presents the number of posts per day for the period May 4, 2012 to September 9, 2012. The posts are from the nostalgia data set utilized by Davalos et al. [6], which employed the Facebook Graph API program [7] to carry out a variety of operations upon the data obtained from Facebook. The data set consisted of public posts that utilized nostalgic/word phrases (such as: do you remember when, down memory lane, flashback, go back in time, good old days, I miss those days, nostalgic, recollect, redolent of, relive the past, reminiscent, those were the days, and when we were younger) [6]. Figure 1 summarizes nostalgic expressions and discussions on Facebook over time, these posts vary depending on the time of the year and peak at times of festivals and events. Longitudinally examining the incidence of nostalgic expressions would be difficult using erstwhile techniques of psychology research. Thus, machine learning techniques and the examination of big data can supplement, enhance and build on previous insights obtained through traditionally employed techniques in psychological research (Figure 1).

Techniques Useful for Big Data Analyses

One of the results of the widespread adoption and advancement of information systems at all different levels of society is the large volumes of data being generated on a regular basis. The term "big data" refers to massive volumes of structured data (such as patient records) and unstructured data (such as patient notes) that are so large that it overwhelms the abilities of traditional database and software systems. Initially, big data consisted mostly of structured data. However, the

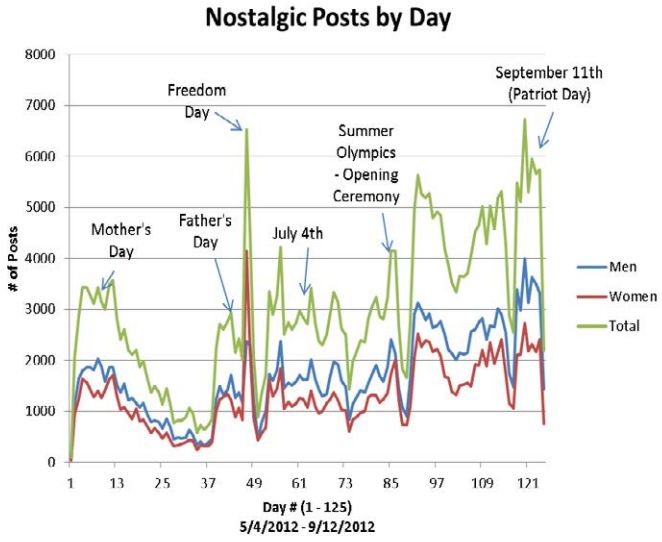

Figure 1: Peaks and troughs of nostalgia expressed on facebook.

***Corresponding author:** Sergio Davalos, Associate Professor of Information Systems, Milgard School of Business, University of Washington Tacoma, 1900 Commerce Street, Tacoma, WA 98402, USA, Tel: 253 692 4658; Fax: 253 692 4523; E-mail: sergiod@uw.edu

Received October 01, 2015; **Accepted** November 19, 2015; **Published** November 26, 2015

Citation: Davalos S, Merchant A, Rose G (2015) Using Big Data to Study Psychological Constructs: Nostalgia on Facebook. J Psychol Psychother 5: 221. doi:10.4172/2161-0487.1000221

growth of unstructured data generated in online interactions such as social media (twitter, Facebook), online shopping, email, user forums, blogs, and text messages now constitutes the majority (75%) of the big data being generated [8]. According to IBM, 2.5 exabytes - that's 2.5 billion gigabytes (GB) - of data were generated every day in 2012. As a consequence big data analysis now often uses characteristics such volume, velocity, and variety to describe dynamic, constantly changing streams of data. Due to the limitations of traditional data processing systems, several platforms, both commercial and open sources, have been developed to specifically deal with large amounts of data quickly and efficiently.

Instruments, such as surveys, questionnaires, and interviews, used in clinical settings are focused and generally directed toward more specific research questions. As a consequence, these traditional tools are most effective in examining predefined research questions and testing hypotheses. Recent tools such as text mining, in contrast, are particularly effective for examining user generated content (UGC) in a natural setting found on online social media platforms such as Facebook, twitter, blogs, user forums, and user reviews. These new forms of data can reveal novel ideas, themes, topics and insights not revealed through previous techniques. Effective content analysis, moreover, can discover and uncover phenomenon and information not previously considered. The volume of data naturally created in social media contexts, such as Facebook; provide an excellent context for discovery oriented research. Additional insights obtained from emerging techniques in big data analysis is analogous to the importance of including interpretivist and naturalistic inquiries to study consumer culture and gauge human behavior in conjunction with traditional research methods, like surveys and experiments [9]. Similarly, in the context of psychotherapy, Mahrer and Boulet [10] suggest that the use of observational or naturalistic research together with experimental testing of hypotheses can provide unique insights. Keeping in mind these assertions we call for the use of text-mining and other techniques suggested in this paper for analyzing big data along with traditional research methods. Employing multi-method research programs will further nurture the development of robust epistemologies incorporating information missed through pure positivist research techniques.

Since big data analysis involves massive volumes of data, appropriate content analysis tools are needed. Text mining and text analysis methods can provide effective content analysis techniques [11]. Text mining refers to techniques that enable the processing of text based data. Text mining tasks include text categorization, key word or phrase analysis, clustering of text (word, phrase, paragraph, or document level), topic/concept/entity extraction, and conducting sentiment analysis. Text is often referred to as corpus, documents, and terms. Corpus refers to the entire collection of documents. A document can be a title, a sentence, a phrase, a paragraph, or many paragraphs. A term is typically a word. Additionally, most text mining methods use a "bag of words" approach where the word order is not relevant. We will focus on these methods in this paper.

One commonly used text mining method is LIWC. LIWC analysis can be used to analyze text to determine the writer's different emotions, thinking style, social concern, and grammatical category of speech [12,13]. The words are mapped to language categories that capture people's social and psychological states. Key word and phrase analysis can identify terms by frequency rank or by special metrics such as TF-IDF (term frequency – inverse document frequency), which measures the relative information value of a term and not just the frequency. This is a weighting that is based on the assumption that the more often a term occurs in a document, the more it is representative of its content. However, the more documents in which a term occurs, the less discrimination that term provides. For example, the word "the" occurs many times in documents and most likely occurs in all documents. Whereas the word "depressed" will most likely occur only in documents related to depression. The general formula for the calculation of the TF-IDF [14] for each term is:

TF = (number of times the term occurs/ number of terms in a document)

IDF = log (number of documents/number of documents the term occurs in)

*TF-IDF = TF*IDF*

Another useful method is singular value decomposition (SVD) [15], a dimensionality reduction technique similar to principal component analysis, which is used to rank terms. With these measures, other forms of analysis can be undertaken, such as cluster analysis to identify terms, phrases, or documents that cluster together. For instance, we could determine which terms cluster together for nostalgic documents or examine if documents cluster together on the basis of stages of depression. Collocation analysis identifies terms that occur near each other with a higher than chance frequency [16]. This is useful for understanding the relationships between terms. N-gram analysis [17] can identify a special case of collocation – the contiguous sequence of n terms. Cluster analysis can also be used, in addition to these techniques. Figure 2 below shows a cluster analysis of the words using the Facebook nostalgic posts databases. The graphical representation

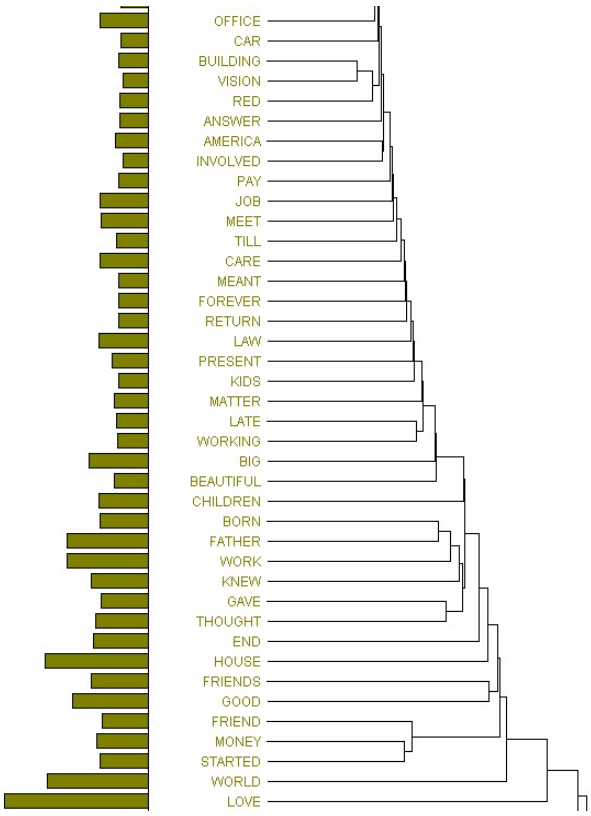

Figure 2: Example of a dendrogram generated for nostalgic facebook posts.

is known as a dendrogram [18]. The bar chart on the left indicates the relative importance of each term.

Another effective text mining tool is topic modeling, which identifies topics within a collection of documents. A topic model is a type of statistical model [19]. We assume that a document is about a particular topic, or set of topics, and as a consequence, we expect certain words to appear in the document more or less frequently based on the topic. For instance, "school" and "teacher" will appear more often in documents about education while "drug" and "treatment" will appear in documents about medical treatment. Terms such as "the" and "and" will appear equally in both and such terms are ignored on the basis of their SVD or TF-IDF scores. Topic modeling can be used to determine what people are talking about from a wide range of topics such as cyber terrorism, health care, the Emmys, or presidential elections. While text mining can be effective in exploring, discovering, and identifying general topics.

A more focused approach is needed when examining the occurrence of a specific type of content such as nostalgia [6], emotion [20], or depression [21] in online social media. An effective tool for this is a lexicon or dictionary of key words. A lexicon is then used to identify all text documents (user posts in the case of Facebook) that contain words in the lexicon. Recently Davalos et al. [6] developed a nostalgic lexicon using thirteen nostalgic key words and phrases, and found significant evidence of nostalgic expressions in Facebook conversations. They also employed cluster analysis (using procedures discussed earlier) to discover themes of nostalgic longing related to politics, life stories, historical events (man on moon and Gandhi), spirituality, appreciating life, romanticism and fun. Nostalgic posts were also analyzed using LIWC procedures (also discussed earlier) and were found to be more reflective, more emotional (than non-nostalgic posts), and frequently included both positive and negative emotions, which is consistent with the deep, often bittersweet character of nostalgia.

Conclusion

In this paper we have highlighted the importance of using big data to comprehend complex psychological constructs such as nostalgia. We have also discussed some methods and techniques for big data analysis which can be helpful to psychology researchers. We further call for more research on emotional constructs within the context of social media vehicles (such as Facebook, Twitter, Linked In, etc.), which will expand our understanding of psychological constructs in everyday life. Using big-data studies along with experimental research will help build robust explanations of how humans behave in an on-line connected world.

References

1. Zhou X, Sedikides C, Wildschut T, Gao DG (2008) Counteracting loneliness: on the restorative function of nostalgia. Psychol Sci 19: 1023-1029.
2. Merchant A, LaTour K, Ford JB, LaTour M (2013) How strong is the pull of the past: measuring personal nostalgia evoked by advertising. Journal of Advertising Research 53: 150-165.
3. Batcho KI (2007) Nostalgia and the emotional tone and content of song lyrics. Am J Psychol 120: 361-381.
4. Braun-LaTour KA, LaTour MS, Zinkhan G (2007) Using childhood memories to gain insight into brand meaning. Journal of Marketing 71: 45-60.
5. Loveland KE, Smeesters D, Mandel N (2010) Still preoccupied with 1995: the need to belong and preference for nostalgic products. Journal of Consumer Research 37: 393-408.
6. Davalos S, Merchant A, Rose G, Lessley B, Teredesai A (2015) 'The Good Old Days': An Examination of Nostalgia in Facebook Posts. International Journal of Human-Computer Studies 83: 83-93.
7. Facebook Graph API. https://developers.facebook.com/docs/graph-api.
8. Wall M (2015) Big Data: Are you ready for blast-off?
9. Arnould EJ, Thompson CJ (2005) Consumer Culture Theory (CCT): Twenty Years of Research. Journal of Consumer Research 31: 868-882.
10. Mahrer AR, Boulet DB (1999) How to do discovery-oriented psychotherapy research. J Clin Psychol 55: 1481-1493.
11. Miner G, John E IV J, Hill T, Nisbet R, Delen D, et al. (2012) Practical text mining and statistical analysis for non-structured text data applications. Elsevier/Academic Press, Waltham, MA.
12. Pennebaker JW, Francis ME, Booth RJ (2001) Linguistic Inquiry and Word Count (LIWC): LIWC2001. Lawrence Erlbaum Associates, Inc., Mahwah, NJ.
13. Tausczik YR, Pennebaker JW (2010) The psychological meaning of words: LIWC and computerized text analysis methods. Journal of Language and Social Psychology 29: 24-54.
14. Ramos J (2003) Using tf-idf to determine word relevance in document queries. In: Proceedings of the first instructional conference on machine learning.
15. Albright R (2004) Taming Text with the SVD. SAS Institute Inc.
16. Mello RA (2002) Collocation analysis: A method for conceptualizing and understanding narrative data. Qualitative research 2: 231-243.
17. Bespalov D, Bai B, Qi Y, Shokoufandeh A (2011) Sentiment classification based on supervised latent n-gram analysis. In: Proceedings of the 20th ACM international conference on Information and knowledge management, ACM.
18. Fitzpatrick E (2007) Corpus linguistics beyond the word: corpus research from phrase to discourse. Rodopi.
19. Blei DM (2012) Probabilistic topic models. Communications of the ACM 55: 77-84.
20. Mohammad SM, Turney PD (2013) NRC Emotion Lexicon, National Research Council of Canada. Technical Report.
21. De Choudhury M, Gamon M, Counts S, Horvitz E (2013) Predicting Depression via Social Media. In ICWSM.