Potential Applications of Sentiment Analysis in Educational Research and Practice – Is SITE the Friendliest Conference?

Matthew J. Koehler (mkoehler@msu.edu)
Spencer Greenhalgh (greenha6@msu.edu)
Andrea Zellner (zellnera@msu.edu)

Michigan State University

Abstract: Despite the widespread use of sentiment analysis by many disciplines, it has been a largely underused tool in educational contexts. The purpose of this paper is to explore some potential uses for sentiment analysis in educational settings and to present a sample study using the approach. Using sentiment analysis, we compare the "friendliness" of two educational technology conferences and use these data to answer the question: Is SITE the friendliest conference? We then expand the discussion to consider how education researchers and practitioners may fruitfully use sentiment analysis.

"Sentiment is wherever you go" - *Mitch Albom*

"It is as healthy to enjoy sentiment as to enjoy jam" - *Gilbert K. Chesterton*

Introduction

The late 1990s and early 2000s marked a turning point in the research and practical application of automated natural language processing by computers (Pang & Lee, 2010). This turning point came about, in part, because of rapid changes in computer processing power, advances in machine learning techniques, and increased attention to the role that computers could play in processing language. Newfound applications included automated or assisted dictation and transcription; detailed analysis of written text; and more intelligent (or "natural") user interfaces ranging from phone help menus to the computer interfaces that eventually evolved into today's Siri and Cortana.

Education researchers and practitioners have been late to embrace these natural language processing technologies. One exception, perhaps, has been in the use of automated techniques to analyze and evaluate writing. While the move to automated essay scoring has not been without controversy (Perelman, 2014), the need to administer large-scale assessment schedules in response to K-12 education reforms and policies has led to more investment in these types of automated systems. Researchers at the University of Memphis, for example, have developed a sophisticated text cohesion tool called Coh-Metrix that has been used successfully to analyze texts and determine the difficulty of those texts for human readers (Graesser, McNamara, & Kulikowich, 2011; Hongwei, 2012; McNamara & Graesser, 2011; McNamara, Louwerse, McCarthy, & Graesser, 2010). Additionally, researchers, testing companies, technology companies, and publishing companies have put a variety of automated essay tools to the test through competitions and large-scale validation, thereby demonstrating the promise of automated essay scoring, especially when the focus is on issues of text cohesion, grammar, sentence structure, and text complexity (Shermis, 2014).

The focus of this paper, however, is on an under-utilized approach in education called sentiment analysis or, sometimes, opinion mining. In this approach, computers code language to determine if a piece of text is objective (i.e., is neutral) or subjective (i.e., expresses an opinion); if computers code text as subjective, they then determine whether the opinion expressed is positive or negative. The early roots of sentiment analysis can be tracked to efforts in 2001 to analyze the sentiment of financial markets (Das & Chen, 2001; Tong, 2001) and to subsequent efforts by businesses to track opinions about their brands, products, and corporate images (Cambria, Schuller, Xia, & Havasi, 2013).

Sentiment analysis applications have been widely used in fields outside of education. Researchers have firmly demonstrated the utility of sentiment analyses by successfully correlating changes in opinion expressed in social media with social, political, and economic events (Bollen, Pepe, & Mao, 2009, Thelwall, Buckley, & Paltoglou, 2011). For example, sentiment analysis can successfully predict future changes in the stock market (Choudhury, Sundaram, John, & Seligmann, 2008) up to 3-4 days in advance (Bollen, Mao, & Zeng, 2011). Also, researchers have shown that the average "mood" of Facebook or Twitter users in the United States changes with significant news events or major holidays (e.g., Valentine's Day). Sentiment analysis of particular topics on Twitter and Facebook has shown strong correlation to established opinion polls (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Moreover, opinion mining has been used to "nowcast" and "forecast"—that is, monitor and predict—the results of high-stakes elections nationally and internationally (Ceron, Curini, & Iacus, 2015).

In the remainder of this paper we detail a study using sentiment analysis and discuss potential future uses for the technique in education.

Purpose and Research Questions

Despite the widespread use of sentiment analyses in many disciplines, education researchers and practitioners have largely underused this tool. The purpose of this paper is to explore some potential uses for sentiment analysis in educational contexts.

In order to make sentiment analysis and its affordances more concrete, we first present the results of a study aimed at illustrating the new looks at data that sentiment analyses can bring to researchers. Specifically, we use sentiment analysis to analyze text associated with the SITE conference; this conference has often been presented as "the friendly conference," replete with Aloha shirts as a conference norm for attire, polite reviews, and a congenial conference atmosphere. How friendly is it? Is it the "friendliest" conference? We explore these questions using sentiment analysis to examine the tweets from two educational technology conferences held within the last year: SITE and Educause. We compare the average "mood," or valence of the opinion of the tweets about the conference, and measure the "friendliness" (or sentiment) of both conferences. By measuring the friendliness of the conference tweets and comparing conferences, we hope to illustrate the new types of analyses possible in large data sets when tools like sentiment analysis are considered.

Methods

Tweets were collected for two different educational technology conferences between March and October 2014 as summarized in the table below.

CONFERENCE	DATES	HASHTAG	TWEETS COLLECTED
SITE	03.17.14 – 03.21.14	#siteconf	397
Educause	09.29.14 – 10.02.14	#edu14	8649

All tweets containing the official Educause hashtag (i.e., #edu14) were collected during that conference. In the case of the SITE conference, tweets were not collected during the conference but were retroactively retrieved. Because of limitations on the Twitter API at the time of collection, only a selection of #siteconf tweets was available. Accordingly, the 397 tweets for the SITE conference represent a roughly random sample of conference tweets.

Each tweet was "cleaned" by removing URLs, name signifiers (e.g., @johnsmith), special characters (e.g., backslashes and "#"), and retweet information (e.g., "RT"). Each tweet was analyzed for sentiment using Jacob Perkins's Sentiment API (Perkins, n.d.). A score between -1.0 and +1.0 representing the valence of the opinion was assigned to each tweet. Example tweets and the assigned sentiment score are included below.

ORIGINAL TWEET	SENTIMENT SCORE
@spgreenhalgh I enjoyed it. It was nice to finally meet @mete_akca as well. Keep in touch! #siteconf	+ 0.77
Rain looks like it has paused for now at #siteconf	- 0.44
#siteconf How does our customization of mobile devices relate to potential personalization of mobile learning?	+ 0.01
#siteconf Looking forward to @phdslyk 's keynote this morning!	- 0.06

The first three examples show tweets that appear to be appropriately analyzed by the sentiment algorithm. The first is a positive interaction that is scored high at +0.77; the second, with a score of -0.44, is moderately negative in referring to the rain; and the third asks a neutral question, which is scored neutral at +0.01. The fourth example, however, shows that automated techniques are fallible. The tweet is positive ("looking forward"), yet is scored slightly negative.

Results

Mean sentiment scores for the Twitter discourse of conference-goers were computed and analyzed for trends. As shown in Figure 1, each tweet is represented as a color-coded point, with greener points indicating more positive opinions and redder points indicating more negative opinions. There are two panels in the display, one for Educause and one for the SITE conference. Each panel is segmented into separate days of the conference, each of which is represented as a point on the x-axis. Also presented is a "best-fit" curve for the data, with error represented by the grey shaded area.

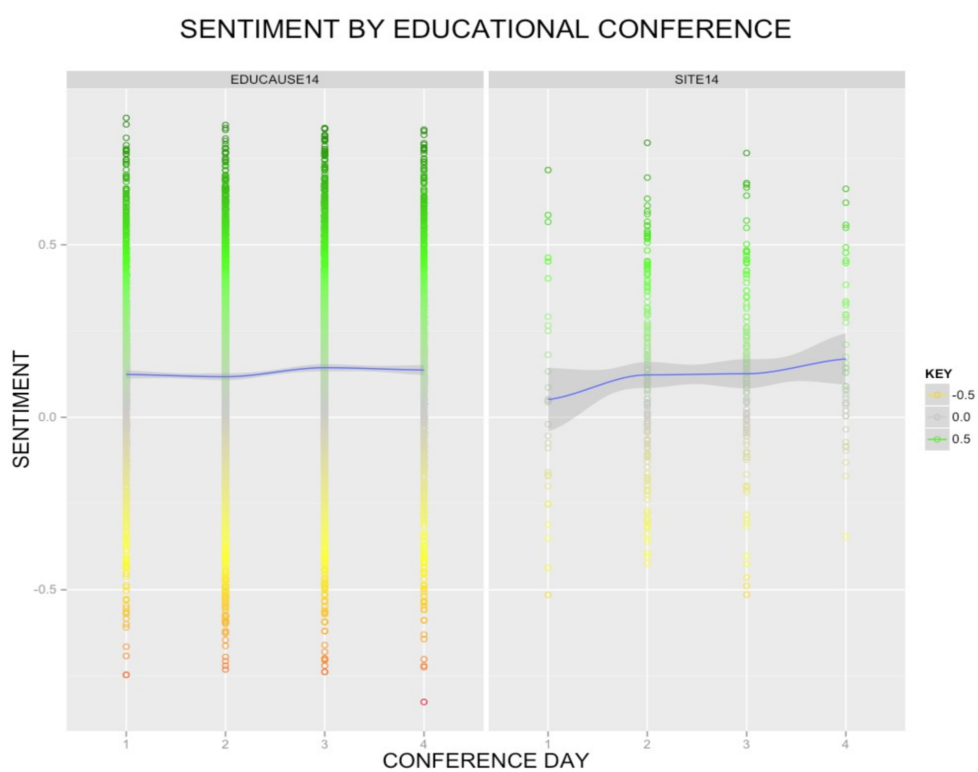

Figure 1: Tweet sentiment by conference and day within conference

Both visual inspection and statistical analysis tell the same story. First, each conference has an overall (mean) positive sentiment [SITE μ=.1240, EDUCAUSE μ=.1295] indicating that there is a net positive sentiment expressed by the conference-goers in their Twitter conversations. Second, there is no overall difference in the mean sentiment rating for the conferences ($F[1,9039] = 0.0128$, $p > .05$); rather, both have statistically equivalent positive opinions represented on Twitter. Third, there is a difference in sentiment as the SITE conference proceeded through the 4 days ($F[3,4039] = 4.840$, $p < .01$). This may be a trend in conferences generally or a trend unique to the SITE conference or even SITE 2014. In particular, the first day was marked by quite a few tweets referring to the bad weather and problems with travel. Mood may have improved as the weather did.

Overall, according to sentiment analysis, SITE is a friendly conference but cannot claim to be the friendliest conference. Clearly there are limits to this type of analysis, as it does not capture all the interactions that go into a successful conference (e.g., the face-to-face and interpersonal interaction). Nonetheless, we hope this example proves useful in demonstrating the potential of the approach as we consider other uses for sentiment analysis in educational settings in the next section.

Looking Ahead – Potential Educational Uses of Sentiment Analysis

Since 2001, sentiment analysis has been extensively used in research and practice in business, sociology, political science, linguistics, and engineering. Very little attention has been given to sentiment analysis as a potential tool for teachers and teacher educators. Like for any new technology, educators' adoption and adaption of sentiment analysis is no simple task, as they must pay close attention to the interaction of the technology with content and pedagogy (Koehler & Mishra, 2009, Mishra & Koehler, 2006). Nonetheless, to begin these explorations, we propose four possible avenues of investigation where sentiment analysis may prove useful to education researchers and practitioners.

Expanding the Study of Social Media

The present study used hashtag searches (e.g., #siteconf, #edu14) to perform sentiment analysis on the conversations conference-goers have about educational technology conferences. This approach could be expanded to broadly examine larger educational issues. For example, just as sociologists and media researchers have studied the Occupy movement using hashtags and sentiment analysis (Wang, Wang, & Zhu, 2013), education researchers could use the same approach to study key issues or movements in education, such as the Common Core (#commoncore), the digital divide (#digitaldivide), or cyber-bullying (#cyberbullying).

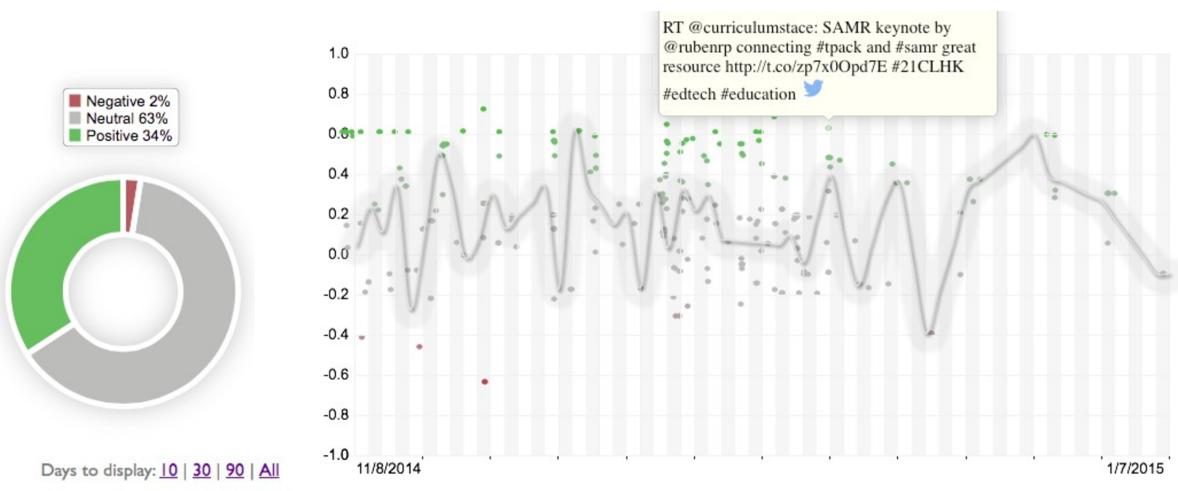

Figure 2: Screenshot of the #tpack mood tracker

For example, Figure 2 depicts a sentiment analysis of the #tpack hashtag over *time* (Koehler, 2015). Each dot on the graph indicates a tweet containing the #tpack tag, indicating a tweet about technological pedagogical content knowledge (Koehler & Mishra, 2007). Each tweet has a sentiment score between +1.0 and -1.0 and is placed on the graph according to its date (X-axis) and sentiment score (Y-axis). Viewers can mouse over any point on the graph to see the original tweet and click on the tweet to view it on Twitter. A grey-colored spline represents the moving average sentiment over time.

Using this tool, viewers can see: a) how active conversation around #tpack is on Twitter over time; b) the range of sentiment represented in a single day and over several days; c) how sentiment varies over time; and d) the average sentiment of the discussion around #tpack on Twitter (at the time of this screenshot, 2% negative, 65% neutral, 34% positive). In some ways, this mirrors the ways in which businesses use sentiment analysis to manage brand opinions. Educators could follow suit by tracking the "brand" of their degree programs, courses, conferences, or anything else they can identify with a hashtag.

Opportunities in Online Learning

Sentiment analysis has practical applications in teaching contexts, especially in online courses where so much of the learning and discussing happens as typed text. One possibility is to monitor in real time the online conversations students have in course discussion forums, discussion groups, or social media channels and to analyze the sentiment evident in this text. In face-to-face discussions, instructors can actively and consistently monitor every conversation in terms of students' motivation, mood, and understanding of material. However, instructors in online course conversations are often "late to the party" because many interactions and postings unfold between instructor logins. Using sentiment analysis, instructors could be notified via the messaging service of their choice when sentiment suddenly changes in a course. This idea is similar to the many "internet mood" meters that use sentiment analyses on tweets (e.g., http://hedonometer.org/), except that it would be a "course mood" meter, capable of real-time alerts. Quick changes in the mood of a course or conversation can be an important moment in a course, and early notification and intervention can be key to instructional fidelity in such instances.

A different approach by Ortigosa, Martín, and Carro (2014) performs sentiment analysis on course-related online communications and organizes data on a student-by-student basis. Instructors can pull up a dashboard of data for each student and retrieve data such as the student's overall mood over time in the course, with particular attention to changes in sentiment. The authors argue that such data enables instructors to make better instructional decisions. For example, assignments or assessments can be tailored to students based upon their current sentiment—students with low motivation or sentiment might be given confidence-building assignments, whereas high performing students could be given more challenging activities. In theory, such information could play a role in many ways, including group formation, course evaluation, and course design and re-design.

Opinion and Survey Research

Previous research has shown that sentiment analysis of particular topics on Twitter and Facebook is strongly correlated to established opinion polls (O'Connor, Balasubramanyan, Routledge, & Smith, 2010). Researchers have also shown that it is possible to determine with some certainty the geo-location of many social-media users based upon what they say in their posts (Baucom, Sanjari, Liu, & Chen, 2013). Combined, these approaches suggests that in some cases, sentiment analysis of a large data set is a fast, reliable alternative to traditional research done through the traditional methods of opinion and survey research. Researchers, including educational researchers, can not only establish the opinion of many users on various topics but also determine who those participants are and where they likely live. Access to large, inexpensive data sets on current issues could be a valuable tool for education researchers. For example, instead of conducting an opinion poll about attitudes towards adopting the Common Core, sentiment analysis on tweets or Facebook posts could serve as a cheap, fast proxy to polling. Combined with data-mining techniques, it is possible to know a lot about tweet and post authors, such that sentiment about the Common Core could be broken down into demographic groups, including region of the country, gender, age, or educational background.

Study of Discourse Communities

Sentiment analysis offers a potential tool in studying the many communities of practice (Eckert, 2006) that exist in educational settings, including professional development communities, informal teacher communities, professional communities, and academic communities. These analyses could inform the study of how communities change and evolve over time, how they differ from site to site (e.g., between teachers from two different districts), or how communities respond to particular events (e.g., introduction of new standards). It is also possible to use sentiment analysis to study individuals and to establish how a single teacher's experience might change as they enter a new community, as they become an established member of that community, and yet again as they become a senior member within the same community.

Conclusion

In this paper we presented an example of how sentiment analysis may be used to study online texts. Specifically, we explored the opinions represented by the tweets associated with two educational technology conferences. Although we did not find any differences between conferences, this example provided the context to discuss how a sentiment analysis is conducted as well as the possible applications of sentiment analysis in education.

Sentiment analysis remains an untapped resource for educational researchers and practitioners alike. Twitter and Facebook represent huge data sets that offer tremendous opportunities to study discourse communities, social issues, educational controversies, and public opinion polling. Moreover, expanding sentiment analysis techniques to instructional websites, professional development communities, and educational resource sites allows further development of approaches to online instruction, professional development, and communities of practice.

References

Baucom, E., Sanjari, A., Liu, X., & Chen, M. (2013). Mirroring the real world in social media: Twitter, geolocation, and sentiment analysis. In *Proceedings of the 2013 International Workshop on Mining Unstructured Big Data using Natural Language Processing* (pp. 61-68). New York: ACM Press.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science, 2(*1), 1-8.

Bollen, J., Pepe, A., & Mao, H. (2009). Modeling public mood and emotion: Twitter sentiment and socioeconomic phenomena. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 450-453). Menlo Park, California: The AAAI Press.

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems, 1*, 15-21.

Ceron, A., Curini, L., & Iacus, S. M. (2015). Using sentiment analysis to monitor electoral campaigns: Method matters—evidence from the United States and Italy. *Social Science Computer Review*, *33*(1), 3-20. doi:10.1177/0894439314521983

Choudhury, M. D., Sundaram, H., John, A., & Seligmann, D. D. (2008). Can blog communication dynamics be correlated with stock market activity? In *Proceedings of the 19th ACM Conference on Hypertext and Hypermedia* (pp. 55–60). New York: ACM Press.

Das, S., & Chen, M. (2001). Yahoo! for Amazon: Extracting market sentiment from stock message boards. In *Proceedings of the 8th Asia Pacific Finance Association Annual Conference (APFA)* (pp. 37-56).

Eckert, P. (2006). Communities of practice. *Encyclopedia of language and linguistics, 2*, 683-685.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223–234. doi:10.3102/0013189X11413260

Hongwei, W. (2012). Coh-Metrix: A computational tool to discriminate writing qualities. *International Education Studies, 5*(2), 204–215.

Koehler, M. J. (2015). #Tpack Mood tracker [web site]. Retreived from http://www.matt-koehler.com/Sentiment/graph.php?days=all&tag=tpack

Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education, 9*(1), 60-70.

McNamara, D. S., & Graesser, A. C. (2011). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In *Applied natural language processing: Identification investigation and resolution* (pp. 188–205). Upper Saddle River, New Jersey: Prentice-Hall International.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*(4), 292–330. doi:10.1080/01638530902959943

Mishra, P., & Koehler, M. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *The Teachers College Record, 108*(6), 1017-1054.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (pp. 122–129). Menlo Park, CA: AAAI Press.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, *31*, 527-541.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1&2), 1-135.

Perkins, J. (n.d.) *Sentiment Analysis API*. http://text-processing.com/docs/sentiment.html

Perelman, L. (2014). When "the state of the art" is counting words. *Assessing Writing, 21*, 104–111. doi:10.1016/j.asw.2014.05.001

Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing, 20*, 53–76. doi:10.1016/j.asw.2013.04.001

Thelwall, M., Buckley, K., & Paltoglou, G. (2011). Sentiment in Twitter events. *Journal of the American Society for Information Science and Technology, 62*(2), 406-418.

Tong, R. M. (2001). An operational system for detecting and tracking opinions in on-line discussion. In *Proceedings of the Workshop on Operational Text Classification* (p. 6).

Wang, C. J., Wang, P. P., & Zhu, J. J. (2013). Discussing Occupy Wall Street on Twitter: Longitudinal network analysis of equality, emotion, and stability of public discussion. *Cyberpsychology, Behavior, and Social Networking, 16*(9), 679-685.