

Case Study: University of Southern California Annenberg Innovation Lab

Gains insight into public sentiment with near-real-time analytics of social media streams

Overview

The need

The University of Southern California (USC) Annenberg Innovation Lab wanted to uncover insights buried in the millions of daily online conversations.

The solution

Through a sentiment-analytics project using IBM big data solutions, USC scholars are capturing and analyzing millions of tweets, Facebook posts and other social media conversations to help uncover trends in near-real time.

The benefit

Demonstrated the impact of a TV ad within a day of airing; helped show the sentiment of debate viewers in near-real time; expected to enable nations to gain early notice of emerging health crises or civil unrest.

On the University of Southern California (USC) campus in Los Angeles, exciting new research is helping businesses, nonprofit organizations and governmental bodies gain new insight from millions of online conversations. The project uses IBM big data solutions—including IBM® InfoSphere® Streams software and IBM BigSheets, a component of IBM InfoSphere BigInsights™ software—to help researchers analyze millions of tweets in near-real time and uncover trends that may have gone unnoticed previously.

“USC’s Annenberg Innovation Lab is what we call a think-and-do tank,” says Professor Jonathan Taplin, director of the USC Annenberg Innovation Lab. “The lab offers a place to understand today’s participatory culture and how we can make cities and society work better.”

Gaining game-changing insight

To understand what insights advertisers could gain from online conversations, the lab launched a research project using IBM BigSheets, a component of IBM InfoSphere BigInsights software, that analyzed social media posts related to new movie releases.



“If you could monitor a malaria epidemic or a civil conflict in Africa by analyzing huge amounts of Twitter data, we think that organizations could see where problems are developing and potentially get ahead of an issue.”

—Professor Jonathan Taplin, Director,
Annenberg Innovation Lab, University of
Southern California

“We were trying to understand: Could you determine how movies would open based on the social buzz?” says Taplin. “What we found was that the ability to understand public sentiment in real time was very predictive of how a movie would open and what advertising worked. For instance, we tracked the DreamWorks’ movie *Puss in Boots*, which had a slower following in the weeks leading up to its release. When they dropped a gigantic TV buy, we saw within two days, *Puss in Boots* became the most talked about movie. We could see immediately that the ad campaign worked.”

In television, this insight could potentially change how networks develop shows.

Taplin explains: “TV today is measured the same way as it was measured in the 1950s. The current ratings system just tells you what channel is turned on in 2,500 homes in America. It tells you nothing about whether the viewers are engaged and what they feel about the programs. What happens when you begin to analyze a million tweets around a piece of programming? We took the *Oscars* and we could see what people thought about the programming and the advertising, every segment of it, in real time. What this means is that producers can actually see where the sentiment turned south in their programs, and then look at the tweets to understand what viewers didn’t like. In some cases, the outcome of a reality show could be changed by audience response.”

Understanding the reaction of debate viewers in near-real time

The business opportunities for media, entertainment and advertising are significant. This insight can give marketers near-real-time insight into what consumers are thinking so they can adjust their strategies and offers when it’s still possible to influence the outcomes.

Solution components

Software

- IBM® InfoSphere® Streams
- IBM BigSheets, a component of IBM InfoSphere BigInsights™

Servers

- IBM BladeCenter® HS22
-

But the social implications of sentiment analytics are even more compelling—potentially influencing everything from election strategies to public policies to emergency response.

“We looked at one Republican [candidate] debate, and each candidate had a sentiment meter,” says Taplin. “It was literally an 800,000 person focus group in real time. You could see as someone made a mistake, the sentiment change on the dashboard in seconds. It was very powerful. During the final Presidential debates, we were analyzing 2,000 tweets per second in real time.”

Monitoring our “global pulse”

Another research project the lab is exploring is how to monitor what the United Nations (U.N.) calls the “global pulse.”

“Twitter and Facebook are used a lot in the Third World,” says Taplin. “If you could monitor a malaria epidemic or a civil conflict in Africa by analyzing huge amounts of Twitter data, we think that organizations could see where problems are developing and potentially get ahead of an issue.”

Creating a platform for stream computing

According to Taplin, one of the biggest challenges in conducting sentiment analytics on social conversations is the amount of data. “How do you analyze it when you’re at the end of a fire hose?” he says.

In the case of analyzing the Presidential debates, the Annenberg Innovation Lab used IBM InfoSphere Streams software, an IBM big data solution, running on IBM BladeCenter® blade servers, to handle this huge amount of data. The software helps process, filter and analyze the millions of Twitter messages and Facebook posts as the data streams in, and uses natural language processing (NLP) capabilities to determine whether each message is positive or negative. Students also use InfoSphere Streams software to archive data, making it easier to search for information from a particular timeframe.

“When you’re looking at a political debate or an Oscars show, you need to analyze a million tweets in a very short period of time. We’ve never had a problem tackling these projects with IBM Infosphere Streams.”

—Professor Jonathan Taplin

“Speed and volume crush most of the analytics products,” says Taplin. “When you’re looking at a political debate or an *Oscars* show, you need to analyze a million tweets in a very short period of time. We’ve never had a problem tackling these projects with IBM Infosphere Streams. It’s never crashed. And that’s been amazing.”

Ease of use of the technology is critical, he says.

“A lot of students are incredibly interested in social media analytics, but they aren’t necessarily math or computer science majors,” says Taplin. “IBM’s big data solutions allow the layperson to work with this data and make it reasonably easy for people to create very rich graphics and beautiful pictures of what’s going on in the world.”

Teaching the software the nuances of language

In every subject area—politics, entertainment, technology, and so on—people often use jargon, and sometimes sarcasm, to convey their ideas and opinions. To help improve the accuracy of sentiment analytics, computers must be trained to learn these nuances of language.

During the first research projects, students annotated the data to help the software more accurately interpret the meaning of each tweet. Moving forward, IBM InfoSphere BigInsights software, another IBM big data offering, will be used to help the system understand the tone of each message.

“We found out early on that almost 70 percent of tweets on politics are sarcastic,” says Taplin. “For example, in one of our big data polls of a Republican candidate debate, one of the tweets was: ‘I’m so happy so-and so threw their tinfoil hat in the ring.’ We used annotations to help the software understand that ‘tinfoil hat’ made it a negative comment versus a positive one. And, at this point, we’re in the 70 percent range [in terms of accuracy], which is better than a lot of other products in the marketplace. We’re in a really nice partnership, where we’re sharing what we’ve learned with IBM and IBM is helping us make our tools better.”

Confirming the integrity of the results

As part of the research project, USC mathematicians and computer scientists are examining the mathematics behind the scenes to confirm the integrity of the results.

“There are hundreds of companies selling ‘black box’ solutions to tell you what the social media sphere is saying,” says Taplin. “But they won’t show you the math and the algorithms behind the predictions, and our fear is that it could taint this very important process. IBM takes a very rigorous approach and is helping our mathematicians and computer scientists understand how it works.”

For more information

To learn more about IBM InfoSphere Streams software, please contact your IBM sales representative or IBM Business Partner or visit the following website: ibm.com/software/data/infosphere/streams

To learn more about IBM big data solutions, visit:
ibm.com/software/data/bigdata

To get involved in the conversation, visit:
www.smartercomputingblog.com/category/big-data

For more information about the USC Annenberg Innovation Lab, visit:
www.annenberglab.com

For more information about the University of Southern California, visit:
www.usc.edu



© Copyright IBM Corporation 2013

IBM Corporation
Software Group
Route 100
Somers, NY 10589

Produced in the United States of America
February 2013

IBM, the IBM logo, ibm.com, BladeCenter, InfoSphere, and BigInsights are trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at ibm.com/legal/copytrade.shtml

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions. It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.



Please Recycle