Analyzing Twitter for trends in online higher education

Janu Verma

Perspective

This summer I completed Coursera's Introduction to Data Science taught by Prof. Bill Howe, University of Washington. As a part of this course, students were encouraged to participate in real-life projects posted on the forums. Prof. Howe invited organizations and individuals to post research and business problems for the students to work on. I worked on a project which involved analyzing twitter contents relevant to higher education. This work was proposed by Coursolve, an organization based in Boston.

Background

Coursolve connects courses with organizations to empowers students to solve real-world problems. In other words, they match nonprofits, social enterprises, small businesses, and other organizations with relevant academic courses so that student's work in course projects can meet these organizations' needs. The goals for this project are as follows:

1. Learn what topics are currently popular in online higher education discussion circles.
2. Investigate how best to leverage social media to disseminate relevant content and promote brand visibility.

Increasing reach on social media will:

- Boost brand recognition among instructors and students (and organizations)
- Offer a venue for major announcements and updates
- Enable Coursolve to gradually become thought leader in the field of higher education, skills development, and connected learning, thereby further expanding their reach and credibility

Also Coursolve can focus on the popular topics to attract more students and organizations to their platform.

Research Questions

1. What are the most common topics that thought leaders in online higher education are discussing?
2. Which tweets in the online higher-ed get the most interactions? What characterizes most successful tweets?
3. What timings for tweets maximize engagement?

Data Set

We made a list of individuals, educational organizations and media houses who actively post content relevant to higher education. An initial list was proposed by Coursolve research group and this list was refined over time by further discussions. Also new thought leaders emerged as we performed our analysis. In addition, we looked for tweets matching some keywords relevant to the topics of interest.
The dataset includes :

1. Tweets from the users: Coursera, Udacity, edX, MITOCW, Futurelearn, Inside Higher Education, IHEtech, Clayton Christensen, Sebastian Thrun, Anant Agarwal, Sanjay Sharma, Roy Pea, Educationweek, The Chronicle of Higher Education, MOOC News Review.
2. Tweets matching the search results (hashtags, keywords) for: mooc, moocs, highered, edtech, elearning.
3. Tweets obtained by filtering the live stream for the above keywords.

The data was collected over a period of 1 month starting early May to 10 June 2013.

Methodology

1. Extract the data from Twitter and cleaning the dataset using Python scripts into a processable format (JSON).
2. Look for popular hashtags in these tweets.
3. Extract times of the day/week when most of the tweets in the data are created.
4. Create a smaller dataset comprising of the tweets getting maximum attention (i.e. with high enough retweet or favorite count).
5. Look for the top hashtags and most popular words in highly successful tweets.
6. Extract the heavily shared links and articles.

Analysis

1. We computed the most frequently occurring hashtags in the full dataset. In addition to the expected hashtags (e.g. edtech, highered, elearnling, mooc, moocs, college, learning, teachers, students, online etc), the output also contains hashtags like (in order of decreasing frequencies):
   edchat, ipad/ipaded, edtechchat, edapp/edapps, k12, infographic, ubtech, stem, ukedchat, mathchat, socialmedia, apps, onlinelearning.
   See the tag cloud - hashtags in all tweets.png

2. We sorted the tweets in dateset by retweet count and created a new dataset consisiting of tweets with more than 30 retweets or favorites (call these popular tweets). Some of these tweets were not relevant to our analysis e.g. tweet by a user having 'edtech' in twitter handle, though the tweet has nothing to do with higher education. We excluded such tweets to make our analysis more predictive.

3. Very popular tweets i.e. tweets with more than 100 retweets or favorites were collected in a different file and studied further. One of these tweets talks about edX going opensource and one is a link to an article in NY-Times.

4. Top hashtags in the most successful tweets are visualized through a word cloud. We see that hashtags involving online chats, apps, iPads occur very frequently.
   See - hashtags in popular tweets.png

5. Word cloud expressing word-count in most engaged tweets in provided. We can see in this word cloud that tweets involving announcements are more successful. Some words which show this are announces, announcing, best, new, great, turns, partnership, starts, offer, officially etc.
   See - wordcount in popular tweets.png

6. We provide a bubble cloud of most popular leaders in the field of online and tech higher education. For this we considered the set of most successful tweets and then looked at the users featuring prominently in these tweets.

7. Timings of tweets are plotted in the charts
   Sunday.png, Monday.png, Tuesday.png, Wednesday.png, Thursday.png

   We collected tweets in the field of online higher education over a period of 5 days and then extracted the hours of the day when the tweets were created. We then constructed these bar graphs for each day. The hours are in accordance with UTC timings. It can be read from the graphs that the online higher education community at twitter are most active during the hours 1300 - 1700 everyday.

Insights

1. In addition to the regularly used hashtags (mooc, edtech, elearning, highered), some other hashtags which we discovered in our analysis are also getting lot of attention.

2. Tweets about the announcements in the online higher education are getting maximum engagement.

3. Tweets about the technological hashtags/keywords like apps, ipads, edapps appear frequently in the most popular tweets

4. Tweets about the online degrees were more engaged. Hashtags and keywords like phd, masters, CS appeared often in most engaged tweets.

5. Tweets about different chat hashtags get a lot of interaction. Some of these, occurring very often in our data, are edchat, edtechchat, ukedchat, and mathchat. Following and participating in these chats can help in increasing visibility.

6. Retweets of thought leaders tend to become more popular. Tweeting interesting content at top leaders which may result in a retweet is a good strategy. This will help the content to reach wider audience. Some of the popular tweets in our data were tweeted at thought leaders and were then retweeted by the leaders.

7. Tweets at the peak times have more chances of being noticed (1300-1700 hours UTC).