Exploring Redditors' Communication Style

Yilang Zhao

Department of Curriculum and Instruction, University of Wisconsin–Madison, USA.

Abstract

This paper explores the communication style of parent and child posts in a popular Reddit channel, r/todayilearn. By implementing epistemic network analysis (ENA), this paper discovers that parent posts on this subreddit channel tend to have more opinions, yet child posts are more likely to present evidence or external supports to back up their statements. Understanding how people communicate in online communities may help establish a better community atmosphere for more productive discussion.

Keywords: Reddit, online community, communication style, epistemic network analysis.

1. Introduction

Reddit, as a popular online discussion community, has a massive amount of discussion participants (over 2 billion comments). It can be interesting to learn how people communicate with each other on Reddit, as learning scientists find that people in informal learning settings may have more effective learning (Miyake & Kirschner, 2014). Reddit has the potential to be a platform where people gain informal learning experiences from communicating with others on various topics. Thus, it is necessary to explore Redditors' communication styles to gain insights about how to establish an online environment that promotes active discussion and possibly informal learning?

2. Theoretical Background

While browsing Reddit threads, people may notice that Redditors' posts have different patterns in terms of their communication styles. Some of them prefer to refer to factual knowledge or post external links, while others are more in favor of expressing their own opinions or explaining something in their own words. What are the differences between facts and opinions?

Corvino (2014) states that the fact/opinion distinction is philosophically ambiguous, and there are three general distinctions that confuse people. First, people think facts are reality and opinions are beliefs that represent reality, but both facts and opinions can represent reality either successfully or unsuccessfully. For example, the sentence that "there is beer in the refrigerator" can either be a statement of fact or an opinion claimed by someone. Second, many people believe that facts are objective, and opinions are subjective. The problem of this distinction resides on the definition of the subjective statement. Subjectivity is mind dependent. For instance, the statement that "God created the earth" can be an object matter to God believers but seems to be a subjective assertion made by God believers to atheists. Finally, people hold the idea that facts are descriptive, but opinions are normative. Nonetheless, not all the opinions are normative. The statement that "a Democrat will win the presidency in 2016" is an opinion but it is not normative. Therefore, Corvino (2014) propose the following definitions to differentiate facts and opinions: A statement of fact is one that has objective content and is well-supported by the available evidence, and a statement of opinion is one whose content is either subjective or else not well supported by the available evidence.

Besides the fact/opinion distinction, it is also essential to know the structure of a Reddit thread. On Reddit, a piece of information can spread to a large number of people in a very short time through a person-to-person process, which is defined as a viral structure (Goel, Anderson, Hofman & Watts, 2015). However, the viral structure is extremely complicated and hard to analyze. This study only concentrates on the separate parent and child posts.

Another reason for choosing separate parent and child posts for analysis is that in Miyake and Kirschner's (2014) collaborative learning model, they claim that interdependency and task cohesion are two elements that contribute to constructive collaborative learning. They state that achieving a learning goal requires the inter-connections among the completion of sub-tasks and those independent sub-tasks have shared commitments and collaborative efforts of a group. On Reddit, the sub-tasks can be the parent and child posts which share a goal of adding new information or discussing the existing information. Thus, it is intriguing to explore the connections between parent and child posts.

Therefore, the research question of this study is: what are the communication styles of the parent and child posts on Reddit?

3. Methods

3.1. Data Collecting and Processing

The dataset was collected from the subreddit channel, r/todayilearn. The original data was the top 50 hottest (most commented) threads on March 6th, 2019, including all the posts of those threads. The total number of posts was 14,050. Half of them were child posts and the other half were their direct parent posts. Since a child post could be a parent post of other posts, there was a considerable number of overlapped lines. Since the analytical tool has a data amount limit, with all the duplicates are removed, 1706 unique parent posts and 5217 unique child posts were kept for later analysis.

3.2. Coding

Six codes are included (see Table 1) in this study, which are Link, Questioning, Explanation, Opinion, Science, and Politics. Codes are created by examining the top 3 hottest threads which have more comments than any other rest threads and then conceptually similar codes are aggregated. Since the dataset in this study has a large volume of data that is impossible for human coders to code, an automatic coding tool, nCoder (www.n-coder.org), is used to code the dataset and all the codes are statistically valid (kappa > 0.65, rho < 0.05).

Table 1. Coding Scheme

Name	Definition	Example	Kappa	Rho
Link	External references that people will request or post to share information.	Well... I stand kinda corrected. Someone conceived of the idea first. However [Tesla was the innovator](https://www.pbs.org/tesla /ins/lab_imlight.html) and deserves most of the credit and should have applied for a patent.	1.00	0.01
Questioning	Confused by the parent post and want additional information.	Still not sure if advertisement for Amtrak	1.00	0.04
Explanation	Suggesting reasons, giving details, and clarifying vagueness.	I remember a similar story, where a university student came to math class late and their professor had an unsolvable or unsolved equation/proof/theorem on the board.	0.89	0.02
Opinion	Expressing personal ideas, feelings, judgements; no matter they are new or not.	I don't get it. Why would anyone think frosted glass would be impossible?	1.00	0.03
Science	Using general scientific knowledge/facts/stories to respond to a parent post.	I mean, radiation does have its place in the medical field...we use it to treat cancers. We use it for x Ray's. We use it to trace through your blood stream for tests....toothpaste and radioactive spam though?	1.00	0.02
Politics	Responses to the parent post that contain political terms/issues.	The government has been doing this for a while but to search for a bomb instead of using it as one	1.00	0.03

3.3. Data Analysis

In this study, Epistemic Network Analysis (ENA) is applied as the analytical tool (www.epistemicnetwork.org). ENA is the tool that measures relationships among codes by quantifying the co-occurrence of codes in discourse (Shaffer, 2017). For this analysis, every single parent and child post is segmented as a conversation. The minimal unit of analysis is a single post. The ENA algorithm in this analysis uses the whole conversation as a stanza.

In the ENA model of this study, networks are visualized using network graphs where nodes corresponded to the codes, and edges reflect the relative frequency of co-occurrence, or connection, between two codes. The results are two coordinated representations for each unit of analysis: (1) a plotted point, which represents the location of that unit's network in the low-dimensional projected space, and (2) a weighted network graph. The positions of the network graph nodes are fixed, and those positions are determined by an optimization routine that minimizes the difference between the plotted points and their corresponding network centroids. Because of this co-registration of network graphs and projected space, the positions of the network graph nodes—and the connections they define—can be used to interpret the dimensions of the projected space and explain the positions of plotted points in the space. Additionally, the weights of those edges represent the degree of connection between two nodes.

4. Results

4.1. Qualitative Analysis

The below is a qualitative excerpt of a parent post and a child post.

Parent post:

There was a Carl Weathers for Governor skit in SNL once, but the only YouTube link I found had been taken down. I think his slogan was: "Carl Weathers, because I was also in Predator."

Child post:

so the ultraviolet light from your headlights is super weird to them. I would not expect that the typical incandescent car headlight would emit much UV light -- the filaments just don't get hot enough. [LEDs also do not typically emit much UV](https://sciencing.com/light-bulbs-not-emit-uv-radiation-15925.html). [HID lights do emit significant amounts of UV](https://en.wikipedia.org/wiki/High-intensity_discharge_lamp), however we then add filters to absorb this, as this UV would be a danger to us. That said, a bright light at night that suddenly appears would indeed be "super weird" to them -- at least until they've experienced a few cars. No need to blame this on UV.

From this qualitative example, the parent post attempts to express an opinion, which is subjective and not well-supported by evidence, though this person tries to use a YouTube video to support his/her claim. However, the selected child post is not only expressing and explaining personal ideas but also providing external links and using scientific terms to justify his statement, which includes both opinions and facts. Thus, the communication style of parent posts is more on the opinion side, while the communication style of child posts is more comprehensive, which usually consists of both facts and opinions.

4.2. Quantitative Analysis

From Figure 1, the positions of the codes along the X-axis are depending on their closeness to a fact or an opinion. *Questioning* uses existing posts as evidence and *Science* has scientific terms, which make them more like facts. *Explanation* and *Opinion* are more subjective and possibly lack of supportive evidence so that they are on the opinion side according to Corvino's (2014) definitions. *Link* and *Politics* can be used for either presenting facts or expressing opinions. That is why they are close to the Y-axis. In addition, the child post (green) has strong connections between the fact codes, and the opinion code. *Explanation* is also tightly linked with the left fact codes. However, the parent post has a different pattern. It is completely based on the opinion codes, *Explanation* and *Opinion*, and there is no connections with the fact codes, *Questioning* and *Science*. A two-sample t-test with the ENA web tool was conducted to examine a generalized result. Along the X axis, the parent posts (mean=0.01, SD=0.22) was statistically significantly different at the alpha= 0.05 level from the child posts (mean=0.00, SD=0.18) with a p-value equal to 0.00 and Cohen's d equal to 0.09.
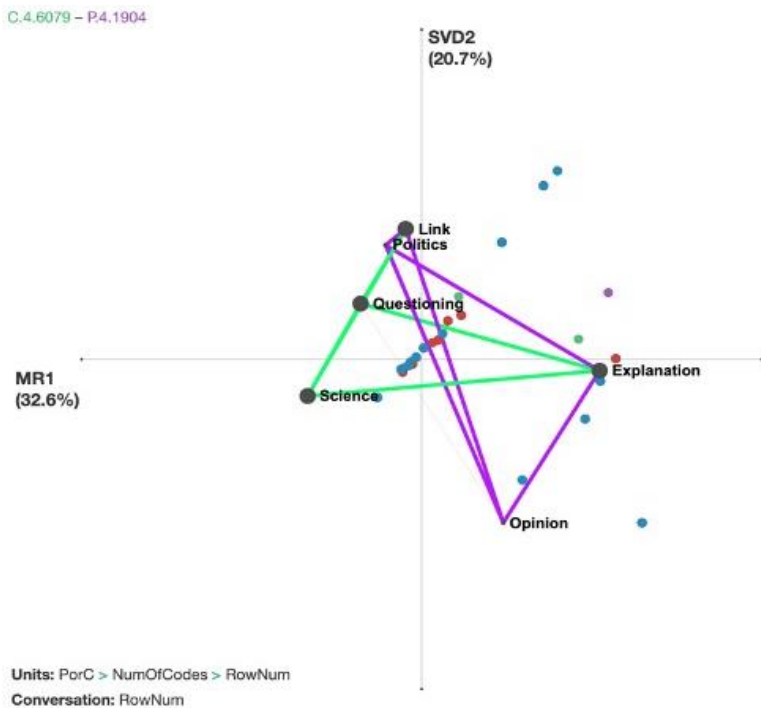

Figure 1. ENA Comparison Plot

5. Discussion

5.1. Finding

The above qualitative and quantitative results reveal that parent posts and child posts on Reddit have different communication styles. The parent posts are often simpler and express an opinion. However, in response to those opinion-based parent posts, the child posts are usually more comprehensive, including both facts and opinions. People tend to have discussions with others by using child posts that are well-supported by sound evidence and contain the replier's own ideas.

The theoretical contribution of this study is to extend understanding of the pattern with which people communicate in discussion communities to promote further collaborative online learning. As for the practical contribution, online discussion participants may understand how to post a more constructive response that provides either facts or opinions with a goal of contributing to a better informal learning environment.

5.2. Limitation

One of the major problems of this study is the Redditors' use of language. The discussion atmosphere of some Reddit channels is extensively informal so that Redditors tend to use casual languages and include many memes which cannot be accurately identified by nCoder. Additionally, the topics on Reddit vary a lot every day. It is impossible to create a coding scheme that covers all the topics. The codes selected in this study may only be effective for analyzing the data collected in the certain subreddit on that day.

References

Corvino, J. (2014). The fact/opinion distinction. The Philosophers' Magazine, (65), 57-61.

Goel, S., Anderson, A., Hofman, J., & Watts, D. (2015). The Structural Virality of Online Diffusion. Management Science, 150722112809007.

Miyake, N., & Kirschner, P. (2014). The Social and Interactive Dimensions of Collaborative Learning. In R. Sawyer (Ed.), The Cambridge Handbook of the Learning Sciences (Cambridge Handbooks in Psychology, pp. 418-438). Cambridge: Cambridge University Press. doi:10.1017/CBO9781139519526.026

Shaffer, D. W. (2017). Quantitative ethnography. Madison, WI: Cathcart Press.