Structured conversations with virtual pen pals to teach empathy for opposing viewpoints

Vincent La and James Peng

Introduction

Currently, the inequity within our society is growing to unprecedented levels (Piketty, 2015). This means that people are often living in communities that are more siloed and thus do not have access to interactions with other people from different communities and perspectives. This breeds a culture of cynicism, discrimination, and xenophobia. By encouraging people to have conversations with others with different experiences, we can teach empathy and expose people to different viewpoints, which foster a deeper sense of community with the world at large. Studies have shown that when people have empathy, they are more likely to display societal engagement, have better communication skills, and develop more positive relationships (Rumble, 2010).

This problem is especially exemplified in today's political climate. Ditto and Koleva describe the differences between American liberals and conservatives in terms of a lack of moral empathy. Often, the difficulty to relate to others' points of view comes from a lack of understanding of the visceral, conditioned responses that are outside of an individual's control. For example, liberals often cannot relate to extreme conservatives' visceral distaste for "nontraditional" sexuality, while conservatives often cannot relate to liberals' strong desires to address prejudice at the expense of personal freedoms. While this moral empathy gap is hard to bridge, Ditto and Koleva argue that recognizing and appreciating these affective differences is a good first step to alleviating the culture war (Ditto, 2011).

We believe that existing solutions aren't sufficient because there is not an existing tool that explicitly aims to connect an individual with someone who has a diametrically opposing viewpoint. Furthermore, we believe that applications of natural language processing can help facilitate conversations, by, for example, screening out harmful or offensive language. In addition, while pen pal programs exist for classroom, we aim to make an application that is open to all ages, not just young students.

Related Work

There have been numerous studies regarding to efficacy of traditional pen-pal relationships, though not necessarily through online applications. For example, Ceprano investigated a semester-long pen-pal project at 18 participating university language arts students, a class of first graders and their teacher, and the university instructors to become co-learners in meaningful, actionable research. The weekly letters the university students

exchanged with a class of first-graders became the springboard for discussions. Insights about writing as derived by all participants were discussed and were accompanied with exemplary samples of children's emergent literacy. The authors found that children's growth in voice as well as the technical aspects of their writing improved. In addition, excerpts from the students' journals show that they were more aware of the effects of mediation on their own writing voice (Ceprano, 1998).

Online applications to facilitate pen-pal relationships at scale have been developed more recently. For example, Lie et al. looked at the efficacy of PenPalSchools, an application that matches classes from schools with students from different geographic regions to discuss topics (Lie, 2018). For example, after the 2016 election, PenPalSchools connected classes from different geographic regions in the United States to discuss election issues in their pen pal correspondences. It was also used in Malaysia, in order to help Malaysian students learn English and alleviate the aversion that many Malaysian students have toward writing. This study explored the potential of using online material, in the English as Second Language (ESL) classroom to develop writing skills with peers around the world via an online collaborative project. Thirty 12-year-old primary school students from Cheras, Selangor were chosen to join PenPalSchools and take part in the online collaborative project. The research used mixed-method design; quantitative data from pre- and post-tests and responses from semi-structured interviews were used to measure the outcome. The post test result reflected the improvement in their writing skill, and it had found out that this educational website could make the writing lessons become more interesting as they could communicate and learn with peers around the world, particularly native speakers.

In addition, keeping users safe is of paramount concern for us. We don't want users to be interacting with potential online bullies which might result in negative experiences for others who legitimately want to learn. To address this, we will consider the use of Natural Language Processing (NLP) to analyze text and detect harmful and offensive language to better facilitate conversation. Automatic detection of offensive language has been done in many other contexts as well. For example, Kou's research examined the use case of detecting harmful language in the context of a popular video game, League of Legends (Kou, 2013). They use machine learning and NLP to detect offensive language within the game chat feature and escalate reports to moderators who can then administer punishments such as bans from the system.

Proposed Work

The tool will be an online application that will be open to all people. To use the tool, a user will first have to login or create a new account. Upon creating a new account, the user will have to fill out a profile composing of age, gender, political status, opinions on controversial topics, among other topics they feel passionate about. The user can then choose in which of the

areas they would want to develop more empathy. The application will then have a matchmaking system that will then match users to other users with different viewpoints who share similar passion about the topic.

Before diving too deeply into developing the rest of the application, we will create mockups of the application first and conduct user research to help us better figure out use cases and inform intricacies of the application. James will likely be in charge of creating mockups and researching the characteristics in which we hope to teach empathy, while Vincent will be in charge of conducting user research.

After matching with a pen pal, we envision a very structured conversation between the two pen pals. For example, the first few lines of the conversation could be used to establish the humanity of the participants. For example, we could ask non-controversial questions such as "What is your favorite TV show?" or "What is your hometown like?" After the pen pals establish a humanistic relationship, we then envision a prompt that dives deeply into the heart of their disagreements. For example, one topic might be regarding the viability of Universal Basic Income. For this topic, a starting question might be: "Would you support universal basic income. Why or why not?". Further questions might be: "How would you support people whose jobs are at high risk of automation?" or "Would you support universal basic income for everyone or just people under a certain income level?". The users should then write thoughtful discussions; we do not envision this application to be a chat application. The responses should have sufficient substance, perhaps we might impose a minimum word count, to facilitate engaging conversations. In order to provide a scaffold for a more informed argument, we also will consider having a library of articles expressing both viewpoints. We envision peppering these articles into the conversation and asking follow-up questions about what the pen pals thought about the article and how it relates to their viewpoints and experience. Afterwards, we could have a survey that asks the users to provide feedback on their interactions and perhaps even leave user reviews in terms of how they felt their conversation went.

Vincent will primarily be in charge of creating initial set of questions for users to get to know each other and facilitate conversation. James will primarily be involved with researching related articles for each topic. For example, for the Universal Basic Income example, Reed's 2016 article and Atkinson's 1996 research are very relevant.

On the backend, both Vincent and James will work on deciding the schema document for the web app, and Vincent will design the database structure for the web app. We will use Django to build the web app.

Other considerations include machine learning and natural language processing to protect the safety of our users; especially to detect harmful and offensive language. Vincent will primarily be in charge of discovery work around viability of ML and NLP in the application. However, there is potential risk here since the majority of the work in our project will be to develop the application. Realistically, since we likely will not have large scale adoption of the application before the semester is over it might be very difficult for us to obtain sufficient

training data to truly build machine learning models that can detect offensive language. Thus, a fall-back approach might be to allow for "moderators" to manually review conversations that get flagged as inappropriate from the user to at least provide punishments (e.g. bans from the application) for users who are particularly flagrant.

Finally, we believe it is very important to have an objective way to evaluate the effectiveness of the application. Although a full-scale study into its efficacy is not feasible within the timeframe of this class, we plan on recruiting pilot participants to test the application and engage in real conversations with pen pals with opposing viewpoints. Our goal is to recruit between 10-20 pilot participants to participate. At the end of the pilot, we'll ask the pilot participants about their experience and how their mindset did or didn't change throughout the process.

Deliverables

The deliverables for this project will include:

Intermediate Milestone #1: For the first milestone, we'll create a detailed schema of the app, describing its full functionality and basic implementation details. Everything -- from the login process to the matchmaking procedure -- will be described in this first milestone document. Additionally, we will include a list of attributes that we want to teach empathy around, such as political party, economic status, among others. We'll develop a library of resources with literature from "both sides," so that we can strategically pepper articles into conversations in order to promote a more informed discussion. Finally, we'll create a set of survey questions that we plan on asking all users about their experience with the app and how to make it better.

Description of Deliverable	This will consist of two files. 1) **Title: Design_For_Web_App.pdf**: a document with a detailed description of the app schema. This will describe each page of the app, detailing all content and functionality, and present abstract code describing the response to user interaction with the page. We will also include screenshots that show a mock-up of the app's design. 2) **Title: Design_For_Database.pdf**: Design of database. This will be a graphic, showing all of the tables in the database, their variables and primary keys, as well as foreign key relations.
Fall-Back Plan on Unpredictable Items	N/A

Intermediate Milestone #2: Our goal for the second milestone is to develop a completely functional version of the app using the Python/Django stack with a SQL database -- except we will limit our natural language processing features, since these may take a little bit more time to develop. The app will be hosted on a website (name to be determined) and completely functional for signups for new users, with an automatic matchmaking algorithm to pair pen pals when possible.

Description of Deliverable	This will be a fully functional website using the Python/Django stack, with all functionality described in the Intermediate Milestone #1 schema document. For this milestone, we'll ignore all NLP functionality due to time limitations, since building the app itself will require a significant amount of work.
Fall-Back Plan on Unpredictable Items	1) Matching algorithm might be hard to implement. If not fully implemented, then we'll describe a method for administrators to match two pen pals on the backend manually. 2) UI Design. Since neither Vincent nor James are experienced in front-end engineering, the UI may look worse than the screenshots described in the schema document. However, all backend functionality should be implemented in this phase.

Final Project: The final project will include two parts. First, we will prepare a fully functional version of the app with the NLP functionality, hosted online and available for usage. The code for this will be available in a private Github repository. Second, we will include a paper that describes the app's functionality and NLP features, as well as results from a small pilot group that used the app. Artifacts such as real conversations between pen pals will be attached as appendices to the final paper.

Description of Deliverable	This will consist of two parts. 1) A fully functional website using the Python/Django stack, with all functionality described in the Intermediate Milestone #1 schema document, including NLP features and matching algorithm implemented. 2) An "academic" paper describing the website's functionality and results from a one week experiment with pilot users.
Fall-Back Plan on	1) NLP might be hard to implement, given the limited amount

Unpredictable Items	of training data. If that's the case, we'll -- in our paper -- describe a plan for manual monitoring of the conversations as well as present an overview of NLP technologies in other social media contexts.

Bibliography

1. Atkinson, A. B. (1996). Public economics in action: the basic income/flat tax proposal. *OUP Catalogue*.
2. Ceprano, M. A., & Garan, E. M. (1998). Emerging voices in a university pen-pal project: Layers of discovery in action research. Literacy Research and Instruction, 38(1), 31-56.
3. Ditto, P. H., & Koleva, S. P. (2011). Moral empathy gaps and the American culture war. Emotion Review, 3(3), 331-332.
4. Kou, Y., & Nardi, B. (2013). Regulating anti-social behavior on the Internet: The example of League of Legends.
5. Lie, W. W., & Yunus, M. M. (2018). Pen Pals Are Now in Your Finger Tips—A Global Collaboration Online Project to Develop Writing Skills.
6. Piketty, T. (2015). About capital in the twenty-first century. American Economic Review, 105(5), 48-53.
7. Rumble, A. C., Van Lange, P. A., & Parks, C. D. (2010). The benefits of empathy: When empathy may sustain cooperation in social dilemmas. *European Journal of Social Psychology*, *40*(5), 856-866.
8. Reed, H., & Lansley, S. (2016). Universal Basic Income: An idea whose time has come?. London: Compass.