Comparing Values and Sentiment Using Mechanical Turk

Thomas Clay Templeton
University of Maryland
4110 Hornbake Building, South Wing
College Park, MD 20742-4345

clayt@umd.edu

Kenneth R. Fleischmann
University of Maryland
4105 Hornbake Building, South Wing
College Park, MD 20742-4345

kfleisch@umd.edu

Jordan Boyd-Graber
University of Maryland
4105 Hornbake Building, South Wing
College Park, MD 20742-4345

jbg@umiacs.umd.edu

ABSTRACT

Human values can help to explain people's sentiment toward current events. In this experiment, we compare people's values with their agreement or disagreement with paragraphs that were classified as either supporting or opposing a specific topic. We found that five value types have statistically significant agreement ($p<0.001$) for both the supporting and opposing paragraphs, in opposite directions. We hope to use these paragraph ratings to train an automatic text classifier to agree or disagree with paragraphs based on a specific value profile.

Categories and Subject Descriptors

I.2.7 [**Artificial Intelligence**]: Natural Language Processing – *discourse*, *text analysis*.

General Terms

Experimentation, Human Factors.

Keywords

Human values, sentiment analysis, text classification.

1. INTRODUCTION

In this paper, we describe an experiment used to gather data that connects people's values to their attitudes toward texts. We use texts centered on a particular topic. This provides interesting information about what values are relevant to debate over a topic. We hope to use the data to inform values-based text classifiers.

2. BACKGROUND

A human value can be understood as "a belief pertaining to desirable end states or modes of conduct that transcends specific situations [that] guides selection or evaluation of behavior, people, and events" [1, p. 20]. As such, behavior is guided by values. Here, we are particularly interested in sentiment [2] toward ongoing debates within popular culture, and the role that values might play in such debates. By understanding the role that values play in such debates, we may better understand our own decision-making, making such decision-making easier to simulate [3].

Amazon's Mechanical Turk provides a convenient platform for crowdsourcing simple tasks. In this paper, we describe an experiment using Mechanical Turk to explore relationships between values and evaluations of opinion-laden text on a topic.

3. METHODS

As a current issue with a large body of opinion-laden text associated with it, we chose the controversy over the proposed community center/mosque in downtown Manhattan. We chose forty-eight paragraphs from editorials found using Google News that represented a wide range of perspectives on the issue. We then included the paragraphs in a Mechanical Turk task in which Turkers were asked to express agreement or disagreement with each paragraph on a scale from 1 to 5. As a requirement for working on the task, we had Turkers complete a survey including demographic items and 21 questions from the Schwartz Portrait Values Questionnaire (PVQ), also using a scale from 1 to 5 [4]. We collected and analyzed responses from 59 Turkers.

In analyzing the data, we first averaged across PVQ questions that addressed the same value type, so that for each paragraph, we had 59 rows consisting of one evaluation of the paragraph and 10 value type scores. For each paragraph, we then split the responses into two groups: "disagree" (scores of 1 or 2) and "agree" (scores of 4 or 5), and took the average value type scores for each group. Next, we coded each paragraph as "pro-" or "anti-" the completion of the mosque project, and compiled averages for all ten value types for each of four groups: {agree, pro}, {disagree, pro}, {agree, con}, {disagree, con}. We used unpaired t-tests to test for statistical significance.
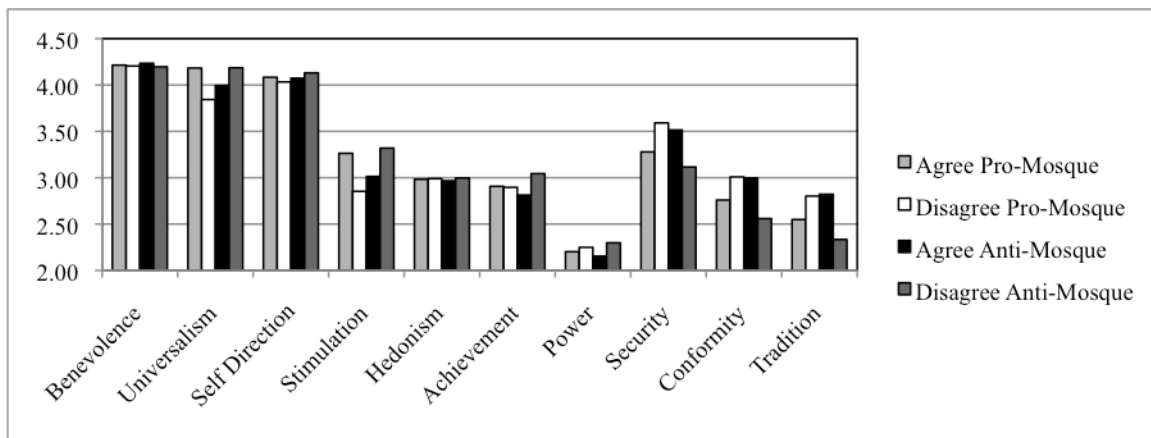
4. RESULTS

PVQ ratings are highly correlated with agreement or disagreement with pro- or anti-mosque paragraphs. Pro-mosque paragraphs were agreed with by people with high universalism and hedonism and disagreed with by people with high security, conformity, and tradition. Anti-mosque paragraphs were agreed with by people with high security, conformity, and tradition, and disagreed with by people with high universalism, stimulation, achievement, and power. Thus, two of the value types aligned completely with pro-mosque sentiment, three of the value types aligned completely with anti-mosque sentiment, and two of the value types aligned partly with pro-mosque sentiment. It is important to note that all of these differences were statistically significant to $p<0.001$.

Table 1 shows the mean PVQ results for individuals who either agreed or disagreed with either pro-mosque or anti-mosque paragraphs. Figure 1 is a column plot of these mean PVQ results.

Table 1. Mean PVQ Results for Individuals Who Agree/Disagree with Pro-/Anti-Mosque Paragraphs (***=p<0.001)

	Benevolence	Universalism	Self Direction	Stimulation	Hedonism	Achievement	Power	Security	Conformity	Tradition
Agree Pro-Mosque	4.21	4.18***	4.08	3.26***	2.98	2.91	2.20	3.28***	2.76***	2.55***
Disagree Pro-Mosque	4.20	3.84***	4.03	2.85***	2.99	2.90	2.25	3.59***	3.01***	2.80***
Agree Anti-Mosque	4.23	3.99***	4.07	3.01***	2.97	2.81***	2.15***	3.51***	3.00***	2.82***
Disagree Anti-Mosque	4.20	4.19***	4.13	3.32***	3.00	3.04***	2.30***	3.12***	2.56***	2.33***

Figure 1. Plot of Mean PVQ Results for Individuals Who Agree/Disagree with Pro-/Anti-Mosque Paragraphs (***=p<0.001)

5. IMPLICATIONS/FUTURE DIRECTIONS

These findings provide initial validation of our hypothesis that value types would be correlated with sentiment toward topics. In future studies we will include additional topics to evaluate the generalizability of this finding. We will also analyze demographic information to evaluate how much of the overall variation in sentiment toward paragraphs can be explained by values. Finally, we will use the annotated paragraphs as training data for automatic text classifiers that model specific human values. Instead of measuring the sentiment of the author of a text, this classifier would provide a reaction to a text, which would be based on a specific value profile. This approach will enable us to model human diversity based on values. Further, we hope that the findings of our future research might advance the state of the art in sentiment analysis by leveraging insights about human values.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Schwartz, S.H. 1994. Are there universal aspects in the structure and contents of human values? *J. Soc. Issues* 50, 4 (Winter 1994), 19-45.

[2] Pang, B. and Lee, L. 2008. Opinion mining and sentiment analysis. *Found. Trends Inform. Retrieval* 2, 1-2 (2008), 1-135.

[3] Fleischmann, K.R., Templeton, T.C., and Boyd-Graber, J. 2011. Modeling diverse standpoints in text classification: Learning to be human by modeling human values. In *Proceedings of the 6th iConference* (Seattle, WA, Feb. 8-11, 2011).

[4] Schwartz, S.H. 2007. Value orientations: Measurement, antecedents, and consequences across nations. In *Measuring Attitudes Cross-Nationally: Lessons from the European Social Survey*, R. Jowell, C. Roberts, and R. Fitzgerald, Eds. Sage, London, UK, 169-203