Measuring Personal Values in Cross-Cultural User-Generated Content

Yiting Shen*, Steven R. Wilson*, and Rada Mihalcea

University of Michigan, Ann Arbor MI, USA
{yiting,steverw,mihalcea}@umich.edu

Abstract. There are several standard methods used to measure personal values, including the Schwartz Values Survey and the World Values Survey. While these tools are based on well-established questionnaires, they are expensive to administer at a large scale and rely on respondents to self-report their values rather than observing what people actually choose to write about. We employ a lexicon-based method that can computationally measure personal values on a large scale. Our approach is not limited to word-counting as we explore and evaluate several alternative approaches to quantifying the usage of value-related themes in a given document. We apply our methodology to a large blog dataset comprised of text written by users from different countries around the world in order to quantify cultural differences in the expression of person values on blogs. Additionally, we analyze the relationship between the value themes expressed in blog posts and the values measured for some of the same countries using the World Values Survey.

Keywords: content analysis · personal values · user-generated content.

1 Introduction

In psychological research, *values* are typically characterized as networks of ideas that a person views to be desirable and important [20]. Psychologists, historians, and other social scientists have long argued that people's basic values influence their behaviors [1, 19]; it is generally believed that the values which people hold tend to be reliable indicators of how they will actually think and act in value-relevant situations [18]. Human values are thought to generalize across broad swaths of time and culture [21], and in fact, recent work suggests that the study of values plays a central role in cross-cultural analyses [11]. Further, values and are deeply embedded in the language that people use on a day-to-day basis [4], and we therefore expect that a strong relationship exists between the values of a cultural group and the type of content that is written about by people from that group.

While values are commonly measured using tools such as the Schwartz Values Survey and the World Values Survey [9] – well established questionnaires that ask

* Equal contributions

respondents to rate value items on a Likert-type scale [21] – it has recently been shown that topic modeling based approaches are another useful way to measure specific values, and can be applied to open-ended writing samples [3]. Even more recently, a lexicon for personal values has been introduced [24], which defines and organizes a set of dimensions related to personal values. This lexicon can be used to quantify the degree to which a text is *about* different value concepts.

Computational methods like these can be used at scale, potentially reaching larger populations. Further, these *observational* approaches do not rely on self-report data, but rather on naturally occurring data produced by authors of texts. However, it is important to measure value content in text that is personal in nature so that we can have more confidence that when value-related terms are used, they are used in connection with the author's own thoughts and beliefs. One possible source of such data is social media such as blogs, where people commonly write things about themselves in ways that might reflect their values. For example, one user in the dataset that we will explore who had a high lexicon score for the value of "Family" writes, "Happy Mother's Day to my dear mom in law, our two daughters, my sister, sister in laws, nieces, aunts, cousins, and many other beautiful women in my life", while another user with a high score for "Hard Work" writes that "Self-discipline can make the difference between an averagely talented person doing something amazing with their lives and a naturally talented person realizing very little of their potential". Based on examples like these, we expect that computational linguistic approaches, like those mentioned above, should be able to capture values in user-generated content. Further, we can use attributes associated with users' profiles to infer aspects of their culture, adding another dimension to this text-based value analysis.

Our goal in this paper is to explore cultural differences in the usage of value-laden language in personal, online user-generated content. While there are many different ways to define culture [5], we use country of residence as one way in which to divide users culturally, based on the notion of National Culture [8], and while there are many types of user-generated content, we focus on blogs because they are an ideal platform for users to write, at length, about the things that are important to them. We use the lexicon for personal values to quantify the degree to which personal blogs reflect various dimensions of value-related content. We experiment with various extensions to the typical "word counting" based approaches for quantifying concept usage in a text with lexicons, and find that more sophisticated semantic matching allows us to better discover documents that are related to the value dimensions from the lexicon. Further, we explore the degree to which conclusions that might be drawn from this type of analysis corresponds to traditional survey-based findings from the world values survey, which also divides respondents across countries

2 Methodology

As we seek to measure values-related content in text, we turn to the hierarchical values lexicon [24] that was created for this very purpose. This resource contains a

hierarchy of concepts that are related to personal values and provides the ability to define categories based on subtrees of this hierarchy. We use the authors' recommended set of 50 value concepts,[1] which includes sets of words related to values such as "Family", "Religion", and "Justice". By examining how frequently words related to these values appear in texts, we aim to get a sense of the types of values that are being discussed.

While we use a pre-constructed lexicon in this study, we first consider how exactly we should use the lexicon to quantify the usage of themes within a given document. In this section, we describe the typical approach used, list some common issues with this approach, and propose and evaluate several solutions to these problems, finally reaching a conclusion about the methodology that we will employ for our cross-cultural analysis.

2.1 Quantifying Concept Usage with Lexical Resources

Typically, a dictionary-based lexical resource, $\mathcal{L}$, contains a list of m concepts $\mathcal{L} = \{\mathcal{C}_0, \mathcal{C}_1, \ldots, \mathcal{C}_m\}$, and each concept contains a list of n patterns that match words which are associated with that concept, i.e., $\mathcal{C}_i = \{p_0, p_1, \ldots, p_n\}$. Often, these patterns are specific strings that must be matched exactly, but they might also include wildcard characters (which we denote using the "*" character) that can match any sequence of characters within a single token, e.g., "happi*" which could match the tokens "happiness", "happily", "happier", and others. The purpose of such a lexicon is to assign m scores to a document, $\mathcal{D}$, one for each of the concepts in $\mathcal{L}$, in a way that accurately captures the degree to which $\mathcal{D}$ is *about* each of the concepts. $\mathcal{D}$ itself is composed of a sequence of k tokens, which, assuming a bag-of-words model, are represented as a multiset $\mathcal{D} = \{w_0, w_1, \ldots, w_k\}$. The most common approach to compute a score, $s_{WF}(\mathcal{D}, \mathcal{C}_i)$, for $\mathcal{D}$ for any concept $\mathcal{C}_i \in \mathcal{L}$ is what we will refer to as the Word Frequency approach:

$$s_{WF}(\mathcal{D}, \mathcal{C}_i) = \frac{|\{w_j \in \mathcal{D} : m(w_j, \mathcal{C}_i) = 1|}{|\mathcal{D}|}$$

where $m(w_j, \mathcal{C}_i)$ returns 1 if at least one pattern in $\mathcal{C}_i$ matches w_j, and 0 otherwise.

Indeed, such count- or frequency-based lexicon scoring has been successfully applied to various domains, such as the measurement of depression-related content [17], the measurement of morals [7], sentiment analysis [23], and various other psychologically relevant word classes [15]. However, there are several potential problems with the Word Frequency approach. The set of words related to a concept are typically well thought-out and do a good job of capturing the "essence" of the concept, but there may be other ways to express this concept that were not included in the lexicon for any number of reasons. Content words are, by nature, open class, and thus new words may come into existence or shift

[1] The words for each category are available from the resource available in the "Values Lexicon" section at http://nlp.eecs.umich.edu/downloads.html

in meaning over time, yet we would still like to be able to quantify their relationship to lexicon themes. On the other hand, there may be words that are *somewhat* related to a concept in a lexicon, and it might be advantageous to be able to capture this. While the pattern-based nature of the paradigm that we have presented does allow for some morphological variation in the terms in a text, we may also want to match words that are semantically similar to those in a concept even if they are morphologically different. Further, simply using a wildcard may lead to some erroneous matches. Continuing our example from above, we would also match the pattern happi* with happing, which, although not a commonly used word in most text corpora, is not related to the intended concept of positive emotion and could lead to false positives. This type of problem is extremely noticeable in short texts, such as a tweets, where the categories assigned to each word contribute to a substantial proportion of the total score for that text. Yet another issue with the pattern matching approach is polysemy. Should the word "father" be more related to the value of "family" or "religion"? This is highly dependent on the context in which this word appears. In this section, we describe and evaluate two alternative approaches that can be used to help ameliorate some of the aforementioned issues with the Word Frequency approach.

Distributed Dictionary Representation

The Distributed Dictionary Representation (DDR) method was introduced to both increase the coverage of lexicon categories, but also to perform matching between categories and documents at a deeper, semantic level than can be achieved using the Word Frequency approach [6]. With DDR, the representation of the words in a category is computed by averaging their word embedding vectors, and this averaged representation is used to represent the concept of this category. That is, given a set of d-dimensional word embeddings $\mathcal{E}_i^{\mathcal{C}} = \{e_0, e_1, \ldots, e_n\}$, one per each pattern in $\mathcal{C}_i$, we compute a single vector representation of $\mathcal{C}_i$ as the mean of all embeddings in $\mathcal{E}_i^{\mathcal{C}}$, and we refer to this averaged vector as $\bar{\mathcal{E}}_i^{\mathcal{C}}$. Similarly, the representation of the given document, $\mathcal{D}$, is computed by averaging the bag of word embedding vectors $\mathcal{E}^{\mathcal{D}} = \{e_0, e_1, \ldots, e_k\}$, one for each word in the text, to get the averaged embedding $\bar{\mathcal{E}}^{\mathcal{D}}$. Importantly, the word embeddings used come from the same vector space, and so each word maps to the same d-dimensional embedding regardless of whether it appears in the $\mathcal{L}$ or $\mathcal{D}$.

Given these averaged embeddings, DDR assigns scores to documents using cosine similarity:

$$s_{DDR}(\mathcal{D}, \mathcal{C}_i) = \frac{\bar{\mathcal{E}}_i^{\mathcal{C}} \cdot \bar{\mathcal{E}}^{\mathcal{D}}}{||\bar{\mathcal{E}}_i^{\mathcal{C}}||||\bar{\mathcal{E}}^{\mathcal{D}}||}$$

Using word embeddings instead of word-counting, DDR is able to capture the concept of a category or a piece of text at a semantic level, which is consistent with the original motivation of many lexicons which were designed to identify the presence of a semantic concept in a document. In this study, we obtain all word embeddings using the FastText[2] model [2], which also has the advantage

[2] While other contextual word embeddings like ELMo [16] do a good job of capturing the meanings of words in specific contexts, lexicons such as the values lexicon that

of using subword information to obtain embeddings for words that were not seen during training.

Unsupervised Context-Based Relatedness Classification

We propose an additional technique that can be applied to an existing lexicon in order to tackle one glaring problem that is not addressed by DDR; namely, that every instance of a word will be counted toward a given category, regardless of whether or not the word present in the document has the same sense as the word in the lexicon. In this approach, which we call Unsupervised Context-Based Relatedness Classification (UCRC), we only count occurrences of tokens have the correct sense, which is inferred from the context. Similar to the Word Frequency approach, UCRC gives scores to documents as:

$$s_{UCRC}(\mathcal{D}, \mathcal{C}_i, \delta) = \frac{|\{w_j \in \mathcal{D} : m'(w_j, \mathcal{D}, \mathcal{C}_i, \delta) = 1|}{|\mathcal{D}|}$$

where the new function $m'(w_j, \mathcal{D}, \mathcal{C}_i, \delta)$ returns 1 if any pattern in ,$\mathcal{C}_i$ matches w_j and the sense of w_j is the same as the sense of the matching pattern based on a context window of size δ. This means that we must determine both the sense of w_j based on the $\frac{\delta}{2}$ previous and $\frac{\delta}{2}$ next words in $\mathcal{D}$, as well as the sense of the pattern p that matched w_j based on the intended meaning of the lexicon, which we allow to be defined manually. Rather than looking for an exact sense-level match, we simplify this by group senses into two categories: lexicon-related and non-lexicon-related. In this case, we can say that we only need to determine if w_j in $\mathcal{D}$ is lexicon-related or not.

To achieve this, we first get a set of possible contexts of the pattern from the WordNet database[14]. We can get the possible contexts of a pattern by getting the usage examples of each synset that contains words which are matched by the pattern. We term the set of possible contexts of the pattern the *context set* for p.

Next, we determine, in the *context set*, what kind of context indicates that the pattern is relevant to the lexicon, and what kind of context indicates that the pattern is not. To complete this step, we need to know whether each synset is related to $\mathcal{L}$ or not. We begin by manually annotating synsets for a subset of patterns (i.e., patterns that match a word in the synsets) that belong to some concept in $\mathcal{L}$. We randomly select 50 patterns, examine all relevant synsets, and label whether or not each synset is related to the notion of personal values. Then, under the assumption that the set of value-relevant synsets are related to one another, we automatically expand the set of lexicon-related synsets to include all synsets with a WordNet path distance that is less than some hyperparamter N, following hypernym and hyponym links when searching across paths. To tune N, we label an additional set of 30 patterns' synsets and measure the F1-score on this test set when labeling all synsets with a path distance $< N$ to be lexicon-related, varying N from 1 to 25. We find the maximum F1-score of 0.747 when

we use is this study do not provide contexts along with the category-specific words, and so further research would be required to determine how to best create, e.g., value-specific dictionary embeddings with ELMo to use within the DDR framework.

$N = 10$. Note that the default approach is to label *every* occurrence of a match as lexicon-related, leading to a high rate of false positives. Indeed, we find that in our sample, when we set $N = 10$, we *only* reduce false positives without introducing additional false negatives.

In the final step of our process, we find a single context from the *context set* of p that is most similar to the context of w_j in $\mathcal{D}$, where context is determined from a sequence of length δ surrounding the word or pattern. The similarity is computed using the cosine similarity between the average FastText embeddings for the two contexts, similar to the scoring function that was used in the DDR method. Finally, we consider the synset that appears in the most similar pattern context, and check to see if that synset was classified as lexicon-related in the previous step. If it is, then we say that the occurrence of w_j in $\mathcal{D}$ is also lexicon-related, and therefore we can count the match toward the score for the category in the lexicon.

2.2 Evaluation of Lexicon Quantification Approaches

Given our two proposed methods for improving the quantification of documents by lexicons, we design a series of evaluations that can be used to determine the viability of these approaches.

Category-Text Matching

First, we aim to determine how well the DDR quantification approach is able to accurately assign scores to documents based on the concepts defined in the lexicon. To do this, we obtain scores for each category in the lexicon across a text corpus in order to find the documents that have high, average, and low scores for each category. To test a category, we select two documents: one that has a high score for that category and another than doesn't. These two documents are presented to a set of judges on Amazon Mechanical Turk[3] who are given the category label and asked to decide which document best expresses the concept described by the label. If the judges can select the correct document significantly more than half of the time, we know that the lexicon is able to identify text that expresses the category being evaluated. There are two settings for this Category-Text Matching: *high-low* and *high-median*. In *high-low*, one of the top q scoring documents is paired with one of the bottom scoring q documents for the category, while high-median pairs this same high-scoring document with one of the q documents surrounding the median scoring document. The latter is a much more difficult version of the task since the judge must determine which of two texts that are related to a concept *most* expresses this concept.

The score for either version of the task is reported as the percentage of judges who correctly selected the high-scoring text. In each HIT, a crowd worker is shown seven pairs of texts, one of which is a randomly inserted checkpoint question based on a Wikipedia article title and contents: the title of the article is shown, and the first paragraph of the article is shown as one choice while the first paragraph of a *different* article is shown as an alternative. HIT are rejected when

[3] https://www.mturk.com

	High-Median		**High-Low**	
	DDR	**WC**	**DDR**	**WC**
Baseline	*50%*	*50%*	*50%*	*50%*
Average	**81.06%**	67%	**97.92%**	72%

Table 1. Accuracy of DDR and Word Frequency (WF) in the High-Median and High-Low settings.

workers are unable to identify the correct article. For our set of documents, we collect posts from Reddit[4] that we expect to contain some value-related content based on their subreddit categorization, such as "/r/family" and "/r/christian". We assign lexicon scores to each post using either s_{DDR} or s_{WF}, and the results are presented in Table 1. We can see that the DDR method does a much better job selecting documents that are actually perceived to be related to the lexicon categories.

Word-Sense Disambiguation

To evaluate the UCRC method as a means of unsupervised word sense grouping, we run first it on the SemCor corpus [12]. The Semcor is a lexical resource where words are annotated in terms of their WordNet synsets. With UCRC we know, for any synset related to a pattern in the lexicon, whether that synset is lexicon-related or not. Therefore, we can simply check whether each lexicon pattern that matches a labeled instance in the SemCor is relevant to the lexicon or not, essentially creating a binary prediction task (in contrast to the sense-level classification that is typically performed on the SemCor dataset). Among 352 text files in the SemCor, there are 1419 in-text lexicon pattern matches, of which 1304 are relevant to the value lexicon, and 115 are not. If the Word Frequency approach is used and every instance is labeled lexicon-related, the F1-score is 0.92. On the other hand, if UCRC is used, the F1-score is 0.97. Among these 115 conceptually unrelated words, the UCRC method is able to detect 74 of them, increasing the specificity from 0% to 64.35%. Overall, UCRC indeed has a higher F1-score and a greatly improved specificity.

Document Ranking

Before considering asking human crowd-workers to do the Category Text Matching for the same set of Reddit posts used before to evaluate the DDR method, we wanted to determine how different the ranking generated by UCRC is from the ranking generated by the standard Word Frequency approach. If there is not much of a difference, then the pairs selected for the Category Text Matching will likely not change and so UCRC would not impact the Category Text Matching score. We hypothesize that the difference in rankings might be small because the number of true negatives (with respect to lexicon-relatedness) is actually quite low in practice. To quantitatively show how those two rankings in order of relevance differ, we run Kendall's τ Rank Correlation Coefficients [10] to compare the ranking using s_{UCRC} and the ranking using the Word Frequency

[4] https://www.reddit.com

based score, s_{WF}, for each category. The closer the coefficient is to 1, the less different the two rankings are from each other (See Table 2 for specific results). We can see that all the coefficients are more than 0.83, and most of them are very close or equal to 1. Based on these results, we do not run the Category-Text Matching evaluation on UCRC, understanding that the results will likely not change. The effect of UCRC is maximized when it is run on the text where many false positives exists, but we do not find this to be the case in the corpora that we explore, and so we do not use UCRC in the following experiments. However, we do recommend the use of UCRC to those using lexicons containing many ambiguous terms or when applying lexicons to corpora containing these kinds of words.

Category	τ	**Category**	τ	**Category**	τ
forgiving	1.0	accepting-others	0.99	emotion	1.0
society	0.96	helping-others	0.94	feeling-good	0.85
significant-other	1.0	achievement	0.98	honesty	0.95
family	1.0	life	0.97	animals	0.98
friends	1.0	purpose	0.99	self-confidence	1.0
career	1.0	perseverance	1.0	dedication	1.0
relationships	1.0	religion	1.0	social	0.96
nature	1.0	learning	0.99	advice	1.0
optimism	0.94	wealth	0.98	gratitude	0.97
siblings	1.0	truth	0.91	order	0.84
health	1.0	respect	0.97	thinking	0.99
creativity	1.0	work-ethic	0.96	marriage	1.0
cognition	0.99	parents	1.0	future	0.99
security	0.97	spirituality	0.92	justice	0.96
hard-work	0.97	autonomy	1.0	art	1.0
responsible	0.98	inner-peace	0.97	children	0.96
helping-others2	0.83	moral	1.0		

Table 2. Kendall's τ Rank Correlation Coefficients for Each Category

3 Data

As a source of a large amount of user-generated content from authors around the world, we collected a corpus of blog posts from the popular platform, Blogger.[5] Since the values lexicon that we are using in this study was developed in the English language, we only consider text written by users from countries that have a large number of English speakers[6] which also have a significant presence on the Blogger platform[7]. As a result, we collect all posts written by a sample of authors from these countries: United States, India, Philippines, Nigeria, United

[5] https://www.blogger.com

[6] Based on estimations provided at https://en.wikipedia.org/wiki/List_of_countries_by_English-speaking_population

[7] At least 1,000 users claim to be from that country.

Kingdom, Canada, Australia, Pakistan, South Africa, New Zealand, Tanzania, Ireland, and Singapore. For each country, we collect a list of blogs written by users from that country[8], and subsequently collect posts from those blogs. We preprocess each blog post them by removing all HTML tags[9], and since we seek text that is personal in nature, we remove any blog posts that do not contain the word "I". Next we perform language identification,[10] and we ignore any documents that are not mostly written in English, since that is the language in which the values lexicon is built. For each document, we compute value scores for the 50 value categories described above using the DDR method with FastText embeddings,[11] and then we average the scores across all documents written by each user, since we wish to avoid unfairly weighting the scores for a country in favor of a few high-producing authors. We only consider authors for which we could retrieve at least 5 posts and when a single user has written more than 100 posts, we randomly sample 100 posts to use as a representation for that user. Finally, we average the scores for all users from a given country in order to get overall scores for each of the 50 values for each of the thirteen countries.

4 Results and Analysis

Table 3 depicts the average value category scores for each of the thirteen countries. In order to emphasize differences across the value categories, the scores for each row were divided by the average score for that row. From this heatmap, we can see that values like "marriage" and "responsibility" were talked about to a higher degree in Nigeria, while values like "life" and "gratitude" were talked about more often in blogs written by users from the Philippines and Singapore. Interestingly, certain countries, like Nigeria, had highever average usage rates of words from all value categories, while others, like India and Pakistan, had lower average usage scores overall. This is not completely surprising due to the intercorrelations between many some of the value theme scores, but it also showcases the following phenomenon: writers of blogs in some countries write in general about things related to a wide range of values, while blogs in other countries more often focus on topics that are not value-related.

Interestingly, as we analyze the various cultural differences in the usage of value-related words, we notice several groups of countries that used words from the value categories to similar degrees, possibly indicating cultural similarities between these countries. In order to explore and emphasize the similarities between countries usage of value themes, we performed a projection of the countries into a 2-dimensional space using T-SNE [13] (Figure 1). Here, we see some regional groupings, such as India and Pakistan, but also some countries that are not close as close to their neighbors in this "values space": USA is much close

[8] We collected these using code from https://github.com/costaspappus/Blogs-Scraper
[9] We use https://www.crummy.com/software/BeautifulSoup/ to clean the HTML.
[10] Using https://github.com/saffsd/langid.py
[11] As the overall results are not expected to change by a noticeable degree based on our evaluation, we opt not to use the UCRC method in the present analysis.

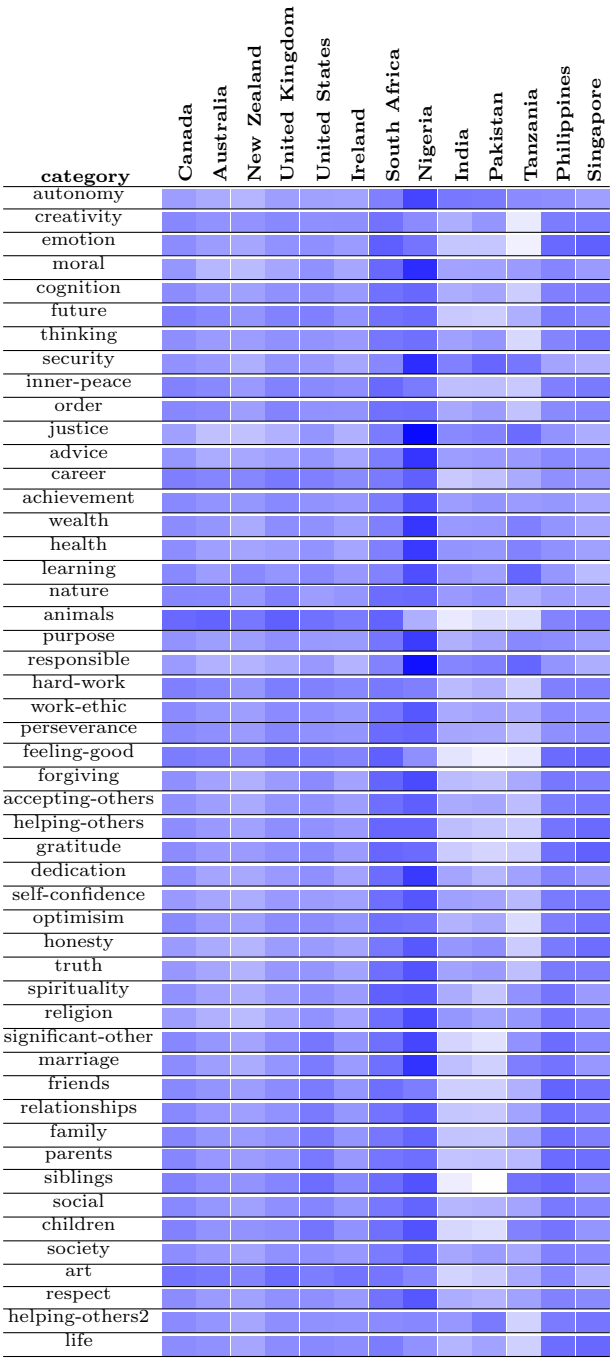

Table 3. Heat map showing normalized average lexicon scores for blog data from thirteen countries.

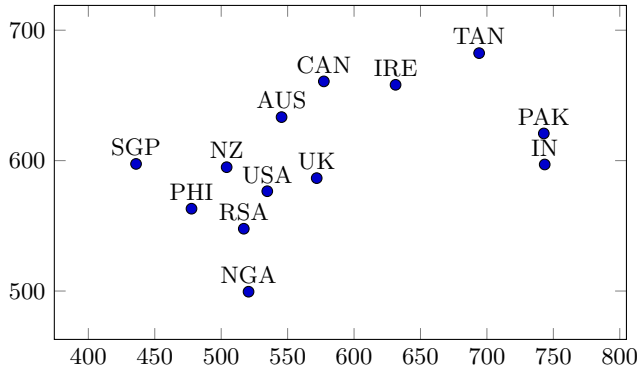

Fig. 1. T-SNE projection of countries' blogger content based on averaged value lexicon scores.

to the UK than it is to Canada, which is closer to countries like Australia and Ireland.

As a final analysis, we seek to compare the scores for some of the value categories from the lexicon with values as measured by the World Values Survey (WVS)[12]. We select a set of questions from the WVS that measure similar concepts to some of the 50 default value categories present in the values lexicon. For each of the questions, we average the results for any countries included in the WVS that are also included in our study, which includes Australia, New Zealand, India, Pakistan, Nigeria, South Africa, the Philippines, Singapore, and the United States. Then, we compute the correlation between the averaged answers to these questions and the average scores for a value lexicon category that is related to the question (Figure 2). While some of the WVS questions have little relationship to the value categories that we might expect, others actually exhibit quite a strong relationship that is even statistically significant with a small sample of measurements. For example, the average score for the "Religion" lexicon category is strongly correlated with people's answers to the question about their membership in a church or other religious organization. Two interesting cases are those of "Security" and "Trust": people from countries with high average lexicon scores for these categories actually reported feeling less secure in their neighborhoods and had less overall trust in other people. These inverse relationships may point to the level of activation of these values as a consequence of the residents' environments. Certain values may be activated in relevant situations [22], and we may be observing cases where people actually talk more about values that they feel are threatened, thus making them more relevant that they are in places where things like security and trustworthiness might be taken for granted.

[12] We use data from round 6 of the WVS, available at http://www.worldvaluessurvey.org/

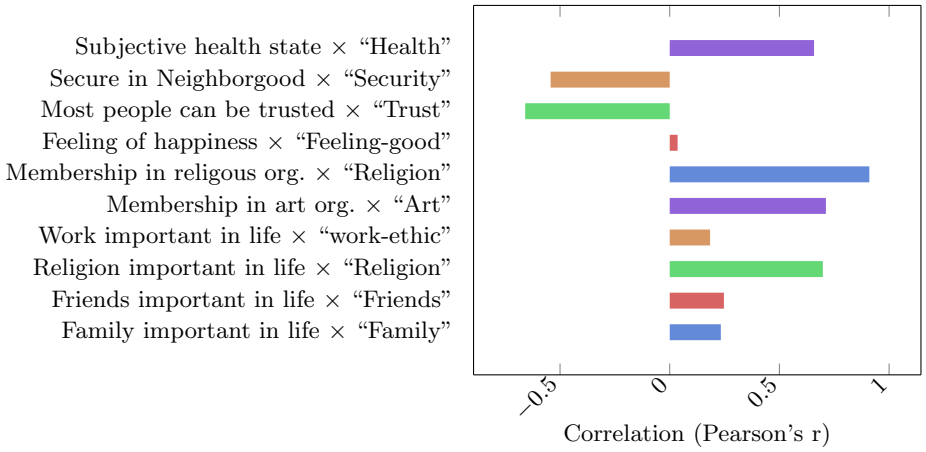

Fig. 2. Country-level correlation between aggregated WVS question responses and value lexicon scores.

5 Conclusions

We have explored our ability to employ these lexicons on a set of blog data written by authors from a range of countries in order to investigate cross-cultural differences in personal values from text. We used a lexicon that was designed to measure expressions of personal values in text data, but rather than just using it "as is", we first explored and evaluated several techniques that can be used to improve the way that we quantify the usage of lexicon themes. We found that both the DDR and UCRC methods have their merits, but for our analyses, we chose the DDR method and applied it to blogs from thirteen countries in order to gather information about the expressions of values in these countries. We used the average value theme scores to group these countries in a low-dimensional space to show which countries share similar value theme usage rates, and we compared the findings obtained using this text-based method with the results from the most recently completed round of the World Values Survey, finding some interesting correlations.

Acknowledgments

This material is based in part upon work supported by the Michigan Institute for Data Science, by the National Science Foundation (grant #1815291), and by the John Templeton Foundation (grant #61156). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Michigan Institute for Data Science, the National Science Foundation, or the John Templeton Foundation.

References

1. Ball-Rokeach, S., Rokeach, M., Grube, J.W.: The Great American Values Test: Influencing Behavior and Belief Through Television. Free Press, New York, New York, USA (1984)
2. Bojanowski, P., Grave, E., Joulin, A., Mikolov, T.: Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606 (2016)
3. Boyd, R.L., Wilson, S.R., Pennebaker, J.W., Kosinski, M., Stillwell, D.J., Mihalcea, R.: Values in words: Using language to evaluate and understand personal values. In: Ninth International AAAI Conference on Web and Social Media (2015)
4. Chung, C.K., Pennebaker, J.W.: Finding values in words: Using natural language to detect regional variations in personal concerns. In: Geographical psychology: Exploring the interaction of environment and behavior, pp. 195–216 (2014). https://doi.org/10.1037/14272-011
5. Faulkner, S.L., Baldwin, J.R., Lindsley, S.L., Hecht, M.L.: Layers of meaning: An analysis of definitions of culture. Redefining culture: Perspectives across the disciplines pp. 27–51 (2006)
6. Garten, J., Hoover, J., Johnson, K.M., Boghrati, R., Iskiwitch, C., Dehghani, M.: Dictionaries and distributions: Combining expert knowledge and large scale textual data content analysis. Behavior Research Methods **50**(1), 344–361 (2018)
7. Graham, J., Haidt, J., Nosek, B.A.: Liberals and conservatives rely on different sets of moral foundations. Journal of personality and social psychology **96**(5), 1029 (2009)
8. Hofstede, G.: Dimensionalizing cultures: The hofstede model in context. Online readings in psychology and culture **2**(1), 8 (2011)
9. Inglehart, R., Haerpfer, C., Moreno, A., Welzel, C., Kizilova, K., Diez-Medrano, J., Lagos, M., Norris, P., Ponarin, E., Puranen, B., et al.: World values survey: Round six-country-pooled datafile 2010-2014. JD Systems Institute, Madrid (2014)
10. Kendall, M.G.: The treatment of ties in ranking problems. Biometrika pp. 239–251 (1945)
11. Knafo, A., Roccas, S., Sagiv, L.: The value of values in cross-cultural research: A special issue in honor of shalom schwartz (2011)
12. Landes, S., Leacock, C., Tengi, R.I.: Building semantic concordances. WordNet: an electronic lexical database **199**(216), 199–216 (1998)
13. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(Nov), 2579–2605 (2008)
14. Miller, G.: WordNet: An electronic lexical database. MIT press (1998)
15. Pennebaker, J.W., Boyd, R.L., Jordan, K., Blackburn, K.: The development and psychometric properties of liwc2015. Tech. rep. (2015)
16. Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., Zettlemoyer, L.: Deep contextualized word representations. arXiv preprint arXiv:1802.05365 (2018)
17. Ramirez-Esparza, N., Chung, C.K., Kacewicz, E., Pennebaker, J.W.: The psychology of word use in depression forums in english and in spanish: Testing two text analytic approaches. In: In the International AAAI Conference on Web and Social Media (2008)
18. Rohan, M.J.: A Rose by Any Name? The Values Construct. Personality and Social Psychology Review **4**(3), 255–277 (2000). https://doi.org/10.1207/S15327957PSPR0403_4

19. Rokeach, M.: Beliefs, Attitudes, and Values., vol. 34. Jossey-Bass, San Francisco (1968)
20. Rokeach, M.: The nature of human values, vol. 438. Free press New York (1973)
21. Schwartz, S.H.: Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries. Advances in Experimental Social Psychology **25**, 1–65 (1992). https://doi.org/10.1016/S0065-2601(08)60281-6
22. Schwartz, S.H.: An overview of the schwartz theory of basic values. Online readings in Psychology and Culture **2**(1), 11 (2012)
23. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M.: Lexicon-based methods for sentiment analysis. Computational linguistics **37**(2), 267–307 (2011)
24. Wilson, S.R., Shen, Y., Mihalcea, R.: Building and validating hierarchical lexicons with a case study on personal values. In: International Conference on Social Informatics. pp. 455–470. Springer (2018)