Toward Three-Stage Automation of Annotation for Human Values

Emi Ishita,[1] Satoshi Fukuda,[1] Toru Oga,[1] Douglas W. Oard,[2]
Kenneth R. Fleischmann,[3] Yoichi Tomiura[1] and An-Shou Cheng[4]

[1] Kyushu University, Fukuoka 819-0395, Japan
[2] University of Maryland, College Park, MD 20742, USA
[3] University of Texas at Austin, Austin, TX 78705, USA
[4] National Sun Yat-sen University, Kaohsiung 804, Taiwan
ishita.emi.982@m.kyushu-u.ac.jp

Abstract. Prior work on automated annotation of human values has sought to train text classification techniques to label text spans with labels that reflect specific human values such as freedom, justice, or safety. This confounds three tasks: (1) selecting the documents to be labeled, (2) selecting the text spans that express or reflect human values, and (3) assigning labels to those spans. This paper proposes a three-stage model in which separate systems can be optimally trained for each of the three stages. Experiments from the first stage, document selection, indicate that annotation diversity trumps annotation quality, suggesting that when multiple annotators are available, the traditional practice of adjudicating conflicting annotations of the same documents is not as cost effective as an alternative in which each annotator labels different documents. Preliminary results for the second stage, selecting value sentences, indicate that high recall (94%) can be achieved on that task with levels of precision (above 80%) that seem suitable for use as part of a multi-stage annotation pipeline. The annotations created for these experiments are being made freely available, and the content that was annotated is available from commercial sources at modest cost.

Keywords: Content Analysis, Human Values, Text Classification, Annotation Cost, Nuclear Power Debate.

1 Introduction

Sentiment analysis has provided an effective and efficient way to measure popular sentiment at Web scale, allowing us to detect trends such as positive sentiment toward a particular politician or negative sentiment toward a specific policy position [12]. However, a much deeper and more interesting question is why people hold the sentiment that they hold. Human values, or what people consider important in life [4], can be used to predict individuals' attitudes [15] and behaviors [13]. Thus, this study adopts the approach of attempting to automate the annotation of human values in documents about contentious public policy issues to provide a richer and more nuanced understanding of popular views toward specific policy issues [3].

Earlier studies of values in public policy debates include Cheng et al.'s analysis of the relationship between values and sentiment in net neutrality public hearings [2] and Verma et al.'s examination of the relationship between trust and human values for online information [17]. These studies were based upon relatively modest corpora, which could be annotated by hand by experts or crowd workers. However, today, millions of people's opinions can be found in social networking services such as Facebook and Twitter, on discussion forums, through reports in the mass media, and in many other places. Making productive use of this wealth of human expression demands, however, that we analyze huge amounts of content efficiently and effectively, at a scale that would be impossible to do by hand. There has been some work on automated content analysis (e.g., [5, 10, 11, 18]), but much remains to be done if we are to make it possible for social scientists to extend their analysis to this scale.

Prior work on automating values classification by Takayama et al. makes two implicit assumptions [14]. First, they assume that the documents to be labeled have already been assembled. While this can be a reasonable assumption for small-scale studies in which content analysis is performed by people, scaling that process up to larger collections requires some level of automation in the determination of which documents should be labeled. Second, they assume that the process of determining which spans of text should be labeled can be conflated with the process of determining which human values label(s) should be assigned to each span. This conflation is reasonable when most sentences should express or reflect some human value(s), as was the case in the collection with which Takayama et al. worked, but there are other settings in which the expression of human values is less frequent.

The approach to automating parts of the content analysis process that we propose in this paper thus consists of three stages; 1) identification of the documents to be labeled, (2) determination of which text spans should be labeled in each document, and (3) annotating of human values for those selected text spans. For the second and third stages, we adopt Takayama's approach of using sentences as text spans, which limits the complexity of the text span selection without a serious adverse effect on the utility of the labeling process as a basis for content analysis, as shown by Cheng et al. [6]. As a result, our first stage is binary (on-/off-topic) topic classification at document scale, our second stage is non-topical binary classification (values present/values absent) at sentence scale, and our third stage, as in Takayama et al, is non-topical multi-label classification for each possible values label, again at sentence scale.

In this paper, we report findings from automatic on/off topic identification using classifiers, as well as determining whether value invocations would be interpreted by an annotator as present or absent within sentences from on-topic articles. Classifiers require significant amounts of annotation content for training, annotations that can be expensive to obtain. We are therefore interested in techniques that make the best use of limited amounts of human annotation effort. We consider two ways of creating training annotations, one in which annotators work together to create the best possible annotations for each document, and the other in which they work along to annotate the largest possible number of different documents for a given level of annotation effort. We then compare classifiers trained with each approach to annotation by plotting learning curves that show how well we can do as the number of annotations increases. We find that in cases where we have limited training data, training a classifier with lower quality, but lower cost, annotations can be a cost-effective choice. We also show, using a more limited set of experiments, that promising levels of precision and recall can be achieved for the value sentence identification task. First, however, we begin by introducing the new corpus of newspaper editorials that we have assembled for these experiments.

2 Nuclear Power Debate Editorial News Corpus

Our focus in this work is on automation of content analysis to identify the role of human values in the nuclear power debate in Japan. The Great East Japan Earthquake occurred on March 11, 2011, damaging the Fukushima Daiichi nuclear power plant. After this disaster, the safety of other nuclear power plants received extensive discussion in the Japanese mass media. This event also reignited the global debate over the safety of nuclear power that dates back to the early days of nuclear power, implicating a broad range of human values, and playing out differently in different countries. We have chosen newspaper editorials as our focus for this study, but other work is also looking at the ways in which human values are expressed in newspaper articles. Most newspaper articles adopt a third-person perspective to report on events and the statements of others, but we have chosen to focus here on newspaper articles because we might expect to find a first-person perspective in editorials.

We performed focused collection to obtain a Japanese Editorial corpus. To create the corpus, we searched for editorials (as labeled by metadata) in the commercially available Mainichi Shimbun CD-ROM [1] from 2011 to 2016 using the query “原発 (abbreviation of nuclear power plant) OR 原子力 (nuclear power)”. This query retrieved 750 documents. There are 242 articles from 2011, 166 from 2012, 124 from 2013, 93 from 2014, 62 from 2015, and 63 from 2016. The number of editorials decreased with the passing of years. However, nuclear power has become somewhat of a meme in Japanese society, often invoked tangentially or as a point of comparison, rather than being the focus of the editorial. It is therefore necessary to separate those documents that we wish to analyze from those that simply happen to use the term. This on-topic/off-topic classification task is our first stage.

To support text classification experiments for this first stage, we performed 10 rounds of annotation. In Round 1 we randomly selected 7 editorials from the first year of the corpus. Two annotators, the first and third authors of this paper, both native speakers of Japanese, then independently annotated each editorial as on topic or off topic. The two annotators then met and created adjudicated on/off topic labels by consensus, formalizing the basis for their decisions as written annotation guidelines. As Table 2 shows, Annotator A, a social science researcher, exhibited perfect agreement with the consensus; Annotator B, an information science researcher who was new to the task, achieved considerably lower agreement with the consensus adjudicated judgments. As shown in Table 1, this process of independently annotating, reaching consensus, and updating the annotation guidelines was then repeated for a total of

10 rounds of progressively increasing size,[1] in each subsequent case using documents selected from across the time span of the corpus. As can be seen in Table 2, Annotator B's decisions came to more closely agree with the consensus as iterations proceeded, ultimately achieving noticeably closer agreement with the consensus than Annotator A in two of the last three rounds.

Table 1. Stratified document samples for the ten-round annotation process.

	2011	**2012**	**2013**	**2014**	**2015**	**2016**	**Total**
Round 1	7						7
Round 2		5	3	2	1	2	13
Round 3	8	4	2	2	2	2	20
Round 4	8	4	2	2	2	2	20
Round 5	7	7	5	5	3	3	30
Round 6	7	7	5	5	3	3	30
Round 7	7	7	5	5	3	3	30
Round 8	33	22	17	12	8	8	100
Round 9	33	22	17	10	8	8	98
Round 10	33	22	17	12	8	8	100
Total	143	100	73	55	38	39	448

Table 2. Cohen's Kappa for each round.

Round	#doc	A vs. B	Adjudicated vs. A	Adjudicated vs. B
1	7	0.588	1.000	0.588
2	13	0.806	0.806	1.000
3	20	0.875	1.000	0.875
4	20	0.494	0.886	0.583
5	30	0.430	0.796	0.605
6	30	0.474	0.735	0.730
7	30	0.437	0.769	0.651
8	100	0.535	0.705	0.816
9	98	0.464	0.730	0.741
10	100	0.477	0.635	0.821

3 On/Off Topic Identification

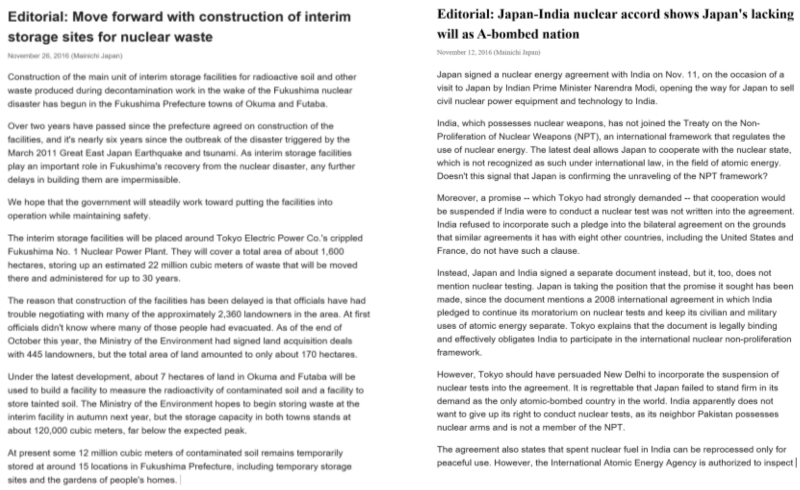
3.1 Annotation Guidelines

The central principle of our annotation guidelines is that a circumstance, a current status, or government policy related to the Fukushima Daiichi nuclear power disaster are on topic content. For examples, editorials describing discussions caused by the disaster, or that expressed value judgments regarding one or more events relating the

[1] Two documents in Round 9 were discovered to be duplicates and removed.

disaster, would be classified as on topic. On the other hand, an editorial about a government or local election would be classified as off topic if it mentioned that issues related to the disaster were important in a campaign, but those issues were not the focus of the editorial. Editorials about restarting some other nuclear power plant, without mentioning the Fukushima Daiichi disaster, would also be off topic.

Figure 1 illustrates some of the challenges of topic classification. The editorial on the left is about the contentious issue of storage of nuclear waste that results from the operation of nuclear power plants. The editorial on the right, by contrast, is about commercial overseas sales of nuclear power technology. Unquestionably, both address topics related to the commercial production of nuclear power, but according to our annotation guidelines, only the editorial on the left is on topic. The focus of the editorial on the right is on limiting the proliferation of nuclear weapons by imposing limitations on how the spent fuel for nuclear power plants can be reprocessed.



Editorial: Move forward with construction of interim storage sites for nuclear waste

November 26, 2016 (Mainichi Japan)

Construction of the main unit of interim storage facilities for radioactive soil and other waste produced during decontamination work in the wake of the Fukushima nuclear disaster has begun in the Fukushima Prefecture towns of Okuma and Futaba.

Over two years have passed since the prefecture agreed on construction of the facilities, and it's nearly six years since the outbreak of the disaster triggered by the March 2011 Great East Japan Earthquake and tsunami. As interim storage facilities play an important role in Fukushima's recovery from the nuclear disaster, any further delays in building them are impermissible.

We hope that the government will steadily work toward putting the facilities into operation while maintaining safety.

The interim storage facilities will be placed around Tokyo Electric Power Co.'s crippled Fukushima No. 1 Nuclear Power Plant. They will cover a total area of about 1,600 hectares, storing up an estimated 22 million cubic meters of waste that will be moved there and administered for up to 30 years.

The reason that construction of the facilities has been delayed is that officials have had trouble negotiating with many of the approximately 2,360 landowners in the area. At first officials didn't know where many of those people had evacuated. As of the end of October this year, the Ministry of the Environment had signed land acquisition deals with 445 landowners, but the total area of land amounted to only about 170 hectares.

Under the latest development, about 7 hectares of land in Okuma and Futaba will be used to build a facility to measure the radioactivity of contaminated soil and a facility to store tainted soil. The Ministry of the Environment hopes to begin storing waste at the interim facility in autumn next year, but the storage capacity in both towns stands at about 120,000 cubic meters, far below the expected peak.

At present some 12 million cubic meters of contaminated soil remains temporarily stored at around 15 locations in Fukushima Prefecture, including temporary storage sites and the gardens of people's homes.

Editorial: Japan-India nuclear accord shows Japan's lacking will as A-bombed nation

November 12, 2016 (Mainichi Japan)

Japan signed a nuclear energy agreement with India on Nov. 11, on the occasion of a visit to Japan by Indian Prime Minister Narendra Modi, opening the way for Japan to sell civil nuclear power equipment and technology to India.

India, which possesses nuclear weapons, has not joined the Treaty on the Non-Proliferation of Nuclear Weapons (NPT), an international framework that regulates the use of nuclear energy. The latest deal allows Japan to cooperate with the nuclear state, which is not recognized as such under international law, in the field of atomic energy. Doesn't this signal that Japan is confirming the unraveling of the NPT framework?

Moreover, a promise -- which Tokyo had strongly demanded -- that cooperation would be suspended if India were to conduct a nuclear test was not written into the agreement. India refused to incorporate such a pledge into the bilateral agreement on the grounds that similar agreements it has with eight other countries, including the United States and France, do not have such a clause.

Instead, Japan and India signed a separate document instead, but it, too, does not mention nuclear testing. Japan is taking the position that the promise it sought has been made, since the document mentions a 2008 international agreement in which India pledged to continue its moratorium on nuclear tests and keep its civilian and military uses of atomic energy separate. Tokyo explains that the document is legally binding and effectively obligates India to participate in the international nuclear non-proliferation framework.

However, Tokyo should have persuaded New Delhi to incorporate the suspension of nuclear tests into the agreement. It is regrettable that Japan failed to stand firm in its demand as the only atomic-bombed country in the world. India apparently does not want to give up its right to conduct nuclear tests, as its neighbor Pakistan possesses nuclear arms and is not a member of the NPT.

The agreement also states that spent nuclear fuel in India can be reprocessed only for peaceful use. However, the International Atomic Energy Agency is authorized to inspect

Fig 1. English examples of on-topic (left) and off-topic (right) editorials (trimmed for display). The actual articles used in the study are in Japanese.

3.2 Experimental Design

The second author of this paper used the resulting annotations to train and evaluate automated text classifiers for the on-off topic identification task. We first used the standard Juman package 7.01 [8] to perform tokenization and morphological analysis for Japanese. We performed feature selection by using only nouns, verbs and adjectives that occur two or more times in the corpus as the term space for each classifier. We used all such terms found in the body of each editorial. The bodies of the editorial contained, on average, 668.2 terms before feature selection, and 323.0 terms after feature selection. We conducted experiments with two types of classifiers. The first was a Support Vector Machine (SVM), a commonly used binary classifier that we implemented using TinySVM [16], for which we tried either a linear or a quadratic kernel (second-order polynomial kernel) function. The second was fastText, a neural network "deep learning" classifier developed by Facebook's AI Research lab [7].[2]

Figure 2 illustrates the evaluation design. To create learning curves, we first randomly shuffle the documents within each of the first 9 annotation rounds. Specifically, we shuffle the 7 documents in Round 1 into a random order, we then shuffle the 13 documents in Round 2 into another random order, and we continue that process through Round 9. We then iteratively train a classifier by incrementally adding training data. Specifically, we train a classifier with just the first document, evaluate that classifier using F_1 (the harmonic mean of precision and recall) using the 100 annotated documents from Round 10 as a held-out evaluation set, then retrain the classifier on the first two documents from Round 1, then evaluate that classifier in the same way, and so on until all 348 documents from Round 1 through Round 9 have been used for training. We repeat that entire process 100 times, with 100 random shuffles of each Round, and then we plot the average F_1 (or accuracy) over those 100 rounds as a learning curve. The evaluation set, which by construction was randomly sampled from the entire collection, is nearly balanced, with 48 on-topic and 52 off-topic editorials. The learning curves for F_1 and accuracy are therefore similar, so for space reasons we show learning curves only for F_1 in this paper.

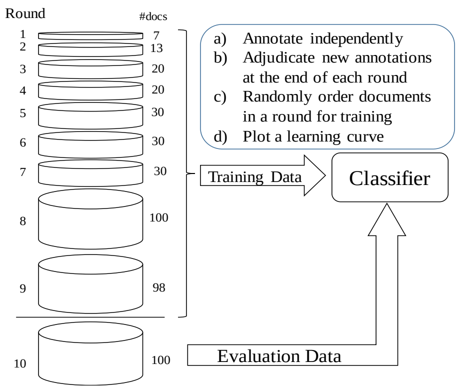


Fig. 2. Experiment design.

We plot learning curves by placing the number of annotations on the horizontal axis and the F_1 value for that number of annotations on the vertical axis. We plot four types of learning curves:

- A: we use Annotator A's annotations for training.
- B: we use Annotator B's annotations for training.
- Hybrid: we alternate between using Annotator A's or Annotator B's annotations for training.
- Adjudicated: we use the adjudicated annotations for training.

Note that creating one adjudicated annotation requires that Annotator A and Annotator B each first independently annotate the same document. In order to plot all four learning curves consistently, we randomly select half of the documents from each

[2] As fastText parameters we selected: -dim 50; -loss negative sampling (ns); -epoch 1000; -wordNgrams 5; -minCount 2.

round when training using Adjudicated annotations, and we plot the resulting values based on the original number of independent annotations. For this reason, the learning curve for Adjudicated annotations are plotted only at even numbers of annotations.[3] Note also that it is not possible to meaningfully train a classifier until at least one on-topic and at least one off-topic editorial are available for training, and with some annotators and some random shuffles that does not happen until several annotated documents have been seen. We therefore suppress the plotting of the learning curve until meaningful F_1 values are available for all 100 random shuffles.

3.3 Results

Figure 3 shows four learning curves, all of which are evaluated using adjudicated annotations from Round 10. These are for a quadratic kernel SVM, which was the better of the two SVM kernel functions that we tried. As can be seen, Annotator A's annotations initially provide the best training data, with Annotator B's annotations becoming more useful for training in the later rounds, where they more closely approximate the adjudicated annotations. Alternating annotation between Annotator A and Annotator B (the Hybrid learning curve) unsurprisingly comes out between the two. Using the Adjudicated annotations, which we reasonably presume are of higher quality than the annotations from either annotator, is not the best choice, at least not until rather late in the process. The reason for this is that the x axis here shows the number of annotations, and to get adjudicated annotations, every document has to be annotated twice. It seems clear that the additional diversity introduced by having different annotators annotate different documents (as in the Hybrid learning curve) yields better results than having different annotators annotate the same documents.

Figure 4 shows learning curves for the same training conditions, in this case evaluated using either Annotator A's independent annotations from Round 10 (left) or Annotator B's (right) independent annotations from Round 10. Unsurprisingly, each annotator's training data is better matched to their own evaluation data, but note that the Hybrid training learning curve still dominates the Adjudicated training learning curve until quite a substantial number of annotations have been obtained. From Figures 3 and 4 we can therefore conclude that the central message of Figure 3, that the Hybrid training strategy works well early in the training process, does not depend on an assumption that our adjudicated annotations are perfect (which they are not).

Figure 5 shows the same four learning curves when using the fastText classifier. As in Figure 3, these are evaluated using the adjudicated judgments from Round 10. The fastText results exhibit a similar pattern to what we observe with the SVM, although the initial learning rate for fastText is somewhat lower. We do see that fastText ultimately does outperform the SVM with Hybrid annotations, but that only happens rather late in the learning process. This is consistent with what we would expect from neural methods, which can achieve good results in data-rich settings. Although not shown, learning curves using Annotator A's or Annotator B's annotations for evaluation show similar patterns.

---

[3] We have also generated learning curves using all of the adjudicated annotations, again plotting only at even values for the number of original annotations. This yields similar results.

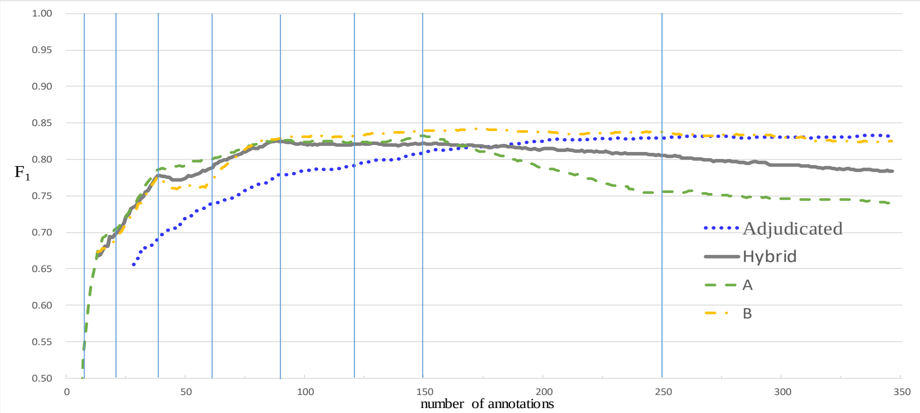


Fig. 3. F_1 for quadratic kernel SVM, 100 held out adjudicated annotations used for evaluation, average of 100 random shuffles within each round (best viewed in color).

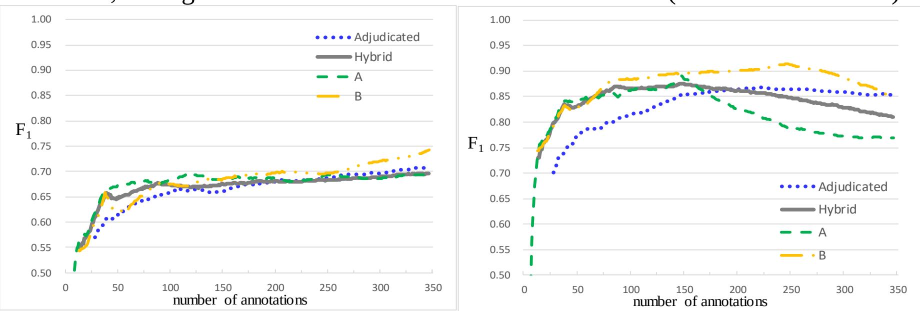


Fig. 4. F_1 for quadratic kernel SVM, 100 held out annotations by Annotator A (left) or Annotator B (right) used for evaluation; average of 100 random shuffles per round.

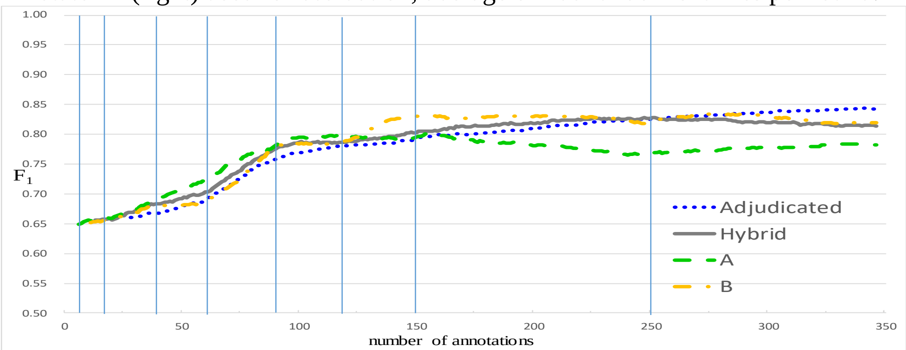


Fig. 5. F_1 for fastText, 100 held out adjudicated annotations used for evaluation, averaged over 100 random shuffles per round.

4 Identifying Value Sentences

Although editorials generally contain more of the author's perspective than news articles, they still include events, facts, and other elements that are present for background. Editorials include, for example, “There are 54 nuclear power plants along the coastline,” “In this time, multiple hypocenters interlocked and caused a huge earthquake,” and “A basic energy plan in Japan that will guide the country's medium- to long-term energy policy will be decided by the Cabinet within the month” (these are English translations from editorials in the corpus). Our interpretation is that these types of sentences do not express or reflect specific values, so we annotate them as “fact sentences.” Here are some examples of what we call value sentences: “It is time to find the appropriate power source for an earthquake-prone country such as Japan and how to live based on it,” “It could be said that securing transparency is the most important issue in selecting new policies,” and “The government and electric power companies need to deepen their political discussion by accepting citizens' anxiety.” For our initial experiments with automatically determining which sentences an annotator would interpret as value sentences, we have built a small training set and conducted a pilot study to characterize classification effectiveness. The same two annotators as above independently reviewed each sentence in a total of 28 on-topic editorials to distinguish between value and fact sentences. Three rounds of independent annotation were used, with adjudication by consensus between the two annotators performed at the end of each round. Table 3 shows the number of documents and sentences in each round, along with agreement statistics before adjudication. A total of 578 of the 762 sentences were annotated as value sentences in the adjudicated annotations.

Table 3. Cohen’s Kappa for each round.

Round	#docs	#sentences	A:B
1	2	91	0.505
2	12	317	0.639
3	14	354	0.640

We then trained a quadratic-kernel SVM classifier on the adjudicated judgments, evaluating it using 10-fold cross-validation at sentence scale. For the first fold we selected the first 686 sentences for training and the remaining 76 for evaluation, repeating a similar process for ten different sets of 76 (or 77) evaluation sentences. As Table 4 shows, this yielded F_1 values near 0.87, regardless of whether we performed feature selecting as above of simply used all of the terms in each sentence as features. Because we plan to perform value labeling only on sentences that have been automatically classified as values sentences, a high level of recall from the stage that finds value sentences is needed. As Table 3 shows, recall values near 0.94 are possible (with feature selection) on this collection.

Table 4. Precision, Recall, and F_1 by SVM with quadratic kernel for finding value sentences.

Features	Precision	Recall	F1
noun, verb, adjective	0.805	0.938	0.867
All	0.818	0.924	0.868

5 Conclusion and Future Work

We have introduced a three-stage model for automating the association of human values with specific passages of text on large-scale collections, and we have shown results for the first two stages: on/off topic detection, and automating the determining of which sentences an annotator would interpret as reflecting values. Prior work has shown that the third stage, automatically assigning values labels to sentences, is possible [6,14]. Moreover, we have shown that when annotation budgets are limited, it can be useful to focus on single-labeled training examples rather than adjudicated training examples that are created by assigning multiple annotators to the same document, and that an SVM classifier can be a better choice than a state of the art neural deep learning classifier. Both of these results are consistent with results that have been reported in other settings (e.g., [9]); our contribution has been to bring these insights to bear in the context of a multi-stage annotation process for human values.

The annotations created for these experiments are being made freely available, and the content that was annotated is available from the publisher at modest cost. Much remains to be done, however. First, our motivation for introducing a stage to predict whether an annotator would interpret a sentence as a value sentence is that we need systems that can operate effectively and efficiently in settings in which values sentences are relatively uncommon. As our sentence counts in Section 4 indicate, however, that is not the case for editorials. We therefore plan to next automate value sentence annotation for news articles, where we expect values sentences to be considerably less common. Second, we of course next need to build classifiers to automatically assign human values labels to the resulting sentences. Preliminary results for that task have already been reported on a collection of 2,100 Japanese news articles on the nuclear power debate for a set of ten human values (effectiveness, human welfare, importance, independence, innovation, law and order, nature, personal welfare, power, wealth) with fairly good recall (between 0.55 and 0.73), but with relatively poor precision (0.05 to .38) [6]. Indeed, it is in part those relatively poor levels of precision when labeling values in news articles that motivates our interest in a multi-stage process. Automating the interpretation of value sentences on that collection of 2,100 news stories will thus be a natural next step in our research.

Acknowledgements. This work has been supported in part by JSPS KAKENHI Grant Number JP18H03495.

References

1. CD-Mainichi Shimbun Data Collection 2011 version; 2012 version; 2013 version; 2014 version; 2015 version; and 2016 version.
2. Cheng, A.-S., Fleischmann, K.R., Wang, P., Ishita, E., Oard, D.W.: The role of innovation and wealth in the net neutrality debate: A content analysis of human values in congressional and FCC hearings. Journal of the American society for information science and technology, 63, 1360-1373 (2012).
3. Fleischmann, K.R.: Information and Human Values. San Rafael, CA: Morgan & Claypool. (2014).
4. Friedman, B., Kahn, P. H. Jr., and Borning, A.: Value sensitive design and information systems. In: P. Zhang and D. Galletta (eds.) Human-computer interaction and management information systems: Foundations, pp.348-372. Armonk, NY: M.E. Sharpe. (2006). DOI: 10.1002/9780470281819.ch4.
5. Grimmer, J., Stewart, B.M.: Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Political Analysis 21, 267-297 (2013).
6. Ishita, E., Oga, T., Takayama, Y., Cheng, A.-S., Oard, D. W., Fleischmann, K. R., Tomiura, Y.: Toward automating detection of human values in the nuclear power debate. In: Proceedings of 80th annual meeting of the association for information science and technology 54(1), pp.714-715(2017). DOI: 10.1002/pra2.2017.14505401127
7. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification, https://arxiv.org/abs/1607.01759, last accessed 2018/9/10.
8. JUMAN (a user-extensible morphological analyze for Japanese), http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?JUMAN, last accessed 2018/9/10.
9. Khetan, A., Lipton, Z. C., Anandkumar, A., Learning from noisy singly-labeled data. In: Proceedings of ICLR 2018. 15p. (2018), https://arxiv.org/abs/1712.04577, last accessed 2018/9/10.
10. Nelson, L.K.: Computational grounded theory: A methodological framework. Sociological methods and research (2017). https://doi.org/10.1177/0049124117729703
11. Nelson, L.K., Burk, D., Knudsen, M., McCall, L.: The future of coding: A comparison of hand-coding and three types of computer-assisted text analysis methods. Sociological methods and research (2018). https://doi.org/10.1177/0049124118769114
12. Pang, B. and Lee, L.: Opinion mining and sentiment analysis. Journal of foundations and trends in information retrieval, 2(1-2), pp. 1-135. (2008). DOI: 10.1561/1500000011.
13. Schwartz, S.H.: Value orientations: Measurement, antecedents and consequences across nations. In: R. Jowell, C. Roberts, R. Fitzgerald, and G. Eva (eds.), Measuring attitudes cross-nationally: Lessons from the European social survey, pp.169-203. London: Sage. (2007). DOI: 10.4135/9781849209458. n9.
14. Takayama, Y., Tomiura, Y., Ishita, E., Oard, D. W., Fleischmann, K. R., Cheng, A.-S.: A word-scale probabilistic latent variable model for detecting human values. In proceedings on ACM international conference on information and knowledge management (CIKM 2014), p.1489-1498 (2014). DOI 10.1145/2661829.2661966
15. Templeton, T.C. and Fleischmann, K.R.: The relationship between human values and attitudes toward the Park51 and nuclear power controversies. In: Proceedings of the 74th annual meeting of the American society for information science and technology, New Orleans, LA. (2011). DOI: 10.1002/ meet.2011.14504801172.
16. TinySVM: Support Vector Machines, http://chasen.org/~taku/software/TinySVM/, last accessed 2018/9/10.

17. Verma, N., Fleischmann, K. R., Koltai, K. S.: Human values and trust in scientific journals, the mainstream media and fake news, In: In: Proceedings of 80th annual meeting of the association for information science and technology 54(1), 426-435 (2017).
18. Yan, J.L.S., McCracken, N., Crowston, K.: Semi-automatic content analysis of qualitative data. In: Proceedings of the iConference, 1128-1132(2014).